BrJAC

Brazilian Journal of Analytical Chemistry

an International Scientific Journal

Photo: Carlos Dias
Cover: André Marcelo de Souza

Special Issue dedicated to Prof. Ronei J. Poppi

July – September 2021 Volume 8 Number 32

XXIII INTERNATIONAL MASS SPECTROMETRY CONFERENCE

RIO DE JANEIRO 2021

IMSC 2021 will cover all aspects of mass spectrometry, from fundamentals to instrumentation and applications, with the most recent findings and innovations and a very selected group of renowned speakers but with a great addition: our "Brazillian seasoning".

The conference is organized with short courses, oral and poster sessions, workshops, awards, commercial exhibitions, as well as very exciting social programs and companies user meetings. All major MS companies are supporting Rio 2021 and will be in the exposition area presenting their latest products and innovations.

We are working very hard to make Rio 2021 the largest and greatest ever IMSC and for it to be an exceptional occasion for you to enjoy science, to grow your network and make new friends and colleagues, to get new ideas for your research, for establishing new collaborations, starting new projects, and more; all in the unforgettably warm, beautiful and charming city of Rio de Janeiro!

Please come, bring your students, invite friends and colleagues and contribute to make Rio 2021 a very successful, most amazing and unforgettable IMSC ever!

You are very welcome to RIO 2021!

Access WEBSITE

Welcome to the

23rd **International Mass Spectrometry Conference** 2021 in Rio de Janeiro

December 12 to December 16

IMSF international mass spectrometry foundation

Sociedade Brasileira de Espectrometria de Massas

Brazilian Journal of Analytical Chemistry

VISÃO FOKKA - COMUNICATION AGENCY

The *Braz. J. Anal. Chem.* (BrJAC) is a peer-reviewed, open access, scientific journal intended for professionals and institutions acting mainly in all branches of analytical chemistry.

ISSN 2179-3425 printed ISSN 2179-3433 eletronic

Scope

BrJAC is dedicated to the diffusion of significant and original knowledge in all branches of Analytical Chemistry and Bioanalytics. BrJAC is addressed to professionals involved in science, technology and innovation projects in Analytical Chemistry at universities, research centers and in industry.

BrJAC is a quarterly journal that publishes original, unpublished scientific articles, reviews and technical notes that are peer reviewed in the double-blind way. In addition, it publishes interviews, points of view, letters, sponsor reports, and features related to analytical chemistry. Once published online, a DOI number is assigned to the paper.

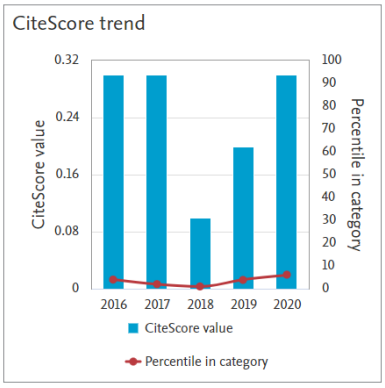
Manuscripts submitted for publication in BrJAC cannot have been previously published or be currently submitted for publication in another journal.

For complete information on ethics and policies on conflicts of interest, copyright, reproduction of already published material and preprints, in addition to the manuscript submission and peer review system, please visit “About us” and Guidelines for Authors at www.brjac.com.br

Indexing Sources

Scopus

CiteScore 2020: 0.3; SJR: 0.131; SNIP: 0.146. CiteScore Tracker 2021: 0.4 Access source details



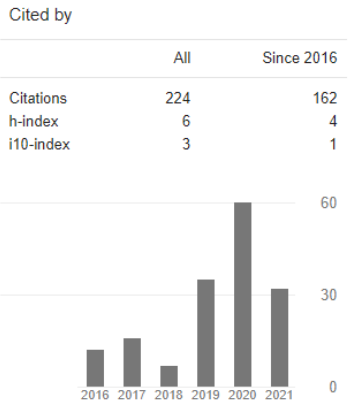
CLARIVATE ANALYTICS
EMERGING SOURCES CITATION INDEX
INDEXED

Expectancy: to be in the SCIE in 2021

CAPES
QUALIS

Expectancy: to be classified in the A3 stratum for the Quadrennial Evaluation in 2022

Google scholar



Production Editor
Silvana Odete Pisani

Publisher
Lilian Freitas
MTB: 0076693/ SP
lilian.freitas@visaofokka.com.br

Advertisement
Luciene Campos
luciene.campos@visaofokka.com.br

Art Director: Adriana Garcia

WebMaster: Daniel Letieri

BrJAC’s website: www.brjac.com.br / Contact: brjac@brjac.com.br
Like BrJAC on Facebook: https://www.facebook.com/brjachem

ABEC BRASIL
BrJAC is associated to the Brazilian Association of Scientific Editors

VISÃO FOKKA
BrJAC is published quarterly by:
Visão Fokka Communication Agency
Av. Washington Luiz, 4300 - Bloco G - 43
13042-105 – Campinas, SP, Brazil
contato@visaofokka.com.br
www.visaofokka.com.br

EDITORIAL ADVISORY BOARD (Continuation)

CONTENTS

CONTENTS

CONTENTS

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 1-5
doi: 10.30744/brjac.2179-3425.editorial.mcbreitkreitz.N32

EDITORIAL

Analytical Quality by Design

Márcia Cristina Breitkreitz

Guest Editor of this special edition on Chemometrics devoted to Prof. Ronei J. Poppi
Professor at the Institute of Chemistry, University of Campinas
Campinas, SP, Brazil

Throughout the 20th century, the notion of 'quality' underwent major changes. The concept of '*Quality by Testing*', i.e. evaluating the quality of a product by testing it for some pre-defined parameters after completing the manufacturing process, started to be replaced by the enhanced approach of *Quality by Design* (QbD). In this concept, idealized by Joseph Juran [1], quality is initially conceived for the product, and *then* the product is manufactured and evaluated to reach that quality. Driven by the need to reduce costs and to encourage companies to improve their understanding of their products and manufacturing processes, pharmaceutical regulatory agencies, through the ICH (International Council on Harmonization), published the ICH Q8 guideline, highlighting the QbD strategy for pharmaceutical development [2]. This represented a significant step in replacing the existing quality paradigms in the conventional manufacturing of pharmaceutical products and paved the way to support development and production activities within a scientific, flexible environment, with a high level of quality, without the need for extensive regulatory surveillance. The higher level of understanding gained during development allows out-of-specification batches to be foreseen and ensures that the desired quality will be obtained at the end of the manufacturing process. The two pillars of the QbD concept are quality risk management and multivariate study of the outcomes (quality parameters) as functions of the inputs (materials, formulation and process inputs). At this point, Chemometrics emerged officially as a strategy to support the pharma regulatory requirements- the beginning of a joyful and long-life union.

Soon analytical researchers realized that the QbD concepts could be readily applied to the analytical method development, since it is also a 'process' with quality requirements to be fulfilled. This generated what is nowadays called '*Analytical Quality by Design'* (AQbD) [3-5]. The development of increasingly efficient analytical methods presents itself as a fundamental aspect in the pharmaceutical industry, since the analysis of drugs in raw materials, intermediates and finished products is necessary in practically the entire process of the development and production of medicines. High Performance Liquid Chromatography (HPLC) is the most commonly used technique in the pharmaceutical industry [6], along with the modern version of Ultra High Performance Liquid Chromatography (UHPLC). Therefore, the majority of efforts in AQbD are directed towards these techniques.

In the same way as with QbD, AQbD is driven by quality risk management and understanding of the quality parameters of the method as functions of the inputs (experimental) variables. To implement the principles of AQbD, one must first define all of the requirements that the method must meet in terms of quality; that is, define the Analytical Target Profile (ATP), which includes both procedure attributes and validation parameters, depending on the purpose of the method and the guidelines being followed. From the ATP, the attributes that are critical for the proper performance of the method should be identified; these are called Critical Quality Attributes – CQA. Some examples of CQA include the adequate retention in the column (measured by the retention factor, k); appropriate resolution among all target peaks in the chromatogram (measured by the resolution R_s or separation factor, α); adequate symmetry for integration

Cite: Breitkreitz, M. C. Analytical Quality by Design. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 1-5. doi: http://dx.doi.org/10.30744/brjac.2179-3425.editorial.mcbreitkreitz.N32

(measured by tailing factor or an equivalent parameter); plate number; number of peaks observed in the chromatogram, etc. In addition, a target value or acceptance limit must be established for each of them.

The next step involves identifying the critical experimental variables that influence these quality attributes, using prior knowledge and risk assessment tools (e.g. Ishikawa diagram), and then assigning a risk to each of them – which can be done either by a formal risk analysis tool such as FMEA (*Failure Mode Effect Analysis*) or through risk analysis tables. An important aspect is the justification of the risk assigned to each variable according to analytical knowledge of the samples being analyzed and the analytical technique used. The high-risk variables are called Critical Procedure Parameters (CPP) or Critical Method Parameters (CMP) and some examples are: column type (chemical composition of the stationary phase and column length), composition of the mobile phase (type of organic modifier, pH of the aqueous phase, ion-pairing reagent type), gradient slope or gradient time, temperature of the column, flow rate, etc.

The QbD strategy recommends the study of the CQA as a function of the CPP in a multivariate way and, for this reason, the tools of Design of Experiments – (DOE) are fundamental in this strategy. This procedure has several advantages over the univariate OFAT (*One-factor-at-a-time*) method, which is normally time-consuming and often does not bring satisfactory solutions, due to interactions of the variables involved (i.e., the response obtained by changing variable A depends on the level of variable B), which can only be identified using multivariate methods. In addition, in the OFAT method, the results are only obtained at the points where the experiments are carried out, not allowing characterization of the entire experimental domain. On the other hand, multivariate methods may allow an understanding of how the CQA changes depending on the variation of the levels of CPP through the estimation of mathematical models which allow the visualization of results by plotting response surfaces, and the prediction of responses within the experimental domain without the need for extra laboratory work.

Since chromatographic method development involves many variables, different DOE methods can be used according to the nature of these variables (statistically independent/non-independent variables), the goal of the experiment and the number of variables involved. As a general practical guide, the primary variables that affect selectivity in liquid chromatography - type of stationary phase, nature of organic modifier (typically methanol (acidic property), acetonitrile (dipolar) or tetrahydrofuran (basic) in reversed phase liquid chromatography) and pH range (acidic, neutral, basic ranges) - should be screened in the first step of method development. Since initial guidance is the objective of this step, economical designs, such as fractional factorial or optimal designs are recommended. Optimal designs are versatile computer-generated designs that set the experiments according to a given pre-defined mathematical criterion, for example, D-, G-, A-, I-optimality criteria. They are especially welcome in situations when constraints need to be imposed on the design region, when a nonstandard model is necessary to adequately explain the response or when the number of experiments should be restrained to estimate a pre-defined model [7].

A second step encompasses the use of design methods to estimate the coefficients of a prediction model, with pre-defined CQA as responses, and temperature, flow rate, gradient time (or slope) and pH (in a narrower range) typically being the input variables. For this purpose, Full Factorial, Central Composite, Box-Behnken, Doehlert and optimal designs can be used. It should be highlighted that Full Factorial designs only allow the estimation of linear models, whereas the other methods mentioned allow higher order coefficients to be estimated. Mixture designs can be used for optimization of the mobile phase composition, considering solvents with different characteristics as the mixture ingredients – the selection can be made using the well-known Snyder's selectivity triangle for liquid chromatography. A detailed description of DOE methods applied to chromatography can be found in many references [8-11].

In order to ensure that quality parameters are achieved for all compounds in the same chromatographic run, a procedure for the simultaneous optimization of several responses can be employed. Even though many 'response functions' attempt to describe the overall separation by a single value, experience shows that the best way is to model individual responses and then combine them using a multiple response optimization method, such as the contour map overlay and the desirability functions of Derringer and Suich [12-13]. These methods allow a very important precept of QbD to be achieved, which is the '*Design Space*'

(DS) or '*Method Operable Design Region'* (MODR), defined as the "multidimensional combination of variables that influence a method and that guarantee the quality of the data produced by it" [14]. According to this definition, the MODR is considered a robust work region, since the experimental variations within this region do not cause changes in the quality attributes of the method. It is worth mentioning that, at this stage, a careful statistical analysis of the mathematical models from the selected DOE that generated the MODR must be done, using, for example, analysis of variance (ANOVA), confidence intervals and analysis of residuals.

The major benefit of building an MODR is to incorporate the robustness into the development, instead of evaluating it in the validation step, as is currently done. This avoids undesirable surprises during validation and may allow changes to be made over the method's lifecycle in a safe way, without the need for regulatory approval [15,16]. An important step of increasing interest is the evaluation of MODR uncertainty limits. Monte-Carlo simulations, Bayesian modelling and bootstrapping are useful tools to accomplish this task and ensure the robustness within the MODR [17,18]. Estimates of confidence intervals for the predicted responses also represent a way to take into account the uncertainty of MODR.

A concrete proof that this approach is the future of analytical development by liquid chromatography is the launching of commercial software for method optimization and data treatment, already connected to or integrated with analytical instruments. The most representative example is Fusion QbD LC method development (S-Matrix, California, EUA) which provides a full range of DOE tools to support liquid and supercritical fluid chromatography method development and which is integrated in the software of major instrument manufacturers, such as Empower (Waters Technologies), Chromeleon (Thermo Scientific) and OpenLab Chem Station (Agilent). Fusion QbD mainly works with optimal designs (other designs are also available) to select the minimum number of runs to obtain information in a two-step approach - Screening and Optimization. The sequence of runs is organized by the software and information is exported to the instrument. After completion, chromatographic peaks should be integrated by the user and the results exported back to Fusion. Then, the software allows a complete statistical data analysis, including model selection and evaluation by ANOVA and residual graphs; response surfaces are built and desirability functions combined with Monte Carlo simulations allow the MODR to be determined with uncertainty boundaries. Also, predicted chromatograms can be plotted in real-time according to the variations made within the studied experimental region. This certainly represents the state-of-the art in the LC method development.

Another outcome that highlights the importance of AQbD is the fact that this strategy has been recognized by official entities all over the world. In June 2019, the British Medicines and Healthcare products Regulatory Agency (MHRA) published the results of a public consultation on the use of AQbD principles to pharmacopeial standards for medicines. This was accompanied by the publication of a technical review of an MHRA project to explore the application of AQbD to a pharmacopeial assay procedure [19]. Responses to that public consultation declared that the MHRA will support and complement the evolution of developing regulatory science, with the application of AQbD concepts [20]. The United States Pharmacopeia (USP) has published a proposed new General Chapter <1220> The Analytical Procedure Lifecycle aiming at describing the current thinking of the USP Validation and Verification Expert Panel which advises the General Chapters – Chemical Analysis Expert Committee with regard to future trends in analytical procedure development, qualification, and continued monitoring. This general chapter article had the purpose of describing an alternative approach to the classical process of analytical validation and subsequent verification and transfer, describing these activities as a continuum and closely interrelated events rather than as discrete actions [21]. This is only possible by using AQbD concepts, in particular modelling of the CQA as functions of the CPP and the establishment of the MODR. A similar document is expected to be published this year by the ICH for public consultation, the ICH Q14 *Analytical Procedure Development guideline*, which is based on the principles of the ICH Q8, i.e. sound scientific and risk-based approval, as well as post-approval change management, now for analytical procedures [22].

CONCLUSIONS AND FUTURE PERSPECTIVES

The strategy of varying the experimental conditions in a univariate way and selecting the "best chromatogram" is becoming unacceptable within the context of modern analytical chemistry and quality environments. The use of DOE and the application of multivariate methods to method development are key elements in the context of the Analytical Quality by Design initiative - these will be strictly necessary tools for analytical development in the future, and can no longer be avoided, especially after being adopted by regulators. The development of methods using the univariate strategy, in addition to being very time-consuming, does not always bring satisfactory solutions, because interactions can occur between the many variables involved, which can only be identified using multivariate methods. In addition, results are only obtained at the points where the experiments were carried out, not allowing an overview of the complete experimental domain. In the current method development strategy, robustness is only tested in the validation step; nevertheless, once this stage is reached, one may realize that the method is not robust even to small fluctuations in the experimental conditions meaning that it is necessary to go back to the development stage. With the proposed strategy, the robustness is naturally incorporated into the method development stage.

Acknowledgments

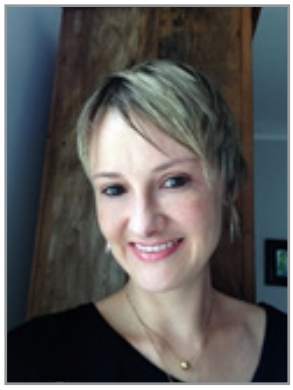
I would like to thank Prof. Dr. Roy Edward Bruns and Prof. Dr. Isabel Cristina Sales Fontes Jardim, my Masters advisors (2004-2007) when I started my studies on DOE applied to chromatography, already knowing it would be something very important in the future. Also thanks to Amanda Guiraldelli Mahr (USP Pharmacopoeia) and Richard Verseput (S-Matrix) for the current partnership in AQbD projects - and very nice discussions.

Finally, I would like to thank and dedicate this Editorial to Prof. Dr. Ronei Jesus Poppi, who was my PhD advisor and became a friend, counselor and example much beyond Chemometrics.

REFERENCES

1. Juran, J. M. *Juran on Quality by Design: The New Steps for Planning Quality Into Goods and Services.* Free Press, New York, **1992,** Chapter 2, p 27.
2. International Council on Harmonization. ICH Q8 (R2) *Pharmaceutical development,* **2009.**
3. Vogt, F. G; Kord, A. S. *J. Pharm. Sci.*, **2011**, *100* (3), pp 797-812 (https://doi.org/10.1002/jps.22325).
4. Orlandini, S.; Pinzauti, S.; Furlanetto, S. *Anal. Bioanal. Chem.*, **2013**, *405*, pp 443–450 (https://doi.org/10.1007/s00216-012-6302-2).
5. Peraman, R.; Bhadraya, K.; Reddy, Y. P. *Int. J. Anal. Chem.*, **2015,** *2015***,** pp 1-9 (http://dx.doi.org/10.1155/2015/868727).
6. Siddiqui, M. R.; AlOthman, Z. A.; Rahman, N. *Arabian J. Chem.*, **2017**, *10*, pp 1409–1421 (http://dx.doi.org/10.1016/j.arabjc.2013.04.016).
7. Johnson, R. T.; Montgomery, D. C.; Jones, B. A. *Qual. Eng.*, **2011**, *23*, pp 287-*301* (https://doi.org/10.1080/08982112.2011.576203).
8. Ferreira, S. L. C.; Bruns, R. E.; da Silva, E. G. P.; dos Santos, W. N. L.; Quintella, C. M.; David, J. M.; de Andrade, J. B.; Breitkreitz, M. C.; Jardim, I. C. S. F.; Neto, B. B. *J. Chromatogr. A*, **2007,** *1158*, pp 2–14 (https://doi.org/10.1016/j.chroma.2007.03.051).
9. Hibbert, D. B. *J. Chromatogr. B*., **2012**, *910*, pp 2–13 (https://doi.org/10.1016/j.jchromb.2012.01.020).
10. Dejaegher, B.; Heyden, Y. V. *J. Pharm. Biomed. Anal.*, **2011**, *56*, pp 141– 158 (https://doi.org/10.1016/j.jpba.2011.04.023).
11. Bezerra, M. A.; Santelli, R. E.; Oliveira, E. P.; Villar, L. S. Escaleira, L. A. *Talanta*, **2008**, *76*, pp 965–977 (https://doi.org/10.1016/j.talanta.2008.05.019).
12. Derringer, G.; Suich R. *J. Qual. Tech.*, **1980**, *12*, pp 214-219 (https://doi.org/10.1080/00224065.1980.11980968).
13. Candioti, L. V.; De Zan, M. M.; Cámara, M. S.; Goicoechea, H. C. *Talanta,* **2014**, *124*, pp 123–138 (https://doi.org/10.1016/j.talanta.2014.01.034).

14. Rozet, E.; Lebrun, P.; Debrus, B.; Boulonger, B.; Humbert, P. *TrAC, Trends in Anal. Chem.*, **2013**, *42*, pp 157 – 167 (https://doi.org/10.1016/j.trac.2012.09.007).
15. Hubert, C.; Houari, S.; Rozet, E.; Lebrun, P.; Hubert, Ph. *J. Chromatogr. A*, **2015,** *1395*, pp 88–98 (http://dx.doi.org/10.1016/j.chroma.2015.03.059).
16. Chatterjee, S. IFPAC Annual Meeting, Baltimore, January 25, **2013**. Available at: http://qbdworks.com/wp-content/uploads/2014/06/AnalyticalQbDIFPAC.pdf [Acessed March 2021].
17. Deidda, R.; Orlandini, S.; Hubert, P.; Hubert, C. *J. Pharm. Biomed. Anal.*, **2018**, *161*, pp 110–121 (https://doi.org/10.1016/j.jpba.2018.07.050).
18. Dispas, A.; Avohou, H. T.; Lebrun, P., Hubert, P.; Hubert, C. *TrAC, Trends Anal. Chem.*, **2018**, *101*, pp 24-33 (https://doi.org/10.1016/j.trac.2017.10.028).
19. Medicines & Healthcare products Regulatory Agency (UK). Consultation: *Application of Analytical Quality by Design concepts to pharmacopoeial standards for medicines*, **2019**. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/806198/AQbD_Consultation_document.pdf [Acessed March 2021].
20. Medicines & Healthcare products Regulatory Agency (UK). Consultation response: *MHRA response and Strategy for the Application of Analytical Quality by Design concepts to pharmacopoeial standards for medicines*, **2020**. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/908703/AQbD_Consultation_Public_response_Final.pdf [Acessed March 2021].
21. Stimuli to the revision process USP Validation and Verification Expert Panel. Available at: http://www.ichq3d.fr/files/supplier/fichiers/2/43(1)%20Stimuli%20to%20the%20Revision%20Process:%20Proposed%20New%20USP%20General%20Chapter:%20The%20Analytical%20Procedure%20Li.pdf [Acessed March 2021].
22. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). ICH Q14. *Analytical Procedure Development and Revision of Q2(R1) Analytical Validation*, **2018**. Available at: https://database.ich.org/sites/default/files/Q2R2-Q14_EWG_Concept_Paper.pdf [Acessed March 2021].



Márcia Cristina Breitkreitz has been working since 2014 as Professor in the Chemistry Institute of the University of Campinas (UNICAMP), coordinating the Laboratory of Pharmaceutical Research and Chemometrics (LabFarQui). She was a senior researcher in Chemometrics from 2006-2007 at the International Institute of Pharmaceutical Research (IIPF, Brazil) and a coordinator of research and development from 2007 to 2009 in the same Institute. She received her PhD in Chemistry from UNICAMP in 2013, under the supervision of Prof. Ronei Jesus Poppi. Her major research interests are related to Chemometrics applied to the development of pharmaceutical products and processes according to the strategy of Quality by Design (QbD); development of analytical methods based on liquid/supercritical fluid chromatography and vibrational spectroscopy (Raman and NIR) for pharmaceuticals. In addition to her research work and teaching (graduate and undergraduate levels), Prof. Marcia is dedicated to courses and consultancy on relevant topics to the pharmaceutical industry, especially related to Chemometrics within the context of Quality by Design.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 6-12
doi: 10.30744/brjac.2179-3425.interview.ajcapdevila

BrJAC

INTERVIEW

Anna de Juan, an internationally recognized researcher on Chemometrics, spoke to BrJAC

Anna de Juan Capdevila
Associate Professor
Faculty of Chemistry at the University of Barcelona, Spain

Professor Anna de Juan's research focuses on the theoretical development of multivariate curve resolution and its application to process analysis, hyperspectral image analysis, and general analytical and bioanalytical problems. She has published more than 140 works (h index = 40) and has given more than 200 presentations at international conferences. She received the 4th Chemometrics Elsevier Award in 2004 and the Kowalski Prize from the *Journal of Chemometrics* for the best applied paper in 2009. She has served on the Editorial Advisory Board of *Chemometrics and Intelligent Laboratory Systems* since 2002 and of *Analytica Chimica Acta* since 2006. Her teaching activity covers undergraduate and graduate topics related to chemometrics and analytical chemistry, and she has been an invited professor for short periods at the Université de Lille (France), the University of Dalhousie (Canada), the Università di Modena e Reggio Emilia (Italy), the Institute of Advanced Studies in Basic Sciences (IABS) Zanjan (Iran), the Universidad de Santa Fe (Argentina), and the Universidad Pontificia de Valparaíso (Chile).

The first question I would like to ask you is: Why science? There are plenty of opportunities and possibilities to make a life. So, what were the factors that lead you to chemistry, what fascinated you?

To be honest, when I was a kid, I was fascinated by many areas of knowledge and arts. I loved science, but also enjoyed learning languages and had a strong artistic inclination that made me do classical dance for many years, for instance. I think I could have been very happy dedicating my life to do language research or to an artistic activity. In secondary school, I discovered chemistry and, to some extent, this is the most "magic" science among all others; it connected a bit with the artistic/fantasy side that I have always had. I met a very special chemistry teacher in high school, doña María, not the best in terms of conceptual explanation, but her passion when talking about chemistry captivated me. Chemistry experiments were also very exciting, because you could "see" in a spectacular way sometimes, the phenomena you were told about. At university, I realized that it was a good choice and I could discover the many real sides of chemistry that always connect with the understanding of any event in nature and life, providing explanations at all possible scales, from the molecular theory to the macroscopic observation.

Cite: Capdevila, A. J. Anna de Juan, an internationally recognized researcher on Chemometrics, spoke to BrJAC. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 6-12. doi: http://dx.doi.org/10.30744/brjac.2179-3425.interview.ajcapdevila

How was the beginning of your career? How did you discover chemometrics and why were you interested in this field? Could you please comment about the time you were at Vrije Universiteit Brussel (Belgium), in Luc Massart's laboratory? At the same time, Prof. Ronei Jesus Poppi was there as a post-doc student. Did you get to know each other as colleagues or even worked together?

When I started my MSc, I realized I could not see myself doing routine work for the rest of my life and decided to start a research career. Professor Enric Casassas influenced significantly this choice. He was a great researcher and a multitalented person, with a strong Renaissance wisdom spirit. With him, I learned how to approach a scientific problem and the importance of conveying appropriately scientific knowledge, writing and presenting science using a precise language and clear and attractive images and explanations. He introduced me to Romà Tauler, a young professor who had been supervised by him, and suggested that I could do "something" with him. The "something" was chemometrics. So, in the early 90s, at Romà's initiative, I started reading the book *Factor Analysis in Chemistry* by Malinowski and to do some basic works with him using factor analysis and target factor analysis to study solvent effects in equilibria. At that time, in 1992, I attended a chemometrics school in Ulvik (Norway), organized by Kvalheim's team, and I attended a COMETT school in Santiago de Compostela in 1993, a great initiative supported by the European Union, where the best chemometric experts in Europe were teaching and offering a wide vision of chemometrics. There, Bernard Vandeginste, a teacher in the two schools mentioned, played a crucial role for my later stage in Massart's laboratory. I arrived at Massart's laboratory in 1995, with fundamentals of factor analysis and a good knowledge of the developments in multivariate curve resolution. That was a very exciting period in my life and served to consolidate the choice of chemometrics as my research field. Massart's lab at the time comprised around 20 people between PhD students, post-doctoral people, and invited professors (the essential role of people like Beata Walczak through many years has to be mentioned) from all over the world. Continuous group meetings were taking place and almost all possible areas within the chemometric field were addressed by someone in the group. Theoretical developments were accompanied by practical, challenging problems brought to the laboratory by the industrial consortium (ChemoAC) associated with the group. My personal contribution was on the study of a pure variable selection method, the needle algorithm, on the design of some new constraints for multivariate curve resolution and on the comparison of multiset analysis with other multi-way algorithms, together with Sarah Rutan, an invited professor at that time. As a scientific and multicultural personal experience, it was a unique year in my life. During my period there, I met Ronei Poppi, who joined the group for a research stage. Paula Fernandes de Aguiar, then a PhD student in Massart's lab, introduced him to us. Although I could not work with him directly, he was the kind of person appreciated by everyone: shy, but extremely gentle and clearly a very hard worker, trying to learn as much as he could to bring it back to Brazil and create some years later not only a group, but a continuous inspiration for the development of chemometrics in his country.

You have given a brilliant contribution in the field of multivariate curve resolution for different purposes and combined with different analytical techniques. What are your lines of research and main interests at the moment? What has changed in your interests since the beginning of your career?

Right now, my main interest is still mainly in multivariate curve resolution both from a theoretical and practical side. In terms of theoretical aspects, I am very much interested in the adaptation of the algorithm to new data structures, modifying the algorithm or the underlying mathematical models used. This implies pursuing the use of incomplete multisets trying to understand better how the different factorizations can be addressed and combined and how these structures and data analysis compare with other approaches developed with the same objective, such as combined tensor and matrix factorizations. Another side is insisting on using the value of the scores and loadings obtained from multivariate curve resolution as seeding information for other data analysis purposes, e.g., heterogeneity studies, segmentation, clinical/food sample profiling, etc. We all use score/loadings from principal component analysis in a very natural

way as compressed expressions of the information in our data, but do not think that meaningful multivariate curve resolution scores and loadings can be equally useful and provide an easier interpretation of the results obtained.

The main applied field now is the application and adaptation of multivariate curve resolution to hyperspectral image analysis. This field has new challenges, such as the active treatment of the spatial information provided by the image, and the data fusion challenges that force us to envision new strategies to cope with the multiscale spatial differences among platforms and the specificities in terms of dimensionality and the underlying model of the spectroscopies that go with them. The massive size of these data is another element, which is common to other kinds of data sets, and that needs to be addressed. The industrial process application is something that we have also tackled in recent years, where sensor fusion for process understanding and process control is a need and a practical introduction of image-based monitoring and control can provide new and more useful results.

For you, what have been the most important achievements in the analytical chemistry recently? Could you briefly comment on recent advances and challenges and how chemometrics could be helpful?

Analytical chemistry is in constant evolution. On one side, the analytical platforms are evolving in incredible ways and provide possibilities to have an integral description of samples from a qualitative and quantitative chemical/molecular point of view, e.g., the latest high resolution mass spectrometry developments, from a multiresolution spatial point of view, e.g., hyperspectral platforms, and from a temporal dimension, all with ultrafast spectroscopies and methods. The role of nanotechnology in analytical chemistry would also deserve a chapter in itself. At the other end, cheap and fast analytical devices allow in situ or in vivo monitoring of processes, samples, and living systems. The problems addressed, e.g., -omics challenges, environmental monitoring, and bioanalytical studies, have gained in complexity and require fusion of very different kinds of information. Chemometrics plays an essential role to combine and interpret reliably all the information provided. Another perhaps philosophical gain has been understanding that we can talk about analytical science, since the multidisciplinary value of analytical chemistry in many neighboring scientific fields is undoubtable nowadays.

Working at the synchrotron ALBA in 2019 with the research group: Sara Piqueras, Adrián Gómez, and Rodrigo Rocha.

Do you believe that the current graduate programs produce quality researchers in the field of analytical chemistry and chemometrics? In your university, is chemometrics taught at the undergraduate and graduate level?

"...in recent times, my impression is that there is a stronger concern in terms of scrutinizing better which concepts need to be conveyed to prepare both researchers and professionals in graduate programs, and a lot of attention is paid to develop other skills, such as the communication of scientific knowledge and development of critical thinking."

To some extent, I think that education programs have always provided quality researchers; otherwise, we would not be where we are now. But in recent times, my impression is that there is a stronger concern in terms of scrutinizing better which concepts need to be conveyed to prepare both researchers and professionals in graduate programs, and a lot of attention is paid to develop other skills, such as the communication of scientific knowledge and development of critical thinking.

In the Faculty of Chemistry at the Universitat de Barcelona, chemometrics is present in both undergraduate and graduate curricula. In the undergraduate program, a compulsory Advanced Analytical Chemistry topic includes classical statistics, design of experiments, and univariate calibration. In the last year of the Chemistry degree, an optional chemometrics topic offers the fundamentals of multivariate analysis, including theoretical and practical work about exploratory, calibration, and classification methods. Team work on case studies based on real data sets is one of the assets of the subject, where a real problem is posed and chemometric knowledge, team work, and decision-making skills are built. In the graduate program linked to the Master of Analytical Chemistry, a subject called Chemometrics and Process Control is offered. The same contents as in the undergraduate option, incorporating multivariate curve resolution and multivariate statistical process control, are offered.

In the lab: Víctor Olmos, Lorenzo Strani, Sanae Benabou, Silvia Concolino, Adrián Gómez, Alba Navarro, and Rodrigo Rocha.

Do you work in collaboration with industry (chemistry, pharmaceutical, etc.)? How do you see this relationship (both for university funding and for the spread of academic knowledge)?

We have been working more often in contact with industries than in the past. We had a very good opportunity under the frame of the European Union project ProPAT (pro-pat.eu), focused on the development of cheap sensors and related process control methodologies for industrial end-users. To be honest, I think that industries and universities still need to learn from each other. We need to understand and try to adjust to the real objectives of the industrial problems, whereas industries need to understand that perhaps a more open view of the problem can help them to find a more complete solution and that knowledge is value and appropriate funding is required. In the last decades, small technological companies have emerged that act as a bridge between the industrial end-user and the academic world. I think they often play a very interesting role because they have a language understandable by both sides and they can be useful to foster collaboration.

Thinking of chemometrics, the value of the discipline is getting recognized, but I have also realized, through attending different PAT-related meetings, that many industries still tend to see engineers as their

natural partners and chemists or chemometricians as more academic researchers. In this sense, there still exists different paradigms in these two worlds – the deterministic way to interpret processes by engineers and the soft experimental monitoring and control linked to chemometrics – which is sometimes seen as not completely reliable. Proving the usefulness and compatibility of both approaches is needed.

You are a woman in science, we know that historically this hasn't been always easy. Along your scientific career, have you faced any gender prejudice?

"I consistently see more young women in science, and there is a very positive movement all over the world to promote the presence of women in science, as a life option as natural for them as any other one."

Actually, I have never suffered such a problem in my home institution. I grew up in a research group where gender was not an issue, and this surely helped me to feel more confident once I moved to other contexts and situations. Generally speaking, I am really happy with the treatment I have received from my colleagues, although I may refer to very few occasions where I felt uncomfortable as a woman and certain situations where I really had to be more assertive than needed to make my point. But it is true that in many scientific conferences, and chemometrics conferences are not an exception, you can still see not many women giving keynotes or plenary lectures, and this is a problem that comes from many years ago. For too long, science has been seen as a masculine territory, and we still drag this historical misconception, joined to the fact of the hidden brilliant women scientists in the past who are just now starting to be rediscovered and granted the credit they have always deserved. In 2020, the Royal Society of Chemistry made a statement on inclusion and diversity in Chemical Sciences (https://www.rsc.org/news-events/articles/2020/jun/id-joint-societies-statement/). These statements mean that much work still needs to be done. However, I am confident, because I consistently see more young women in science, and there is a very positive movement all over the world to promote the presence of women in science, as a life option as natural for them as any other one.

What is MCR GUI? What is your opinion about software interfaces in chemometrics?

MCR GUI is a free, downloadable graphical interface that tries to incorporate all the solid knowledge on the multivariate curve resolution-alternating least squares method that has been developed through the years with the joint effort of many PhD students and researchers, led by Romà Tauler, and with the contributions by Joaquim Jaumot and myself (for the last version of it, please read: J. Jaumot, A. de Juan and R. Tauler. Chemom. Intell. Lab. Sys., 140 (2015) 1 and go to www.mcrals.info for the download). Much credit needs to be given to Joaquim Jaumot, who has been implementing the interface. In our case, the appearance of the interface has promoted a massive increase in the use of the algorithm. It was sharing the knowledge with non-experienced users, providing an additional method to their chemometric toolbox.

I am really in favor of chemometrics software interfaces. Some experts claim that people misuse them and they are the cause of many sorts of mistakes. In my opinion, good interfaces are the best way to encourage people to use chemometrics in their professional or research activities and an invaluable help for teaching purposes. Of course, the use of any tool requires previous preparation and it is the responsibility of the user to follow the necessary education to be able to apply the interface possibilities properly. The only thing I do not like in certain interfaces is the presence of default options. As I mention often to my students, we never work with "default data sets"; they all have their specific characteristics. Interfaces need to offer different options to work and, at the most, suggestions of use, but leave the final decision-making task to the user.

Recently, the coronavirus pandemic has shaken the world in many ways. How do you think this would affect international scientific conferences—are we moving toward the online conferences era? In 2015, you participated in the II Winter School of Chemometrics, which took place at the University of Campinas (UNICAMP), and Prof. Ronei was one of the organizers. What are your memories of that meeting?

Dinner with the speakers and organizers of the II Winter School of Chemometrics (UNICAMP, 2015).

It is clear that coronavirus has affected many aspects in our lives. It has been like a slap in our face reminding us that we are vulnerable and any abuse we do toward our planet and all the nature within will bounce back and bring this pandemic, or there may be others in the future. I do not think all the conference world will move to an online format in the future. Of course, the situation has reinforced these modalities and they are a temporal solution to the problem. I would make a distinction between huge meetings with thousands of people, where the online format or hybrid models are probably very valid options in the future, and the small conferences, very focused on particular fields, e.g., chemometrics, where the direct contact and personal collaboration are invaluable aspects that cannot disappear. This also applies to teaching: We teach online at the university until we can go back again and, if anything, we can envision some kind of hybrid model in the future, but never fully online.

For instance, I cannot imagine losing the pleasure and atmosphere of the live winter school of chemometrics. In the edition where I participated, I was invited with Doug Rutledge, José Manuel Amigo, and Héctor Goicoechea and, as teachers, we still remember the good time we spent there with Márcia Breikreitz, Ronei Poppi, and all other organizers. I have very good memories of that school because of the wonderful organization and, above all, for the huge number of motivated students from all over Brazil and other places in South America. It is one of the few times I have seen so many young people so eager to learn chemometrics. This is the reward of the good education task made by Ronei Poppi and many other teachers in Brazil with respect to the chemometrics discipline.

Participants of the II Winter School of Chemometrics (UNICAMP, Campinas, 2015).

How do you see the connection between these "new" tools and terms that we see in the media nowadays ("big data," "machine learning," "deep learning," "artificial intelligence," etc.) and chemometrics? Are they an extension of currently used chemometrics methods applied to a higher amount of data, or are they part of a separate science?

All eras need their terms, and the evolution of the massive numerical and non-numerical information nowadays has required new labels. It is also true that the concern about the proper interpretation of information exceeds the chemometrics field and people from other domains look for new names—and that is OK. Often, though, when you dig under these labels, you discover good old friends, like machine learning-labeled works that use principal component analysis, and you have to smile. But it is also true that many other old -metrics, e.g., econometrics, psychometrics, have used principal component analysis as well. So, the method does not belong to anyone. There are many tools in data sciences, and the further they go in terms of applications and domains of knowledge the better. We can always claim that we knew many data analysis tools before these new terms were coined, but we can also learn many more in the future, and this is exciting.

In your opinion, what are the skills that an analytical chemist researcher should have that would be vital to the future? What kind of advice would you give to a newcomer to analytical chemistry and chemometrics?

As I mentioned before, the analytical way of thinking will be essential to address many of the current scientific problems in a multidisciplinary way. Looking at the problem as a whole and considering carefully the nature of the question, of the samples involved, have a solid knowledge about the analytical platforms and the possibilities they can offer, have a complete toolbox of data analysis methods to extract reliable conclusions, etc., are all good skills for the future. Analytical chemists are at the interface of many disciplines, and this is a good value.

I could say a similar thing with respect to chemometrics; a chemometrician does not only have the data analysis tools, but also the chemical knowledge to interpret which ones are appropriate depending on the scientific context and which kind of new developments must be done to adapt to new needs in terms of problem complexity or new measurement characteristics. Again, it is at the interface of mathematics/statistics and applied sciences, becoming an essential piece of the game.

If you were starting over today, what would Dr. Anna de Juan say to the young Anna?

Just go for your goals, it will not always be easy, but it is worth it! Do not forget that life is made of many wonderful sides, do not stick only to science, because anything else in your life can make you (and your science) way better. Take care of yourself and of your beloved people. Any minute of love that you could not enjoy with them is a wasted piece of life.

But I also know that young Anna would make her own decisions… as it should be!!!

An outing in nature with calçotada, Aiguamúrcia, mas d'en Ferran. Rodrigo, Víctor, Adrián, Sara, Lorenzo, Sanae, Anna, Alba, Óscar, and Raimundo.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 13-15
doi: 10.30744/brjac.2179-3425.point-of-view-hcgoicoechea

BrJAC

POINT OF VIEW

Past, Present and Future of Multivariate Calibration in Analytical Chemistry

Héctor C. Goicoechea

Full Professor of Analytical Chemistry / Chemometrics
University of Litoral, Santa Fe, Argentina

Booksh and Kowalski, in their outstanding report, stressed that analytical chemistry has its own theory, which we all know as chemometrics. The authors indicated that chemometrics "can be used to specify exactly what information can be extracted from the data produced by any analytical instrument or method" [1]. In accordance with this argument, it can be stated that multivariate calibration (MVC) has historically been applied to analytical chemistry, becoming one of the most important tools of chemometrics [2].

Table I comprehensively shows the classification of the data according to their structures, as well as the corresponding type of calibration that can be built with them. As can be observed, calibration performed with spectra (first-order data), matrices (second-order), and beyond is known as "multivariate calibration". In addition, if the calibration is built with second-order data or higher, it is called "multiway" calibration [3-6].

Table I. Hierarchy of data considering the nomenclature based on the concept of order (for data) and ways (for calibration)

Data order	Zero-order	First-order	Second-order	Third-order	Fourth-order
Structure for one sample	Scalar	Vector	Matrix	Three-dimensional array	Four-dimensional array
Structure for a sample set	Vector	Matrix	Three-dimensional array	Four-dimensional array	Five-dimensional array
Calibration	- One-way - Univariate	- Two-way - Multivariate	- Three-way - Multivariate - Multiway	- Four-way - Multivariate - Multiway	- Five-way - Multivariate - Multiway

Along these lines, we should first ask when the multivariate calibration was started to be implemented. Calibration methods, such as principal component regression (PCR) or partial-least squares regression (PLS), were introduced during the 1970s [7]. From that point onwards, numerous commercial and free software packages have been introduced to the analytical world to perform MVC with focus on PLS calibration. This MVC method has become the routine application of near infrared spectroscopy (NIR)-based analysis, primarily in the food industry and the process of analytical chemistry [2].

Cite: Goicoechea, H. C. Past, Present and Future of Multivariate Calibration in Analytical Chemistry. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 13-15. doi: http://dx.doi.org/10.30744/brjac.2179-3425.point-of-view-hcgoicoechea

The very first paper concerning MVC that reports the modeling of second-order data exploiting the second-order advantage was published in 1978 [8]. The second-order advantage represents the ability to selectively quantitate an analyte that is immersed in a multi-component system, with potential interferents, by only calibrating the pure analyte. This property was first observed in the field of second-order data, but has been extended by multiway data analysis [1]. Hence, it could be said that this is a revolutionary fact in analytical chemistry.

The second question that could be asked is what is the present state of MVC? A quick document search of "multivariate calibration" in *Scopus,* within the article title, abstract and keywords field, reveals that 3,878 documents were reported from 1978 to the present day, of which 1,040 documents were published in the last five years. The numbers indicate growing interest in this topic. Interestingly, few works have reported the modeling of third-order data or higher, meaning that, nowadays, researchers are mainly focusing on second-order data applications.

The final question to formulate is what can be expected from the MVC for the near future? The analyzed literature reveals that current researches are focused on the generation, exploration and modeling of third order-data. It is possible to observe that the concern lies in the generation of data with a high number of sensors in all three modes. Therefore, it can be envisaged that the future of MVC is aiming to consolidate second-order data applications to a wide window of fields and to investigate and optimize new ways to generate more and better higher-order data. On the other hand, this will be associated with the improvement of chemometric models and will deepen the study of the figures of merit, which is necessary to validate the models [9].

As a final thought, I would like to recall the good moments shared with the late Ronei Poppi, in 2006, at UNICAMP, making our first steps in the modeling of second-order data [10]. The objective was the development of a method for the determination of pesticides and their metabolites in wine by second-order high performance liquid chromatography data. By fortune, that goal allowed me to meet a great researcher and an exceptional person.

REFERENCES

1. Booksh, K. S.; Kowalski, B. R. *Anal. Chem*., **1994**, *66*, pp 782A–791A (https://doi.org/10.1021/ac00087a001).
2. Brereton, R. G. *Analyst*, **2000**, *125*, pp 2125–2154 (https://doi.org/10.1039/B003805I).
3. Escandar, G. M.; Olivieri, A. C.; Faber, N. M.; Goicoechea, H. C.; Muñoz de la Peña, A.; Poppi, R. J. *Trends Anal. Chem.*, **2007**, *26*, pp 752–765 (https://doi.org/10.1016/j.trac.2007.04.006).
4. Escandar, G. M.; Goicoechea, H. C.; Muñoz de la Peña, A.; Olivieri, A. C. *Anal. Chim. Acta*, **2014**, *806*, pp 8-26 (https://doi.org/10.1016/j.aca.2013.11.009).
5. Montemurro, M.; Siano, G. G.; Alcaráz, M. R.; Goicoechea, H. C. *Trends Anal. Chem.*, **2017**, *93*, pp 119-133 (https://doi.org/10.1016/j.trac.2017.05.011).
6. Alcaraz, M. R.; Monago-Maraña, O.; Goicoechea, H. C.; Muñoz de la Peña, A. *Anal. Chim. Acta,* **2019**, *1083*, pp 41-57 (https://doi.org/10.1016/j.aca.2019.06.059).
7. Faber. N. M. Views on Multiway Calibration: Its Past and Future. In: Muñoz de la Peña, A.; Goicoechea, H. C.; Escandar, G. M.; Olivieri, A. C. (Eds.). *Fundamentals and Analytical Applications of Multiway Calibration*, Volume 29 of Data Handling in Science and Technology Series. Elsevier, Amsterdam, **2015**, pp 1-5.
8. Ho, C. N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.*, **1978**, *50*, pp 1108–1113 (https://doi.org/10.1021/ac50030a026).
9. Olivieri, A. C. *Chem. Rev*., **2014**, *114*, pp 5358–5378 (https://doi.org/10.1021/cr400455s).
10. Braga, J. W. B.; Bottoli, C. B. G.; Jardim, I. C. S. F.; Goicoechea, H. C.; Olivieri, A. C.; Poppi, R. J. *J. Chromatogr. A*, **2007**, *1148*, pp 200–210 (https://doi.org/10.1016/j.chroma.2007.03.018).

Héctor C. Goicoechea was born in Santa Fe, Argentina, on May 15, 1961. He received his Ph.D. (2000) from the National University of Rosario. After a postdoctoral research at North Dakota State University, USA, he joined the University of Litoral, Santa Fe (2004), where he is Full Professor of Analytical Chemistry/Chemometrics. He is a fellow (Senior Researcher) of the National Research Council of Argentina (CONICET), and has founded a research group (LADAQ). His works are focused on the development of analytical methods based on spectroscopy, separations and electrochemistry coupled to chemometrics. He has published more than 210 papers in well-known international journals, as well as two books and 8 book chapters, and has supervised thirteen Ph.D. theses.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 16-21
doi: 10.30744/brjac.2179-3425.letter-dnrutledge

BrJAC

LETTER

Multiblock Methods in Analytical Chemistry

Douglas N. Rutledge[1,2]

[1]*Université Paris-Saclay, INRAe, AgroParisTech, UMR SayFood, 75005, Paris, France*
[2]*National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia*

Chemometrics, and multivariate data analysis in particular, has become a significant component of Analytical Chemistry as can be seen from the NGrams plot in Figure 1 for the word “chemometrics”. This is because of the need to have mathematical methods capable of extracting the pertinent information from the ever-increasing amounts of data generated by modern instruments. Usually, these multivariate data analysis methods are concerned with the exploratory analysis of a single data matrix, as in PCA, or with relating one explanatory matrix to another descriptive matrix, as in regression methods such as PCR and PLS, or discriminant methods, such as FDA and PLS-DA.

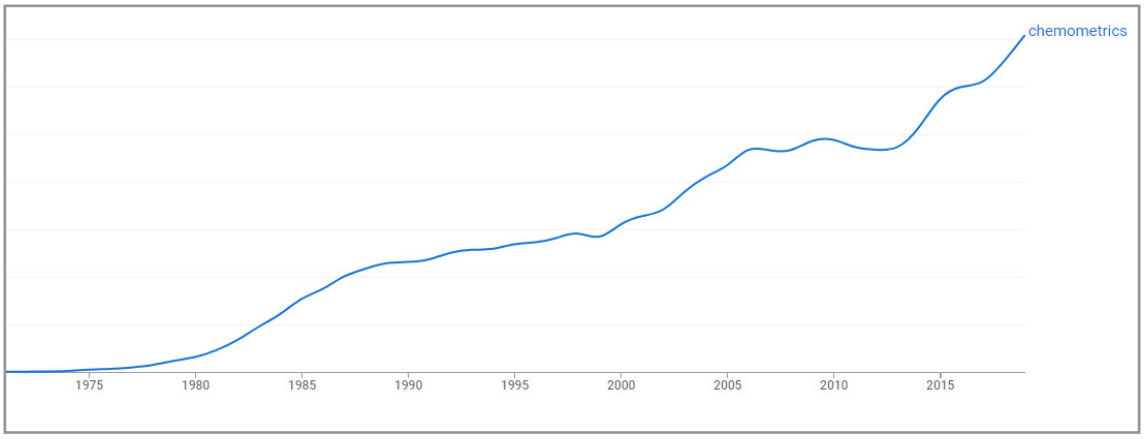


Figure 1. Evolution of the usage of the word “chemometrics” as given by NGrams (https://books.google.com/ngrams).

Recently however, there has been a trend towards analyzing many matrices simultaneously, the data being in the form of blocks of variables describing the same individuals. This trend of multiblock analysis (Figure 2) is the result of two forces: the availability of a wide range of very different instrumental techniques, and a paradigm shift towards a holistic study of complex systems. This is the case, for example, for -omics data, where combining or fusing data from different instruments can result in a better characterization of the individuals under study than that which is possible using each source of information separately.

Cite: Rutledge, D. N. Multiblock methods in Analytical Chemistry. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 16-21. doi: http://dx.doi.org/10.30744/brjac.2179-3425.letter-dnrutledge

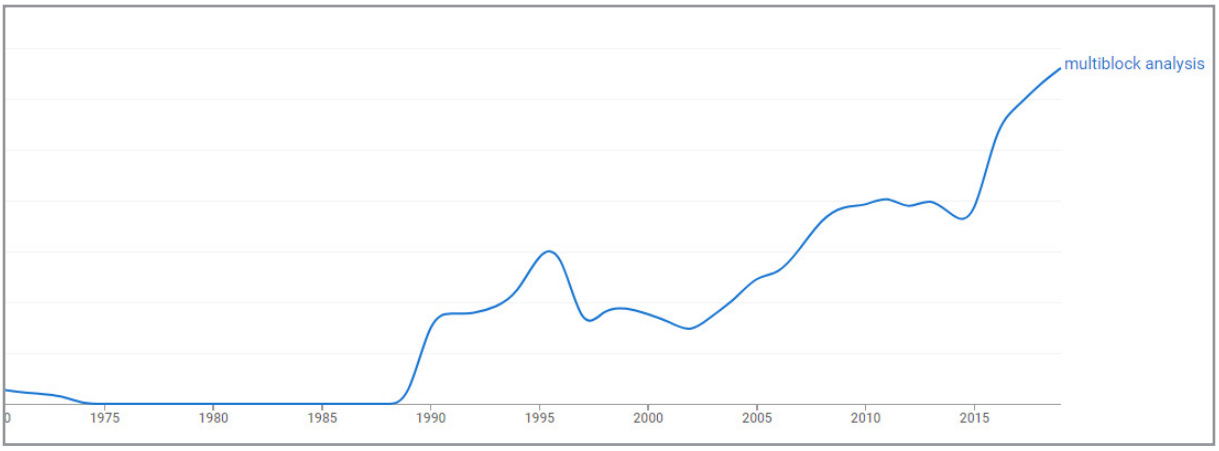


Figure 2. Evolution of the usage of the term “multiblock analysis” as given by NGrams (https://books.google.com/ngrams).

Much work has been done recently on the development of multiblock data analysis methods in order to treat this sort of data. Multiblock methods were initially used in sensometrics for the analysis of sensory data. For example, when a set of samples are evaluated by a group of judges, each using several (possibly different) descriptors, it is interesting to know more than just the relations between the samples but also which judges consider the samples as similar and which evaluate them differently. In chemometrics, the aim is to highlight the relationships between various blocks of variables. Early discussions of multiblock methods can be found in Smilde, Westerhuis and de Jong [1], Westerhuis et al. [2], and Qin et al. [3]. A very recent overview of the many different multiblock methods and their applications can be found in the book “Data fusion methodology and applications” [4].

The ComDim multiblock data analysis method

The main characteristic of multiblock methods is that they extract global components corresponding to the directions of greatest dispersion of the individuals, common to the multidimensional spaces defined by each of the data blocks. Although, as can be seen in [4], there are many interesting methods available to do multiblock, multivariate data analysis, I will concentrate in the following on a particular algorithm simply because I know it well and I think it works very well.

“Common Components and Specific Weights” (CCSWA or ComDim) is an unsupervised, multi-block (or multi-table) data analysis method developed by Qannari et al. [5-7], in the context of sensory profiling for the simultaneous analysis of several data tables describing the same individuals. It has since been widely applied in chemometrics, for example: simultaneous analysis of Mid Infrared (MIR) and Fluorescence spectra of cheeses [8], chemical and sensory characteristics of wines [9], the fatty acid composition of edible oils evaluated by combining Near Infrared (NIR) and Ultraviolet-Visible (UV-Vis) spectroscopy and Gas Chromatographic (GC-FID) data [10], interpretation of NIR and NMR spectral data, quality parameters and sensory properties of Brazilian coffees [11], monitoring surface water quality using physico-chemical, microbiological and 3D Fluorescence data [12], characterizing structural changes in a semisolid pharmaceutical formulation by NIR spectroscopy and Raman imaging [13], discrimination of commercial Yerba mates by combining HPLC, phytochemical composition, antioxidant activity, Visible and NIR spectroscopy, colorimetry and electronic nose data [14], coupling data from 3 laser-induced breakdown spectroscopy (LIBS) detectors to sort geological materials from caves [15], combining NMR and MIR spectra with stable isotope data to differentiate organically- and conventionally-produced tomatoes [16], combining 1H NMR and 13C NMR spectra with stable isotope data to differentiate organically- and conventionally-produced milk [17], combining NMR, MIR and Isotope Ratio MS data to discriminate tomato varieties [18].

ComDim and two of its extensions have also been applied in metabolomics, for example in a large-scale, multi-instrument inter-laboratory study [19], using ComDim and OPLSDA to combine positive- and negative-mode ElectroSpray Ionisation data from an UHPLC-TOF/MS system [20], using ComDim and OPLSDA to evaluate the therapeutic potential of a series of 83 flavonoid derivatives by relating five blocks of physicochemical properties to their affinity toward P-glycoprotein, and to differentiate a series of 60 human cancer cell lines by combining transcriptomic, metabolomic and proteomic data [21].

The ComDim method consists in determining a common space for all the data tables, with each matrix having a specific contribution ("salience") to the definition of each orthogonal direction of this common space. The components are iteratively extracted so as to correspond to the maximum amount of variance that is common to the largest number of tables. Each table is first normalized so that larger tables do not automatically have more influence in the calculation.

An iterative process is used to estimate the contribution ("salience") of each block to each CC. A significant difference in the saliences of the blocks for a given CC reflects their different contributions to the construction of that common dimension.

The coordinates of the observations on the ComDim directions are the 'Global Scores' and the contributions of the variables within each of the normalized tables are the 'Scaled Loadings'.

In the original algorithm, each CC is the first normed scores vector of a weighted sum of scalar matrices calculated from all the data tables as shown in Figure 3, in the simplest case of two centered and normalized data blocks, $\mathbf{X}_1$ and $\mathbf{X}_2$. A weighted sum $\mathbf{W}_G$ of the samples-based variance–covariance matrices, $\mathbf{W}_i = \mathbf{X}_i \times \mathbf{X}_i^T$, is calculated using an initial weighting, or salience, of $\lambda_i = 1$ for all tables. The vector of scores of the first normed Principal Component is extracted from $\mathbf{W}_G$ as an initial estimate of the first Common Component (CC). The salience, λ_i, of each block **W**i is then recalculated from these scores. The estimations of the Global Scores and saliences are optimized by iterative recalculation until convergence. Each original matrix $\mathbf{X}_i$ is then deflated, and the procedure is repeated for the calculation of the second CC, and so on.

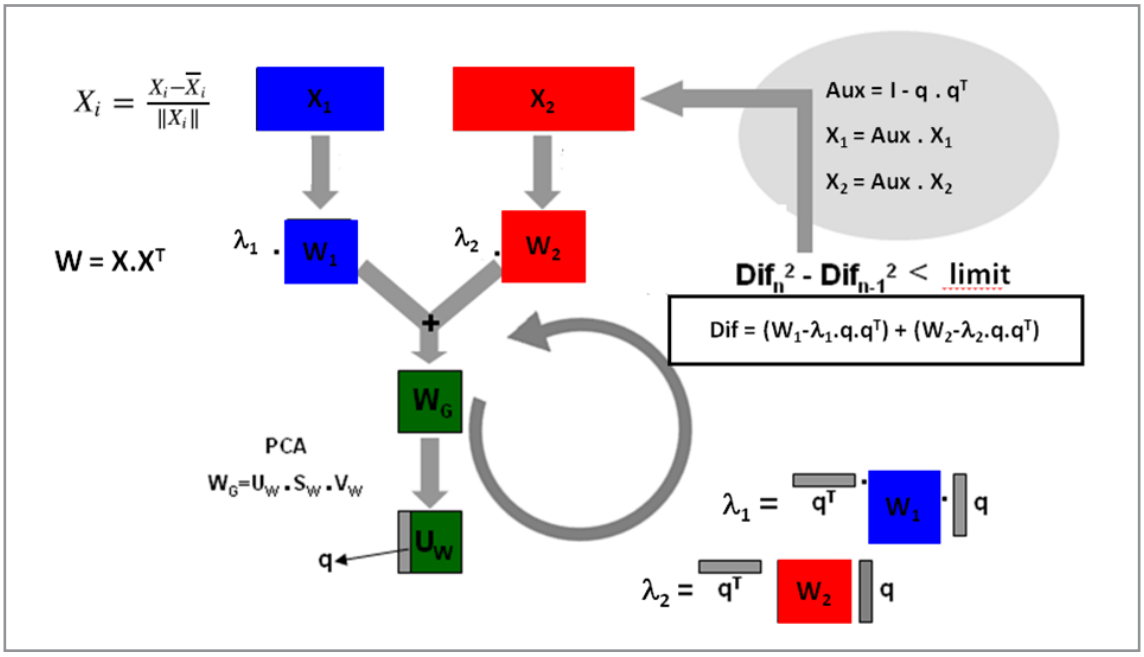


Figure 3. Schema of the original ComDim algorithm in the case of two data blocks.

Extensions of ComDim

As mentioned above, there have been a number of extensions and adaptations of the ComDim algorithm, such as replacing the PCA step by a PLS regression, or a discriminant analysis using PLSDA or OPLSDA [20,21].

Another entirely different predictive method, P-ComDim, has been developed by Qannari et al. [22] and even extended as a Path-Modelling method [23] which is useful if all the blocks are assumed to have a specific pattern of directed relations among them reflecting, for instance, a chain of influence.

In a way similar to the ANOVA–PCA method proposed by Harrington et al. [24], AComDim is an adaptation of ComDim to identify significant factors and interactions in an experimental design [25,26].

Software

A Graphical User Interface for multiblock data analysis (MB-GUI) [27] has been developed to make the implementation of multi-block data analysis easier, so that it can also be done by practitioners with no programming skills. The GUI can be downloaded from (https://github.com/puneetmishra2/Multi-block.git) and can be either installed to run in the MATLAB environment or as a standalone executable program.

The program covers a range of tasks such as multi-block data pre-processing, visualization, exploration, predictive modelling, variable selection and multi-block analysis for data fusion. The article also includes a list of other free software resources available for multi-block data analysis.

CONCLUSION

The acquisition of multi-modal data in order to have a more complete understanding of the characteristics of complex systems is becoming widespread. This has necessitated the development of new algorithms to perform multiblock data analyses. The existence of these new tools is now having the interesting effect of producing a positive feedback, leading to even more multi-modal analyses. We are only at the beginning of this revolution.

To conclude, I would like to point out that much of the progress in this field is a result of the heritage of Ronei Jesus Poppi, through the inspiration that he has given to so many young chemometricians.

REFERENCES

1. Smilde, A. K.; Westerhuis, J. A.; de Jong, S. *J. Chemom*., **2003**, *17* (6), pp 323–337, (https://doi.org/10.1002/cem.811).
2. Westerhuis, J. A.; Kourti, T.; MacGregor, J. F. *J. Chemom*., **1998**,*12*, pp 301–321 (https://doi.org/10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S).
3. Qin, S. J.; Vallee, S.; Piovoso, M. J. *J. Chemom*., **2001**, *15* (9) pp 715–742 (https://doi.org/10.1002/cem.667).
4. Cocchi, M. (Ed.) *Data Fusion Methodology and Applications*, ISBN: 978-0-444-63984-4, Elsevier, Amsterdam, **2019**.
5. Qannari, E.; Wakeling, I.; Courcoux, P.; MacFie, H. J. H. *Food Quality and Preference*, **2000**,*11*, pp 151–154 (https://doi.org/10.1016/S0950-3293(99)00069-5).
6. Hanafi, M.; Qannari, E. M. Nouvelles propriétés de l'analyse en composantes communes et poids spécifiques, *Journal de la Société Française de Statistique*, **2008**,*149*, pp 75–97.
7. Cariou, V.; Jouan-Rimbaud-Bouveresse, D.; Qannari, E. M.; Rutledge, D. N. ComDim Methods for the Analysis of Multiblock Data. In: Cocchi, M. (Ed.), *Data Fusion Methodology and Applications*, ISBN: 978-0-444-63984-4, **2019**, pp 179-204.
8. Mazerolles, G.; Devaux, M. F.; Dufour, E.; Qannari, E. M.; Courcoux, P. *Chemom. Intell. Lab. Syst.*, **2002**, *63*, pp 57-68 (https://doi.org/10.1016/S0169-7439(02)00036-9).
9. Blackman, J.; Rutledge, D. N.; Tesic, D.; Saliba, A.; Scollary, G. R. *Anal. Chim. Acta*, **2010**, *660*, pp 2-7 (https://doi.org/10.1016/j.aca.2009.10.062).

10. Rosa, L. N.; de Figueiredo, L. C.; Bonafé, E. G.; Coqueiro, A.; Visentainer, J. V.; Março, P. H.; Rutledge, D. N.; Valderrama, P. *Anal. Chim. Acta*, **2017**, *961*, pp 42-48. (https://doi.org/10.1016/j.aca.2017.01.019).
11. Baqueta, M. R.; Coqueiro, A.; Março, P. H.; Mandrone, M.; Poli, F.; Valderrama, P. ComDim for multi-block data analysis of Brazilian coffees: Identification of metabolites and application to the interpretation of the near infrared and 1H NMR spectral data, quality parameters and sensory properties, SLACA 2019, Anais do 13º Simpósio Latino Americano de Ciências de Alimentos, ISSN: 2447-2840, **2019**, *4*, 114673.
12. Daou, C.; El Hoz, M.; Kassouf, A.; Legube, B. *Water*, **2020**, *12* (6), 1673 (https://doi.org/10.3390/w12061673).
13. Mitsutake, H.; de Géa Neves, M.; Rutledge, D. N.; Poppi, R. J.; Breitkreitz, M. C. *J. Chemom*., **2020** (https://doi.org/10.1002/cem.3288).
14. Vieira, T. F.; Makimori, G. Y. F.; dos Santos Scholz, M. B.; Zielinski, A. A. F.; Bona, E. *Food Analytical Methods*, **2020**, *13*, pp 97–107 (https://doi.org/10.1007/s12161-019-01520-9).
15. Hohmann, M.; Monakhova, Y.; Erich, S.; Christoph, N.; Wachter, H.; Holzgrabe, U. *J. Agric. Food Chem.*, **2015**, *63*, pp 9666–9675 (https://doi.org/10.1021/acs.jafc.5b03853).
16. Hohmann, M.; Monakhova, Y.; Erich, S.; Christoph, N.; Wachter, H.; Holzgrabe, U. *J. Agric. Food Chem*., **2015**, *63* (43), pp 9666–9675 (https://doi.org/10.1021/acs.jafc.5b03853).
17. Erich, S.; Schill, S.; Annweiler, E.; Waiblinger, H.-U.; Kuballa, T.; Lachenmeier, D. W.; Monakhova, Y. B. *Food Chem*., **2015**, *188*, pp 1–7 (https://doi.org/10.1016/j.foodchem.2015.04.118).
18. Monakhova, Y. B.; Hohmann, M.; Christoph, N.; Wachter, H.; Rutledge, D. N. *Chemom. Intell. Lab. Syst*., **2016**, *156*, pp 1–6 (https://doi.org/10.1016/j.chemolab.2016.05.006).
19. Martin, J. C.; Maillot, M.; Mazerolles, G.; Verdu, A.; Lyan, B.; Migne, C.; Defoort, C.; Canlet, C.; Junot, C.; Guillou, C.; et al. *Metabolomics*, **2015**, *11*, pp 807-821 (https://doi.org/10.1007/s11306-014-0740-0).
20. Boccard, J.; Rutledge, D. N. *Anal. Chim. Acta*, **2013**, *769*, pp 30–39 (https://doi.org/10.1016/j.aca.2013.01.022).
21. Boccard, J.; Rutledge, D. N. *Anal. Chim. Acta*, **2014**, *813*, pp 25–34 (https://doi.org/10.1016/j.aca.2014.01.025).
22. El Ghaziri, A.; Cariou, V.; Rutledge, D. N.; Qannari, E. M. *J. Chemom*., **2016**, *30*, pp 420–429 (https://doi.org/10.1002/cem.2810).
23. Cariou, V.; Qannari, E. M.; Rutledge, D. N.; Vigneau, E. *Food Quality and Preference,* **2018**, *67*, pp 27–34 (https://doi.org/10.1016/j.foodqual.2017.02.012).
24. Harrington, P. d. B.; Vieira, N. E.; Espinoza, J.; Nien, J. K.; Romero, R.; Yergey, A. L. *Anal. Chim. Acta*, **2005**, *544* (1–2), pp 118-127 (https://doi.org/10.1016/j.aca.2005.02.042).
25. Jouan-Rimbaud-Bouveresse, D.; Pinto, R. C.; Schmidtke, L. M.; Locquet, N.; Rutledge, D. N. *Chemom. Intell. Lab. Syst.*, **2011**, *106* (2), pp 173-182 (https://doi.org/10.1016/j.chemolab.2010.05.005).
26. Guisset, S.; Martin, M.; Govaerts, B. *Chemom. Intell. Lab. Syst.*, **2019**, *184*, pp 44-63 (https://doi.org/10.1016/j.chemolab.2018.11.006).
27. Mishra, P.; Roger, J-M.; Rutledge, D. N.; Biancolillo, A.; Marini, F.; Nordon, A.; Jouan-Rimbaud-Bouveresse, D. *Chemom. Intell. Lab. Syst.*, **2020**, *205*, 104139 (https://doi.org/10.1016/j.chemolab.2020.104139).

Douglas Neil Rutledge is Emeritus Professor at AgroParisTech in Paris, where he was until 2017 Director of the Analytical Chemistry Laboratory. His research interests cover chemometrics, spectroscopy, chromatography and their application to the characterization of a wide range of materials, including agricultural, food, forensics, and pharmaceuticals. His work has resulted in the publication of over 200 refereed journal papers, 420 conference communications and 20 book chapters mainly on the development and application of novel chemometrics methods. He received the prestigious prize “Celestino da Costa-Jean Perrin” from the Portuguese Foundation for Science and Technology (FCT).

Prof. Rutledge’s collaboration with Brazil started in 2008 when two former PhD students of Prof. Ronei Poppi, Patricia Valderrama and Paulo Henrique Março, came to do a Postdoc in his laboratory. This visit led to a very fruitful long-distance collaboration with them after their return to Brazil. In 2013, thanks to a CAPES project that they set up, Prof. Rutledge was invited to spend a month in Brazil, participating in a series of conferences and giving lectures in universities and research centres in Rio de Janeiro, Curitiba, Campo Mourão, Dois Vizinhos, Campinas and Maringá. This visit gave him the opportunity to meet Prof. Ronei Poppi, as well as two others of his former PhD students, Prof. Dr. Márcia Cristina Breitkreitz and Dr. Andre Marcelo Souza, who are both now important members of the Brazilian Chemometrics community.

Prof. Poppi, Prof. Breitkreitz and Dr. Souza prepared a research and teaching project within the program of “*Cátedras Franco-Brasileiras*” in the State of São Paulo in partnership with the General Consulate of France. This project was selected in 2016, and Pr. Rutledge was invited to spend a month teaching and developing research projects at the Chemistry Institute of the University of Campinas – UNICAMP.

Since then, he has been invited back to Brazil several times and has also had the opportunity to welcome PhD students, Postdocs and Professors to visit his Laboratory in Paris.

These collaborations have been extremely fruitful. Much of it would not have been possible without the role played by Prof. Ronei Poppi as a teacher, a mentor and a major force in Chemometrics.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 22–44
doi: 10.30744/brjac.2179-3425.RV-30-2021

BrJAC

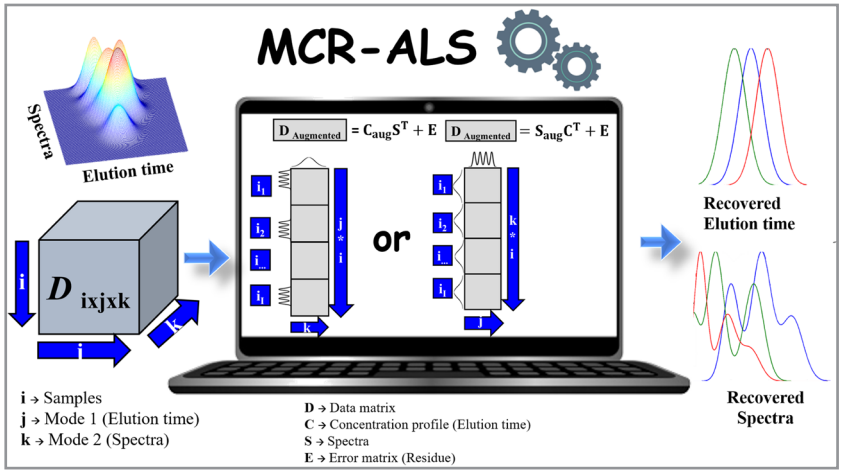
REVIEW

Multivariate Curve Resolution Alternating Least Squares Applied to Chromatographic Data: From the Basics to the Recent Advances

Igor Miranda Santana[1], Márcia C. Breitkreitz[1], Licarion Pinto[2]*

[1]*Departamento de Química Analítica, Instituto de Química, Universidade Estadual de Campinas (UNICAMP), Rua Josué de Castro, s/n, 13083-970, Campinas, São Paulo, Brazil*

[2]*Departamento de Química Fundamental, Universidade Federal de Pernambuco (UFPE), Av. Jornalista Anibal Fernandes, s/nº - Cidade Universitária, 50740-540, Recife, Pernambuco, Brazil*



This revision presents applications of multivariate curve resolution alternating least squares (MCR-ALS) applied to chromatographic data. Initially, the fundamentals and recent advances of the MCR-ALS method will be presented. Several critical issues such as data organization, advantages of the modelling, constraints, evaluation of ambiguity and the use for mathematical separation is discussed. An extensive revision of the papers on MCR-ALS applied to chromatographic data reported up to 2020 is presented. A practical example of an innovative application of cholesterol lowering drugs using supercritical fluid chromatography (SFC) is described highlighting important aspects of the method. At the end, a list of links to MCR-ALS algorithms and graphical interfaces developed in Matlab, R and Python 3 is provided.

Keywords: Multivariate curve resolution, Chromatography, statins, multiway, mathematical separation.

INTRODUCTION

Over the last decade the number of published manuscripts describing the use of multivariate curve resolution with alternating least squares (MCR-ALS) for chromatographic data analysis has increased, especially after the first time the word “chroMATHography” was used to refer to the mathematical separation that can be achieved using curve resolution methods [1–3].

Cite: Santana, I. M.; Breitkreitz, M. C.; Pinto, L. Multivariate Curve Resolution Alternating Least Squares Applied to Chromatographic Data: From the Basics to the Recent Advances. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 22–44. doi: http://dx.doi.org/10.30744/brjac.2179-3425.RV-30-2021

Submitted 01 March 2021, Resubmitted 14 June 2021, 2nd time Resubmitted 25 June 2021, Accepted 25 June 2021, Available online 21 July 2021.

Since the first paper of curve resolution analysis was published in 1971 [4] and the first work of multivariate curve resolution [5] was published in 1984, the chromatographic instrumentation evolved to make it easier and faster the acquisition of high amount of spectral data synchronized to the chromatographic elution. Multiway data, which encompasses second or higher order data can be obtained in several different ways by synchronizing two or more acquisition modes [6]: Excitation emission 3D molecular fluorescence [7–10], kinetic process followed by an ultraviolet or molecular fluorescence detector [11–14], to name a few. The automated chromatographic instrumentation simplifies the synchronization with a spectrometer detector which makes the process of acquiring higher order data easier. For example, liquid chromatography coupled to photodiode array detector (LC-PDA) [3,6,15–18], liquid chromatography coupled to fluorescence detector (LC-FLU) [17,19–23], liquid or gas chromatography coupled to mass spectrometer (LC-MS or GC-MS) [17,24–29], multidimensional liquid or gas chromatography (MDLC or MDGC) [17,26–28,30–33] and kinetic process followed by a chromatographic analysis [34] are some of the numerous way to obtain higher order data using a chromatographic system.

When the signal of one or more analytes are mathematically unmixed from interferents using a curve resolution method, pure profiles of these analytes are obtained [1], which can be used for many purposes such as calibration [3,6,15–17,20,23], pattern recognition [24,25,30] or even to evaluate the purity of the peak [34,35]. This mathematical separation is possible due to the second order advantage [6,17,20,23,36,37] and provides several benefits for the analytical method development: (1) clean-up and sample preparation protocols may be significantly reduced in time and complexity; (2) full chromatographic separation is no longer a requirement to perform an accurate quantification; (3) baseline profile can be retrieved as an extra profile and may be mathematically removed from the modelling; (4) the same elution can be used to analyze different types of samples which facilitates the development of multiproduct analysis; (5) the results of validation regarding the figures of merit can be enhanced by the increase of synchronized detection modes; (6) better accuracy can be achieved even for highly complex samples, (7) it is possible to analyze coeluted samples with identical spectra profile [3,6,15–17,19,20,24,29,30,32,34,37–39].

MCR-ALS is a flexible algorithm that presents several advantages on modelling chromatographic data. Many implemented constraints make the MCR-ALS more adaptable to different systems and can be used to efficiently retrieve pure signals of the mixture constituents. In this context, this paper provides a systematical review of the fundamental and recent publications of MCR-ALS algorithm applied to chromatographic data. Illustrative examples will be used to highlight different applications and advantages of replacing the physical-chemical separation by the mathematical separation. In addition, a new application of MCR-ALS on supercritical fluid chromatography (SFC) is reported.

THEORY OF MULTIWAY ANALYSIS

This section introduces the nomenclature, data organization, the main problems of chromatographic data for multiway modelling, and the main advantages of multiway modelling with MCR-ALS of chromatographic data.

The instrumental data can be categorized in accordance with the quantity, type and organization of data used for the analysis. In this context, the univariate approach is used when a single result is used to characterize a sample and the multivariate approach when more than one result is used for the same purpose. Univariate modelling does not require much computational effort, so it is frequently implemented in chromatographic systems for identification or quantification. The major disadvantage of this type of modelling is that it requires complete separation of the analyte from the matrix constituents since information from the interferences cannot be taken into account in this approach. This can be time-consuming and might require a large volume of solvents, especially for the analysis of complex samples, and even so an interference may still preclude the analysis. On the other hand, multivariate analysis simplifies laboratory work, however it demands higher computational effort, especially for big data sets and an experienced analyst to evaluate the quality of the model built before its use for predictions. Nevertheless, the ability to identify (first order advantage) and even to make accurate predictions in the presence of interferences

(second order advantage) make the use of multivariate modelling worthy [37,38]. MCR-ALS in the context of chromatographic analysis displays the advantages mentioned for second and higher order data [6,17,37,38,40,41]. When there is more than one variable used to build the model the data set is called multivariate, if these variables are acquired with more than one instrument source the multivariate data set can be called multiway. In this sense multivariate is a general concept that includes multiway.

Multiway data is obtained when more than one detection mode is synchronized and the number of these modes are related to the order of the data. Therefore, second order data - also called three-way data - can be built when two acquisition modes are synchronized; third order - also called four-way data - when three modes are synchronized and so on. Figure 1 relates the data to the order; it can be seen that when the sample information is a scalar, vector, matrix or a three-way tensor, the sample dataset are respectively described as zeroth, first, second and third order data [37,38,40].

Second and higher order data present the second order advantage and some authors highlight that the sensitivity is increased with the number of acquisition modes [6,17,37,38,40]. The increase in sensitivity is more evident from second to third order data than from third to higher order data, therefore the increase in sensitivity is described as the third order advantage [37,38]. Useful information to achieve the second order advantage is obtained when another mode is analyzed [34], however so far, no special advantage was related to the modelling of fourth order data or higher [37].

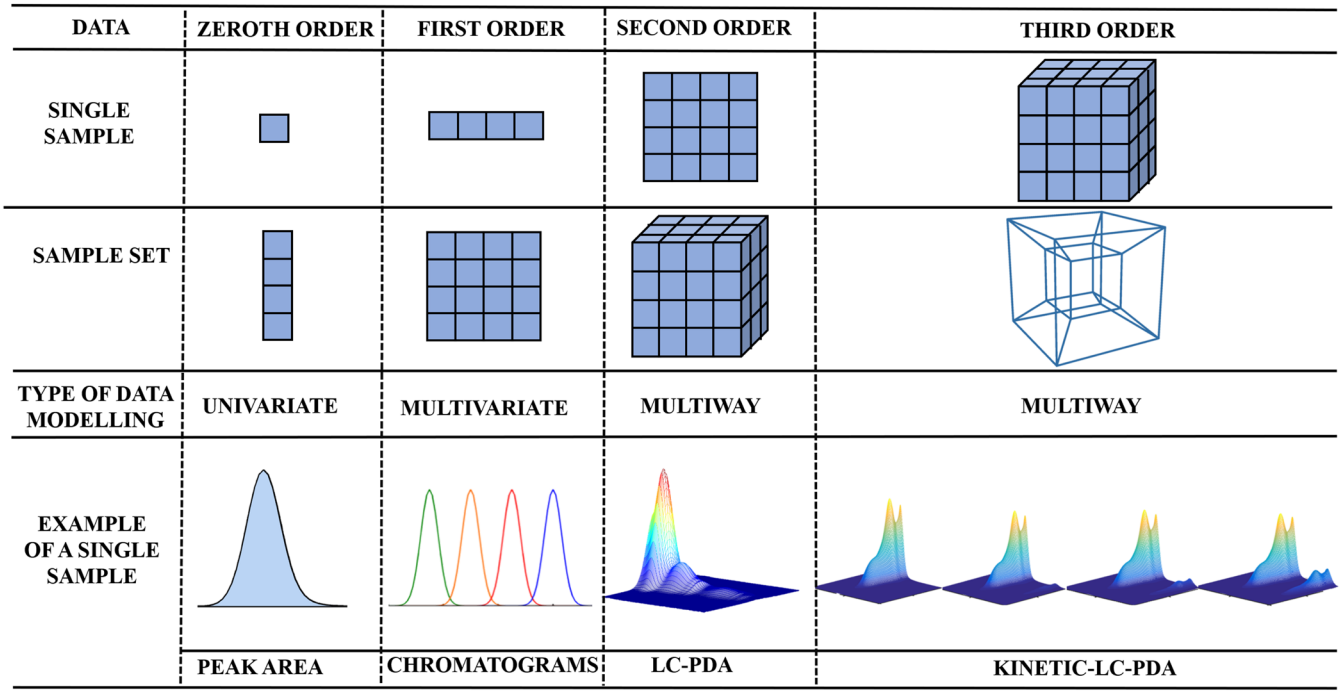


Figure 1. Examples and organization of data type related to data order. [Reprinted (adapted) from *Anal. Chim. Acta*., **2014**, *806*, pp 8-26 (https://doi.org/10.1016/j.aca.2013.11.009). Authors: Graciela M. Escandar, Héctor C. Goicoechea, Arsenio Muñoz de la Peña, Alejandro C. Olivieri. Title: Second- and higher-order data generation and calibration: A tutorial. With permission from Elsevier.]

THEORY OF MCR-ALS

Multivariate curve resolution with alternating least squares is a bilinear algorithm used for curve resolution purposes [1,5,42–45]. It fits particularly well to multiway chromatographic data because each sample is individually bilinear due to the synchronization of the detectors. However, the tensor built with more than one injection is not trilinear or quadrilinear due to retention time shifts and peak shape changes caused by differences in mass transfer process inside the column [1,6,17,40].

It is important to say that most chemical data sets are bilinear and a few present trilinearity, therefore bilinear algorithms are suitable for most of them. The great disadvantage between bilinear and trilinear models remains in the uniqueness advantage that is achieved only when trilinear models are applied. Several papers describe how to achieve a feasible result with MCR-ALS using different strategies [1,6,17,40,46].

Identifying the number of components: chemical rank analysis

Before applying MCR-ALS, the number of components present in the system under evaluation should be determined, since the results depend on this parameter, similarly to other curve resolution algorithms. The adequate number of components to start with is ideally equal to the number of chemical constituents of the sample being analyzed. However, instrumental features such as background and noise make this task harder and generally it is necessary to include another profile to explain the baseline especially when gradient is used in a chromatographic elution. There are several ways to estimate the correct chemical rank of a dataset, but principal component analysis (PCA) and single value decomposition (SVD) are the most used for this task [43,47–50]. Other approaches such as morphological score [51] and subspace comparison [52] can be more adequate when the instrumental noise level is high such as for GC-MS data.

It is important to evaluate if there is any breaking mode or rank deficiency due to completely overlapped profiles [15,34,53,54]. The easiest way is augmenting the data individually for each acquisition mode, if the number of estimated components differ it is an indicative that there is a breaking mode or a rank deficiency problem. Consider LC-PDA data where there are three chemical compounds without noise or baseline drift. If the data is augmented in the elution mode and the estimated number of components is lower than three this indicates that there are at least two spectra that present strong overlap. On the other side, if the data is augmented in spectral mode and the number of components is lower than three this indicates that there are at least two completely overlapped peaks. In this last situation it is possible to estimate more than three components, and it indicates that there are retention time shifts or peak shape changes between runs, resulting in a trilinear breaking mode. MCR-ALS is capable of dealing with trilinear breaking mode due to retention time shift [3,16,18,55] or strong overlap of profiles [15,53,54] by augmenting the matrix in the mode with this problem, however it is not capable of dealing with both simultaneously without any preprocessing step [15]. In four-way analysis the third acquisition mode can provide information to overcome both problems, as will be discussed in details in the illustrative examples.

MCR-ALS

MCR-ALS is a bilinear soft modelling algorithm in which the main goal is to resolve a mixture of signals. Since 1971 [4] until now [1] this algorithm has constantly evolved to overcome or at least minimize some limitations. The application of MCR-ALS goes from first-order data to multiway analysis [3,20,34,43,56,57], however the main application field relies on the analysis of chromatographic data [2,41]. In general, MCR-ALS can be mathematically described by the Equation 1:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{T} + \mathbf{E} \qquad \text{Equation 1}$$

where **D** is the original data matrix, **C** is the augmented matrix (generally retention times), **S** is the non-augmented matrix (generally the spectra) and **E** is the residual matrix.

Let us consider a second order data acquired by a system consisting of a Chromatographic coupled to a Detector that registers a spectrum for more than one sample. The original data set is a tensor sized *i x j x k*, where *i* represent the sample, *j* the elution time and *k* the ultraviolet spectral dimension modes. Once MCR-ALS is a bilinear algorithm, before modeling, the original sample data must be column-wise augmented. Since there are two acquisition modes, retention times and spectra, the tensor can be column-wise augmented along the retention time (*ij x k*) or along the spectra for different samples (*ik x j*). Figure

2 illustrates this augmentation where each line of the augmented matrix (**D**) consists of spectra acquired in different retention times for a set of samples (Figure 2a) or the retention times registered in each wavelength for a set of samples (Figure 2b).

Regardless of the configuration, the non-augmented mode must be reproducible between samples and must provide differences between the constituents for feasible resolution of mixture signals and the mode containing overlapped signals should be the augmented one [1,3,15,20,34,43,53,54]. Therefore, if the augmented mode is the elution time (a) the spectra of the analyte must not vary between injections and if the elution time presents retention time shift and/or peak shape changes, this will not affect the modeling. On the other side, when the augmented mode is the spectra (b) the retention time should not present retention time shifts or peak shape changes and the signals may be correctly resolved if there are slight differences between retention time even for coeluted constituents (resolution lower than 1.5) with identical spectra [15].

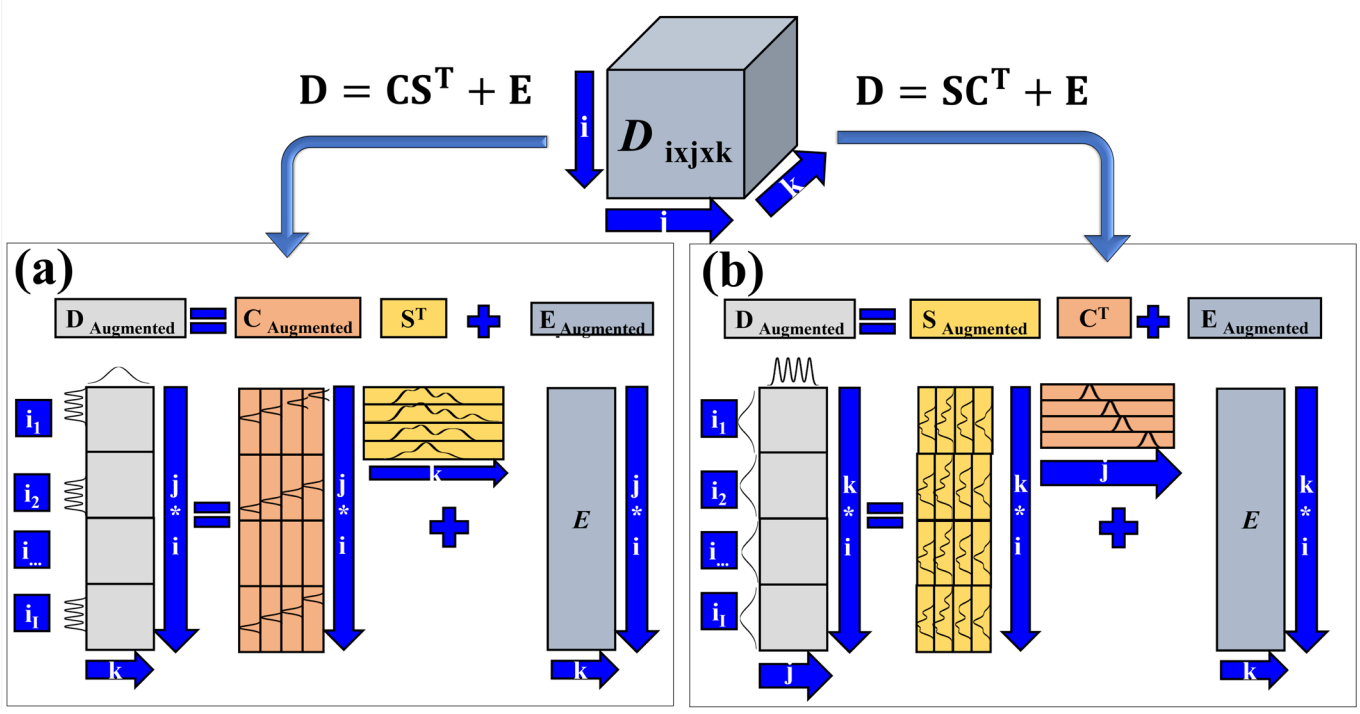


Figure 2. Illustrative scheme of a sample set registered by a chromatographic system coupled with a single detector system augmented in (a) retention time and (b) spectra from a 3D tensor. [Reprinted (adapted) from *Anal. Chim. Acta*, **2020**, *1133*, pp 77-87 (https://doi.org/10.1016/j.aca.2020.08.008). Authors: Licarion Pinto, Isabel C.S.F. Jardim, Douglas N. Rutledge, Márcia C. Breitkreitzb. Title: Multiblock modelling on the study of the kinetic degradation of rosuvastatin calcium in the presence of retention time shifts and rank deficiency. With permission from Elsevier.]

The iterative procedure of MCR-ALS algorithm is basically described in the following steps [1,43,59]:

1. Organization of the data in a matrix where the augmented mode is the one that breaks trilinearity or present extreme overlapped profiles;
2. Estimation of the number of constituents that should be resolved;
3. Initial estimate of **C** or **S**;
4. Rebuild of the **D** augmented matrix and adjustment of **C** and **S** with least squares;
5. Repeat step 4 until convergence, which is achieved by the iterative changes of **C** and **S** matrices until a predefined criterion is established (change between old and new profiles lower than 0.1% or a previous defined number of iterations).

After the convergence, the quality criteria of lack of fit (LOF) and explained variance (R^2) is analyzed to evaluate if the rebuilt **D** matrix is similar to the original.

The initial estimation is ideally as close as possible to the pure signal of each constituent that MCR-ALS is trying to retrieve. There are three common approaches to estimate the initial guess of **C** and **S**: (1) Evolving Factor Analysis, (2) Simple-to-use Interactive Self-modeling Mixture Analysis (SIMPLISMA) or (3) using a known pure spectrum of the target constituent.

The first strategy (EFA) consists of performing a PCA on the elution time adding a new spectrum at a time and it investigates the rise and decay of the components [60,61]. In this way, it is possible to identify where each constituent initiates the elution and where it ends; the theory fits well for evolutional data such as chromatography and kinetic data. Therefore, it is not recommended to apply on spectral signals. The second strategy -SIMPLISMA - searches for the purest signals and uses them as the initial estimation. This strategy can be applied in the elution time or in the spectral direction, but it is recommended to apply in the direction where there is a variable with higher purity –which is normally the spectral direction. In cases where the spectra of the constituents are identical, the best direction to search for a pure variable is the elution time [1,62,63]. The third strategy to initialize the C or S matrix is using the known pure profile, in chromatographic calibration data it is common to have the spectral profile of the analyte and the analyst may use this information to initialize the iterative procedure. However, if the spectra or concentration profiles of all constituents are known and provided as initial estimation, this algorithm turns into a classical least square (CLS) and the step five is not performed. For complex samples generally it is not possible to know all the constituents, therefore in these cases the known spectra are inputted as the initial profile and the other profiles may be better estimated during the ALS procedure. Other approaches of peak purity as Orthogonal Projection Approach (OPA) can be used, but it is not as common as these three above mentioned [62,64].

The high number of applications of MCR-ALS algorithm relies on the flexibility given by the fourth and fifth iterative steps, where the constraints are applied [1,43,44,46,59].

Constraints of MCR-ALS

The constraints are considered the “art” of MCR-ALS, once they will guide the resolution to a chemically interpretable profile and reduce the ambiguity. Constraints reduce the possible solutions of a bilinear curve resolution decomposition and consequently reduce the ambiguity of the results. There are several constraints that can be used and are implemented in ALS optimization in MCR-ALS GUI interface [43,44]. The most important are: (1) non-negativity, (2) unimodality, (3) multilinearity, (4) correlation, (5) correspondence among components, (6) closure, (7) hard modelling and (8) local rank. Multilinear, correlation, correspondence among components and hard modelling constraints can only be applied to the augmented mode. The other constraints can be applied in both modes.

Other strategies that can be used to reduce the ambiguity are dividing the modelled region to reduce the complexity of the results, which consequently increase the sensitivity [65] and using multiblock analysis.

All the constraints can be applied individually to each individual constituent and may differ among samples. Non-negativity implies that negative values are not acceptable for the mode where it is implemented; this is the most common constraint applied to chemical data. However, it is important to previously analyze the data, it is not possible to apply non-negativity to data that intrinsically present negative values such as circular dichroism [66] or negative baseline profiles [67]. In the first case, non-negativity cannot be used, but in the second case it is possible to retrieve a single profile where non-negativity is not applied to explain the negative baseline or even correct the baseline before MCR-ALS modeling.

Unimodality limits multimodal profiles. It is common to be implemented for elution profiles, where each constituent is expected to present a gaussian shape, i.e., unimodal. Therefore, it is not recommended to implement this constraint to spectra profiles, once they may present more than one maximum for the same constituent.

Depending on the multilinear constraint [68] implemented, a bilinear model is built for the augmented or super augmented matrix and then an appropriate refolding of each of the augmented profiles is performed using a Kronecker product. Once the model is primordially bilinear, ambiguity may be present in the results, even applying multilinear constraint [46]. However, this constraint implemented in MCR-ALS GUI makes it easier to visualize third order data.

In the correlation constraint, the concentration values calculated for the calibration sample subset are correlated with their known nominal concentration values during ALS optimization by linear regression [69]. The prediction is then performed for the test samples and the scores are displayed as concentration values, skipping additional calculations to predict the analyte concentration.

Correspondence among components [43,47,48] is applied by providing information to the algorithm where the analyte and/or interferents can be retrieved. There are cases, such as calibration with standards, where it is known that the interference is not present in the calibration sample set. Therefore, it is possible to inform the algorithm not to recover the interferent profile on calibration samples because it is known that they are not present there. This constraint is generally applied to calibration samples prepared with standards, once the composition of real samples is not known.

Closure [43] is a constraint that limits the total concentration of the constituents. This constraint follows the mass balance principle and avoids misleading results such as conversion higher than 100% for a reaction or the initial concentration of the reagents lower than the real concentration. This can be applied by simply limiting the sum of all constituents to a known total concentration.

Hard modelling [43,70,71] is maybe the constraint that has a higher impact to reduce the ambiguity. The resolved profiles are forced to follow a predefined equation. This modeling can be a hybrid of hard and soft modeling by using an equation where the coefficients are adjusted by the ALS optimization process in addition to other constraints. Due to the impact of reducing the number of possible results, the ambiguity is decreased and there is a greater probability of resolving highly overlapped signals.

Local rank, also called equality constraint [43,47,48,72] can be applied when the spectra or concentration profile of the target constituent is known. When standards are used for calibration, it is possible to inform the algorithm of the spectral profile of the analyte that should be resolved. This constraint is different from the initialization with the known spectra, once you may inform only the target profile and the other profiles can be inputted as “not a number” (NaN). In the case where the spectra are informed with local rank constraint the algorithm is forced to recover that profile, and the NaN data is free to retrieve the profile that best reduces the residues on ALS optimization. In both cases, the initial profile is previously defined. Remember that NaN is not allowed as an input for the initial profiles but can be used in equality constraint to receive any number along the iteration process. Therefore, the use of the known spectra is better applied with the strategy of the local rank constraint than at the initialization step.

Data set organization

In some cases, even the constraints cannot deal with some difficulties that come from the data acquisition or the lack of synchronization between the detection modes. In some situations, these difficulties can generate a non-multilinear data structure [17,23,34]. In these situations, the flexibility of MCR-ALS algorithm makes it possible to reorganize the data set in some ways that is possible to achieve correct deconvolution results. For second order data it is possible to use any array organization shown in Figure 2. For third order data it is possible to organize in other ways, such as a super augmented matrix or in a multiset data.

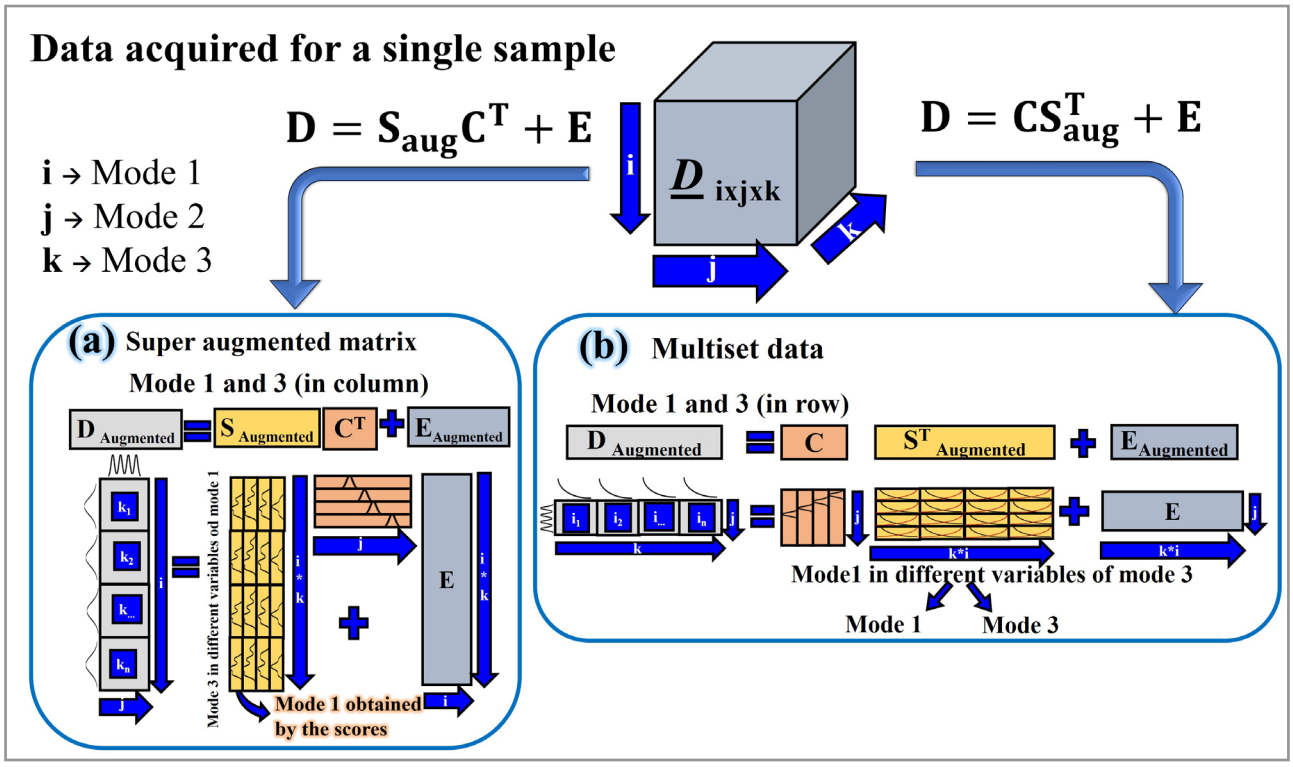


Figure 3. Data organization of a single sample registered by 3 detection modes. Mode 2 is represented as a chromatographic mode, mode 1 and 3 can be other detection modes. The data can be organized in a super augmented matrix (a) or in a Multiset data (b). [Reprinted (adapted) from *Anal. Chim. Acta*, **2020**, *1133*, pp 77-87 (https://doi.org/10.1016/j.aca.2020.08.008). Authors: Licarion Pinto, Isabel C.S.F. Jardim, Douglas N. Rutledge, Márcia C. Breitkreitzb. Title: Multiblock modelling on the study of the kinetic degradation of rosuvastatin calcium in the presence of retention time shifts and rank deficiency. With permission from Elsevier.]

As discussed before, the MCR-ALS models are based on bilinear data, therefore before modelling a third order data, it must be augmented or unfolded. Figure 3 shows an example where a single sample can be augmented in a super augmented matrix (Figure 3a) and unfolded in a multiset data structure (Figure 3b). A super augmented matrix solves problems in the augmented mode such as dependence between the acquisition modes [17,23] when the non-augmented mode presents differences between the analytes. The multiset data structure is related to a modelling of more than one unfolded detection (non-augmented) mode that must present differences between the analytes in one or both modes. Once in the multiset data structure more than one acquisition mode is used, it is easier for the algorithm to find differences between the analyte and interferences and the results of the deconvolution would be less affected by ambiguity, even if rank deficiency is present in the data set [34].

Ambiguities of MCR-ALS

Ambiguities are the biggest fragility of MCR-ALS and they can be reduced by applying constraints. The higher the number of constraints, the higher is the reduction of ambiguity. However, it is not possible to affirm that the used constraint will completely eliminate the ambiguity of the bilinear modeling [44,46,73–82].

In MCR-ALS, ambiguity occurs when there is more than one feasible result for **C** and **S** (Equation 1). Basically, there are three different ambiguities in MCR-ALS: (1) Position Ambiguity, (2) Intensity Ambiguity and (3) Rotational ambiguity. The first one is related to the position of the resolved signal changing between the matrix columns. The second one is related to the differences in intensity between spectra and elution time profiles. Both are easily solved by keeping constant the position of the analyte over the iterations and normalizing the non-augmented profile, respectively. The third one is related to the differences in elution

and/or recovered spectra profiles and cannot be completely solved, but it can be minimized using the constraints [43,44,46]. Figure 4 illustrates these three ambiguities calculated for a simulated LC-PDA data for a sample with a mixture of three constituents.

As can be seen in Figure 4, position and intensity ambiguities do not modify the original shape of the signals, but on the other hand, rotational ambiguity does. By visual comparison of the real and resolved signals, it can be seen that both negative and non-unimodal peaks are possible solutions when no constraint is applied. Besides, a small signal between variables 100 to 150 on resolved spectra is present for the second and third eluted analytes whereas the real profile does not present this signal. Both problems can be solved by applying appropriate constraints.

MCR-BANDS [44,74] solution to calculate the extension of the ambiguities consists of multiplying **C** and **S** (Equation 1) by an invertible matrix. This can be mathematically expressed by Equation 2 [44], where **T** is any invertible matrix, which is:

$$\widehat{\mathbf{D}} = \mathbf{C}_{\mathbf{old}}\mathbf{S}_{\mathbf{old}}^{\mathbf{T}} = (\mathbf{C}_{\mathbf{old}}\mathbf{T}^{-1})(\mathbf{T}\mathbf{S}_{\mathbf{old}}^{\mathbf{T}}) = \mathbf{C}_{\mathbf{new}}\mathbf{S}_{\mathbf{new}}^{\mathbf{T}} \qquad \text{Equation 2}$$

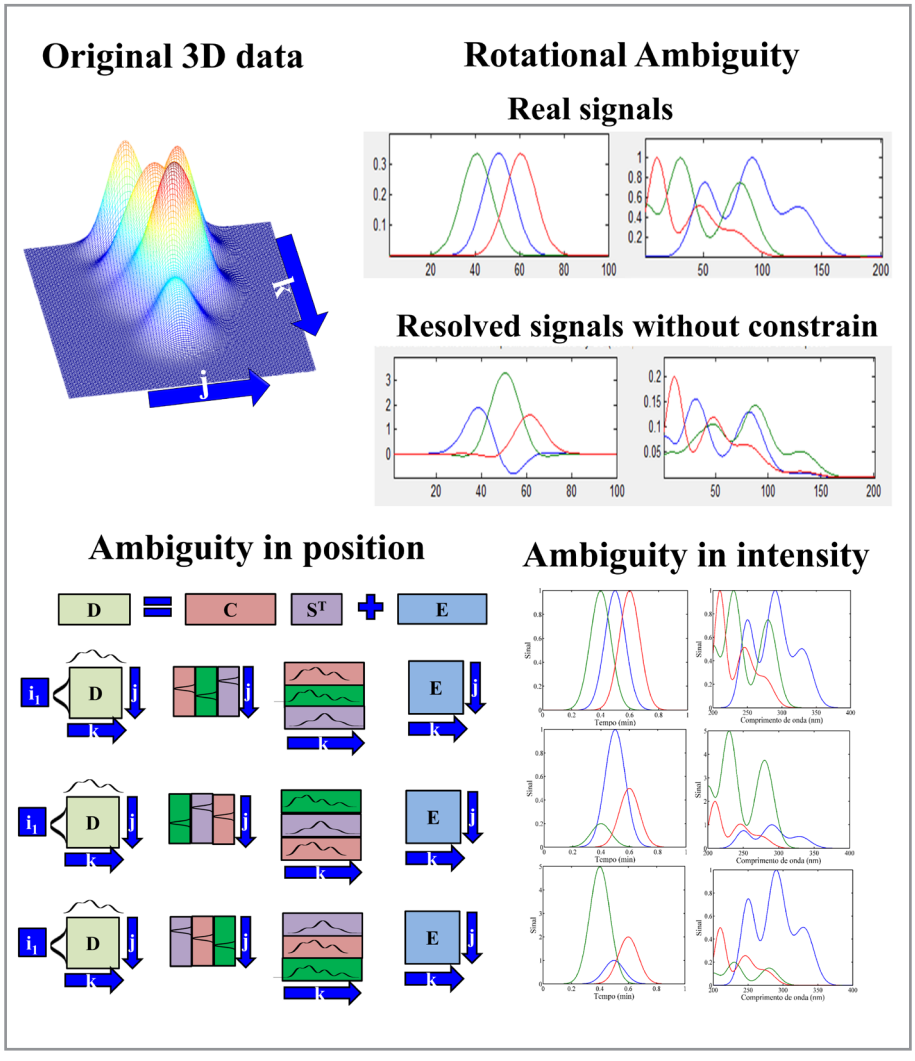


Figure 4. LC-PDA simulated data for a mixture of 3 constituents and examples of position, intensity, and rotational ambiguities. [Reprinted (adapted) from *Anal. Chim. Acta*., **2014**, *806*, pp 8-26 (https://doi.org/10.1016/j.aca.2013.11.009). Authors: Graciela M. Escandar, Héctor C. Goicoechea, Arsenio Muñoz de la Peña, Alejandro C. Olivieri. Title: Second- and higher-order data generation and calibration: A tutorial. With permission from Elsevier.]

The number of possible T matrices is used to calculate the extension of the rotational ambiguity and solutions can be restricted by using constraints. This procedure, implemented in MCR-BANDS, consists in calculating the relative contribution of each constituent to the whole mixture based on the Frobenius norm (f) [44]. The procedure is based on finding T matrices that give maximum (fmax) and minimum (fmin) values of Frobenius norm (relative contribution) for each constituent of the mixture. If the difference between fmax and fmin is near zero, it is said that the rotational ambiguity is low and it is higher as this difference increases.

Recently, a strategy implemented apart from MCR-ALS interfaces can be used to evaluate the extension of the rotational ambiguity using area of feasible solutions (AFS) [73,81–83]. This strategy consists of calculating the feasible solutions for each constituent profile and expressing it as an area. The great disadvantage of this strategy is the limitation of the number of constituents that is possible to simultaneously evaluate, since only 4 constituents can be depicted simultaneously in a three-dimensional space [73,81–83]. A manuscript that compares the evaluation of MCR-BANDS strategy to FACPACK's was published recently [74].

In this paper, the authors try to project the extreme results (fmin and fmax) in the AFS space and it is reported that the values are in the edge of the AFS region. Once with MCR-BANDS strategy it is possible to evaluate the rotational ambiguity for any number of constituents, it is said by the authors that MCR-BANDS provides a suitable and simpler solution to evaluate the extension of rotational ambiguity [74].

Figures of merit

It is necessary to use some metrics to evaluate the quality of a calibration method and to compare it with others. Analytical figures of merits are important metrics used for this purpose. Olivieri used IUPAC's definition of each calibration metric to propose an extension of the well-established univariate figures of merit to first order multivariate and multiway calibration methods [38]. This approach to calculate the multiway figure of merits can be used not only to MCR-ALS but also to other multiway algorithms.

Sensitivity is a crucial parameter to calculate uncertainties and others metrics, and allows the comparison between others methods. Univariate sensitivity is calculated as the slope of the variation of the instrument signal for the change of a concentration unit of the analyte. It is possible to calculate when there is no interference present in the analysis. For first order data the analyte is modelled jointly with other overlapped signals and it is possible to estimate the fraction of the total signal that is related to the analyte, called Net Analyte Signal (NAS). For multiway analysis, where second order advantages are aimed, each sample can present different interference so the contribution of the analyte signal to the sensitivity is calculated by uncertainty propagation [6,38]. The equation for the calculation of all figures of merits were detailed by Olivieri at reference [38]. In this reference [38] it was compared multivariate modeling from first to fourth order and it is shown that the sensitivity increases with the order.

From sensitivity it is possible to calculate the limit of detection (LD) and quantification (LQ) using the analytical sensitivity (ɣ, sensitivity divided by error) and the type 1 and type 2 error for 95% confidence levels. The LD and LQ are respectively calculated as 3.3 ɣ-1 and 10 ɣ-1 in terms of concentration level [38]. Some software's implement the figure of merit calculations using graphical interfaces, as will be discussed further on [47,48].

COMMON PREPROCESSING AND CHALLENGES ON MODELING CHROMATOGRAPHIC DATA

Prior to the use of any multiway algorithm, one needs to know important characteristics of the data to perform adequate modeling, most of them are related to the multi-linearity of the data. In a multiway chromatography data, multi-linearity is the ability to mathematically represent the original multiway data as a linear function of profiles, related to each mode, and the scores that represent the concentration of the target constituent. In this sense, whenever it is possible to represent second order data as a three linear function, this data is called trilinear, while third order data is called quadrilinear, and so on [23,38].

In the chromatographic field, this implies that both the elution and spectral profiles present the same shape and maximum for a target constituent. This is a reasonable statement for spectra profiles, but not for the elution times. Due to the mass transfer particularities described by the Van Deemter equation, the retention time for an analyte is barely reproducible even for well-adjusted chromatography instruments, especially for samples in which the ionic force varies, runs are longer and large columns are used. Therefore, each sample of second order chromatographic data is bilinear but the set of several samples is not trilinear. The ability of MCR-ALS to deal with trilinearity break fits particularly well the chromatographic data treatment needs, indeed many publications on second and higher order data use this algorithm to deal with retention time shifts and peak shape changes [2,3,6,16,18,19,21,31,33,35]. Even so, there are cases where the alignment before MCR-ALS modeling is required because of rank deficiency [15,34] or to use trilinear-based algorithms [55].

Finally, baseline correction can be used, however is less common than alignment in chromatographic data treatment since, if reproducible, baseline correction can be recovered by curve resolution algorithms as a profile. Derivatives and other preprocessing that changes the profile are preferably avoided because they bring difficulties in analyzing the pure analyte profiles.

LITERATURE EXAMPLES

MCR-ALS finds fields of application wherever a curve resolution is needed, such as overlapping peaks (see Table I). A search for recent works combining the words “chromatography” and “multivariate curve resolution-alternating least squares” suggests that MCR-ALS is being successfully applied to a myriad of chromatographic problems. Table I shows some selected recent researches with different goals. In general, MCR-ALS was used in chromatographic elution mode (as evidenced by “time-mode augmented” approach) due to the retention time shifts and peak shape changes, resulting in the break of trilinearity, as discussed before. Detection by DAD is preferred because of its simplicity and quantitative ability, but MS is more potent for selective qualitative/fingerprinting studies such as in metabolomics. In addition, fluorescence detection is also employed because of its superior sensitivity and selectivity over DAD, when applicable. One study used surface enhanced Raman spectroscopy (SERS) as detection mode with thin-layer chromatography (TLC), whose overlapping spots where resolved my MCR-ALS. Capillary electrophoresis (CE) with both DAD and MS was also exploited in combination with MCR-ALS. Still, most data involve either GC or HPLC, but recently supercritical fluid chromatography (SFC) was used coupled to DAD for carotenoid analysis in the presence of interferents.

In subsequent paragraphs, the use of MCR-ALS will be exemplified based on five selected papers to describe each data strategy for MCR-ALS application, i.e., time-mode augmented and spectral-mode augmented (non-trilinear data), followed by the modeling of third-order data with non-quadrilinearity problem. Finally, a final section will be dedicated to report a SFC-DAD application from our group.

Table I. Compilation of papers that used MCR-ALS for chromatographic analysis

Chromatographic technique	Target compounds	Objective	MCR-ALS approach	Other chemometrics techniques used	Ref.
LC-DAD	Pesticides in vegetable samples	Quantitative analysis of pesticides in vegetables without physical removal of interferences	Time-mode augmented	—	[3] [16]
	Epoxidized fraction in olive oil	Cultivar-based classification of olive oils from profile decomposition of the epoxidized fraction	Spectral-mode augmented	NPLS-DA, SIMCA, RF	[84]
	Anti-tumor drugs in biological samples	Simultaneous analysis of drugs in biological samples even when severe time shifts and background interferences occur	Time-mode augmented	ATLD, ATLD-MCR	[85]
	Polyphenols in tea samples	Dataset deconvoluted for identifying components that enables modeling for plucking seasons of green tea	Time-mode augmented	SVM, PLS-DA	[86]
	Degradation profile of lowering-cholesterol drug	Handling the both time elution shifts and similar spectra with third-order advantage	Kinect time and spectral-modes unfolded and retention time augmented		[34]
	Preservatives in facial mask	Fast elution and simple pretreatment for quantitative analysis of preservatives using multivariate approaches	Time-mode augmented	ATLD	[87]
LC-LFD	PAHs in tea samples	Strategy to handle non-quadrilinear data for analysis of PAHs in tea leaves	Dependent time- and excitation-modes super augmented		[20]
	Organic pollutants in water samples	Photoinduced fluorescence to detect organic pollutants in water samples and to quantify with third-order data treatment	Time-mode augmented		[88]
	Quinolones in animal tissues	Comparison of two multi-way strategies to simultaneously quantificate antibiotics in animal tissues with both second- and third-order data	Time-mode augmented	U-PLS/RBL, U-PLS/RTL	[21]
LC-MS	Metabolites of zebrafish exposed to pesticide	To assess metabolite changes due to exposure to chlorpyrifos pesticide	Time-mode augmented	ASCA	[89]
	Untargeted lipids in rice samples	To assess lipidomic profile of rice under heat and hydric stresses	Time-mode augmented	PCA, ASCA, PLS-DA	[90]
	Polyphenols in Chinese propolis	Quantitative analysis of polyphenols in the presence of co-elution and interferences in complex matrices	Time-mode augmented	ATLD	[91]
LCxLC-DAD	Furanocoumarins in apiaceous vegetables	To improve quantification of overlapped peaks by MCR-ALS combining LCxLC resolution to the detection power of ^{1}D LC	Time-mode augmented		[92]

Table I. Compilation of papers that used MCR-ALS for chromatographic analysis (Continuation)

Chromatographic technique	Target compounds	Objective	MCR-ALS approach	Other chemometrics techniques used	Ref.
LCxLC-MS	Untargeted metabolites in rice samples	To assess metabolite changes of rice as a function of watering and harvesting time factors	Time-mode augmented	ASCA, PLS-DA	[30]
GC-MS	Untargeted metabolites in *Daphnia magna* (crustacean)	To identify metabolites through retention index and MS database and to evaluate the influence of environmental factors	Time-mode augmented	PCA, ASCA, PLS-DA	[93]
	Terpane, hopane and sterane (petroleum hydrocarbon)	To investigate the sources of petroleum pollution in a port region		PCA	[94]
	Volatile compounds in illicit drugs	Untargeted analysis of impurities present in illicit methamphetamine drugs	Time-mode augmented		[95]
	Complex perfume and essential oil blends	Investigation of characteristics of ylang-ylang oils used in perfumes, identification of oils used in blends and quantitative analysis by resolving total chromatogram average mass spectra	Time-mode augmented	PCA, RF	[96]
GCxGC-MS	Untargeted metabolites in lettuce samples	Investigation of the effects caused in lettuce morphology in reason of exposure to water contaminants through metabolite analysis	Time-mode augmented	PLS	[97]
CE-DAD	Quinolones in porcine blood	Quantitative analysis of fluoroquinolones when peaks are overlapped	Time-mode augmented		[98]
CE-MS + LC-MS	Untargeted metabolites	Fused data of two chromatographic techniques for metabolomic analysis of two different conditions for yeast growth	Time-mode augmented		[25]
SFC-DAD	Carotenoids and coenzime Q10 in palm oil	Quantitative analysis of bioactive compounds in palm oil in the presence of interferences via green CO2-based chromatography	Time-mode augmented		[99]
TLC-SERS	Chemical components of beer samples	Fingerprinting analysis for classification of Pilsner beers according to their origins	Time-mode augmented	PCA, ICA	[100]

MCR-ALS with augmented elution mode

Interferents may hinder a univariate quantification even if the extraction step is properly done. Due to non-selectivity of the extraction method, a cleanup step might be required, as for example in the QuEChERS method for pesticide extraction. This procedure could result in lower recovery of the analytes and therefore lower detection. To overcome this issue, Sousa and co-workers [3] proposed a replacement of cleanup step by a chemometric approach with MCR-ALS using HPLC-DAD for separation and detection. The quantification of pesticides was carried out in vegetable samples -tomato, carrot, beet and lettuce. Without a cleanup step, interferents emerged along with analytes.

The data matrix was column-wise augmented (elution mode) for calibration and prediction sample set so that MCR-ALS algorithm could retrieve the pure profiles of each pesticide from a new bilinear matrix even in the presence of co-eluting interference. With this configuration, peak alignment was unnecessary. In order to reduce the rotational ambiguity, the data matrix was divided in four regions containing respectively one, two, one and three analytes based on their elution windows. Also, taking each region individually required less processing efforts and enhanced the selectivity and sensitivity metrics. The spectral profiles retrieved by MCR-ALS for the analytes provided similarities over 0.99 with standards and the area under elution curve retrieved by MCR-ALS provided scores proportional to spiked concentrations, which resulted in reasonable recoveries.

Resolving coeluted peaks where the spectra are different is a classical case where MCR-ALS is used. However, other cases where spectra are identical can be also resolved by MCR-ALS, as discussed below.

MCR-ALS with augmented spectral mode

Pinto and co-workers [15] reported the separation of five biogenic amines found in fish samples with HPLC-DAD. The amines were derivatized to yield a chromophore for detection mode with dansyl chloride, resulting in very similar spectra. The univariate approach for quantification would require a time –consuming chromatographic run to achieve baseline separation of all compounds. To perform mathematical separation by MCR-ALS the authors [15] performed an isocratic elution with high amount of organic modifier providing a faster run, shorter re-equilibration time and subsequent high-throughput analysis, and better detectability due to narrower peaks. As expected, strong coelution between analytes and interference was observed in a less-than-4 minutes run. Due to spectral similarity, the authors reported the use of MCR-ALS by row-wise augmentation of the data matrix. In order to accomplish that, the elution mode needed to be previously aligned with the icoshift algorithm. After calibration and validation with a test sample set, MCR-ALS method was applied in spiked fish samples, obtaining recoveries ranging from 88 to 99%. Moreover, this method was applied to the degradation analysis of fish samples over twelve days and one of the analytes (histamine) was found above the recommended limit.

MCR-ALS to solve non-quadrilinearity in third order data

As mentioned before, MCR-ALS can be useful to extract information from higher order data. Carabajal *et al.* [20] applied the MCR-ALS to a LC-FLU system with a non-quadrilinearity type 4. They were concerned on accumulated polycyclic aromatic hydrocarbons (PAHs) in tea leaf samples. In an ideal situation, the data matrix obtained from LC-FLU system is a third order/four-way data with samples, elution time and both excitation and emission wavelengths. However, this type of non-quadrilinearity rises from the fact that elution time and excitation modes depend on each other due to the way the data was generated. The authors built a third order LC-FLU data by injecting the same sample changing the excitation wavelength, this system led to a non-quadrilinear type 4 data [20].

In order to solve this problem, the authors proceeded by concatenating both dependent data. As result, the new matrix data contained a three-way array with elution and excitation modes in a so-called super-augmented matrix, while emission mode remained non-augmented. During the constraint input implementation, augmented matrix with time elution and excitation modes were separated to apply unimodality in the former, then reunited. This new matrix is assumed to be bilinear in such a way that

MCR-ALS can be now applied to retrieve the pure profiles successfully. The developed MCR-ALS method applied to the four PAHs could be used to monitor these contaminants in tea samples. Also, the strategy adopted (LC-FLU-MCR-ALS) was quite faster than the reference method by GC-MS, providing greater analytical frequency.

The third order modeling with non quadrilinear type 4 data was possible to be solved without alignment in this case because the signals were different. When it comes to the same system with rank deficiency, the modeling is even more challenging, but can be solved by a previous alignment or using a multiblock strategy [15,34]. To highlight this case, another third order data involves a kinetic profile along elution time and spectra profiles [34].

In a pharmaceutical environment, a description of potential degradation products of the commercialized drug during its shelf life is required by regulatory agencies. However, a usual obstacle is finding the optimal chromatographic conditions for the separation of the active ingredient and its degradation products due to their chemical similarity. On the other hand, because of chemical similarity, the spectral profiles are also very similar, which gives rise to rank deficiency when using MCR-ALS. Pinto and co-workers [34], proposed the use of a third order setup to solve the problem of coelution and spectral similarity in a sample containing atorvastatin and its major degradation products. LC-DAD data was acquired in different degradation times, which provided information for the MCR-ALS modelling. The data was also non quadrilinear type 4, for the same reasons of the previous example.

Using a rapid elution (less than 3 minutes) MCR-ALS was capable of retrieving both chromatographic and spectral profiles, even in the presence of rank deficiency. Besides, since the kinetic profile was present in the data matrix, another profile was retrieved describing the evolution of the concentration of species over the time. In this work the authors used a multiblock strategy, since the kinetic profile gives additional selective information about the system. The major advantage of this strategy is that peak alignment was not necessary to solve rank deficiency, retention time shifts and non quadrilinear type 4 problem simultaneously.

MCR-ALS applied to supercritical fluid chromatography data

Supercritical fluid chromatography (SFC) is a separation technique that uses dioxide carbon (CO_2)-based mobile phases above CO_2 critical pressure. Given the non-polar character of the CO_2, an organic modifier such as methanol can be employed to promote elution of polar analytes. In recent years, SFC has regained attention in many fields such as pharmaceutical, natural products and bioanalytical, to mention a few. In addition to SFC usually operates as normal-phase liquid chromatography, the use of an additional organic solvent provides complementary separation compared to the reversed-phase liquid chromatography (RPLC). Recently, solving co-eluting peaks with MCR-ALS has been applied on supercritical fluid chromatography data. Guedes *et al.* [99] used SFC with CO_2-ethanol mobile phase on a C18-stationary phase to separate bioactive compounds from palm oil, namely beta-carotene, lycopene, coenzyme Q10 and lutein. They studied the influence of temperature, pressure, and concentration of ethanol on retention of these compounds by a design of experiments approach. The complex matrix of palm oil required MCR-ALS analysis for the accurate quantification of each compound in the presence of the palm oil interferences.

In another study using SFC, a method for simultaneous analysis of lowering-cholesterol drugs (all commercially available statins and ezetimibe) was developed. After column screening, the stationary phase 1-aminoanthracene (1-AA) was the only column able to slightly separate simvastatin and lovastatin [101]. Separation was attempted by a design of experiments approach to assess the influence of pressure, temperature, and modifier content. However, experimental conditions were unable to obtain acceptable resolutions for this couple of peaks.

In this section, we dealt with these co-eluting peaks on SFC to mathematically separate them by applying MCR-ALS. An obvious advantage was dismissing laborious method development to have a baseline separation of all analytes. The chromatographic conditions were: 10.3 MPa of backpressure, 40 °C in the

column oven, flow rate of 1.5 mL min^{-1}, mobile phase consisting of CO_2 (channel A) and methanol:water 95:5 v/v (channel B), gradient ranging from 0 to 30% of B in 5 minutes. Chromatographic data were acquired in a rate of 20 Hz and 1.2 nm of spectral resolution. In this system, there is a region where two overlapped analytes present identical spectra (1 - simvastatin and 2 - lovastatin) and another region where three overlapped analytes present different spectra (3 - ezetimibe, 4 - rosuvastatin and 5 - fluvastatin) (Figure 5A). These overlapped peaks were solved by spectral and retention time augmented matrices, respectively. For the first region (the one with a rank deficiency) a peak alignment with icoshift was first carried out and a spectral augmented (Saug) matrix was used to build the MCR-ALS model (Figure 5B). The second region with different spectral profiles had the elution mode augmented, Caug (Figure 5C).

MCR-ALS algorithm retrieved both spectral and elution profiles and allowed the pseudo-univariate calibration by using area under spectra and chromatographic peak profiles (scores), respectively. This procedure resulted in a linear relationship between MCR-ALS scores and standard concentration (ranging from 10 to 200 mg L^{-1}). Validation samples resulted in relative errors of prediction (REP) for actual concentration ranging from 4,39% (ezetimibe) to 12,73% (rosuvastatin), meeting criteria for acceptable values according to Horwitz equation, except for rosuvastatin [102].

This work was exploratory in the use of SFC-DAD using a mixture of drug analytes, but it shows that SFC is posed as a potential separation technique along LC and GC for complex matrices. Indeed, SFC faced a rise in its applications in the last decade and the combination with chemometric tools such as MCR-ALS might be of choice for whom desires an orthogonal technique for RPLC, a replacement of toxic solvents used in NPLC and to practice a greener and faster separation due to the CO_2-based mobile phases properties. As a conclusion, SFC-MCR-ALS seems to be a good option to solve co-elution problems and interferences when shorter time analysis and lower solvent waste are desired.

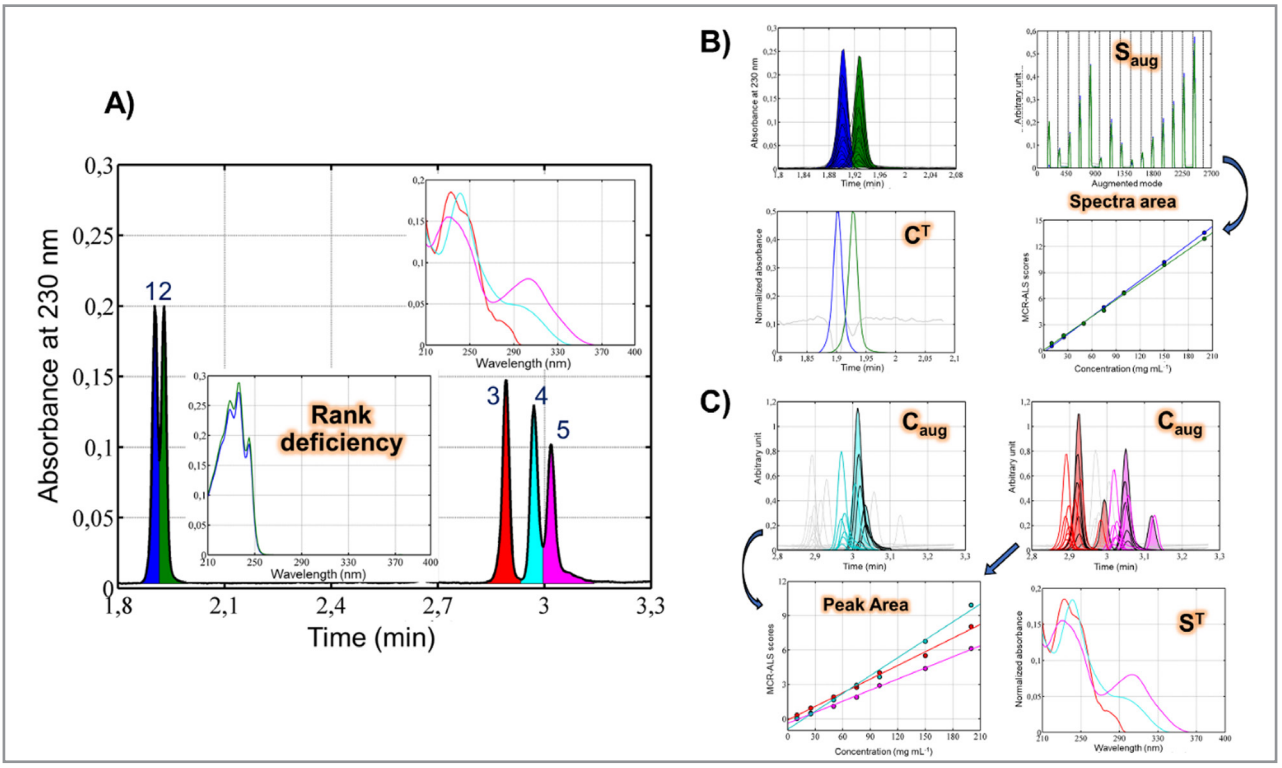


Figure 5. Chromatogram with peaks referring to simvastatin (1), lovastatin (2), ezetimibe (3), rosuvastatin (4) and Fluvastatin (5). Data management was divided in two regions concerning rank deficiency (peaks 1+2) and different spectra (3+4+5) (A). The data of the first region was augmented in the spectral mode (B) and the data from the second region was augmented in the chromatographic mode (C).

SOFTWARE

There are many algorithms and toolboxes to implement MCR-ALS for data analysis, some of them are summarized in Table II. The application and the functionality of these algorithms and toolboxes are briefly commented below, for more details the readers are guided to the references. These algorithms and software can be used for many purposes, at Table II highlights only the most important ones.

The first graphical interface for MCR-ALS was developed and implemented in 2005 by Tauler and coauthors [59]. In its latter version, MCR-ALS GUI was implemented in MATLAB environment [43], and used to resolve mixture signals from any kind of analytical data, not only chromatography. Some extensions were included in MCR-ALS basic software, the main ones are MCR Bands [44] used to evaluate the ambiguity of the profiles and ROIMCR software [103] used to select regions of chromatography mass spectrometry data used to metabolomics analysis.

MVC2 GUI [47] and MVC3 GUI [48] are graphical user interfaces implemented in the MATLAB environment for three and four-way data analysis that used the profiles resolved by MCR-ALS to perform a calibration. Both interfaces are unique when it comes to calculating the figures of merit for second and third order calibration. Both interfaces are constantly updated to include new important features. The analyst can also export the recovered profile and see each one for individual samples. The data files must be in txt, commonly exported by most instruments, which eliminates the step of transforming this datafile to MATLAB file.

MCRC software [64] is a graphical interface developed in the MATLAB environment that is based in MCR-ALS GUI. The major advantage is that signal preprocessing such as smooth and baseline correction, other methods to evaluate the chemical and local rank analysis and others initialization algorithms are implemented in a single interface that communicate with MCR-ALS GUI.

Most of the interfaces were initially developed in the MATLAB environment, however, in the past 5 years others MCR-ALS interfaces were developed in free language such as R [50,104,105] and Python 3 [49,106]. Differently from other interfaces that can be used for modelling any type of data, these algorithms and interfaces aim at a specific goal. Most of them were developed targeting metabolomics data generated by LC-MS, GC-MS or GCxGC-MS, except from OCTAVVS that was developed to analyze spectroscopy data. Although implementing this algorithm in free language is an important step to increase the visibility and applicability of the algorithm, these interfaces do not use all the constraints that are available in the implemented interfaces for MATLAB. Therefore, these interfaces are better suited for metabolomics analysis than other chromatographic applications.

These interfaces include compressions such as wavelet transform or dividing the data into regions before modelling, which is important in metabolomics. Another strategy that is implemented in MCR-ALS GUI and OCTAVVS is selecting the region of interest (ROI) to search for m/z values that are not related to baseline.

The extension of rotational ambiguity can be calculated using MCR-BANDS and/or FACPACK software, both implemented in MATLAB environment, but with different fundamentals to identify the ambiguity [44,83].

Table II. Free software and toolbox for multivariate curve resolution and rotational ambiguity analysis

Name	Implementation	Main Purpose	Available at	Ref.
MCR-ALS GUI	MATLAB	Curve resolution	http://www.mcrals.info/	[1,43, 59]
MVC2 GUI	MATLAB	Calibration	http://www.iquir-conicet.gov.ar/descargas/mvc2.rar	[47]
MVC3 GUI	MATLAB	Calibration	http://www.iquir-conicet.gov.ar/descargas/mvc3.rar	[48]
MCRC Software	MATLAB	Curve resolution	http://sharif.edu/~h.parastar/MCRC%20Software.rar	[64]
RMet	R	Metabolomics analysis	https://github.com/SUTChemometricsGroup/RMet	[104]
Alsace Package	R	Metabolomics analysis	https://www.bioconductor.org/packages/release/bioc/html/alsace.html	[50]
ADAP-GC 4.0	R	Metabolomics analysis	http://www.du-lab.org/	[105]
MARS2	Python 3	Curve resolution	https://github.com/mapancsu/MARS2	[106]
OCTAVVS	Python 3	Curve resolution	https://pypi.org/project/octavvs/	[49]
MCR Bands	MATLAB	Ambiguity analysis	http://www.mcrals.info/	[44]
FACPACK	MATLAB	Ambiguity analysis	http://www.math.uni-rostock.de/facpack/Downloads.html	[83]

OUTLOOK AND CONCLUSIONS

In the present review, the fundamentals and important aspects of MCR-ALS such as constraints, ambiguities and rank deficiency issues were presented for the analysis of chromatographic data. The importance of multivariate mathematical separation was discussed, especially for the analysis of complex samples. Since modern analytical chemistry provides different possibilities of acquiring information about the samples under study, the data organization was also addressed in this review. Recent advances and significant applications were described in an extensive revision of the papers published up to 2020.

It is interesting to note how MCR-ALS evolved over the past five decades. In the last decade much effort was made in two fields: (1) develop strategies to analyze metabolomic data such as the ROI and multiblock analysis, and (2) evaluate the extension of rotational ambiguity where MCR-BANDS and AFS tools can be highlighted. It is expected that the use of MCR-ALS will continue evolving in the near future and become part of the commercial software for chromatography.

Conflicts of interest

All the authors declare no conflict of interest.

Acknowledgements

Licarion Pinto thank the Brazilian Funding Agencies, FACEPE 14/2019 INOVA IAM process number APQ-0437-1.06/19 and the “Instituto Nacional de Ciência e Tecnologia de Bioanalítica (INCT- Bioanalítica)”. Igor Miranda Santana is grateful to CNPq for the PhD scholarship (131217/2017-8). Marcia Cristina Breitkreitz thanks FAPESP (process number 2016/05636-9) and INCT-Bioanalítica.

REFERENCES

1. de Juan, A.; Tauler, R. *Anal. Chim. Acta.*, **2021,** *1145*, pp 59-78 (https://doi.org/10.1016/j.aca.2020.10.051).
2. Amigo, J. M.; Skov, T.; Bro, R. *Chem. Rev.*, **2010**, *110* (8), pp 4582-4605 (https://doi.org/10.1021/cr900394n).
3. Sousa, E. S.; Pinto, L.; de Araujo, M. C. U. *Microchem. J.*, **2017**, *134*, pp 131-139 (https://doi.org/10.1016/j.microc.2017.05.017).
4. Lawton, W. H.; Sylvestre, E. A. *Technometrics*, **1971**, *13* (3), pp 617-633 (https://doi.org/10.2307/1267173).
5. Osten, D. W.; Kowalski, B. R. *Anal. Chem.*, **1984**, *56* (6), pp 991-995 (https://doi.org/10.1021/ac00270a029).
6. Escandar, G. M.; Goiocoechea, H. C.; de la Peña. A. M.; Olivieri, A. C. *Anal. Chim. Acta.*, **2014**, *806*, pp 8-26 (https://doi.org/10.1016/j.aca.2013.11.009).
7. Mounier, S.; Nicolodelli, G.; Redon, R.; Milori, D. M. B. P. *Spectrochim. Acta, Part A,* **2017**, *177*, pp 79-85 (https://doi.org/10.1016/j.saa.2017.01.017).
8. Santos, M. C. D.; Azcarate, S. M.; Lima, K. M. G.; Goicoechea, H. C. *Microchem. J.*, **2020**, *155*, pp 104783 (https://doi.org/10.1016/j.microc.2020.104783).
9. Yin, X.-L.; Gu, H.-W.; Liu, X.-L.; Zhang, S.-H.; Wu, H.-L. *Spectrochim. Acta, Part A,* **2018**, *192*, pp 437-445 (https://doi.org/10.1016/j.saa.2017.11.047).
10. Silva, A. C.; Pinto, L.; Gomes, A. A.; de Araujo, M. C. U. *J. Braz. Chem. Soc.*, **2019**, *30* (2), pp 398-405 (https://doi.org/10.21577/0103-5053.20180189).
11. Fernández, L. P.; Brasca, R.; Goicoechea, H.; Culzoni, M. J. *Microchem. J.*, **2020**, *159*, pp 105315 (https://doi.org/10.1016/j.microc.2020.105315).
12. Girón, A. J.; Durán-Merás, I.; Espinosa-Mansilla, A.; de la Peña, A. M.; Cañada, F. C.; Olivieri, A. C. *Anal. Chim. Acta*, **2008**, *622* (1-2), pp 94-103 (https://doi.org/10.1016/j.aca.2008.05.079).
13. Carabajal, M. D.; Arancibia, J. A.; Escandar, G. M. *Talanta*, **2017**, *165*, pp 52-63 (https://doi.org/10.1016/j.talanta.2016.12.030).
14. Fernández, C.; Callao, M. P.; Larrechi, M. S. *Talanta*, **2013**, *117*, pp 75-80 (https://doi.org/10.1016/j.talanta.2013.08.004).
15. Pinto, L.; Diaz Nieto, C. H.; Zon, M. A.; Fernández, H.; de Araujo, M. C. U. *Anal. Chim. Acta*, **2016**, *902*, pp 59-69 (https://doi.org/10.1016/j.aca.2015.10.043).
16. Sousa, E. S.; Schneider, M. P.; Pinto, L.; de Araujo, M. C. U.; Gomes, A. A. *Microchem. J.*, **2020** *152*, pp 104301 (https://doi.org/10.1016/j.microc.2019.104301).
17. Escandar, G. M.; Olivieri, A. C. *J. Chromatogr. A*, **2019**, *1587*, pp 2-13 (https://doi.org/10.1016/j.chroma.2019.01.012).
18. Moreira, E. D. T.; Pinto, L.; Gomes, A. A.; Goicoechea, H. C.; de Araujo, M. C. U. *J. Braz. Chem. Soc.*, **2015**, *26* (8), pp 1573-1582 (https://doi.org/10.5935/0103-5053.20150125).
19. Montemurro, M.; Pinto, L.; Véras, G.; Gomes, A. A.; Culzoni, M. J.; de Araujo, M. C. U.; Goicoechea, H. C. *Talanta*, **2016**, *154*, pp 208-218 (https://doi.org/10.1016/j.talanta.2016.03.078).
20. Carabajal, M. D.; Arancibia, J. A.; Escandar, G. M. *Talanta*, **2018**, *189*, pp 509-516 (https://doi.org/10.1016/j.talanta.2018.07.017).
21. Anzardi, M. B.; Arancibia, J. A.; Olivieri, A. C. *Chemom. Intell. Lab. Syst.*, **2020**, *199*, 103972 (https://doi.org/10.1016/j.chemolab.2020.103972).
22. Vidal, R. B. P.; Ibañez, G. A.; Escandar, G. M. *Anal. Chem.*, **2017**, *89* (5), pp 3029-3035 (https://doi.org/10.1021/acs.analchem.6b04720).
23. Montemurro, M.; Siano, G. G.; Alcaráz, M. R.; Goicoechea, H. C. *Trends Anal. Chem.*, **2017**, *93*, pp 119-133 (https://doi.org/10.1016/j.trac.2017.05.011).
24. Gorrochategui, E.; Jaumot, J.; Lacorte, S.; Tauler, R. *Trends Anal. Chem.*, **2016**, *82*, pp 425-442 (https://doi.org/10.1016/j.trac.2016.07.004).

25. Ortiz-Villanueva, E.; Benavente, F.; Piña, B.; Sanz-Nebot, V.; Tauler, R.; Jaumot, J. *Anal. Chim. Acta*, **2017**, *978*, pp 10-23 (https://doi.org/10.1016/j.aca.2017.04.049).
26. Bos, T. S.; Knol, W. C.; Molenaar, S. R. A.; Niezen, L. E.; Schoenmakers, P. J.; Somsen, G. W.; Pirok, B. W. J. *J. Sep. Sci.*, **2020**, *43* (9-10), pp 1678-1727 (https://doi.org/10.1002/jssc.202000011).
27. de Godoy, L. A. F.; Hantao, L. W.; Pedroso, M. P.; Poppi, R. J.; Augusto, F. *Anal. Chim. Acta*, **2011**, *699*, pp 120-125 (https://doi.org/10.1016/j.aca.2011.05.003).
28. Bahaghighat, H. D.; Freye, C. E.; Gough, D. V.; Synovec, R. E. *J. Chromatogr. A*, **2019**, *1583*, pp 117-123 (https://doi.org/10.1016/j.chroma.2018.11.027).
29. Parastar, H.; Radović, J. R.; Jalali-Heravi, M.; Diez, S.; Bayona, J. M.; Tauler, R. *Anal. Chem.*, **2011**, *83* (24), pp 9289-9297 (https://doi.org/10.1021/ac201799r).
30. Navarro-Reig, M.; Jaumot, J.; Baglai, A.; Vivó-Truyols, G.; Schoenmakers, P. J.; Tauler, R. *Anal. Chem.*, **2017**, *89* (14), pp 7675-7683 (https://doi.org/10.1021/acs.analchem.7b01648).
31. Pérez-Cova, M.; Tauler, R.; Jaumot, J. *Chemom. Intell. Lab. Syst*., **2020**, *201*, 104009 (https://doi.org/10.1016/j.chemolab.2020.104009).
32. Prebihalo, S. E.; Berrier, K. L.; Freye, C. E.; Bahaghighat, H. D.; Moore, N. R.; Pinkerton, D. K.; Synovec, R. E. *Anal. Chem.*, **2018**, *90* (1), pp 505-532 (https://doi.org/10.1021/acs.analchem.7b04226).
33. Porter, S. E. G.; Stoll, D. R.; Rutan, S. C.; Carr, P. W.; Cohen, J. D. *Anal. Chem.*, **2006**, *78* (15), pp 5559-5569 (https://doi.org/10.1021/ac0606195).
34. Pinto, L.; Jardim, I. C. S. F.; Rutledge, D. N.; Breitkreitz, M. C. *Anal. Chim. Acta*, **2020**, *1133*, pp 77-87 (https://doi.org/10.1016/j.aca.2020.08.008).
35. Dwight, S.; Rutan, R.; Venkatramani, S. C.; Cook, C. J.; Daniel, W. *LCGC North Am.*, **2018**, *36* (4), pp 248-255. Available from: https://www.chromatographyonline.com/view/peak-purity-liquid-chromatography-part-ii-potential-curve-resolution-techniques [Acessed 01 March 2020].
36. Montemurro, M.; Siano, G. G.; Alcaráz, M. R.; Goicoechea, H. C. *Trends Anal. Chem.*, **2017**, *93*, pp 119-133 (https://doi.org/10.1016/j.trac.2017.05.011).
37. Olivieri, A. C. *Anal. Chem.*, **2008**, *80* (15), pp 5713-5720 (https://doi.org/10.1021/ac800692c).
38. Olivieri, A. C. *Chem. Rev.*, **2014**, *114* (10), pp 5358-5378 (https://doi.org/10.1021/cr400455s).
39. Grisales, J. O.; Arancibia, J. A.; Castells, C. B.; Olivieri, A. C. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, **2012**, *910*, pp 78-83 (https://doi.org/10.1016/j.jchromb.2012.04.018).
40. Wu, H.-L.; Wang, T.; Yu, R.-Q. *Trends Anal. Chem.*, **2020**, *130*, 115954 (https://doi.org/10.1016/j.trac.2020.115954).
41. Arancibia, J. A.; Damiani, P. C.; Escandar, G. M.; Ibañez, G. A.; Olivieri, A. C. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, **2012**, *910*, pp 22-30 (https://doi.org/10.1016/j.jchromb.2012.02.004).
42. Tauler, R.; Barceló, D. *Trends Anal. Chem.*, **1993**, *12* (8), pp 319-327 (https://doi.org/10.1016/0165-9936(93)88015-W).
43. Jaumot, J.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2015**, *140*, pp 1-12 (https://doi.org/10.1016/j.chemolab.2014.10.003).
44. Jaumot, J.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2010**, *103* (2), pp 96-107 (https://doi.org/10.1016/j.chemolab.2010.05.020).
45. Ruckebusch, C.; Blanchet, L. *Anal. Chim. Acta*, **2013**, *765*, pp 28-36 (https://doi.org/10.1016/j.aca.2012.12.028).
46. Vidal, R. B. P.; Olivieri, A. C. *Anal. Chim. Acta*, **2019**, *1078*, pp 8-15 (https://doi.org/10.1016/j.aca.2019.06.038).
47. Olivieri, A. C.; Wu, H.-L.; Yu, R.-Q. *Chemom. Intell. Lab. Syst.*, **2009**, *96* (2), pp 246-251 (https://doi.org/10.1016/j.chemolab.2009.02.005).
48. Olivieri, A. C.; Wu, H.-L.; Yu, R.-Q. *Chemom. Intell. Lab. Syst.*, **2012**, *116*, pp 9-16 (https://doi.org/10.1016/j.chemolab.2012.03.018).
49. Troein, C.; Siregar, S.; De Beeck, M. O.; Peterson, C.; Tunlid, A.; Persson, P. *Methods Protocol.*, **2020**, *3* (2), 34 (https://doi.org/10.3390/mps3020034).

50. Wehrens, R.; Carvalho, E.; Fraser, P. D. *Metabolomics*, **2015**, *11*, pp 143-154 (https://doi.org/10.1007/s11306-014-0683-5).
51. Shen, H.; Stordrange, L.; Manne, R.; Kvalheim, O. M.; Liang, Y. *Chemom. Intell. Lab. Syst.*, **2000**, *51* (1), pp 37-47 (https://doi.org/10.1016/S0169-7439(99)00066-0).
52. Shen, H.; Liang, Y.; Kvalheim, O. M.; Manne, R. *Chemom. Intell. Lab. Syst.*, **2000**, *51* (1), pp 49-59 (https://doi.org/10.1016/S0169-7439(00)00054-X).
53. Culzoni, M. J.; de Llanos, A. M.; De Zan, M. M.; Espinosa-Mansilla, A.; Cañada-Cañada, F.; de la Peña, A. M.; Goicoechea, H. C. *Talanta*, **2011**, *85* (5), pp 2368-2374 (https://doi.org/10.1016/j.talanta.2011.07.086).
54. Culzoni, M. J.; Goicoechea, H. C.; Ibañez, G. A.; Lozano, V. A.; Marsili, N. R.; Olivieri, A. C.; Pagani, A. P. *Anal. Chim. Acta*, **2008**, *614* (1), pp 46-57 (https://doi.org/10.1016/j.aca.2008.03.013).
55. Parastar, H.; Akvan, N. *Anal. Chim. Acta*, **2014**, *816*, pp 18-27 (https://doi.org/10.1016/j.aca.2014.01.051).
56. Pinto, L.; Stechi, F.; Breitkreitz, M. C. *Microchem. J.*, **2019**, *146*, pp 202-209 (https://doi.org/10.1016/j.microc.2019.01.014).
57. Parastar, H.; Shaye, H. *RSC Adv.*, **2015**, *5* (86), pp 70017-70024 (https://doi.org/10.1039/C5RA10658C).
58. Olivieri, A. C.; Escandar, G. M.; Goicoechea, H. C.; de la Peña, A. M. *Fundamentals and Analytical Applications of Multiway Calibration*. Elsevier, **2015**.
59. Jaumot, J.; Gargallo, R.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2005**, *76* (1), pp 101-110 (https://doi.org/10.1016/j.chemolab.2004.12.007).
60. Keller, H. R.; Massart, D. L. *Chemom. Intell. Lab. Syst.*, **1991**, *12* (3), pp 209-224 (https://doi.org/10.1016/0169-7439(92)80002-L).
61. Maeder, M. *Anal. Chem.*, **1987**, *59* (3), pp 527-530 (https://doi.org/10.1021/ac00130a035).
62. Sánchez, F. C.; van den Bogaert, B.; Rutan, S. C.; Massart, D. L. *Chemom. Intell. Lab. Syst.*, **1996**, *34* (2), pp 139-171 (https://doi.org/10.1016/0169-7439(96)00020-2).
63. Windig, W.; Guilment, J. *Anal. Chem.*, **1991**, *63* (14), pp 1425-1432 (https://doi.org/10.1021/ac00014a016).
64. Jalali-Heravi, M.; Parastar, H.; Kamalzadeh, M.; Tauler, R.; Jaumot, J. *Chemom. Intell. Lab. Syst.*, **2010**, *104* (2), pp 155-171 (https://doi.org/10.1016/j.chemolab.2010.08.002).
65. Rodríguez-Cuesta, M. J.; Bosqué, R.; Rius, F. X. *Anal. Chim. Acta*, **2003**, *476* (1), pp 111-122 (https://doi.org/10.1016/S0003-2670(02)01360-0).
66. Navea, S.; Tauler, R.; de Juan, A. *Anal. Chem.*, **2006**, *78* (14), pp 4768-4778 (https://doi.org/10.1021/ac052257r).
67. Kuligowski, J.; Quintás, G.; Tauler, R.; Lendl, B.; de la Guardia, M. *Anal. Chem.*, **2011**, *83* (12), pp 4855-4862 (https://doi.org/10.1021/ac2004407).
68. Malik, A.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2014**, *135*, pp 223-234 (https://doi.org/10.1016/j.chemolab.2014.04.002).
69. Antunes, M. C.; Simão, J. E. J.; Duarte, A. C.; Tauler, R. *Analyst*, **2002**, *127* (6), pp 809-817 (https://doi.org/10.1039/b200243b).
70. Mallat, E.; Barceló, D.; Tauler, R. *Chromatographia*, **1997**, *46*, pp 342-350 (https://doi.org/10.1007/BF02490871).
71. de Juan, A.; Maeder, M.; Martínez, M.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2000**, *54* (2), pp 123-141 (https://doi.org/10.1016/S0169-7439(00)00112-X).
72. Van Benthem, M. H.; Keenan, M. R.; Haaland, D. M. *J. Chemom.*, **2002**, *16* (12), pp 613-622 (https://doi.org/10.1002/cem.761).
73. Schröder, H.; Sawall, M.; Kubis, C.; Jürß, A.; Selent, D.; Brächer, A.; Börner, A.; Franke, R.; Neymeyr, K. *Chemom. Intell. Lab. Syst.*, **2017**, *163*, pp 55-63 (https://doi.org/10.1016/j.chemolab.2017.02.002).
74. Zhang, X.; Zhang, Z.; Tauler, R. *Talanta*, **2019**, *202*, pp 554-564 (https://doi.org/10.1016/j.talanta.2019.05.002).

75. Dadashi, M.; Abdollahi, H.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2017**, *162*, pp 203-213 (https://doi.org/10.1016/j.chemolab.2017.01.009).
76. Zhang, X.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2015**, *147*, pp 47-57 (https://doi.org/10.1016/j.chemolab.2015.08.005).
77. Tauler, R. *Anal. Chim. Acta*, **2007**, *595* (1-2), pp 289-298 (https://doi.org/10.1016/j.aca.2006.12.043).
78. Vosough, M.; Mason, C.; Tauler, R.; Jalali-Heravi, M.; Maeder, M. *J. Chemom.*, **2006**, *20* (6-7), pp 302-310 (https://doi.org/10.1002/cem.1022).
79. Tauler, R. *J. Chemom.*, **2001**, *15* (8), pp 627-646 (https://doi.org/10.1002/cem.654).
80. Tauler, R.; Smilde, A.; Kowalski, B. *J. Chemom.*, **1995**, *9* (1), pp 31-58 (https://doi.org/10.1002/cem.1180090105).
81. Sawall, M.; Neymeyr, K. *Anal. Chim. Acta*, **2017**, *960*, pp 40-52 (https://doi.org/10.1016/j.aca.2016.11.069).
82. Lakeh, M. A.; Abdollahi, H.; Gemperline, P. J. *Anal. Chim. Acta*, **2020**, *1105*, pp 64-73 (https://doi.org/10.1016/j.aca.2020.01.022).
83. Sawall, M.; Neymeyr, K. *Anal. Chim. Acta*, **2014**, *828*, pp 17-26 (https://doi.org/10.1016/j.aca.2014.04.026).
84. Jiménez-Carvelo, A. M.; Cruz, C. M.; Olivieri, A. C.; González-Casado, A.; Cuadros-Rodríguez, L. *Talanta*, 2019, 195, pp 69-76 (https://doi.org/10.1016/j.talanta.2018.11.033).
85. Chang, Y.; Wu, H.; Fang, H.; Wang, T.; Ouyang, Y.; Sun, X.; Tong, G.; Ding, Y.; Yu, R. *Talanta*, **2021**, *224*, 121798 (https://doi.org/10.1016/j.talanta.2020.121798).
86. Peng, T.; Yin, X.; Gu, H.; Sun, W.; Ding, B.; Hu, X.; Ma, L.; Wei, S.; Liu, Z.; Ye, S. *Food Chem.*, **2021**, *347*, 128959 (https://doi.org/10.1016/j.foodchem.2020.128959).
87. Yin, X.; Gu, H.; Jalalvand, A. R.; Liu, Y.; Chen, Y.; Peng, T. *J. Chromatogr. A*, **2018**, *1573*, pp 18-27 (https://doi.org/10.1016/j.chroma.2018.09.019).
88. Vidal, R. B. P.; Olivieri, A. C.; Ibañez, G. A.; Escandar, G. M. *ACS Omega*, 2018, *3* (11), pp 15771-15779 (https://doi.org/10.1021/acsomega.8b02439).
89. Gómez-Canela, C.; Prats, E.; Piña, B.; Tauler, R. *Environ. Pollut.*, 2017, *220*, pp 1231-1243 (https://doi.org/10.1016/j.envpol.2016.11.010).
90. Navarro-Reig, M.; Tauler, R.; Iriondo-Frias, G.; Jaumot, J. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.*, 2019, *1104*, pp 148-156 (https://doi.org/10.1016/j.jchromb.2018.11.018).
91. Long, W.; Wu, H.; Wang, T.; Dong, M.; Yu, R. *Microchem. J.*, **2020**, *157*, 105003 (https://doi.org/10.1016/j.microc.2020.105003).
92. Cook, D. W.; Burnham, M. L.; Harmes, D. C.; Stoll, D. R.; Rutan, S. C. *Anal. Chim. Acta*, **2017**, *961*, pp 49 - 58 (https://doi.org/10.1016/j.aca.2017.01.047).
93. Garreta-Lara, E.; Campos, B.; Barata, C.; Lacorte, S.; Tauler, R. *Sci. Total Environ.*, **2018**, *610-611*, pp 602-612 (https://doi.org/10.1016/j.scitotenv.2017.05.190).
94. Azimi, A.; Bahktiari, A. R.; Tauler, R. *Environ. Pollut.*, **2018**, *243*, pp 374-382 (https://doi.org/10.1016/j.envpol.2018.08.073).
95. Shekari, N.; Vosough, M.; Heidar, K. T. *J. Sep. Sci.*, **2017**, *40* (6), pp 1318-1326 (https://doi.org/10.1002/jssc.201601313).
96. Lebanov, L.; Tedone, L.; Ghiasvand, A.; Paull, B. *Talanta*, **2020**, *219*, 121208 (https://doi.org/10.1016/j.talanta.2020.121208).
97. Hurtado, C.; Parastar, H.; Matamoros, V.; Piña, B.; Tauler, R.; Bayona, J. M. *Sci. Rep.*, **2017**, *7*, 6546 (https://doi.org/10.1038/s41598-017-06773-0).
98. Teglia, C. M.; Cámara, M. S.; Vera-Candioti, L. *Electrophoresis*, **2017**, *38*, pp 1122-1129 (https://doi.org/10.1002/elps.201600475).
99. Guedes, L. S.; Santana, C. C.; Rutledge, D. N.; Pinto, L.; Jardim, I. C. S. F.; de Melo, L. V.; Beppu, M. M.; Breitkreitz, M. C. *Can. J. Chem. Eng.*, **2021**, pp 1-16 (https://doi.org/10.1002/cjce.23969).

100. Soares, F. L. F.; Ardila, J. A.; Carneiro, R. L. J. Raman Spectrosc., **2017**, *48*, pp 943-950 (https://doi.org/10.1002/jrs.5168).
101. Santana, I. M.; Jardim, I. C. S. F.; Breitkreitz, M. C. *J. Sep. Sci.*, **2020**, *43* (22), pp 4234-4242 (https://doi.org/10.1002/jssc.202000702).
102. US Food and Drug Administration. *Guidelines for the Validation of Chemical Methods in Food, Feed, Cosmetics, and Veterinary Products* (3rd Ed.). Silver Spring, MD, USA: FDA, **2019**.
103. Gorrochategui, E.; Jaumot, J.; Tauler, R. *BMC Bioinf.*, **2019**, *20*, Article number 256 (https://doi.org/10.1186/s12859-019-2848-8).
104. Moayedpour, S.; Parastar, H. *Chemom. Intell. Lab. Syst.*, **2019**, *194*, 103866 (https://doi.org/10.1016/j.chemolab.2019.103866).
105. Smirnov, A.; Qiu, Y.; Jia, W.; Walker, D. I.; Jones, D. P.; Du, X. *Anal. Chem.*, **2019**, *91* (14), pp 9069-9077 (https://doi.org/10.1021/acs.analchem.9b01424).
106. Ma, P.; Li, M.; Lu, H.; Zhang, Z. *Chemom. Intell. Lab. Syst.*, **2019**, *191*, pp 12-20 (https://doi.org/10.1016/j.chemolab.2019.05.010).

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 45-61
doi: 10.30744/brjac.2179-3425.AR-38-2021

BrJAC

ARTICLE

Data Mining, Machine Learning, Deep Learning, Chemometrics

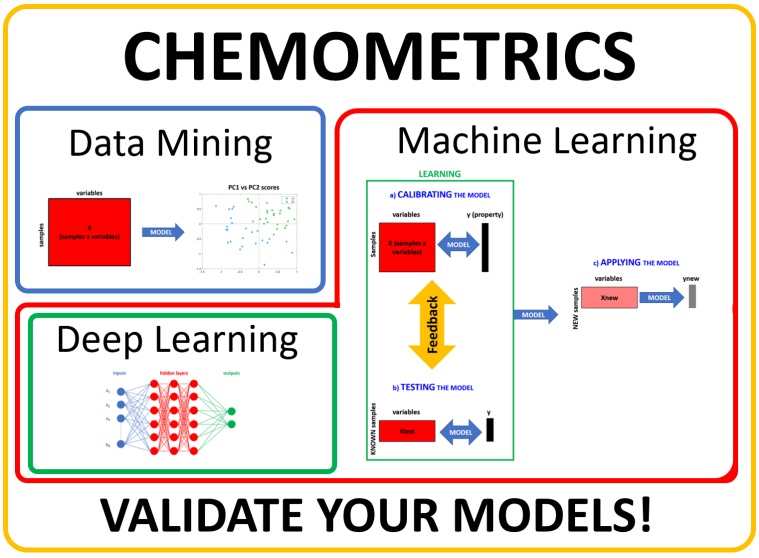
Definitions, Common Points and Trends

(Spoiler Alert: VALIDATE your models!)

José Manuel Amigo

IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

Department of Analytical Chemistry, University of the Basque Country UPV/EHU, P.O. Box 644, 48080 Bilbao, Basque Country, Spain



Concepts like Machine Learning, Data Mining or Artificial Intelligence have become part of our daily life. This is mostly due to the incredible advances made in computation (hardware and software), the increasing capabilities of generating and storing all types of data and, especially, the benefits (societal and economical) that generate the analysis of such data. Simultaneously, Chemometrics has played an important role since the late 1970s, analyzing data within natural science (and especially in Analytical Chemistry). Even with the strong parallelisms between all of the abovementioned terms and being popular with most of us, it is still difficult to clearly define or differentiate the meaning of Machine Learning, Data Mining, Artificial Intelligence, Deep Learning and Chemometrics. This manuscript brings some light to the definitions of Machine Learning, Data Mining, Artificial Intelligence and Big Data Analysis, defines their application ranges and seeks an application space within the field of analytical chemistry (a.k.a. Chemometrics). The manuscript is full of personal, sometimes probably subjective, opinions and statements. Therefore, all opinions here are open for constructive discussion with the only purpose of Learning (like the Machines do nowadays).

MOTIVATION

I have spent the last 20 years analyzing data from many different analytical sources and many different scientific fields. From hyphen and hypernated chromatographic data [1] to all types of spectroscopies and different analytical metrics, they are specialized in hyperspectral image analysis [2]. Curiously, in recent years, I have been asked increasingly more often if what I do is Machine Learning, Data Mining, or even

Cite: Amigo, J. M. Data Mining, Machine Learning, Deep Learning, Chemometrics – *Definitions, Common Points and Trends* (Spoiler Alert: VALIDATE your models!) *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 45–61. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-38-2021

Deep Learning. My answer is always the same: I apply the mathematical procedure that I need to solve the problem I am dealing with if I need one at all. I completely acknowledge that this answer is quite open to interpretation. However, it could be summarized by one small sentence. **What I do is Chemometrics**.

During the last few years, there has been quite a lot of confusion in the literature with the terms Machine Learning, Data Mining, Deep Learning and Artificial Intelligence. I have observed a dangerous trend in the scientific literature towards their usage, highlighting the powerful benefits that these methodologies have in the data, using them as if the most important part of applying Machine Learning, for instance, was the algorithm we use. I have also observed that every analytical data issue could be solved by just constructing more complex algorithms, including more non-linear parameters and without placing more attention on the quality of the data being used.

Therefore, I thought that writing a manuscript such as this one could serve to 1) differentiate the terms mentioned beforehand, 2) highlight the most important facts of building a multivariate model and 3) give some tips and tricks to new students who want to start applying any mathematical model in their analytical data.

This manuscript is divided into 4 sections. The first will discuss the three most important aspects of data analysis: the data, the reference values and the model. I will comment on the structure of the data, the importance of relying on yor reference values, and the meaning of a mathematical model. The second and third parts define the terms Data Mining, Machine Learning, Deep Learning and Artificial Intelligence and puts them into the framework of Chemometrics. The manuscript will finish with a fourth section containing “take-home” messages. This is, arguably, the most important section of this manuscript, as it collects a series of advice that I have been using during my career and advised my students.

The manuscript is written in informal English to arrive at the audience in a more straightforward way. The manuscript is primarily addressed to Chemometricians; nevertheless, it would be a mistake not to open it up to anyone who analyzes data of any kind. It does not explain every term exhaustively or make a comprehensive revision of the literature on those matters. For that, some (but not many) keynote references are provided. Therefore, the reader is encouraged to check those references to find a more comprehensive explanation of specific models.

This manuscript contains several critical (therefore, subjective) opinions and suggestions that are completely open for discussion. Most of them give the impression of being too obvious; they are so obvious that sometimes we forget to pay attention to them. That is why I consider that highlighting them is important. I hope the readers find this manuscript interesting, bearing in mind that constructive criticism is more than welcome, always with the purpose of learning (as machines do nowadays).

THE DATA (X), THE REFERENCE (Y) AND THE MODEL. GI-GO

The reader would probably expect the manuscript to start with the definitions of Machine Learning, Data Mining, etc. Nevertheless, let me start the manuscript with the most important part in applying mathematical models to data: the data.

The data are just a mere collection of information containing relevant information and noise (i.e. not relevant information). Nevertheless, the most important aspect for success when applying any data analysis strategy is to have GOOD data **X** and, if needed, GOOD reference values **Y**. We know it as the GI-GO (Garbage In – Garbage Out) truism [3]:

IF THE DATA DO NOT CONTAIN ANY INFORMATION RELATED TO WHAT YOU WANT TO MEASURE AND/OR IF THE REFERENCES ARE NOT RELATED TO WHAT YOU WANT TO MEASURE, YOU WILL NOT OBTAIN GOOD RESULTS REGARDLESS OF THE MODEL USED

The GI-GO truism is, by far, the major cause of the frustration in the data analysis procedures. We blame the algorithms most of the time, forgetting that the algorithm will not find the information if it is not in the data. One of the biggest mistakes that we can commit is hypothesizing that the model will give

the solution we are looking for. On the contrary, the solution must be in the data and its correlation with the reference values. The model is, and will always be, the tool that helps us to find data patterns or the correlation between the data and the reference (if it exists).

The major issue here is that sometimes those patterns/correlations are difficult to find because they represent a small amount of variance/co-variance in the data. This is there where different algorithmic approaches can be tested, but always after being completely sure that we fully understand the nature (structure) and origin of our data.

KNOW YOUR DATA AND, IF ANY, YOUR REFERENCE

The data

One of the premises to start with is the fact that the analytical information that we measure is (or can be seen as) multivariate. That is, for one sample, many variables/observations are normally collected. We usually want to compare samples assuming that the differences or similarities between them will be found in the variables (or groups of variables) that we measure. In other words, we want to obtain useful information and get rid of the noise.

DATA = INFORMATION + NOISE

For this to be accomplished, we need to know three important data features: the nature (structure), amount and quality.

The nature (structure) of the data

Knowing the nature of our data will help us to choose 1) the appropriate pre-processing methods and 2) the subsequent models. A normal arrangement of the data is in the shape of a matrix (Figure 1a-d). This matrix **X** is normally composed of samples in rows and variables in columns (even though some disciplines prefer the transposed version with samples in columns and variables in rows, we will keep the classical nomenclature that is commonly used in Chemometrics) [3]. Nevertheless, there are many ways in which **X** can be constructed; even with the same apparent dimensions, different scientific instruments might provide data with a completely different structure. For instance, the cases presented in Figure 1a-d, where four matrices **X** are presented with the same dimensions but a completely different structure. The data in Figure 1a represents the typical data coming from any spectroscopic device, where the spectrum at N variables has been collected for each of the mth samples. Instead, the data in Figure 1b represents a single sample where the intensity level of a property has been measured in the pixels located in positions M and N (i.e. a monochannel picture). Even the data coming from the same instrument can be arranged/handled in different ways. For instance, when the chromatogram of a sample is measured, we can study the chromatogram variation between samples (Figure 1c) or construct a table where the integrated area of the peaks composes the variables.

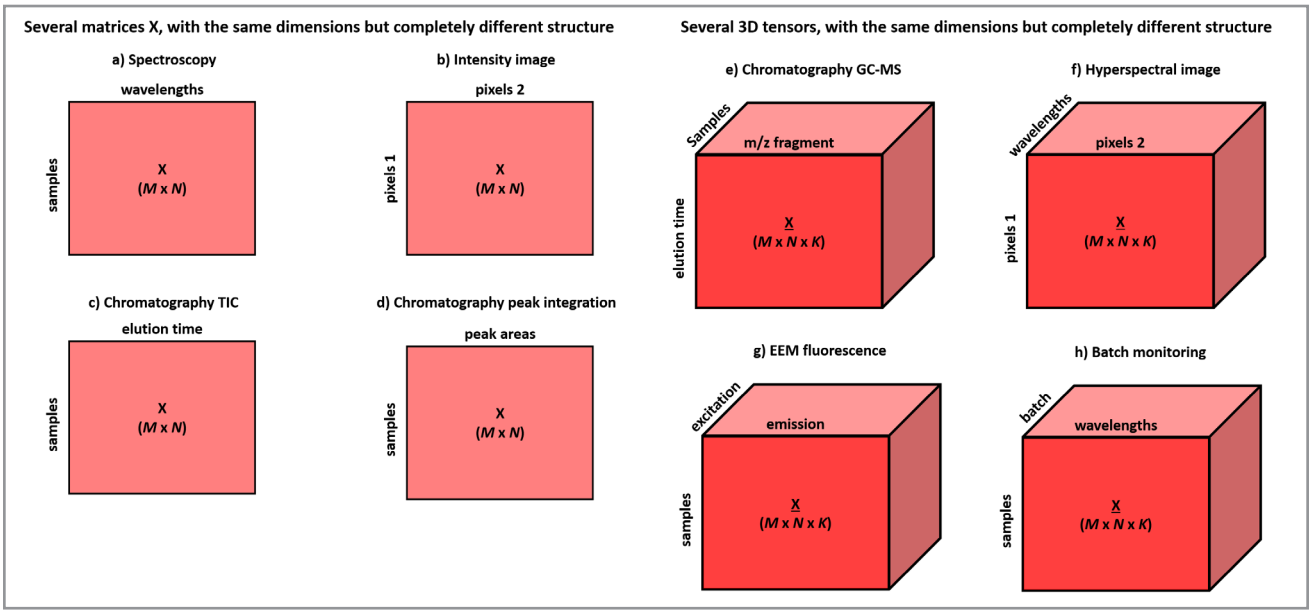


Figure 1. Different structure of datasets having the same dimensions. TIC, total Ion Chromatogram. GC-MS, Gas Chromatography-Mass Spectroscopy.

This difference in the nature of the signal has a strong impact, especially in the pre-processing and variable normalization steps that need to be applied before data analysis, where the relationship between the variables plays a fundamental role. Apart from being continuous, discrete or categorical in the columns, the variables can also be correlated in the rows. A toy example constructs a matrix considering of two continuous variables like pH and temperature. Those two variables, in principle, are independent. Therefore, there is no relevance in constructing a [pH Temperature] matrix or [Temperature pH] matrix. Now, we should think about spectral variables. When a spectrum is measured, for instance, in Near Infrared, there is a correlation between wavelength1, wavelength2 and wavelength3. Therefore, constructing a matrix such as [wavelength1 wavelength3 wavelength2] will have an impact on the pre-processing applied to that matrix because the correlation/continuity between variables has been broken.

Of the four examples, the image (Figure 1b) does not fulfil the requirement of (sample x variables) structure. Indeed, that matrix only represents information relative to one sample. Therefore, and depending on the aim, several steps should be taken beforehand (vectorization, features extraction, etc.). Considering the spectral and chromatographic profile matrices (Figure 1a and c), one can argue that the data structure is the same (or, at least, very similar). Nevertheless, the data source is completely different, and, therefore, the issues coming from the measurement must be addressed properly. The issues normally found in chromatograms (baseline drifts, peak misalignment, normalization by standard peaks, etc.) are not the same as the issues normally found in spectroscopy. Consequently, both scenarios need different pre-processing methods.

The issue can be further complicated if we consider that we might have data that contain more than 2 directions. Figure 1e-h show the situation dealing with 3D datacubes (a.k.a. tensors). There is a third direction in the data that expresses either another analytical variable or another measurement condition. For instance, the data coming from chromatographic measurements where the detector is a multichannel detector normally leads to datacubes where there is an extra spectral dimension. In a hyperspectral measurement, one sample is normally visualized as a datacube. The measurements in one sample give a matrix composed of the emission spectra at different excitation wavelengths in multiway fluorescence [4,5].

Another typical example is the data collected from different batches in factories (Figure 1h). All of these are datacubes. Nevertheless, their structure is completely different, and here we might have to take the initial decision to treat the data as such (by using multiway models [4,5]) or unfold the data in such a way that we set it as a matrix that fulfils the previous statement in the relationship between the samples and the variables.

In the case of chromatographic datacubes, there are normally two alternatives: either use multiway methods with a cube **X** (samples x elution time x spectral direction) (and thus assuming specific relationships between the samples and now the two variable directions) or unfolding the cube in different ways: (samples*elution time x spectral direction) or (samples x elution time*spectral direction). The same issue is found in, for example, Excitation – Emission fluorescence. In the case of hyperspectral images, a previous step of unfolding is normally preferred. Therefore, the datacube **X** (X x Y x wavelengths) is normally treated as a matrix where each pixel is considered a sample, obtaining **X** (X*Y x wavelengths).

The reader will find in the literature a plethora of options for arranging data regarding its structure, the aim of the analysis and the benefits/drawbacks of arranging the data in different ways, which makes this a bit cumbersome. Unfortunately, there is no clear answer in this regard. Nevertheless, I could recommend that the reader perfectly knows the data, the nature of the relationship between the samples and the variables, and the plausibility of applying different strategies that might arrive at the same result.

The amount and quality of data

One recurrent question during my lessons is how many samples are needed to build a good model (either for exploring data or building a regression or a classification model). The answer is easy:

YOUR MODEL NEEDS AS MANY SAMPLES AS NECESSARY TO CERTIFY THE RELIABILITY AND REPRESENTATIVENESS OF YOUR MODEL AND WHATEVER YOU WANT TO EXPRESS WITH YOUR MODEL

There are scenarios where what we want to measure is clearly expressed in the data that are being measured. For instance, measuring proteins in barley with Near Infrared (NIR) has been a classical tool, almost since NIR instruments were invented. The signal of the protein band is normally well expressed in the NIR region, the concentration of protein is high enough to have a good signal-to-noise (S/N) ratio, and the classical interferences in NIR like water (moisture) are not present in such high amounts, so they do not affect the NIR signal to a great extent. Therefore, it is the amount of data that plays the game and affects quality.

The quality of the data can be assessed by controlling different parameters of the measurement scenario in the measured signal:

- The instrumental noise: Data will contain noise. It is a fact. Nevertheless, even though different methodologies can minimize that noise, it must not be higher than the signal being measured.
- The composition of the sample and the plausible interferences in the signal that we are looking for: Measuring multivariate data means that not all of the variables are useful for answering the analytical question. Moreover, the signals that will give us the answer (spectral bands, chromatographic peaks, etc.) may be strongly influenced by another chemical (or physical) compound that could be in the sample matrix.
- The influence of the environmental conditions: Measuring in laboratory conditions is sometimes completely different from measuring in more uncontrolled conditions.
- The correlation with the property that we want to measure: In regression and classification, the close connection of the data with the reference value is essential.

Many of the previously mentioned issues can be partially minimized by developing a proper protocol in the calibration of the instrument and the inclusion of reference compounds to normalize the data. Also, having a good Design of the Experiment [6] and a perfectly optimized analytical method will help us to understand the plausible confounding factors that we might have when designing our experiment.

The reference values

When regression or classification is needed, the role of the reference values is essential to identifying reliable correlations between them and the data **X** through the model, since having good data **X** is not sufficient to obtain a good model. The reference values **Y** can be obtained in many different ways, depending on the aim of the experiment. However, they can be summarized into two major blocks: regression and classification.

Regression

When a regression model is made, the reference value normally comes from a standardized analytical procedure that the proper regulatory agencies have approved; sometimes, it comes by being the most accepted procedure by the analytical community. Regardless of the precedence, we must be aware that the reference values contain an analytical error and a calibration range, together with a limit of detection and quantitation that is utterly linked to the data **X**.

Sometimes, the reference value error can be neglected if the error of the data **X** is larger. However, assuming this statement without verifying it might lead to a wrong interpretation of the result. The normal procedure to verify the error of the reference values is to make repeated measurements of the same sample and ascertain that the variance (standard deviation of the mean) between the replicates is within certain confidence levels. These concepts come from classical analytical chemistry procedures. Nevertheless, developing such protocols is sometimes time-consuming, or there is not enough budget to perform as many as we would like.

Classification

Classification is directly linked to the assignation of the belonging of one sample to one class, several classes, or none, depending on the classification strategy [7]. Therefore, the reference values are normally given by a categorical indexation of **Y**, where an inter-correlation between the different columns in **Y** is expected or, at least, assumed. In many scenarios, the **Y** class is imposed by applying certain thresholds to continuous variables (Temperature < 20 is cold, 21 < Temperature < 30 is mild, Temperature > 30 is hot). Alternatively, there are also occasions where the classes are assigned by using sensory panels. Then, the final class is assessed by an average of the grades given by a certain number of judges. In these cases, the assignation of classes comes with a similar analytical error defined beforehand in the regression scenario, being crucial to understanding the performance of the future model.

Many different strategies can be implemented when dealing with classification problems. Let me start with the simplest one; that is, the two-class problem. Figure 2a shows a case where the samples want to be separated by color. There are two plausible arrangements for the matrix **Y**. The first arrangement is to construct **Y** in the shape of a column vector containing arbitrary values assigned to each class. In this case, the number assigned to each class is completely irrelevant since all of the statistic parameters, and figures of merit will be constructed based on the relative difference between the predictions. The second strategy consists of constructing a matrix with two columns, where 1 and 0 are normally used to denote belonging and not belonging, respectively. Both strategies are equally good in the two-class case, and the results obtained from both strategies will be the same.

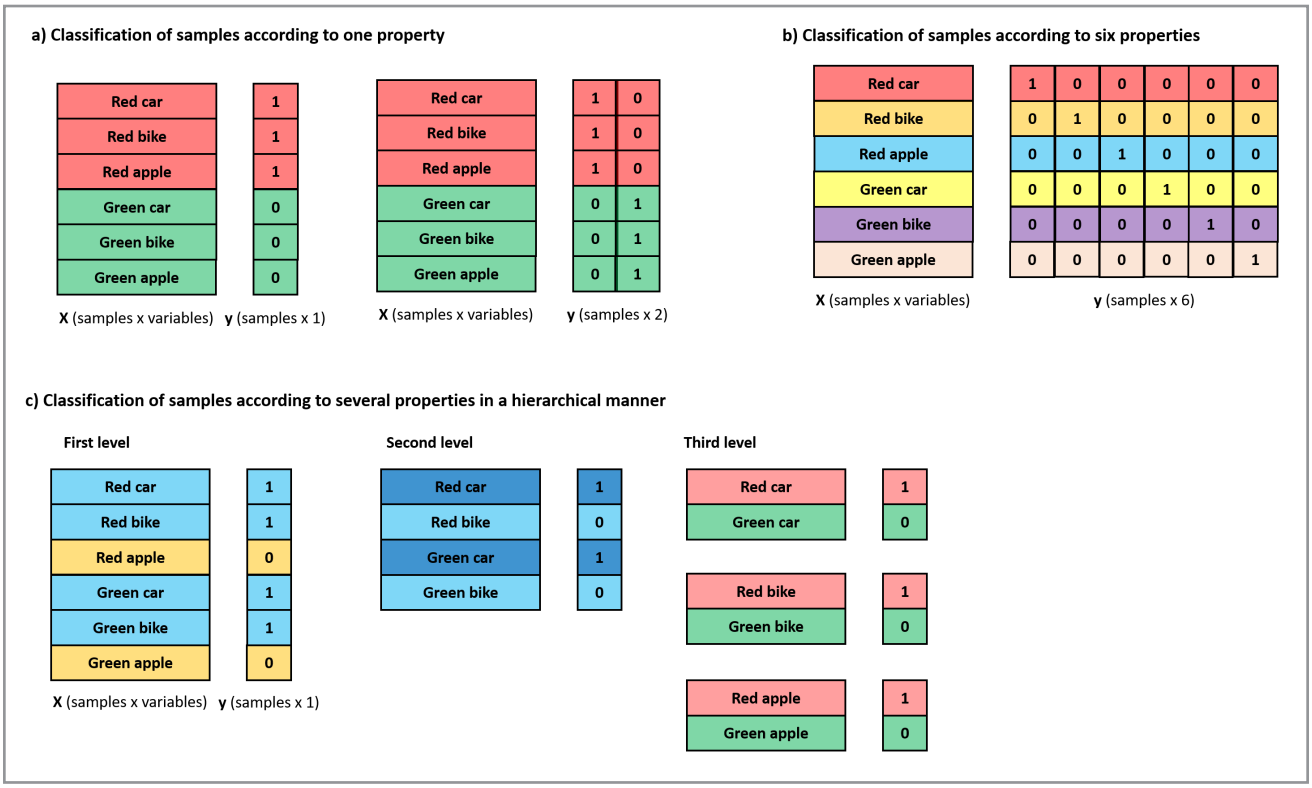


Figure 2. Different classification strategies that can be adopted to classify the same samples in different classes.

Nevertheless, it is important to make a statement that will be important for any classification strategy (not only for the two-class case). As the reader has probably noticed, the toy example in Figure 2 does not seem to be as realistic as we could expect. The samples are, indeed, quite different. There are red and green cars, red and green bikes and red and green apples. This is, indeed, an example that makes it clear that samples can be classified as we need them to be classified. It will depend on the strong correlation between what we want to classify and what we are actually measuring. Obviously, if we want to separate the samples by a different color, we must be sure that the data **X** that we obtain/measure is directly or indirectly related to the color of the samples. We normally hypothesize that our classification problem will be solved by performing certain measurements. Also, as indicated, this is just a hypothesis. Despite the many efforts made with our models, we cannot classify the samples by color measuring the number of wheels.

IF DATA X DO NOT REFLECT OR CONTAIN INFORMATION RELATED TO THE Y CLASS, NO MODEL WILL BE ABLE TO CLASSIFY THEM CORRECTLY

The previous statement is, of course, valid for the regression problem. Despite seeming obvious, it is one of the major sources of frustration; this is quite understandable. We construct our classification/regression hypothesis based on the problem. Nevertheless, the final decisions to solve the problems are normally drafted by the instruments that we have available. Also, we have to adapt, hoping to find those little pieces of information in our data that correlate with the class.

When the classification problem involves more than two classes, we might follow different strategies. The first is to build a classification model with all of the plausible classes. In our bike-car-apple example, this is represented in Figure 2, where we have 6 different and somehow independent classes. In this case, the matrix **Y** is normally constructed following a logical pattern of a correlated/not correlated (a.k.a. yes/no) strategy where there are as many columns as classes. Here, it is extremely important to follow the methodology indicated in the figure since we have to ascertain the equality of the influence for each class

in the model. There might be the temptation to assign a number to each class (if we have 6 classes, **Y** will be a vector with a number from 1 to 6). This is a big mistake. All classification models are normally built following a strategy in which a regression model is first developed, and then some thresholds are applied. Therefore, we will be committing the incorrect assumption that the distance between class 1 and class 4 is larger than the distance between class 1 and class 2, just because 4 is larger than 2.

Apart from that, and after constructing the correct **Y** matrix, normally filled with zeros and ones, we must be completely sure, as before, that the data **X** will contain enough information in the variables that will make the generation of a classification model for 6 classes at the same time possible.

It might be that the variance given in the data **X** by the different classes is not comparable or evenly distributed. Therefore, there is another strategy that facilitates the task of classifying the samples by using what is called a hierarchical model strategy. In a hierarchical strategy (Figure 2c), the problem is split into three minor classification models to solve the classification issue hierarchically. The samples will be classified first by the most important class (the class giving more variance in the data **X**) and then by other classes. However, this strategy is still based on the fact that no matter how many levels you think of, if the variables in data X are not collected, the information will not work either.

The model

Now yes! After understanding the importance of the data **X** and the reference **Y**, it is the perfect time to talk about the model a little bit. What is a model, and what is the difference between that and the word "modelling"?

A model, in mathematics, is the set of parameters and operations that fits the value of a dependent variable (y) to an independent variable (x). Easy [3]:

$$\mathbf{y = f(x)}$$

That relationship can be different (linear or non-linear) and with varying complexity. Nevertheless, it can be directly extrapolated to the concept of multivariate models. The main difference is that in a multivariate model, we have multivariate data. Therefore, given a set of samples measured by independent variables **X** (M x N) and a property **y** (M x 1), the model that establishes the correlation between **X** and **y** is:

$$\mathbf{y} = b_1x_1 + b_2x_2 + \ldots + b_nx_n = \mathbf{bX + e}$$

where **b** is what we normally call the regression vector and
e is the vector (M x 1) containing the residuals.

Of course, the previous equation is a mere visualization of a multivariate model applied to a regression (or even classification). The multivariate model must be considered the relationship between **X** and **y** and all of the pre-processing strategies, variables selected, and optimization parameters needed to predict/forecast the behavior of new **y** samples. This is the main aim of a multivariate model, to predict the value of y in new samples that have not been included in the construction of the model. This is partially true because there are situations where we do not want to correlate our data **X** with any **y**. We just want to study the patterns (points in common and trends) in the data **X**.

Pattern recognition models are the ones that, given one data matrix **X,** aim to study the correlations and differences (variance) between the samples reflected by the variables or groups of measured variables. The workhorse method is, undoubtedly, the principal component analysis (PCA) model [8,9]. PCA will indeed decompose our data matrix **X** into a set of so-called scores and a set of so-called loadings as indicated below:

$$\mathbf{X} = \mathbf{TP}^{T} + \mathbf{E}$$ (by the way, this is one of the essential equations in data analysis)

where **T** (M x F) is the score, **P** (N x F) is the loading with a superscript denoting the transposed value of the matrix, and **E** (M x N) is the residuals. F denotes the number of principal components or, in other words, the number of different independent sources of variance (variability) of the data.

In PCA, the model is, strictly speaking, the loading and pre-processing and normalization strategies used. Nevertheless, we tend to include the scores in the team, which it is acceptable to do.

Many algorithms obtain relationships between **X** and **y** and determine the trends in a single matrix **X** [10]. Nevertheless, in the end, what counts is that the relationships obtained give us useful information. In this regard, it is of utmost importance to highlight one fact:

THE MODELS WORK IN LOCAL SCENARIOS IF THE DATA ONLY CONTAIN LOCAL INFORMATION

The models that we construct are based on the information that we provide (the data **X** and the **Y** reference values, again!). It all depends on how good and comprehensive this information is. Therefore, what we can normally ensure is that our models work under certain conditions and certain limitations. Let me use a simple example. A classification model can be constructed in order to differentiate the origin of red wines. First of all, we might want to avoid expressions like “my model can classify different red wines”. This could be true, but then we have to ensure that our **X** matrix includes all red wines in the world. The model arrives as far as the richness of the data with which it is built. Also, we must be careful with expressions like “This model is always better than this one”. Well, not in a multivariate perspective and not working in local situations. The correct sentence should be: This model is better than this other in measuring a specific property in a specific sample using a specific instrumental device and under specific instrumental and environmental conditions. Generalization has the risk of giving the wrong impression that certain models will always overcome other models, and that is not true. We work on local situations unless, as I said, we ensure that our data **X** and reference values **Y** are representative enough of the problem.

DEFINITIONS

Having perfectly understood (I hope) that three actors are playing an important role in multivariate data analysis, now is the time to define the terms that bring us to this manuscript. It is also the moment to think back on everything that has been said before and put it into the framework of the definition of each term.

Data Mining

Data Mining can be defined as a set of methods used to extract usable information from large raw data sets. It should be noticed that this definition implies that the usable information is already in the data. Nevertheless, the complexity of the data and the multivariate (multiway) nature of the data means that we are unable to find useful information without powerful mathematical tools. Basically, data mining aims to separate the grain from the hay or find patterns that already exist in the data; however, they are hidden due to a large number of samples and variables, the noise of the data, or the difficulty in linking more than two variables at the same time in a univariate fashion (one variable at a time).

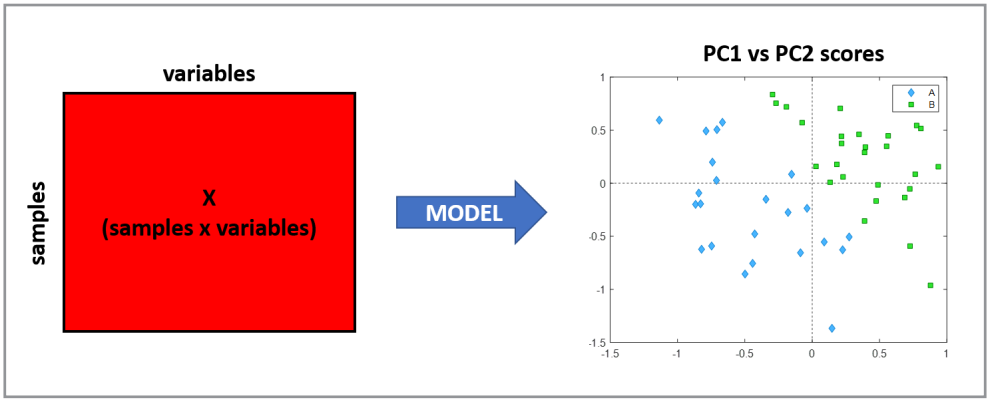


Figure 3. Data Mining applied to a matrix X. A PCA model has been applied, and the score scatter plot between PC1 and PC2 is shown.

One important fact of the Data Mining definition is that it does **NOT** include any procedure of Learning. Therefore, the methods or algorithms used for data mining are of an **unsupervised nature** (Figure 3). This narrows down the choices since any algorithm that involves Learning (i.e. supervision) should be kept aside.

DATA MINING = UNSUPERVISED

Under the umbrella of Data Mining, we can find methods like dendrograms, clustering or PCA (among others). One thing that must be clear is that Data Mining methods are not, *per se*, classification methods. In the literature, we can find works using, for instance, PCA for classification purposes. This might be a common language error since we tend to mention that PCA has classified the samples into some groups when we explain the results of a PCA model (it happens to me quite often). Nevertheless, the word "classify" and the word "group" have two separated meanings. Classification implies supervision (as we will see further). One general action, however, could be applying PCA or dendrograms and then setting some thresholds to group the samples into different classes. In any case, setting the threshold is an operation made *a posteriori* in an attempt to group samples.

Machine Learning

Machine Learning is normally defined as a series of methods that learn from the data to make or construct a model that can make informed decisions based on what is learned. This definition directly implies that the model/algorithm needs to learn. The learning procedure makes the algorithm reliable enough to predict any property in new data that has not been used for Learning. Here is where the supervised methodologies must be used.

MACHINE LEARNING = SUPERVISED

Indeed, the step of Learning could be substituted by the word Training. Therefore, any Machine Learning algorithm (but really, any) is composed of two steps: calibrating and testing the model, that is, the learning (training) step and applying the model to new samples (Figure 4).

Calibrating the model

Given a calibration data matrix **Xcal** and a reference value (in Regression or Classification) **Ycal**, the calibration step involves all of the necessary operations of data pre-processing, data normalization, variable selection, and the removal of outliers, among others, with the only purpose of finding the best conditions for obtaining the highest correlation (co-variance) between **Xcal** and **Ycal** (Figure 4a). That is as simple and as complicated as it appears. Once the optimal conditions have been found, the model is

set, and it could be used to predict whatever property is needed. Nevertheless, this step of calibration is completely useless without the next step, testing the model.

Testing (validating) the model

One of the main drawbacks of all multivariate models is that, since they are based on projection operations and are not parameterized, they need to be tested. Testing a model is used to verify that the model created in the calibration step can predict the outcome.

A GOOD MODEL IS NOT THE ONE THAT BETTER CALIBRATES, BUT THE ONE THAT BETTER PREDICTS. WITHOUT VALIDATION, THE MODEL IS COMPLETELY USELESS. VALIDATE YOUR MODELS!

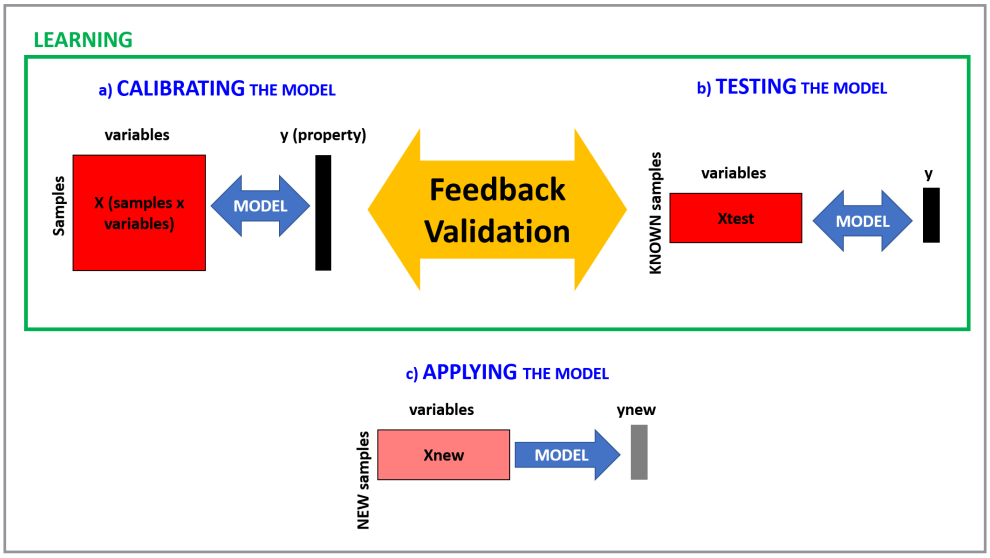


Figure 4. Machine Learning procedure showing the three stages of a) calibration, b) testing validation and c) prediction.

Let us create a matrix **Xcal** (20 x 200) of observations and a **Y** matrix **Ycal** (20 x 1) of a property that we want to measure, both of them composed of normally distributed random numbers. Now, let us apply a common multivariate regression model like Partial Least Squares (PLS). One of the key points of optimizing a PLS model is to ascertain the proper number of Latent Variables (LVs) [10]. This number can be optimized by following how the error (RMSE) and the R2 change when more LVs are included in the models. The results are represented in black in Figure 5. As can be observed, the error of the model drastically decreases with the number of LVs (Figure 5a) to such a point that the error obtained with 4 LVs is zero. Observing the predicted values obtained for different LVs (black lines in Figure 5b-d), it is clear that a perfect regression model is obtained at 4 LVs with the error at zero value and the R2 at 1. Even being the perfect calibration model, it is completely useless.

Let us create another matrix **Xval** (10 x 200) of observations and a **Y** matrix **Yval** (10 x 1) of a known property. Those two matrices are also created with normally distributed random numbers, so we are completely sure that the calibration and the validation matrices span the same space. That is, we can use the calibration models created previously to predict a **val** in such a way that we can calculate the difference between the obtained **val** and the known **Yval**. The results are displayed in red in Figure 5.

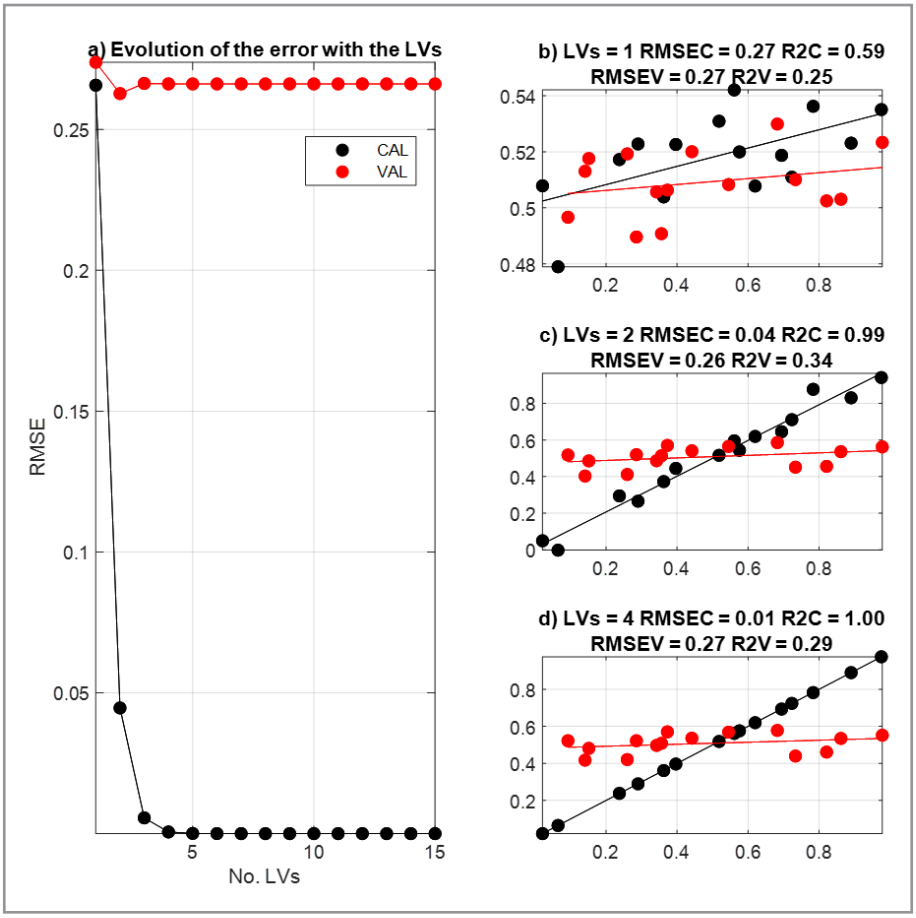


Figure 5. Applying PLS to quantify random Y values from random X values.

Having a completely perfect calibration model does not mean, at all, that this model is optimal for predicting external samples. As shown in Figure 5, it does not matter how many LVs the calibration model contains. What matters is that the real error of my model, regardless of the number of LVs, is around 0.25. This error, compared with the mean and standard deviation values of **Xval** (0.5 and 0.25, respectively), indicates that the model is predicting new samples giving values of **val** like a completely random number. The step of validation/testing the model is crucial to ensure that the model is actually working properly. There are many strategies for validating models (being cross-validation and external test set, the most common ones) [11]. It takes some time to match the most adequate validation methodology since it also involves checking the performance with different pre-processing methods, variable selection, etc. However, one thing is certain; it is necessary to:

VALIDATE YOUR MODELS!

Once the model is properly validated, we can be sure that it will be able to predict external samples with the accuracy and reliability of the validation step (Figure 4c).

Semi-supervised models

There is a family of models that falls between unsupervised and supervised modelling. They are the semi-supervised models. Semi-supervised modelling is a general approach that combines a small set of well-known labelled data (well-known class belonging) with a relatively large amount of unlabeled data (data without a pre-assumption of the class). They are especially useful when labelling data is instrumentally difficult or expensive. Semi-supervised models are normally used in classification scenarios to profit the

similarity between samples in the unsupervised data with the well-labelled samples of the supervised data. In this manner, several assumptions can be made about the belonging of the unsupervised data to different classes reflected in the supervised data. This is done by establishing more or less complicated boundaries to create clusters based on different distances between the samples in the variable space.

Deep Learning

A sub-set of Machine Learning methods is comprised of the Deep Learning (DL) algorithms. Deep Learning algorithms are also a sub-set of the well-known artificial neural networks (ANN) when the usage of multilayer structures (hidden layers) is preferred since they can handle more than one problem at the same time to give a unique answer [12]. Deep Learning algorithms are mostly based on the well-known Deep Neural Networks (DNN) and Convolutional Neural Networks (CNN). ANN and CNN have a basic structure of inputs (the data matrix **X**), hidden layers composed of the so-called neurons and an output layer of responses. As said before, the main difference between DL networks and ANN is the complexity of the connection between the hidden layers (Figure 6). This complexity in the connections allows the feature extraction from the raw data independently, without pre-processing or pre-arranging it.

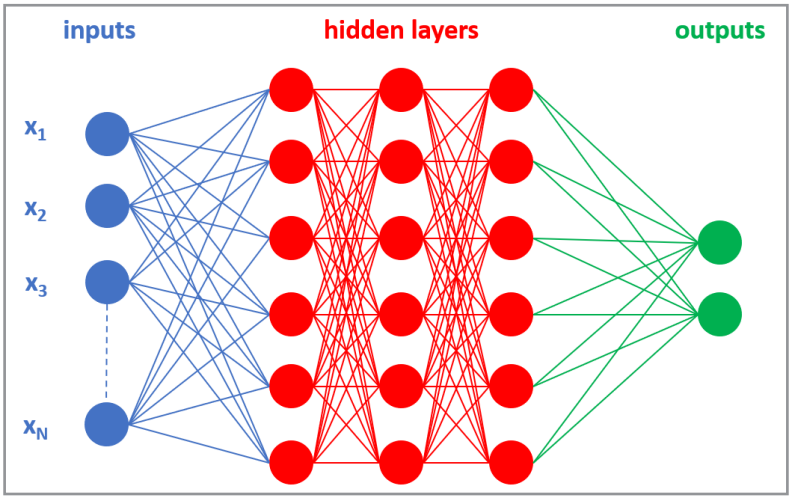


Figure 6. An example of the architecture of a Deep Convolutional Neural Network with three hidden layers.

This statement, as incredible as it appears, is not exempt from drawbacks. It should be remembered that whatever information or feature the Neural Network can extract from the raw data **MUST BE CONTAINED** in the data. The DNN learns from the data. Therefore, DNNs need an overwhelming amount of data (very comprehensive databases) in order to be able to look for the proper information and solve highly non-linear issues. Besides, due to the complexity in the structure and increasing the hidden layers, the validation (Learning) needs to be extensively and exhaustively performed to ensure that the network is not overfitting the solution. DNN trend to overfit since they are designed to model the minimum amount of variance/co-variance in the data.

Artificial Intelligence

Strictly speaking, the definition of Artificial Intelligence is the intelligence demonstrated by machines. Translated into our scenario, Artificial Intelligence is the umbrella that covers all of the previous definitions (Figure 7); that is, the application of Machine Learning, Data Mining and Deep Learning to data. Without entering into more detail, Artificial Intelligence also covers the possibility that the algorithms will be able, in the near future, to perform logical reasoning and interaction in order to improve the outcomes of the models.

However, as always, the major bottleneck in this is the availability of valid data. In a straight analogy, I always say that the most powerful machine that applies Artificial Intelligence continuously is the human brain. The human brain can analyze the analytical information that is continuously received from the analytical instruments (eye, ear, touch, taste and smell) and, in a complex procedure, offer answers or responses that will be as accurate as the information stored in the database (the memory). Together with the senses, the human brain works so well because it is continuously being trained with information and learning procedures. We normally call it education. In data analysis, this is called Training (a.k.a. Learning).

WHERE IS CHEMOMETRICS?

The term Chemometrics was coined in 1972 by Svante Wold and Bruce Kowalsky. The most accepted definition of Chemometrics refers to the chemical discipline that uses mathematical, statistical, and other derived methods employing formal logic to (a) design or select optimal measurement procedures and experiments and (b) provide maximum relevant chemical information by analyzing chemical data [3,13,14].

Another definition could be that Chemometrics is the application of Artificial Intelligence (therefore, Data Mining, Machine Learning, Artificial Neural Networks, and Deep Learning) to data coming from Natural Systems (Figure 7). Therefore, it turns out that we have been talking about Chemometrics from the very first line of this manuscript.

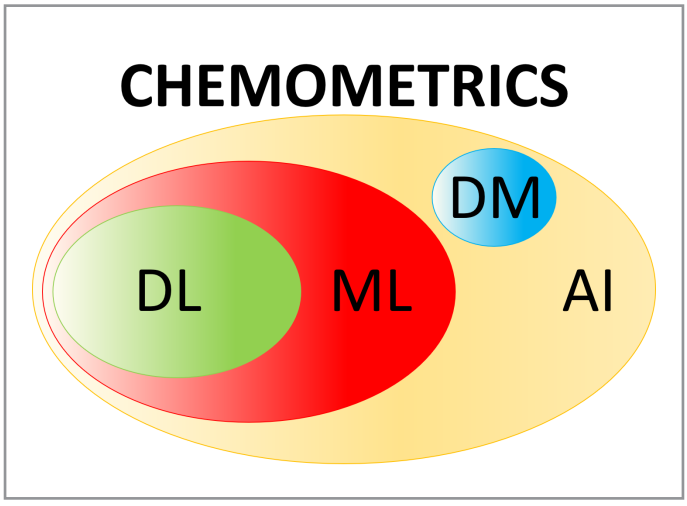


Figure 7. Chemometrics in perspective. AI, Artificial Intelligence; DM, Data Mining; ML, Machine Learning; DL, Deep Learning.

The next question would be what makes us, Chemometricians, different from Machine Learning. The straight answer would be nothing. But this statement might be too straight. As I just said, Machine Learning is part of Chemometrics. Nevertheless, I think that pertinent explanations should be given from a different perspective. Let me say that a Scientist is a person that collects and analyzes data. Therefore, the need to analyze data has been always there. For instance, PCA was proposed at the beginning of the 20th century [15,16], more than 100 years ago. Nevertheless, 100 years ago, there were not the computing capabilities that we have nowadays. We also did not have the advanced analytical instruments that we have in our laboratories or the data storage capabilities that we have now. Nevertheless, there were researchers that needed to extract complex relationships from matrices. This was not exclusively done in the field of Analytical Chemistry, but in all scientific fields. It then merged the -metrics scenarios creating new concepts like Psychometrics [17], Environmetrics [18], Econometrics [19], etc. All of them have one thing in common: the use of mathematical tools to extract relevant information from the data. Nevertheless, all of them have a strong difference: **THE DATA**.

The data coming from the previously mentioned disciplines might have certain similarities, and it could be analyzed with the same methodologies. Nevertheless, the nature, precedence and behavior of the data

are specific to the scientific scenario where the data are obtained. For example, in Chemometrics and Environmetrics, we can study temporal trends to understand the pollution of a river, as, in Econometrics, they can study temporal trends to understand the evolution of certain values in the stock market. The great difference is that a Chemometrician will better understand the environmental data than an Econometrician, and vice-versa.

DATA ARE NOT JUST DATA. THEY HAVE AN ORIGIN, A MEANING, AND A STRUCTURE THAT MUST BE FULLY UNDERSTOOD TO INTERPRET THE OUTCOMES FROM DATA ANALYSIS

Nevertheless, nowadays, there is a dangerous trend to disparage all -metric words favoring the "Data Science" words. I am a bit worried about the vast amplitude of the words "data science" and the little importance that sometimes we give to the word data. Sometimes it seems that the more programming skills you have, the better a data scientist you are. Also, we sometimes forget the "data" on the road.

ONLY BY UNDERSTANDING THE SCIENTIFIC PROBLEM AND THE COMPLEXITY OF THE DATA, WILL WE BE ABLE TO CHOOSE A PROPER DATA ANALYSIS METHODOLOGY, IF IT IS NEEDED. MACHINE LEARNING, DATA MINING, DEEP LEARNING, AND CHEMOMETRICS ARE COMPLETELY USELESS IF WE DO NOT GIVE THE DATA THE DESERVED ATTENTION

TAKE-HOME MESSAGES ABOUT HOW TO APPLY CHEMOMETRICS

During the development of this manuscript, I have included some of the references that I consider a "must-read" to understand what Chemometrics can do in your data and the benefits and major pitfalls of applying Chemometrics in analytical data [20]. To finalize this "short" manuscript, I want to give you some take-home messages about different global aspects that you should consider when applying Chemometrics to your data. They can be seen as a summary of the main concepts and statements made before.

- **A good outcome might not necessarily be a perfect model**. Always remember the GI-GO truism. Before doing anything, check the quality and information that your data can provide, as well as the purpose of the data and how it was obtained. Otherwise, we could finish with any of the following issues:
  a) Wrong hypothesis: Hoping that Chemometrics can find what is not in the data.
  b) Wrong design of the experiment: Hoping that the measurements contain variation that is not contemplated in the design of the experiment.
  c) Overuse of resources: Measuring in one instrument just because it is available in the laboratory.

  There might be temptations to use mathematical models to correct poorly designed experiments or even experimental problems with the data. Even though it might be plausible and useful in some situations, this might not be the best option since biasing the models to correct artifacts that could be easily corrected beforehand might lead to misinterpretations and underfitting situations in the prediction or interpretation of new data with the biased model.

- **Maybe you do not need fancy algorithms to solve the issue. Maybe you do not need an algorithm at all to solve the issue.** Do not even think of having data to apply Chemometrics! Think about whether your problem, with the data you have, requires Chemometrics. Think about the structure of your data, its quality, and the quality of the reference you will use. Sometimes, the reference has an associated error that is larger than the error of the data.

- **Question everything:** Chemometrics does not give you absolute answers. Chemometrics gives you validated answers in local scenarios. How big and comprehensive is the "local" scenario? As big and comprehensive as your data (database). Sentences like "This algorithm is better than this one" are only true in the conditions where those algorithms were tested. Also, you should be careful with "rules of thumb". For example, "the minimum number of PCs is set with PCs whose eigenvalue is larger than 1".

That is not true. There are situations where you will need to check into PCs whose eigenvalue is smaller than 1 but still explain interesting sources of variability in your data.

- **The Illuminati of the software:** Use the software that you can/want/like, but please, use it right. Also, always verify that the software actually does what it claims to do. This goes for commercial Machine Learning packages and for functions/libraries that you can use in software like R, Python or Matlab (among others, of course).

- **The Seven Commandments in Chemometrics:** I always tell my students that Chemometrics (and many aspects of our daily life) is mostly based on the following Commandments:
  1. Think
  2. Be patient
  3. Know your data and your goal
  4. Keep it simple
  5. VALIDATE
  6. Question everything
  7. The chemistry/physics prevails over the algorithm

FINAL COMMENTS

When applying Chemometrics to my data, I normally follow my own standardized protocol, which implies the following steps:

1) Apply the parsimony principle (Occam's razor). The simpler model, the better.
2) Check the raw data. That means to understand the data, plot the data in different manners.
3) Think about plausible artifacts and methods to solve/minimize/avoid them (pre-processing).
4) Start with the simplest methods, and draw figures. If you are using projections methods, plot the score and loading plots, regression lines, etc.
5) Be careful with assessing that a sample is an outlier.
   - In explorative/unsupervised scenarios, consider outliers as extreme samples that do not necessarily have the wrong samples.
   - In Regression and classification, check the proper figures of merit of the model. Using an exploratory method, for instance, PCA, might lead to issues when removing samples whose variance is different than expected, but their co-variance with the property you want to measure is correct.
6) GO TO LINEAR MODELS! At least, at the beginning. If the issues cannot be fixed with linear models, go step by step, increasing the complexity of the model.
7) Optimize your models. Iterate from step 1 until step 5, checking plausible pre-processings, variable selections, etc.
8) **VALIDATE**. A good model is not the one that best calibrates but the one that best predicts.
   - Test your model with the own data (cross-validation).
   - Test your model with completely external data (test, external prediction, etc.).
   - Tune/change/re-parameterize your model if needed (go from Step 2 to step 6, optimizing the figures of merit).
9) Test all plausible (and coherent) combinations of pre-processing, normalization, and variable selection methods applied to the models and then, if needed, re-parameterize the model.
10) Save all of the results. It is extremely important to keep track of your steps and to understand what is wrong or right in the obtained models.

CALL IT MACHINE LEARNING, CHEMOMETRICS OR WHATEVER YOU WANT. USE WHATEVER SOFTWARE YOU WANT OR NEED. BUT PLEASE, ALWAYS BE AWARE OF WHAT YOU ARE DOING TO YOUR DATA AND WHY YOU ARE DOING IT. AND, OF COURSE, VALIDATE YOUR MODELS!

REFERENCES

1. Amigo, J. M.; Skov, T.; Bro, R. *Chem. Rev*., **2010**, *110*, pp 4582–4605 (https://doi.org/10.1021/cr900394n).
2. Amigo, J. M. *Hyperspectral Imaging*, Elsevier, **2019**.
3. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verberke, J. (Eds.). *Handbook of Chemometrics and Qualimetrics: Part A*, **1997** (https://doi.org/10.1016/S0922-3487(97)80056-1).
4. Bro, R. *Chemom. Intell. Lab. Syst*., **1997**, *38* (2), pp 149–171 (https://doi.org/10.1016/S0169-7439(97)00032-4).
5. Bro, R. *Crit. Rev. Anal. Chem*., **2006**, *36* (3-4), pp 279–293 (https://doi.org/10.1080/10408340600969965).
6. Leardi, R. *Anal. Chim. Acta*, **2009**, *652*, pp 161–172 (https://doi.org/10.1016/j.aca.2009.06.015).
7. Ballabio, D.; Consonni, V. *Anal. Methods*, **2013**, *5*, pp 3790-3798 (https://doi.org/10.1039/c3ay40582f).
8. Wold, S.; Esbensen, K.; Geladi, P. *Chemom. Intell. Lab. Syst*., **1987**, *2* (1-3), pp 37-52 (https://doi.org/10.1016/0169-7439(87)80084-9).
9. Smilde, A. K.; Bro, R. *Anal. Methods*, **2014**, *6*, pp 2812–2831 (https://doi.org/10.1039/c3ay41907j).
10. Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst*., **2001**, *58* (2), pp 109–130 (https://doi.org/10.1016/S0169-7439(01)00155-1).
11. Westad, F.; Marini, F. *Anal. Chim. Acta*, **2015**, *893*, pp 14–24 (https://doi.org/10.1016/j.aca.2015.06.056).
12. Marini, F.; Bucci, R.; Magrì, A. L.; Magrì, A. D. *Microchem. J*., **2008**, *88* (2), pp 178-185 (https://doi.org/10.1016/j.microc.2007.11.008).
13. Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.; Walczak, B.; Tauler, R. *Anal. Bioanal. Chem*., **2017**, *409*, pp 5891-5899 (https://doi.org/10.1007/s00216-017-0517-1).
14. Brereton, R. G.; Jansen, J.; Lopes, J.; Marini, F.; Pomerantsev, A.; Rodionova, O.; Roger, J. M.; Walczak, B.; Tauler, R. *Anal. Bioanal. Chem*., **2018**, *410*, pp 6691-6704 (https://doi.org/10.1007/s00216-018-1283-4).
15. Pearson, K. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **1901**, *2* (11), pp 559-572 (https://doi.org/10.1080/14786440109462720).
16. Hotelling, H. *Journal of Educational Psychology*, **1933**, *24* (6), pp 471-441 (https://doi.org/10.1037/h0071325).
17. Borsboom, D. *Psychometrika*, **2006**, *71*, Article Nº 425 (https://doi.org/10.1007/s11336-006-1447-6).
18. Hunter, J. S. 1 Environmetrics: An emerging science. In: *Handbook of Statistics*. Elsevier, **1994**, *12*, pp 1-7 (https://doi.org/10.1016/S0169-7161(05)80003-3).
19. Farebrother, R. W.; Common, M. S. *Journal of the Royal Statistical Society. Series A (General)*, **1978**, *141* (3), pp 417-418 (https://doi.org/10.2307/2344828).
20. Kjeldahl, K.; Bro, R. *J. Chemom*., **2010**, *24* (7-8), pp 558-564 (https://doi.org/10.1002/cem.1346).

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 62-77
doi: 10.30744/brjac.2179-3425.AR-09–2021

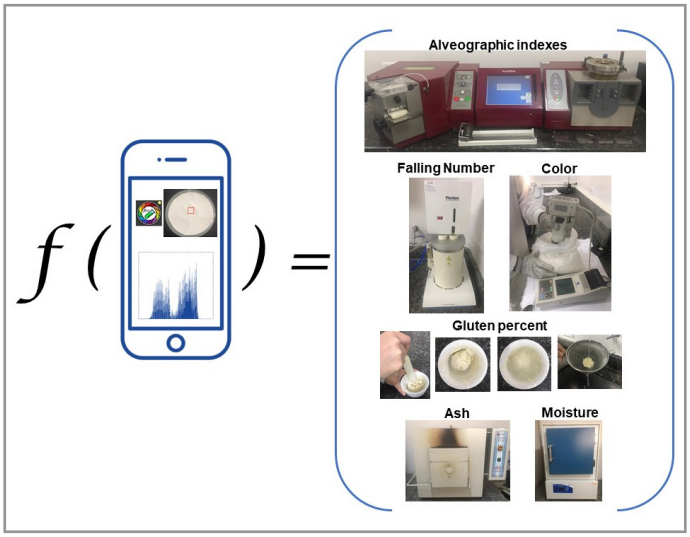
BrJAC

ARTICLE

Digital Images and Multivariate Calibration in the Determination of Rheological and Quality Parameters of Wheat Flour

Leticia Darlla Cordeiro, Patrícia Valderrama*

Universidade Tecnológica Federal do Paraná (UTFPR), Via Rosalina Maria dos Santos, 1233, Postal Code 87301-899, Campo Mourão, PR, Brazil



In this work, the quality and rheological parameters such as moisture, ash, color, falling number, gluten, and alveographic indexes (W, P/L, and I/E) were determined in wheat flour samples provided from the industrial process through digital images and multivariate calibration, based on partial least squares (PLS) regression. The models were validated by the evaluation of the figures of merit such as accuracy, the inverse of analytical sensitivity, adjust, linearity, relative prediction deviation, limits of detection, and quantification, indicating that the proposal is feasible and can be used in the industrial routine analysis as an alternative to the methods highly dependent on the analyst perception, and extremally time-consuming.

Keywords: gluten; falling number; alveographic indexes; PLS; validation.

INTRODUCTION

Wheat flour is a food matrix that containing proteins, carbohydrates, amino acids, dietary fiber, fat, water, minerals, vitamins [1], and it is the basic constituent, for example, of bread, pastries, biscuits, cakes, and pasta [2].

This food matrix arouses interest from the analytical viewpoint, based on legislation and other situations, as fraud, and contamination. According to the Brazilian legislation the Normative Instruction nº 8 (Technical Regulation of Wheat Flour Identity and Quality) of Agriculture, Livestock and Supply Ministry, the moisture content of wheat flour must be up to 15%, being wheat flour classified as Type 1, Type 2 and whole, according to the maximum ash content of 0.8%, 1.4% and 2.5% and protein content of at least 7.5%, 8.0%, and 8.0%, respectively [3].

Cite: Cordeiro, L. D.; Valderrama, P. Digital Images and Multivariate Calibration in the Determination of Rheological and Quality Parameters of Wheat Flour. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 62–77. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-09-2021

Submitted 16 January 2021, Resubmitted 14 April 2021, Accepted 22 April 2021, Available online 19 May 2021.

For quality and identity parameters, the traditional methods for moisture, color, ash, gluten percent, falling number, and alveographic indexes are based on the American Association of Cereal Chemists (AACC) [4]. On the other hand, it is possible to highlighting some alternative methods such as in the simultaneous determination of ash content and protein by using near-infrared (NIR) reflection spectroscopy, and diffuse reflectance infrared Fourier Transform (FT) spectroscopy [5,6]; in the determination of total protein and wet gluten [7], total phenolic content [8], fatty acid [9], and some other quality and rheological parameters [10] by using NIR spectroscopy; to predict the geographic origin of wheat flour by NIR spectroscopy [11]; for prediction of deoxynivalenol (DON) contamination [12], in the discrimination of talcum powder and benzoyl peroxide in wheat flour [13], and to detect low-level peanut powder contamination of whole wheat flour [14] by NIR hyperspectral imaging; in the quantification of total phenolics and ferulic acid using ultraviolet and visible (UV–Vis) spectrophotometry [15]; for quantification of gluten by FT-Raman spectroscopy [16]; for quantitative detection of adulterants with Raman hyperspectral imaging [17]; to investigate the correlations existing between the GlutoPeak indices and the conventional rheological parameters [2,18].

Notwithstanding some analytical techniques employed in the cited examples are expensive (as hyperspectral imaging) or demand sample preparation (as UV-Vis spectrophotometry). Furthermore, even with the relative cheapness of the NIR spectroscopy (especially the hand-held equipment), a digital image based on a smartphone it is even cheaper and no requires sample preparation, became possible the raw sample evaluation provided from the process.

The color of food is the first quality parameter evaluated by consumers and is thus a critical factor for acceptance of the food item by the consumer [19]. Concerning the wheat flour color, it can be influenced, for example, for wheat variety, aging, milling practices, and bleaching, in which the color ranging from brownish-grey to creamy yellow to the whitest white. The wheat flour color reflects the chemical composition and processing history of each flour batch influencing nutritional value and safety [20]. On the other hand, the color of any pixel of the image of an object can be registered using three color sensors per pixel through a digital camera. In this sense, each sensor captures the intensity of the light in the red (R), green (G), or blue (B) spectrum, respectively, is the often-used color model [19]. The variables R, G, and B can assume values between 0 and 255. Then, each sample can generate a vector with dimensions 1 × 768 (256 possible values for R, G, and B variables placed side by side in that order) [21]. These signals are then used as descriptor variables in multivariate calibration models and were applied in different food matrix evaluation such as pesto sauce [22], commercial carbonated soft drinks [23,24], ripening bananas [25], potato chips [26], freeze-dried açai [27], and grape juice [28]. Regarding wheat flour, to the best of our knowledge, was not found paper that describes the application of image analysis coupled with multivariate calibration for evaluation of rheological and quality parameters in this matrix. Due to this, this work aimed to investigate digital images obtained from an iPhone for the determination of moisture, color, ash, gluten percent, falling number, and alveographic indexes in genuine samples of wheat flour coming from a process of a milling wheat industry.

MATERIALS AND METHODS

Samples

A total of 100 genuine wheat flour samples was utilized in this work. The samples were from a wheat flour industry located in the Paraná state, Brazil, and manufactured by this industry from December 2019 to January 2020. All the samples were genuine from the industrial process.

Multivariate calibration

A multivariate calibration through partial least squares (PLS) regression is proposed to correlate digital images with rheological and quality parameters: moisture (%), color (L*), ash (%), gluten percent (%), Falling Number (s), and alveographic indexes such as W (that represent the gluten ‘strength’ and is given by mechanical work, 10^{-4} J), P/L (the tenacity/extensibility ratio which expresses the mass balance, mm), and IE (elasticity index, %).

The PLS' mathematical descriptions are presented in several manuscripts [29,30] and not detailed here. For the PLS model development, the **X** matrix (digital images – **X** block) was correlated to a **y** vector (**y** block), which contained the rheological and quality parameters determined by the reference methods in a PLS1 correlation (Details concerning **X** and **y** blocks below). The models were performed in MATLAB R2007B (The MathWorks Inc., Natick, USA) and PLS-Toolbox 5.2.

A total of 75 samples were employed in the calibration step, whereas 25 samples were used in the external validation step, all of them selected by the *kenston* algorithm [31]. Data were mean-centered, and the number of latent variables (LVs) was chosen based on the Root Mean Square Error of Cross-Validation (RMSECV) through continuous blocks of 5 samples. The variance explained in the **y** block was also considered on chosen the number of LVs. It was as well considered the outliers evaluation based on leverage, unmodeled residuals in **X**, and **y** blocks [32-35].

PLS models were validated by the determination of the parameters of merit such as accuracy, fit, linearity, residual prediction deviation, the inverse of analytical sensitivity, and limits of detection and quantification [35,36], according to the equations presented in Table I.

Table I. Equations for parameters of merit

Parameters	Equations
Accuracy	$\mathrm{RMSEC} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n_c - nVL}}$ and $\mathrm{RMSEP} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n_v}}$
Sensitivity	$SEN = \frac{1}{\|\|\mathbf{b}\|\|}$
Analytical sensitivity	$SEN\_A = \frac{SEN}{\delta x}$
Analytical sensitivity^{-1}	$SEN\_A^{-1} = \frac{1}{SEN\_A}$
Limit of detection	$LD = 3.3\ \delta x \|\|\mathbf{b}\|\| = 3.3\ \delta x \frac{1}{SEN}$
Limit of quantification	$LQ = 10\ \delta x \|\|\mathbf{b}\|\| = 10\ \delta x \frac{1}{SEN}$
RPD_{cal}	$RPD_{cal} = \frac{DP_{cal}}{RMSECV}$

y_i is the reference value of sample *i*; $\hat{y}_i$ is the predicted value of sample *i*; n_c is the number of samples in the calibration set; n_v is the number of samples in the external validation set; nVL is the number of latent variables. **b** is the regression coefficients vector; δx is estimative for the instrumental noise; DP_{cal} is the standard deviation of reference values in the calibration set; RMSECV is the root mean square error of cross-validation; Observation: In the equation for RMSEC, it is employed "n_c - nVL + 1" when the data are mean-centered.

*Images acquire (**X** block)*

An iPhone 7 Plus (Apple Inc.) with 12 megapixels resolution, automatic High Dynamic Range (HDR), without flash, was used to capture the images at the moment when the samples arrived from the industrial process to perform the reference analysis on the industrial laboratory.

The distance from the camera to the wheat flour sample (20 g in a petri dish with 5 cm diameter) was 20 cm and the smartphone was kept in fixed support. The same lighting conditions (2 led lamps with 20 watts each in a 32 m^2 environment, temperature around 20 °C +/- 3 °C) were kept during image acquisition.

The images were captured by the PhotoMetrix® [37] app in the form of the histogram whose generates a response vector related to the channels R (red), G (green), and B (blue). A total of six replicates were done for each sample, and a mean was used for the multivariate model development. So, the R, G, and B vectors (in which for each vector the dimensions were 1 × 768, i.e., 256 possible values for R, G, and B variables placed side by side in that order) were organized into a matrix (100x768).

*Reference analysis (**y** block)*

The reference methods for rheological and quality parameters determination were according to the American Association of Cereal Chemists (AACC) [4] and summarized in Table II. It is important to highlight the time dispended in the analysis of some parameters, as well as high dependence on the analyst perception.

Table II. Reference methods for rheological and quality parameters of wheat flour

Parameters	Description
Moisture	Moisture content followed the AACC 44-15.02 method, where the samples were dehydrated in an oven at 130 °C +/- 2 °C, until constant weight. Analysis time around 4 hours.
Color	Luminosity (L*) followed the AACC 14-22.01 method by using a Minolta colorimeter equipment.
Ash	Ash followed the AACC 08-02.01 method where the samples were subjected to the dry matter combustion in a muffle at 600 °C +/- 5 °C. Analysis time = 5 hours.
Gluten percent	Gluten percent follower the AACC 38-12.02 method where the wheat flour sample (5 g) is weighed and mixed with 10 mL of 5% aqueous sodium chloride solution. The dough was rest for 30 minutes, and then, water was added until it was covered and the dough rest again for another 30 minutes. The agglomerate obtained is washed with water over a 100-mesh sieve, pressing lightly with hands. Washing continued until the water was no longer whitish. The remaining dough was weighed, and gluten percentage calculated by [gluten (%) = dry gluten mass (g).100/sample mass (g)]. Analysis time = 60 minutes.
Falling Number (FN)	FN followed the AACC 56-81B method and the result is obtained utilizing an instrument (PerkinElmer model 1310), based on principles of viscosimetry, that determines the amylolytic activity of the wheat flour. Analysis time = 15 minutes.
Alveographic indexes	W, P/L, and IE followed the AACC 54-30.02 method by using an Alveograph Chopin equipment (model 171), where 250 g of wheat flour sample were weighed and 135 ml of 2.5% saline solution was added. The dough was homogenized for 8 minutes. The total time from the start to the end of alveographic analysis is around 60 minutes.

RESULTS AND DISCUSSION

The R, G, B vectors obtained for the wheat flour samples are shown in Figure 1(A). For each vector, 768 variables were obtained (i.e., 256 possible values for R, G, and B placed side by side in that order). The R and G regions were removed due to in this region no relevant information is present (i.e., zero intensity). Figure 1(B) presents the B vectors of wheat flour samples, that were organized into a matrix (100x256) and used to build the models.

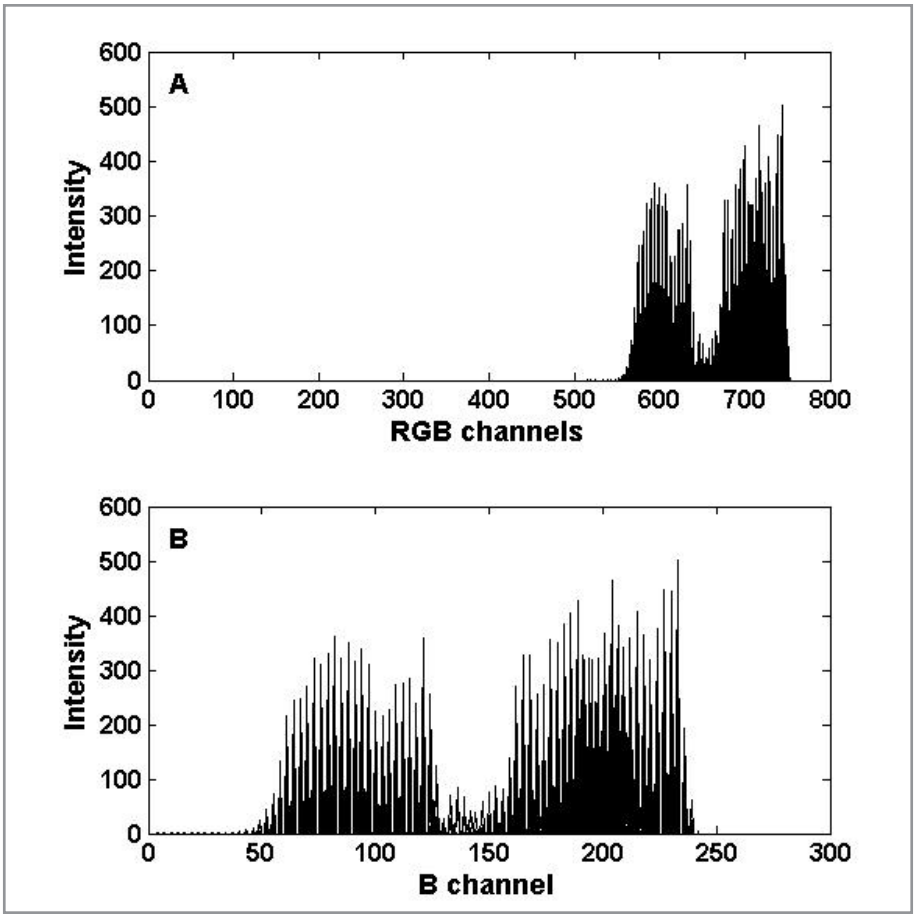


Figure 1. R, G, B vectors of wheat flour samples.

After splitting the data into calibration and external validation samples, in order to improve the model's quality, the first step in model development was outlier evaluation. Outliers were evaluated only one round and removed based on leverage, unmodeled residuals in **X**, and **y** blocks.

The optimum model dimension necessary to retain a significant variance in the data resulted in the number of LVs and parameters of merit present in Table III.

The Root Mean Squared Error of Calibration (RMSEC) and Prediction (RMSEP) are accuracy indicators. Those values report the closeness of agreement between the reference value (obtained by reference analysis – Table 1S in the supplementary material) and the value found by the PLS model. Despite the high number of LVs in some cases, the RMSEC and RMSEP values are also indicators of the properly chosen, since these parameters presented close results in each model [6,38]. Also, some modeled parameters such as gluten present a series of subjective steps, which can contribute to this demand for a higher number of LVs.

The achieved accuracy results for FN, W, P/L, and gluten are lower than those obtained through models based on NIR/PLS [7,10,16]. On the other hand, related accuracy was obtained for ash in this proposal and on a model based on MIR/PLS [5].

Another way to verify the models' accuracy is the adjust [32,33,35], represented by the plot of the rheological and quality parameters determined by reference analysis against those determined by the image/PLS model (Figure 2, and correlation coefficient in Table I). Comparing to the literature [6,7], the results for the correlation coefficient achieved with digital images/PLS are similar and higher when compared to models developed from NIR/PLS to determine ash and gluten, respectively in wheat flour.

The steps in PLS model development become the inverse of analytical sensitivity (analytical sensitivity^{-1}) more suitable for evaluating the sensitivity of this multivariate calibration model. This parameter of merit allows the establishment of a minimum difference, which is discernible by the model in the modeled range of the interesting property [33,35,36]. Based on this, it is possible, to distinguish wheat flour samples with a difference of 0.1140%, 0.2775 (L*), 0.0395%, 3.1409 s, 0.4270%, 11.2570x10^{-4} J, 0.1580 mm, and 0.6780% for moisture, color, ash, FN, gluten, W, P/L, and IE, respectively.

Limits of detection and quantification for the PLS models show coherent results with the measured quantities (modeled range) and the achieved accuracy. Therefore, the PLS models are appropriate, for example, to detect until around 0.38% and quantify 1.14% of moisture in wheat flour, and these results are adequate for this quality parameter since it must be up to 15%, according to the Brazilian legislation [3], and the modeled range for moisture was from 12.40 to 14.40%. So, the model presents low limits of detection and quantification.

The predictive capacity of the PLS models was evaluated through the residual prediction deviation (RPD), which results from 1.5 to 2.4 are considered satisfactory [39]. In the evaluation of the RPD values for the calibration sets, all the models presented results inside that range. These results suggest that PLS models have an excellent forecasting capability, which is fundamental when the multivariate models can be applied to quality and process control [32].

The models' linearity was evaluated by the residuals plot (Figure 1S in Supplementary material). The Jarque-Bera statistical test was employed to confirm the linearity of the developed models. In this test, when the JBSTAT values are lower than those of CRITICAL, it is considered that the residues exhibit a random behavior at a 95% significance, indicating linearity for the PLS models [32,33]. For the developed models all of them presented JBSTAT values inferior to those of CRITVAL confirming that the errors present a random behavior and that the data fit on a linear model as PLS. Furthermore, to identify the most significant variables in the B channel that contribute for the model's development, the variable importance in projection (VIP) scores were calculated (Figure 2S in Supplementary material). The VIP scores are computed based on the scores samples, regression coefficients, and weight of the variables, considering the number of LVs involved in the model. VIP scores result greater than one is generally used as a criterion to identify the most significant variables [40]. In general, the B channel variables around 55 – 120, and 155 – 240 were important for most of the models.

The results obtained for the parameters of merit showed that the proposed method based on image/PLS can be feasible. Furthermore, this approach can contribute to the wheat flour industry especially in terms of quickness, analysis time, besides cost reduction.

Table III. Parameters of merit

Parameters of merit		Moisture (%)	Color (L*)	Ash (%)	FN (s)	Gluten (%)	W (10^{-4} J)	P/L (mm)	IE (%)
Accuracy	RMSEC	0.2217	0.4862	0.0563	5.6754	0.6783	20.7134	0.2369	1.3364
	RMSEP	0.2802	0.5090	0.0718	5.3664	0.7194	26.1266	0.2923	1.2197
Correlation coefficient		0.7738	0.7615	0.7833	0.7442	0.7823	0.7276	0.7653	0.7636
Number of LVs		12	11	13	10	20	10	10	11
Range modeled		12.40-14.40	90.23-93.40	0.40-0.84	337-386	24.72-31.01	228-386	0.51-2.98	50.87-62.40
Analytical sensitivity^{-1}		0.1140	0.2775	0.0395	3.1409	0.4270	11.2579	0.1580	0.6780
Limit of detection		0.3761	0.9158	0.1302	10.3651	1.4091	37.1512	0.5214	2.2374
Limit of quantification		1.1397	2.7753	0.3946	31.4095	4.2700	112.5795	1.5800	6.7801
RPD_{cal}		1.9468	1.9072	1.9696	1.8573	1.8605	1.7997	1.9371	1.9089
Linearity (JBSTAT/ CRITVAL)		1.1985 /3.6234	0.6917 /3.7156	1.4226 /2.8784	1.2758 /3.2985	3.2985 /5.8862	0.8245 /2.8784	1.3765 /3.4158	1.0047 /3.6234

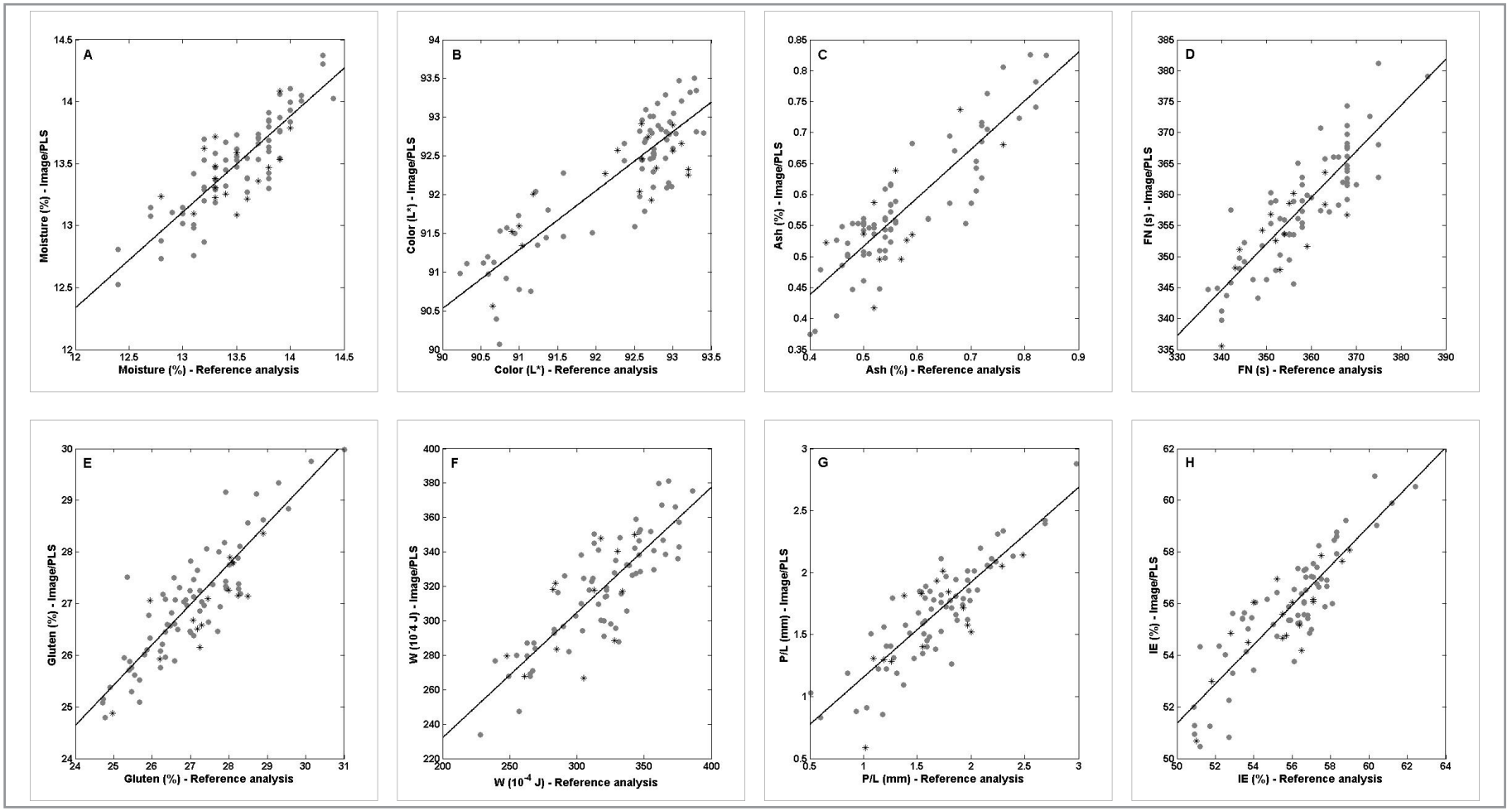


Figure 2. Adjust. (A) Moisture. (B) Color. (C) Ash. (D) FN. (E) Gluten. (F) W. (G) P/L. (H) IE. (•) calibration samples. (∗) external validation samples.

CONCLUSIONS

Image analysis coupled with the PLS model is suitable to determine rheological and quality parameters in the wheat flour samples provided from the industrial process. The models were validated by the parameters of merit showing promising and feasible results. The property values achieved for accuracy, predictive capacity, limits of detection, and quantification indicates that the image/PLS models can be an alternative to reference analysis performed currently by the industry. Moreover, the image/PLS models present several advantages compared to the reference methods. They use a small quantity of wheat flour sample, which is non-destructive, does not generate residues during the analysis, besides the quickness and lower dependence on the analyst perception.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors are grateful to the industry that provide the samples. Patrícia Valderrama acknowledges CNPq [process 306606/2020-8], and Professor Ronei Jesus Poppi for his knowledge, friendship, confidence, and empathy... a Professor and Friend that will be missed in this world.

REFERENCES

1. Sindhu, R.; Shiburaj, S.; Sabu, A.; Fernandes, P.; Singhal, R.; Mathew, G. M.; Nair, I. C.; Jayachandran, K.; Vidya, J.; Vandenberghe, L. P. S.; et al. Enzyme Technology in Food Processing: Recent Developments and Future Prospects. In: Knoerzer, K.; Muthukumarappan, K. (Eds.). *Innovative Food Processing Technologies*. Elsevier, Oxford, **2021**, Vol. 3.12, pp 191–215 (https://doi.org/https://doi.org/10.1016/B978-0-12-815781-7.00016-0).
2. Marti, A.; Ulrici, A.; Foca, G.; Quaglia, L.; Pagani, M. A. *LWT - Food Sci. Technol.*, **2015**, *64*, pp 95–103 (https://doi.org/https://doi.org/10.1016/j.lwt.2015.05.029).
3. Instrução Normativa 8/2005. Regulamento técnico de identidade e qualidade da farinha de trigo. Ministério da Agricultura, Pecuária e Abastecimento, Brazil. **2005**. Available at: https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/legislacao-1/normativos-cgqv/pocs/instrucao-normativa-no-8-de-02-de-junho-de-2005-farinha-de-trigo/view [Accessed March 2021].
4. AACC Approved Methods of Analysis, American Association of Cereal Chemists, 11th edition, Cereals & Grains Association, St. Paul, MN, U.S.A. **2000**.
5. Ferrão, M. F.; Davanzo, C. U. *Anal. Chim. Acta*, **2005**, *540*, pp 411–415 (https://doi.org/10.1016/j.aca.2005.03.038).
6. Ferrão, M. F.; Carvalho, C. W.; Müller, E. I.; Davanzo, C. U. *Ciência e Tecnol. Aliment.*, **2004**, *24*, pp 333–340 (https://doi.org/10.1590/s0101-20612004000300005).
7. Chen, J.; Zhu, S.; Zhao, G. *Food Chem.*, **2017**, *221*, pp 1939–1946 (https://doi.org/10.1016/j.foodchem.2016.11.155).
8. Tian, W.; Chen, G.; Zhang, G.; Wang, D.; Tilley, M.; Li, Y. *Food Chem.*, **2021**, *344*, pp 128633 (https://doi.org/10.1016/j.foodchem.2020.128633).
9. Jiang, H.; Liu, T.; Chen, Q. *Infrared Phys. Technol.*, **2020**, *109*, pp 103423 (https://doi.org/10.1016/j.infrared.2020.103423).
10. Barbon Junior, S.; Mastelini, S. M.; Barbon, A. P. A. C.; Barbin, D. F.; Calvini, R.; Lopes, J. F.; Ulrici, A. *Inf. Process. Agric.*, **2020**, *7*, pp 342–354 (https://doi.org/10.1016/j.inpa.2019.07.001).
11. González-Martín, M. I.; Wells Moncada, G.; González-Pérez, C.; Zapata San Martín, N.; López-González, F.; Lobos Ortega, I.; Hernández-Hierro, J.-M. *Food Chem.*, **2014**, *145*, pp 802–806 (https://doi.org/10.1016/j.foodchem.2013.08.103).
12. Zhao, T.; Chen, M.; Jiang, X.; Shen, F.; He, X.; Fang, Y.; Liu, Q.; Hu, Q. *Infrared Phys. Technol.*, **2020**, *109*, 103426 (https://doi.org/10.1016/j.infrared.2020.103426).

13. Fu, X.; Chen, J.; Fu, F.; Wu, C. *Biosyst. Eng.*, **2020**, *190*, pp 120–130 (https://doi.org/10.1016/j.biosystemseng.2019.12.006).
14. Zhao, X.; Wang, W.; Ni, X.; Chu, X.; Li, Y.-F.; Lu, C. *Biosyst. Eng.*, **2019**, *184*, pp 55–68 (https://doi.org/10.1016/j.biosystemseng.2019.06.010).
15. Tian, W.; Chen, G.; Gui, Y.; Zhang, G.; Li, Y. *Food Control*, **2020**, 107691 (https://doi.org/10.1016/j.foodcont.2020.107691).
16. Czaja, T.; Mazurek, S.; Szostak, R. *Food Chem.*, **2016**, *211*, pp 560–563 (https://doi.org/10.1016/j.foodchem.2016.05.108).
17. Lohumi, S.; Lee, H.; Kim, M. S.; Qin, J.; Cho, B.-K. *Biosyst. Eng.*, **2019**, *181*, pp 103–113 (https://doi.org/10.1016/j.biosystemseng.2019.03.006).
18. Zawieja, B.; Makowska, A.; Gutsche, M. *J. Cereal Sci.*, **2020**, *91*, 102898 (https://doi.org/10.1016/j.jcs.2019.102898).
19. León, K.; Mery, D.; Pedreschi, F.; León, J. *Food Res. Int.*, **2006**, *39*, pp 1084–1091 (https://doi.org/10.1016/j.foodres.2006.03.006).
20. Hardy, A. C.; Cole, P. I.; Ricker, C. W. *Ind. Eng. Chem. Anal. Ed.*, **1929**, *1*, pp 151–152 (https://doi.org/10.1021/ac50067a018).
21. Valderrama, L.; Março, P. H.; Valderrama, P. *J. Chemom.*, **2020**, *34*, e3265 (https://doi.org/10.1002/cem.3265).
22. Foca, G.; Masino, F.; Antonelli, A.; Ulrici, A. *Anal. Chim. Acta*, **2011**, *706*, pp 238–245 (https://doi.org/10.1016/j.aca.2011.08.046).
23. Botelho, B. G.; Assis, L. P.; Sena, M. M. *Food Chem.*, **2014**, *159*, pp 175–180 (https://doi.org/10.1016/j.foodchem.2014.03.048).
24. Godinho, M. S.; Pereira, R. O.; Ribeiro, K. O.; Schimidt, F.; Oliveira, A. E.; Oliveira, S. E. *Quim. Nova*, **2008**, *31*, pp 1485–1489 (https://doi.org/10.1590/S0100-40422008000600039).
25. Mendoza, F.; Aguilera, J. M. *J. Food Sci.*, **2004**, *69*, pp E471–E477 (https://doi.org/10.1111/j.1365-2621.2004.tb09932.x).
26. Edreschi, F. P.; Mery, D.; Mendoza, F.; Aguilera, J. M. *J. Food Sci.*, **2004**, *69*, pp E264–E270 (https://doi.org/10.1111/j.1365-2621.2004.tb10996.x).
27. Caramês, E. T. dos S.; Baqueta, M. R.; Conceição, D. A.; Pallone, J. A. L. *Food Res. Int.*, **2021**, *140*, 109792 (https://doi.org/10.1016/j.foodres.2020.109792).
28. Beltrame, K. K.; Gonçalves, T. R.; Março, P. H.; Gomes, S. T. M.; Matsushita, M.; Valderrama, P. *Aust. J. Grape Wine Res.*, **2019**, *25,* pp 156-160 (https://doi.org/10.1111/ajgw.12387).
29. Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta*, **1986**, *185*, pp 1-17.
30. Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.*, **2001**, *58*, pp 109–130 (https://doi.org/10.1016/S0169-7439(01)00155-1).
31. Kennard, R. W.; Stone, L. A. *Technometrics*, **1969**, *11*, pp 137–148 (https://doi.org/10.2307/1266770).
32. Baqueta, M. R.; Coqueiro, A.; Valderrama, P. *J. Food Sci.*, **2019**, *84*, pp 1247-1255 (https://doi.org/10.1111/1750-3841.14617).
33. Beltrame, K. K.; Souza, A. M.; Coelho, M. R.; Winkler, T. C. B.; Souza, W. E.; Valderrama, P. *J. Braz. Chem. Soc.*, **2016**, *27*, pp 1527-1532 (https://doi.org/10.5935/0103-5053.20160031).
34. American Society for Testing Materials. ASTM E1655-17. *Standard Practices for Infrared Multivariate Quantitative Analysis*. ASTM, PA, USA, **2017**.
35. Santos, D. A.; Lima, K. P.; Março, P. H.; Valderrama, P. *J. Braz. Chem. Soc.*, **2016**, *27*, pp 1912-1917. (http://dx.doi.org/10.5935/0103-5053.20160071).
36. Ferreira, M. M. C. *Quimiometria: conceitos, métodos e aplicações.* Editora da Unicamp, Campinas, SP, **2015**. Chapter 4, p 388.
37. Helfer, G. A.; Magnus, V. S.; Böck, F. C.; Teichmann, A.; Ferrão, M. F.; Da Costa, A. B. *J. Braz. Chem. Soc.*, **2017**, *28*, pp 328–335 (https://doi.org/10.5935/0103-5053.20160182).

38. Santos, D. A.; Coqueiro, A.; Gonçalves, T. R.; Carvalho, J. C.; Bezerra Jr., J. S.; Matsushita, M.; Oliveira, C. A. L.; Março, P. H.; Valderrama, P.; Ribeiro, R. P. *J. Braz. Chem. Soc.*, **2020**, *31*, pp 1883-1890 (https://dx.doi.org/10.21577/0103-5053.20200082).
39. Botelho, B. G.; Mendes, B. A. P.; Sena, M. M. *Quim. Nova*, **2013**, *36*, pp 1416–1422 (https://doi.org/10.1590/S0100-40422013000900023).
40. Vitale, R.; Bevilacqua, M.; Bucci, R.; Magrì, A. D.; Magrì, A. L.; Marini, F. *Chemom. Intell. Lab. Syst.*, **2013**, *121*, pp 90–99 (https://doi.org/10.1016/j.chemolab.2012.11.019).

SUPPLEMENTARY MATERIAL

Table S1. Reference values obtained by AACC methods

Sample number	Moisture (%)	Color (L*)	Ash (%)	FN (s)	Gluten (%)	W (10^{-4} J)	P/L (mm)	IE (%)
1	14.0	92.68	0.43	355	26.21	391	1.54	59.0
2	14.0	92.68	0.43	355	26.21	391	1.54	59.0
3	13.8	92.94	0.49	349	25.36	303	2.17	57.8
4	13.3	90.53	0.67	368	27.42	283	1.27	51.2
5	13.2	92.85	0.54	366	25.41	344	1.86	53.7
6	12.7	90.91	0.71	339	27.91	294	1.61	56.9
7	13.2	90.75	0.66	368	25.54	357	0.92	52.5
8	13.2	90.90	0.71	356	27.45	240	0.72	45.3
9	13.4	93.30	0.46	368	25.94	364	1.99	56.3
10	14.1	92.75	0.56	350	25.81	346	1.86	60.3
11	13.8	90.94	0.76	342	26.57	290	1.37	51.7
12	13.9	92.63	0.50	352	26.34	322	2.69	56.1
13	13.6	92.75	0.50	356	25.81	291	2.92	53.4
14	13.1	92.78	0.50	363	28.39	330	1.79	55.7
15	13.1	93.28	0.52	368	24.73	361	1.86	57.1
16	12.8	90.32	0.82	344	27.18	269	0.92	50.9
17	13.0	91.57	0.62	356	27.46	267	1.93	53.6
18	13.7	92.28	0.68	352	27.25	334	2.33	57.1
19	14.0	92.37	0.54	342	25.47	313	1.72	56.7
20	13.7	92.81	0.56	366	25.67	332	1.52	53.4

Table S1. Reference values obtained by AACC methods (Continuation)

Sample number	Moisture (%)	Color (L*)	Ash (%)	FN (s)	Gluten (%)	W (10^{-4} J)	P/L (mm)	IE (%)
21	13.2	92.95	0.45	368	24.91	363	2.15	57.4
22	13.6	91.21	0.71	351	27.58	272	1.39	54.1
23	13.3	92.57	0.42	344	25.67	265	1.97	54.0
24	14.3	91.95	0.53	375	28.02	328	1.87	56.8
25	13.6	91.14	0.77	340	27.18	328	1.38	52.8
26	13.9	92.59	0.54	404	28.26	318	1.51	57.5
27	13.3	91.20	0.78	345	27.08	322	1.65	51.2
28	13.5	92.59	0.54	404	28.26	318	1.51	57.5
29	13.3	92.57	0.42	344	25.67	265	1.97	54.0
30	14.0	92.37	0.54	342	25.47	313	1.72	56.7
31	13.9	92.63	0.50	352	26.34	322	2.69	56.1
32	13.5	92.75	0.50	368	26.66	376	1.77	55.9
33	13.0	92.70	0.50	358	26.61	265	1.28	60.3
34	13.8	93.22	0.47	381	27.09	375	1.25	57.3
35	13.7	92.60	0.52	351	24.97	254	2.77	51.8
36	13.7	90.74	0.76	355	28.29	329	0.93	51.2
37	13.1	93.11	0.52	358	24.78	255	2.98	51.7
38	13.3	93.11	0.58	399	27.06	285	1.68	60.1
39	13.2	92.18	0.52	349	28.82	308	2.29	58.6
40	13.2	92.70	0.51	351	27.36	308	1.62	56.3
41	13.8	92.94	0.55	362	26.29	249	2.68	58.3
42	13.8	92.96	0.55	355	26.26	321	1.98	57.7
43	13.6	92.60	0.53	359	24.72	331	2.23	56.5
44	13.9	93.00	0.56	385	27.49	375	1.09	58.6
45	13.3	93.20	0.56	353	24.93	364	1.93	55.5
46	13.9	92.92	0.55	367	28.89	376	1.54	58.3
47	13.5	91.04	0.76	354	28.03	261	1.09	46.7

Table S1. Reference values obtained by AACC methods (Continuation)

Sample number	Moisture (%)	Color (L*)	Ash (%)	FN (s)	Gluten (%)	W (10^{-4} J)	P/L (mm)	IE (%)
48	13.4	90.58	0.85	343	28.24	248	1.19	43.8
49	13.7	92.92	0.54	391	27.71	312	1.47	62.4
50	13.9	91.18	0.77	344	28.11	349	0.96	56.4
51	13.4	92.57	0.69	375	27.01	423	1.77	52.8
52	13.5	90.84	0.80	359	27.99	304	1.89	53.7
53	13.0	91.00	0.66	356	27.74	257	0.60	50.9
54	13.9	93.04	0.50	353	25.27	355	2.39	57.3
55	13.1	90.67	0.82	342	28.72	263	1.63	50.9
56	13.6	90.65	0.67	360	28.31	300	1.02	51.0
57	13.5	92.74	0.55	353	26.88	303	1.72	57.4
58	13.4	90.70	0.81	340	27.25	309	1.03	52.7
59	13.3	91.15	0.72	355	27.24	337	1.82	56.3
60	13.7	91.37	0.73	358	27.91	337	1.59	61.2
61	14.0	92.98	0.47	362	25.87	368	1.57	55.2
62	13.8	91.35	0.72	354	27.88	228	0.51	53.5
63	13.3	92.90	0.49	370	26.91	388	1.63	56.6
64	13.3	90.84	0.72	363	27.30	357	1.18	57.8
65	15.1	93.20	0.59	344	25.94	377	2.48	55.5
66	13.5	93.08	0.48	364	25.40	386	1.58	58.2
67	14.0	92.65	0.54	368	26.51	329	1.56	53.9
68	13.7	91.15	0.77	345	27.78	341	2.19	56.6
69	13.8	92.74	0.55	357	26.22	347	1.55	56.6
70	13.8	92.74	0.55	357	26.22	347	1.55	56.6
71	13.2	91.24	0.73	340	28.24	299	1.92	52.9
72	14.0	91.57	0.62	341	27.01	333	1.72	56.4
73	13.2	90.60	0.79	348	27.94	268	1.57	52.7
74	13.8	92.74	0.56	360	25.91	325	1.81	57.0

Table S1. Reference values obtained by AACC methods (Continuation)

Sample number	Moisture (%)	Color (L*)	Ash (%)	FN (s)	Gluten (%)	W (10^{-4} J)	P/L (mm)	IE (%)
75	14.0	92.73	0.54	373	26.72	373	1.14	55.8
76	13.3	90.59	0.84	375	28.49	343	0.85	57.1
77	13.5	92.65	0.55	386	26.99	263	1.82	55.2
78	13.6	92.91	0.54	351	26.98	366	1.21	57.0
79	14.4	93.30	0.40	368	29.55	316	2.18	58.3
80	13.5	93.00	0.55	368	26.48	346	2.03	56.8
81	14.1	92.60	0.50	368	29.3	283	2.30	55.8
82	13.2	90.83	0.59	347	30.15	286	1.21	52.2
83	12.8	92.72	0.48	368	27.29	305	1.74	57.3
84	13.8	93.01	0.50	368	28.26	311	2.09	52.9
85	12.8	92.50	0.45	368	27.29	346	1.31	58.8
86	13.9	92.90	0.41	368	26.39	316	1.43	61.2
87	13.1	90.99	0.71	358	27.91	337	1.59	61.9
88	13.5	92.80	0.46	354	27.09	304	2.25	56.7
89	13.0	90.23	0.72	337	28.09	239	1.07	48.0
90	13.3	91.00	0.76	344	28.90	282	1.26	56.5
91	12.9	92.74	0.51	365	26.82	339	1.67	55.0
92	14.3	92.70	0.52	368	31.01	344	2.06	60.4
93	13.3	90.59	0.84	375	28.49	343	0.85	57.1
94	12.7	92.60	0.48	368	26.34	305	2.92	54.7
95	13.8	93.00	0.53	363	28.95	284	3.48	56.0
96	13.4	90.93	0.70	342	28.3	265	1.19	51.6
97	13.5	91.05	0.79	345	29.83	313	1.55	55.2
98	13.9	92.12	0.57	368	27.55	401	2.00	59.4
99	12.4	93.40	0.47	358	26.59	320	1.97	58.1
100	12.4	93.40	0.47	358	26.59	320	1.97	58.1

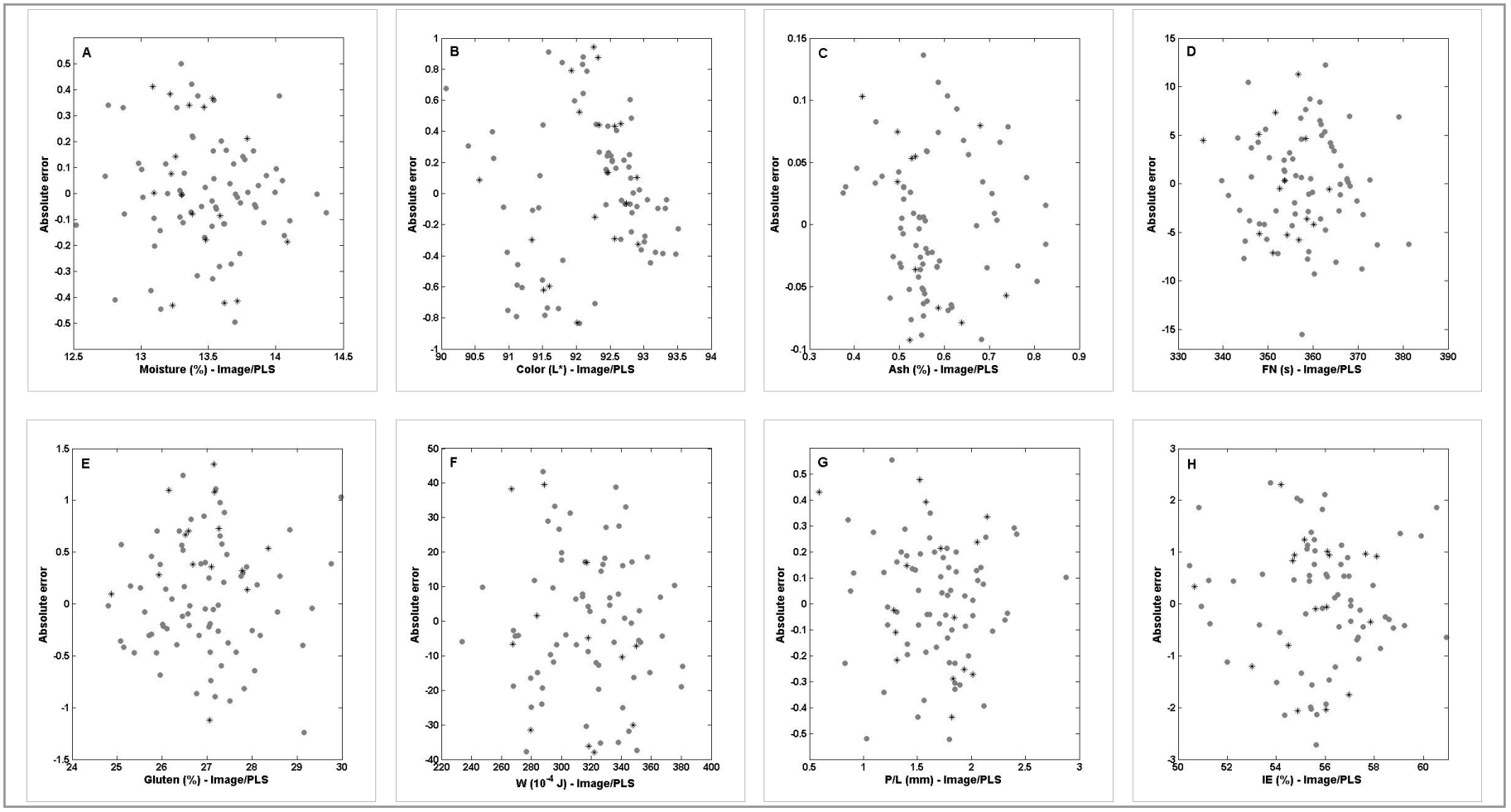


Figure 1S. Residuals. (A) Moisture. (B) Color. (C) Ash. (D) FN. (E) Gluten. (F) W. (G) P/L. (H) IE. (•) calibration samples. (∗) external validation samples.

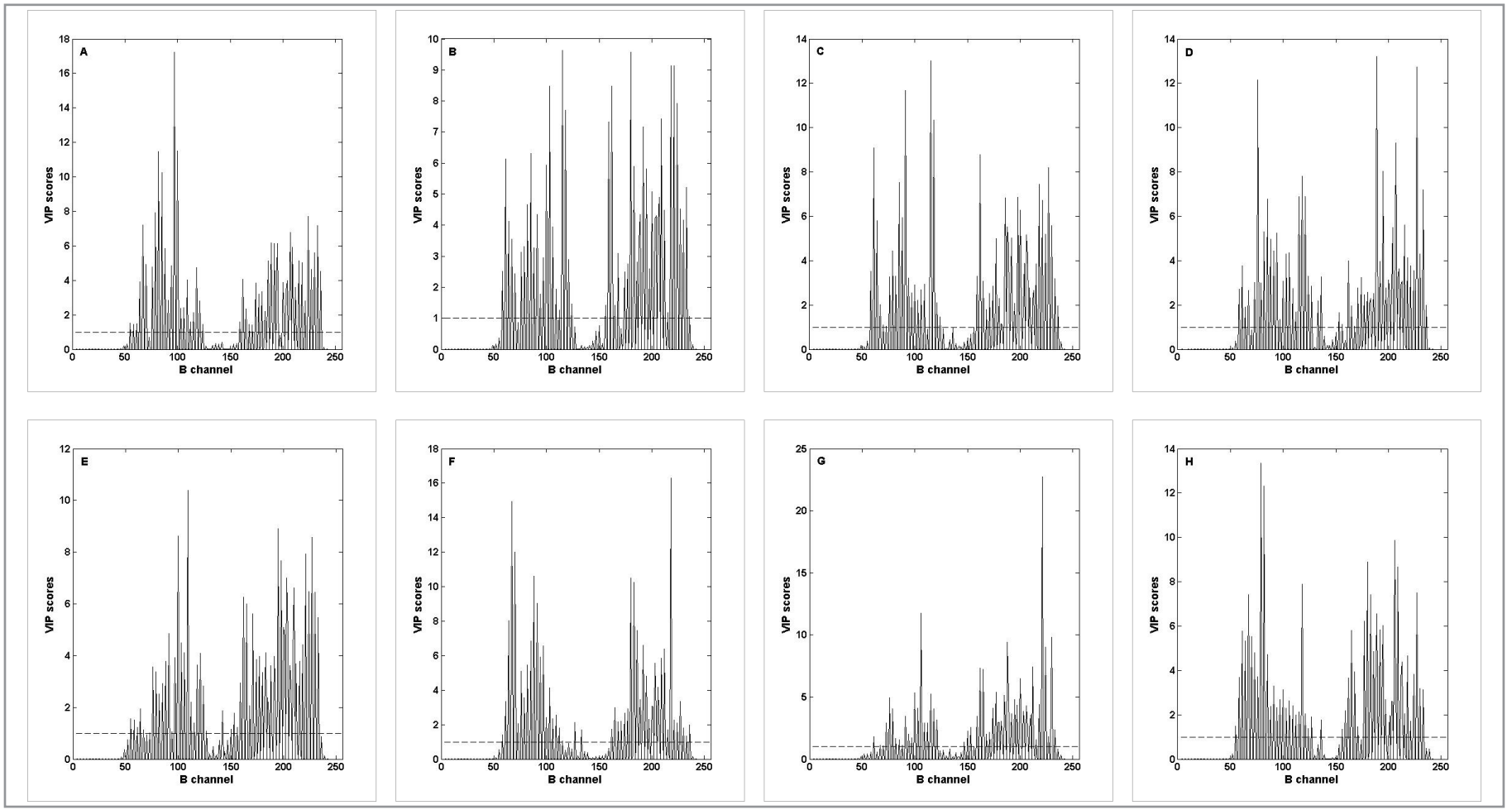


Figure 2S. VIP scores. (A) Moisture. (B) Color. (C) Ash. (D) FN. (E) Gluten. (F) W. (G) P/L. (H) IE.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 78–90
doi: 10.30744/brjac.2179-3425.AR-10-2021

BrJAC

ARTICLE

Monitoring Mineral-Associated Organic Matter in Tropical Pastures using Near Infrared Spectroscopy

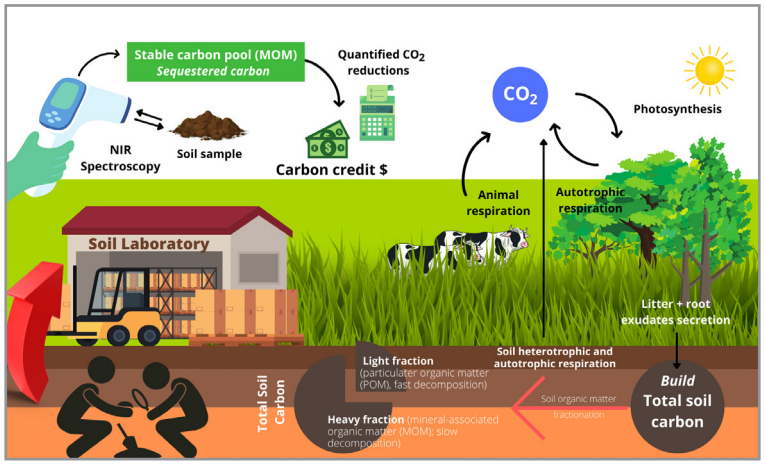
Felipe Bachion de Santana[1], André Marcelo de Souza[2*], Fabiano de Carvalho Balieiro[2], Mirelly Mioranza[3], Renato Aragão Ribeiro Rodrigues[2], Ronei Jesus Poppi[1]†

[1]*Instituto de Química, Universidade Estadual de Campinas, Caixa Postal 6154, CEP 13084-971, Campinas, SP, Brazil*

[2]*Empresa Brasileira de Pesquisa Agropecuária (Embrapa Solos), R. Jardim Botânico, 1024, Rio de Janeiro, RJ, CEP 22460-000, Brazil*

[3]*Universidade Federal Fluminense, CEP 24020-141, Niterói, RJ, Brazil*

† In memoriam (1961 – 2020)



In order to confirm that the mitigation of greenhouse gas emissions could indeed be achieved by farmers, determinations of soil organic carbon (SOC) in total and stabilized fractions are essential, proving the effectiveness of the sustainable practices adopted by the farmer. In this sense, this study proposes an analytical methodology based on near infrared spectroscopy (NIRS) and partial least squares regression (PLSR) as an alternative for the measurement of stabilized and total soil organic carbon (SOC) in agricultural production systems. A set of 122 samples of four different land uses were sampled and physically fractionated. Total organic carbon for both fractions (mineral-associated stable fraction and particulated organic matter - labile fraction) were determined by the dry combustion method. In order to eliminate the extra step regarding soil fractionation to determine the stabilized SOC content, this study investigated different strategies to build the regression model based on partial least squares regression for the determination of the stabilized SOC from the total soil fraction. Two strategies presented the same accuracy as the reference method used to determine the stabilized SOC content in stabilized fraction, with root mean square error in validation of 1.47 g/dm^3. These results indicate that both strategies proposed can determine simultaneously the total and stabilized SOC from the total soil fraction, thus eliminating the fractionation phase.

Cite: de Santana, F. B.; de Souza, A. M.; Balieiro, F. C.; Mioranza, M.; Rodrigues, R. A. R.; Poppi, R. J. Monitoring Mineral-Associated Organic Matter in Tropical Pastures using Near Infrared Spectroscopy. *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 78–90. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-10-2021

Submitted 18 January 2021, Resubmitted 20 March 2021, Accepted 26 April 2021, Available online 21 May 2021.

Keywords: soil stabilized fraction; soil organic carbon; near infrared spectroscopy; multivariate calibration

INTRODUCTION

The ABC Plan (*Agricultura de Baixa Emissão de Carbono* - Low Carbon Agriculture) is part of a strategy adopted by the Brazilian Government for reducing greenhouse gas (GHG) emissions from agriculture. This policy aims to encourage the farmers to adopt technologies and sustainable production systems to ensure GHG emissions reduction [1]. The main goals of the ABC Plan are to expand the integrated crop-livestock-forest systems to 4 million hectares, and the area with zero-tillage to 8 million ha, rehabilitate 15 million ha of degraded pastures, planting more 3 million ha of forests and increase the adoption of biological nitrogen fixation in more 5.5 million ha by 2020.

In order to attest that the mitigation of GHG emissions could indeed be achieved by farmers, soil organic carbon (SOC) testing and monitoring would become an essential part of the process to prove the effectiveness of the adoption of such sustainable practices that support the program. By this way, the demand for SOC analysis would significantly increase and an analytical methodology to measure the additionality of SOC sequestration of any technology or project adopted has become a key factor for the implementation of ABC Plan.

Soil organic matter (SOM) has been characterized by chemical and physical fractionation. Both tries to minimize the heterogeneity of the soil organic components in terms of composition, availability to soil biota and environmental functions [2]. The choice of fractionation methods depends on the study objectives. Chemical fractionation is based in different solubility of humic substances to alkaline and acid solutions and normally is used to chemical characterization of specific organic soil compounds or fraction (mainly humic acid fraction) and land use impacts, while physical fractionation is used to quantify and describe the compartments of SOM linked to nutrient cycling and C sequestration [2,3]. The relationship between SOM and soil physical constituents could be studied at two level of organization: SOM associated with mineral particles (sand, silt, and clay) – particle-size fractionation and SOM associated with macro and microaggregates – aggregate fractionation approach [4].

In the tropics, despite the higher SOM turnover, the particulate organic matter (POM) pools (more labile) make up a relatively small proportion of SOM in comparison to organo-mineral (clay in special) fraction and the mean residence time, or the half-life of C (in years) increase as fraction size decreases [4]. In other words, the lower C-turnover of mineral-organic associated matter (clay + silt associated fraction) comprises the most stable compartment of soil.

Considering SOC analysis for fertility purposes, the standard sample preparation procedures that have been practiced for decades (based on Walkley–Black or dry combustion) are able to assess the bulk organic carbon in the soil sample, without considering the proportion of these labile and stable fractions [5,6]. This means that new steps regarding to soil fractionation are necessary to determine the labile and stable C.

However, all reference methods to fractionate soil organic carbon are cost intensive and laborious. The combination of an extra soil fractionation step with traditional method of SOC analysis (Walkley–Black or dry combustion) would result in an expensive, time-consuming and waste generation ~ 600,000 L/year in Brazil [5]; consequently, restricts their use on a large scale [5,7].

This procedure is very contradictory, since on the one hand the GHG emissions could be mitigated by the adoption of the ABC Plan, on the other hand more pollution could be generated by increasing demand for SOC analysis that would negatively impact the environment. All these factors could eventually restrict the implementation of ABC Plan, requiring drastic changes in the methodologies for SOC analysis.

The implementation of near infrared spectroscopy (NIRS) technologies could become an effective solution to overcome these methodological issues that have been blocking the ABC Plan from reaching the next level. Several papers and documents that have been published over the last decade, such as those produced by FAO [8], recommend the use NIRS for SOC analysis [9–12].

In a previous study, de Santana and collaborators [5] compared the accuracy and precision in SOM determinations by the Walkley–Black and NIR methodology using a vast vis-NIR soil spectral library (~ 43,000 samples). The methodology developed was validated by a proficiency assay for fertility laboratories coordinated by Embrapa soils (Brazilian Agricultural Research Corporation). The results obtained shows that the vis-NIR spectroscopy can determinate the SOM content in Brazilian soils with the same accuracy and precision than the Walkley–Black method [5]. It is important to highlight that the SOC analysis was performed on the total fraction of the sample (bulk soil).

In other study, Jaconi and collaborators [12] evaluated and optimized the use of NIRS to predict SOC content in fractions with the constraint that the carbon fractions (labile and stabilized) should add up to 100% (total organic carbon content). Through employing log-ratio transformation, it was demonstrated that the NIRS models could be suitable to predict soil carbon fractions with high accuracy in a wide range of soil types and land-use types.

Following the studies developed by the group, the present study aims to contribute to the discussion regarding analytical methodology to quantify the most stable fraction of SOC, contributing to the implementation of ABC Plan. In this paper we proposed an effective, fast, cheap, and sustainable methodology without extra sample preparation to quantify the mineral associated C content in soils under sustainable management employing NIRS and partial least squares regression (PLSR). In addition, another innovation of this article is the use of different strategies to improve the SOC stabilized determinations from the NIR spectra of bulk soil.

MATERIAL AND METHODS

Study site

The experimental area is in a thematic center of Embrapa (Brazilian Agricultural Research Corporation, Embrapa), Embrapa Dairy Cattle, Coronel Pacheco's Municipality, Minas Gerais State, Brazil, (21°33′22″S, 43°06′16″W, an altitude of 410 m). According to the Köppen climate classification, the region fits the Cwa type (mesothermal), with a well-defined dry season (autumn-winter) and rainy season (spring-summer) [13]. Annual rainfall (average of 30 years) is around 1,600 mm, of which approximately 90% occurs in the summer season. Annual temperatures varied from 24.3 to 31.7 ºC (maximum) and 11.4 to 19.7 ºC (minimum). The soil class of the experimental area is "*Latossolo Vermelho Amarelo distrófico*" according to the Brazilian Soil Classification System [14], with a clay texture which corresponds to *Typic hapludox* according to the Soil Survey Manual [15]. The local relief is characterized by mountainous topography with a slope of approximately 30%.

Land uses and characterization

A set of 122 samples of four different types of land were collected from (1) silvopastoral system (SPS), (2) managed pasture of *Brachiaria decumbens* (brachiaria managed pasture -BMP), (3) degraded pasture of brachiaria (BDP), and (4) native vegetation (NV). The SPS was established in November 1997 and it is composed with the tropical perennial grass *brachiaria*. The tree component of SPS was composed by the species of *Eucalyptus grandis* Hill ex Maiden, *Acacia mangium* Willd (*acacia*) and *Mimosa artemisiana* Heringer & Paula (*mimosa*), all of them planted alternately in the North/South direction (planting row) aiming to prevent surface erosion. In the planted row, seedlings were planted with 3 m of spacing.

The objective of nitrogen-fixing species (acacia and mimosa) in SPS is to provide shade or shelter to the animals and nitrogen biologically fixed (N_2) and other nutrients, to the eucalyptus and brachiaria. The BMP maintain reasonable grass productivity due to appropriate grazing regime. The history of the use of correctives and fertilizers used in this area is organized in the Table I.

Table I. Management history of the Silvopastoral System in Coronel Pacheco - MG

Year	Management
1997	At the moment of the introduction of the trees, the *guandu beans* (*Cajanus cajan*) were planted, among the wooded strips, in order to be incorporated into the soil, as green manure, before the flowering of the plants.
1998 - 2000	The pastures remained closed to guarantee the initial growth of the tree species.
2001 - 2002	The pastures were used for the grazing of dry cows, respecting periods of occupation of 5 days and resting of 45 days.
2003 - 2010	The grazing management was carried out under rotational stocking and without fertilization.
2011 - 2013	Under continuous stocking and with application of 64 kg/ha of N and K_2O and 16 kg/ha of P_2O_5 per year, divided into two equal applications during the summer.
2013 - 2016	There were no additional fertilizer or corrective applications, characterized by a low input and extensive system (low rate of grazing and lenient grazing). By this way we guarantee the sustainability of the pasture after 18 years of implantation.

The DPB was characterized to presents low productivity for animal grazing due to high weed infestation, bare soil and/or soil erosion. The NV belongs to the ecosystem of the Atlantic Forest [16] without any exploitation or anthropic interference at least in the last 40 years.

Soil sampling design

The soil sampling strategy was designed for two main purposes: soil organic matter physical fractionation and to study the soil C distribution in the soil profile. As all the systems (SPS, BMP, BDP and NV) were immersed in a mountainous landscape we divided them into three sub-areas considering the slope of the terrain: upper, mid and lower third of the slope.

In the SPS, the tree component has a great influence on soil chemical and physical attributes. To cover these influences, we select three different positions to collect soils samples in SPS: inside the ranks of trees, tree border (under canopy influence) and the middle of two ranks. Under BMP, BDP and NV three and under SPS nine trenches were opened. The soil samples were collected in seven different depths in each area for assessment of 0-10, 10-20, 20-30, 30-40, 40-60, 60-80 and 80-100 cm of soil layers.

Soil sample preparation

All soil samples were oven dried at 40 °C for 48 h, then a rubber mallet was used to break the soil clusters, and for the granulometry standardization the soils were sieved using a 2 mm sieve [6].

Physical fractionation

The bulk soil samples were physically fractionated according to the procedure of Cambardella & Elliott (1992) [2] with some adaptations in the sample weight [17]. Briefly, 20 g of soil were placed in plastic bottles and 60 mL of NaOH 1.0 g L^{-1} was added. The mixture was shaken for 16 h in a horizontal shaker, at 160 oscillations min^{-1}. Then, the entire content was placed on a 53 μm sieve and washed with a weak jet of demineralized water. The material retained on the sieve is defined as particulate organic matter (POM) (> 53 μm) known as labile fraction; on the other hand, the material that passed the sieve is defined as the mineral-associated organic matter (MOM), also known as stabilized fraction. Both fractions were dried at 60 °C.

Soil organic carbon analysis

After the physical fractionation step, the soil samples were grounded to a fine powder (<0.149 mm) to facilitate the weighing process. The soil organic carbon in both fractions (POM and MOM) and bulk soil (total fraction) was analyzed by the dry combustion method using a CHNS Perkin Elmer 2400.

Acquisition of NIR spectra

The NIR spectra in reflectance mode were registered using a Perkin Elmer Fourier Transform spectrophotometer, Spectrum 100N, equipped with integrating sphere and rotation accessory, in the range of 1000 - 2500 nm, 0.5 of spectral resolution and 64 scans. Before each analysis, the background was obtained using the Spectralon® present in integrating sphere.

Chemometrics analysis

Spectral preprocessing

In order to reduce the baseline variation and enhance the spectral features, all soil spectra were pre-processed using smoothing and first derivative, based on the Savitzky-Golay algorithm, with a window size of 11 points and second-order polynomial fitting [18].

Principal component analysis

Principal component analysis (PCA) is a classical data analysis technique, which can compress information from many interconnected variables and project these new variables in a low-dimensional subspace defined by a few principal components (PCs). The original dataset (**X**) is decomposed in two matrices **P** and **T** as illustrated in Equation 1. The orthonormal loading matrix **P** contain information about the weight of the original variables in the PC space and the orthogonal score matrix **T** are the new coordinates of these samples in the PC space.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{T} + \mathbf{E} \qquad \text{Equation 1}$$

where **E** is the part non-modeled of **X** (residual matrix) [19].

Partial least squares regression (PLSR)

Partial least squares regression (PLSR) is the most widely used calibration method used in quantifications from NIR spectral data [20]. The PLSR performs a multivariate regression using the **X** block (NIR spectra) and property of interest **y** (C in different fractions), by projecting the **X** block in a low-dimensional subspace, called latent variable space [21]. To obtain feasible results in validation set, it is necessary correctly choose the number of latent variables and remove the outlier's samples in calibration set [21]. The optimum number of latent variables can be selected analyzing the fewest number of latent variables which present minimal difference in relation to the minimum value of root mean squares errors of cross-validation - RMSECV [22]. In this study, due to the large number of samples, we used the venetian blinds cross-validation [23].

The outliers in calibration step should be evaluated in **X** block and **y** block, while in validation set the outliers must be evaluate only in **X** block. Samples with high Hotelling T^2 and residuals in spectral data (Q-statistics) at a significance level of 5% can be considered outliers in **X** block, while in **y** block samples with prediction error superior than ±3 × RMSEC (root mean square error in calibration) can be considered outliers. In this last case, it would be interesting to re-analyze the reference values for this sample. The Hotelling T^2 value is related to leverage, which measures the distance of the sample from the center of the data, while Q residuals is related to unmodelled NIR spectra. The American Society for Testing and Materials, ASTM E1655-17 [24] recommends that a maximum of 3 rounds should be performed to removes outlier samples in calibration step.

External parameter orthogonalization (EPO)

The EPO method aims to remove the effect of external parameters in spectra. In this study we adapted the EPO to remove the labile SOC fraction from the total soil fraction (bulk). The EPO method will remove the labile SOC fraction (POM) from the bulk by projecting the NIR spectra into a portion of spectral space which is orthogonal to POM spectra [25,26].

The EPO method assumes that spectra matrix **X** can be decomposed into two subspaces, a useful subspace related to target chemical responses and a parasitic component related to the part of the spectra that is not related to the target chemical responses as illustrated in Equation 2

$$\mathbf{X} = \mathbf{XU} + \mathbf{XQ} + \mathbf{R} \qquad \text{Equation 2}$$

where $\mathbf{R}$ is an independent residual matrix.

The initial step to build the EPO model is to calculate (labile fraction/POM), which is the difference between the spectra of total and the stabilized fractions (MOM). In the next step, the PCA model is build using and are obtained the scores and the loadings , and using these loadings we can calculate the scores (for new samples (validation samples). By multiplying is obtained , which is the part of the soil spectra related to POM fraction. Considering that is irrelevant, it is possible to calculate the stabilized fraction by subtracting the original spectra from (POM fraction) [27]. Finally, the PLSR model is built using the matrix (spectra of MOM fraction) and the SOC content in MOM fraction.

Strategies to build the PLSR models

The 122 soil samples (all set) were firstly organized in ascending order of concentration and then for each group of three samples in sequence, the first two were selected to be used in calibration set (66.67%) and the remaining one was used in the validation set (33.33%). Two references PLSR model (Classical strategy) were built, and four different strategies were evaluated to determine the SOC content in MOM fraction from the NIR spectra of bulk soil (total fraction). These strategies were proposed to avoid the extra step regarding to soil fractionation.

The classical strategies to determine SOC content in the bulk soil and MOM fractions were named Model 1 and Model 2 (Figure 1). In the Model 1, the matrix ($\mathbf{X}_1$) was measured on soil total fraction (bulk). About 67% of these samples were used in calibration ($\mathbf{X}_{cal1}$) and 33% of these samples were used in validation ($\mathbf{X}_{val1}$). Reference values ($\mathbf{Y}_{cal1}$ and $\mathbf{Y}_{val1}$) were the SOC measured in the soil total fraction. In the Model 2, the matrix ($\mathbf{X}_2$) was measured on MOM fraction. Again about 67% of these samples were used in calibration ($\mathbf{X}_{cal2}$) and 33% of these samples were used in validation ($\mathbf{X}_{val2}$). Reference values ($\mathbf{Y}_{cal2}$ and $\mathbf{Y}_{val2}$) was the SOC measured in the MOM fraction.

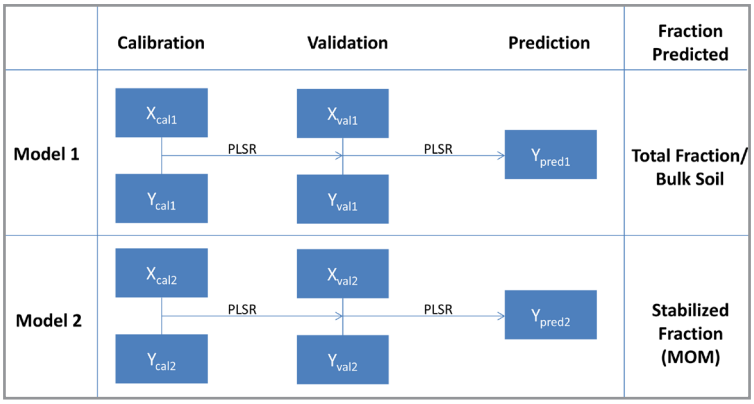


Figure 1. Classical strategy to build PLS models in both subsets: total and stabilized fractions.

The four strategies evaluated here to determine the SOC content in MOM fraction ($\mathbf{Y}_2$) from the bulk soil spectra ($\mathbf{X}_1$) were named Model 3, Model 4, Model 5, and Model 6, as represented in Figure 2. In the Model 3, the calibration samples ($\mathbf{X}_{ca11}$ and $\mathbf{Y}_{cal1}$) were the spectra measured on soil total fraction and the SOC measured in the soil total fraction, in other words is the calibration of Model 1. While the validation samples ($\mathbf{X}_{val2}$ and $\mathbf{Y}_{val2}$) were spectra measured on MOM fraction and the SOC measured in the MOM fraction, that is the validation of Model 2.

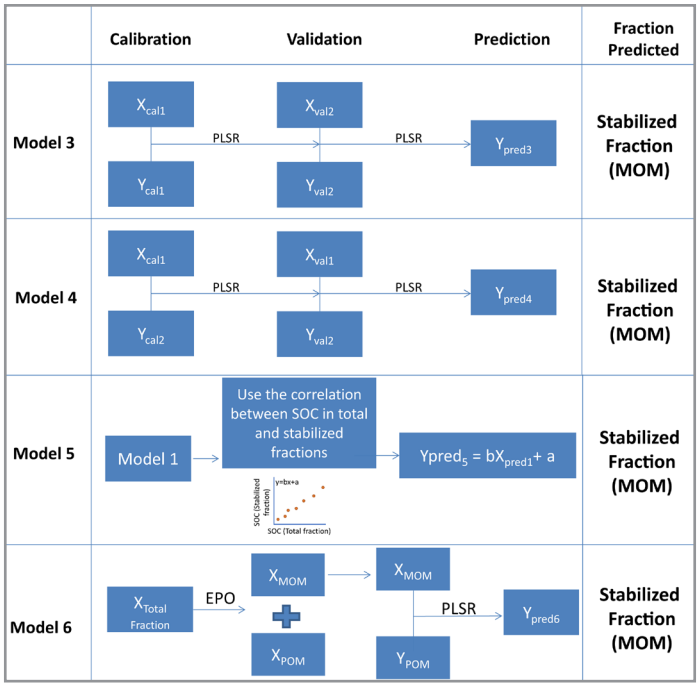


Figure 2. Strategies adopted to determine the SOM content in the MOM fraction from the bulk soil spectra.

In Model 4, the calibration dataset was composed by the spectral data set $\mathbf{X}_{cal1}$ (obtained on soil total fraction) and the $\mathbf{Y}_{cal2}$ (SOC measured in the MOM fraction), the validation dataset is formed by the same logic, using $\mathbf{X}_{val1}$ and $\mathbf{Y}_{val2}$. The predicted value, $\mathbf{Y}_{pred4}$, correspond to the SOC content in MOM fraction.

In Model 5, it was used the Model 1 to determine $\mathbf{Y}_{pred1}$, and then, a linear correlation between the reference values of stabilized $\mathbf{Y}_{cal2}$ and total soil organic carbon $\mathbf{Y}_{cal1}$ were used to determine $\mathbf{Y}_{pred5}$, which is the SOC in MOM fraction. Finally, in the Model 6, it was used the EPO methodology to decompose the $\mathbf{X}_{Total\ fraction}$ in to $\mathbf{X}_{MOM}$ and $\mathbf{X}_{POM}$, then it was used the $\mathbf{X}_{MOM}$ fraction to build the PLSR model. To build the EPO model, about 30% of the calibration samples were used.

RESULTS AND DISCUSSION

The average of the NIR spectra of the total, POM and MOM fractions of soil are shown in Figure 3. It is possible to observe that the spectra of total and MOM fractions are very similar, while the spectra of POM fraction are very different due to majority presence of sand [28]. The spectra of total and MOM fractions show essentially absorptions at 1200-1500, 1800–2000 and 2200–2400 nm, common to most soil NIR spectra already reported in several publications [5,28]. The NIR spectra contain fewer absorptions bands when compared to MIR spectroscopy, due to the broad and overlapping bands, making it difficult to interpret. However, NIR spectra are influenced by the information of organic and inorganic components present in the soil, including SOC content.

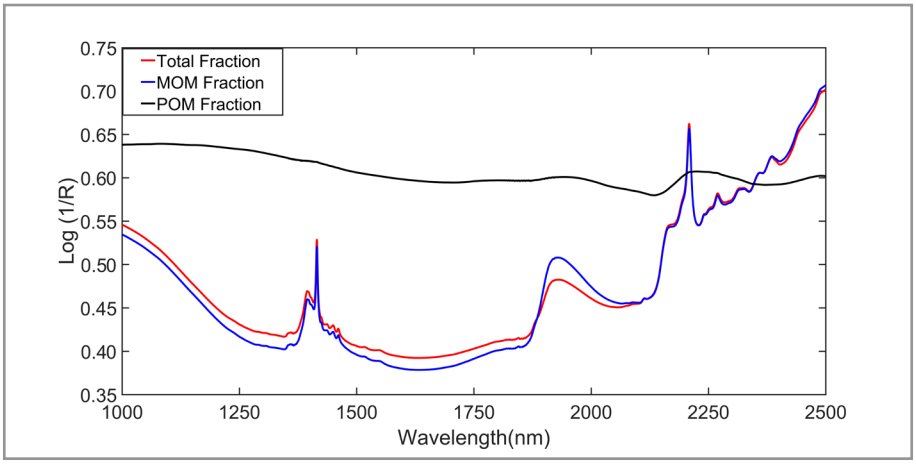


Figure 3. NIR spectra of total, MOM and POM fractions.

A visual inspection along the full spectrum suggests that there is no difference in the NIR spectra of the total and MOM fractions, but a refined comparation between these 3 fractions can be performed through the analysis of scores values of PCA [20]. The scores values of PCA are presented in Figure 4, showing a clear differentiation between POM and others fractions along the PC1 axis, while the PC2 shows the difference between the MOM and total fraction.

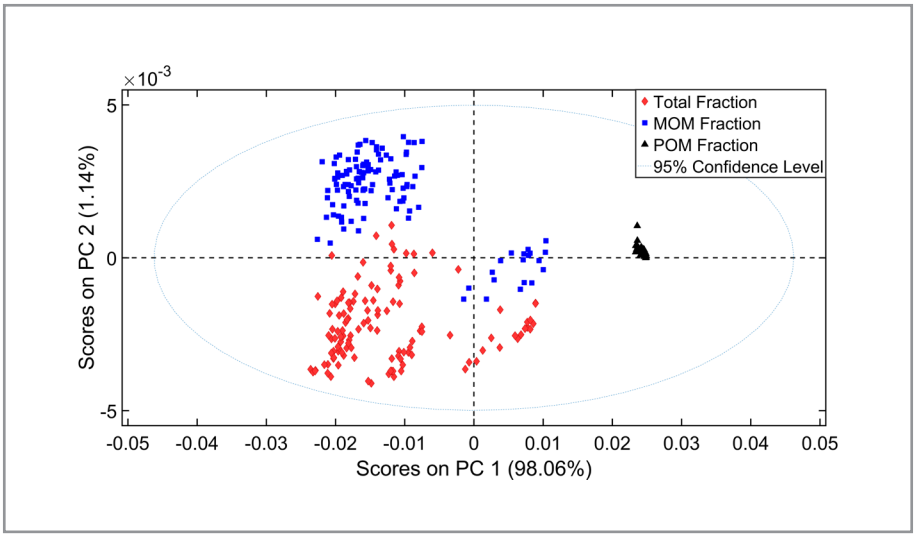


Figure 4. Scores values of PCA from total, MOM and POM soil fractions.

The scores of the total and MOM fractions are not grouped, indicating that these spectra are different and consequently a PLSR model build using the NIR spectra of the total fraction is not able to directly determine the content of SOC in MOM fraction. To solve this problem, we used the four strategies previously described, thus eliminating the additional soil fractionation step.

Figure 5 presents the scatter plots showing the reference versus predicted values by each strategy proposed (Models 1 to 6), the accuracy parameters and the number of outliers samples obtained for each model are organized in Table II. The efficiency in determine the stabilized SOC by each strategy will be compared to the PLSR model build using the NIR spectra of stabilized fraction and the SOC content in this fraction (Model 2).

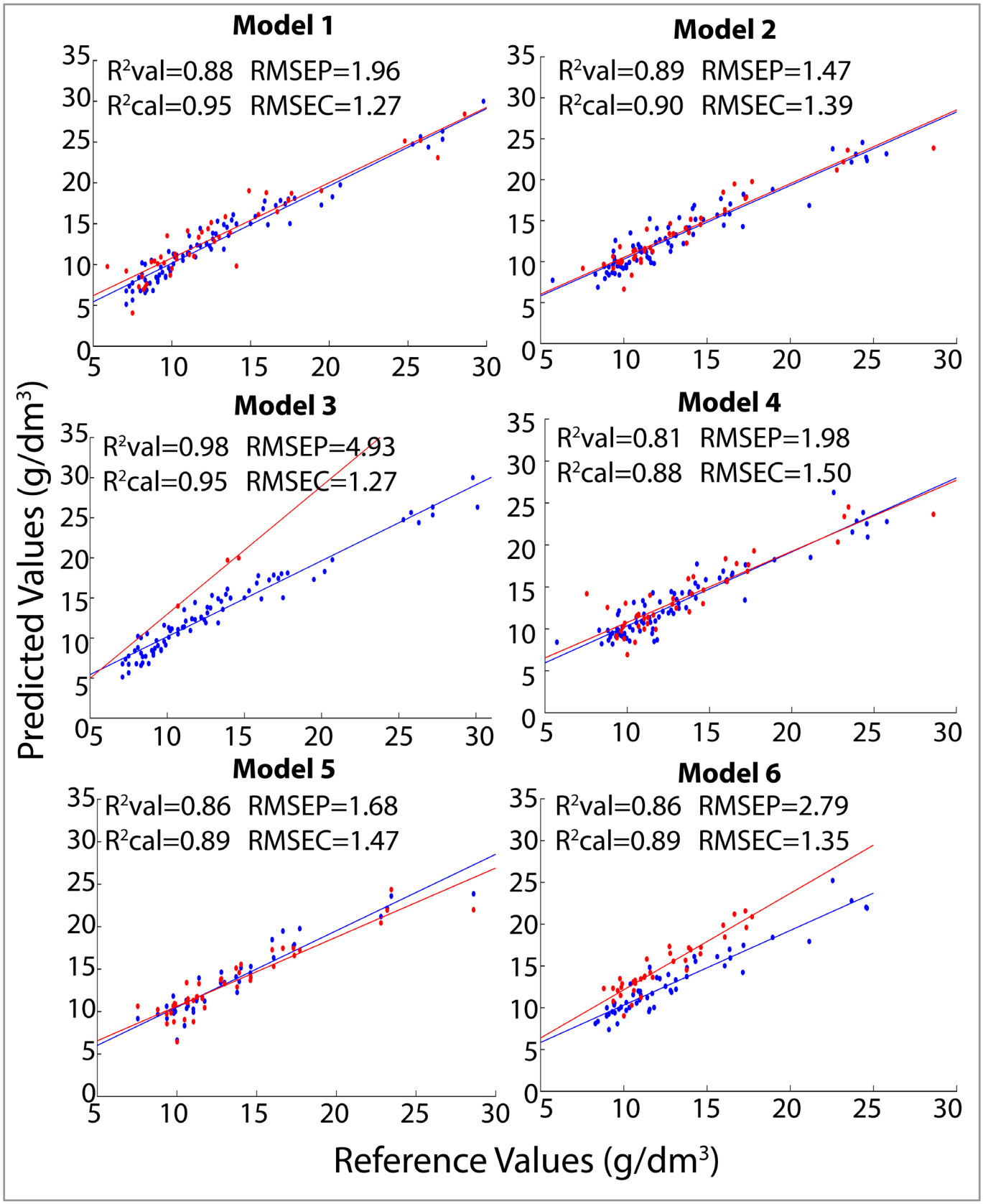


Figure 5. Scatter plots of reference versus PLSR predict values for each reference model and all strategies proposed.

Table II. Accuracy parameters and excluded samples for the strategies proposed

Strategy		No. of samples	RMSE (g/dm^3)	R^2	No. excluded samples
Model 1	Calibration	81	1.27	0.95	1
	Validation	41	1.96	0.88	0
Model 2	Calibration	81	1.39	0.90	1
	Validation	41	1.47	0.89	0
Model 3	Calibration	81	1.27	0.95	1
	Validation	41	-	-	38
Model 4	Calibration	81	1.50	0.88	1
	Validation	41	1.98	0.81	1
Model 5	Calibration	81	1.47	0.89	1
	Validation	41	1.68	0.86	0
Model 6	Calibration	54	1.35	0.89	0
	Validation	41	2.79	0.86	5

It is observed that the strategy Model 3 shows a poor performance since 38 of the 41 validation samples were considered outliers by the PLSR model. This result was already expected, because as observed by the PCA scores of Figure 4, the spectra of SOC in MOM fractions are significantly different, resulting in outlier samples.

The strategies Model 4 and Model 5 presented slopes of the calibration (blue) and validation (red) lines equivalent, as well as the values of RMSEC and RMSEP, indicating a good concordance between the calibration and validation sets. Also, these two strategies showed results very similar to the Model 2, reference model to determine the stabilized fraction. The Model 6 presented the RMSEP value almost twice times higher than Model 2, indicating lower accuracy compared to the reference model.

However, it is noteworthy that the comparison of the regression models should not be based only in a visual inspection on the scatter plots of reference versus predict values and in the accuracy parameters. A statistical test must be applied to verify if there are significant differences between the strategies proposed (Models 3 to 6) and the reference model (Model 2).

Many authors compare the regression models using an F-test, this test consisted in divided the $RMSEP_A$ by $RMSEP_B$ and then compare the value with an F-distribution (Fisher–Snedecor distribution) with n_A and n_B degrees of freedom [29]. However, this test is conservative, requiring the independence of prediction errors between models, which cannot be guaranteed since in general we can have tendency (positive or negative) in both models. By this way, in this study we used the randomization test suggested by H. van der Voet [29], to compare the errors (based on RMSEP) in the determinations of the stabilized SOC by each strategy in the validation set. The main advantage of this test is that assumptions about normality or homoscedasticity of the data are not required (distribution-free). A MATLAB code for applying the randomization test in binary comparations can be found in [30], more details about this test can be found in [29,30].

The randomization test was performed by binary comparisons between each strategy proposed and the reference model - Model 2. There is no reason to compare the strategy Model 3 with the reference calibration model, since more than 90% of the samples were excluded from the validation

set, thus only 3 binary comparisons were performed. The randomization test was carried out with a significance level of probability of 0.05, the null (H_0) and alternative (H_1) hypothesis evaluated were:

H_0: $RMSEP_{strategy} = RMSEP_{Reference\ calibration\ model}$ (accuracy is equal);

H_1: $RMSEP_{strategy} \neq RMSEP_{Reference\ calibration\ model}$ (accuracy is not equal).

The results obtained for each randomization test are organized in Table III. The strategies Model 4 and Model 5 used to determine the stabilized SOC from the NIR spectra of total fraction shows the same predictive accuracy (p-value > 0.05) to the reference PLSR model built using the NIR spectra of MOM fraction. In other words, the strategies Model 4 and Model 5 are able to determine stabilized SOM from the NIR spectra of bulk soils with the same accuracy of the reference model, thus eliminating the extra step regarding soil fractionation to determine the stabilized SOC. It is worth mentioning that the Model 6, which uses the EPO method to remove the liable fraction from the total soil fraction spectra, did not present results as good as the reference model, however it is observed that this strategy was able to minimize the prediction errors in stabilized SOC from the NIR spectra of total soil fraction.

Table III. p-values of randomization t-test for all binary comparations

Binary comparation	p-value for validation set
Model 4 x Model 2	0.12
Model 5 x Model 2	0.38
Model 6 x Model 2	$1x10^{-5}$

CONCLUSIONS

This study presented four different strategies to determine the stabilized SOC content using the NIR spectra of total soil fraction (bulk soil) and multivariate calibration model based on PLSR. Among the four strategies proposed two of them (Models 4 and 5) could determine the stabilized SOC content with the same accuracy of the PLSR reference model, thereby eliminating the extra soil fractionation step that is essential to the monitoring of the carbon sequestration potential by agroecosystems that adopt ABC Plan technologies.

Based on the results obtained and all the advantages reported in this study, this environmentally friendly methodology can contribute for the implementation of the National Plan for Low Carbon Emission in Agriculture (ABC Plan), since the methodology could reduce the cost, time required, and wastes generated in SOC determinations.

In Memoriam

On the 25th of April, 2020, prof. Ronei Jesus Poppi passed away having the same destiny as millions of people who have lost their lives along the sanitarian crisis that afflicted the world. Ronei was a good friend, a shy and sweet person who was always surrounded by his students. His influence is felt on the field he has devoted his career. He will be sorely missed by his family, friends and the worldwide scientific community.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors thank “Instituto Nacional de Ciência e Tecnologia de Bioanalítica” (INCTBio), “Conselho Nacional de Desenvolvimento Científico e Tecnológico” CNPq, Brazil, [465389/2014-7; 303994/2017-7; 307434/2020-6], “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (CAPES, Brazil, Finance Code 001) and “Fundação de Amparo à Pesquisa do Estado de São Paulo” [FAPESP, Brazil, 2014/508673] for financial support.

REFERENCES

1. Ministério da Agricultura, Pecuária e Abastecimento. *Plano Setorial de Mitigação e Adaptação às Mudanças Climáticas para Consolidação da Economia de Baixa Emissão de Carbono na Agricultura – PLANO ABC*. Brasília: MAPA/ACS, **2012**. Available at: http://www.agricultura.gov.br/assuntos/sustentabilidade/plano-abc/plano-abc-agricultura-de-baixa-emissao-de-carbono [Accessed 16 January 2021].
2. Cambardella, C. A.; Elliott, E. T. *Soil Sci. Soc. Am. J.*, **1992**, *56*, pp 1-777 (https://doi.org/10.2136/sssaj1992.03615995005600030017x).
3. Roscoe, R.; Machado, P. L. O. de A. *Fracionamento Físico do Solo em Estudos da Matéria Orgânica*. Embrapa agropecuária Oeste, Dourados, MS, **2002**.
4. Feller, C.; Beare, M. H. *Geoderma*, **1997**, *79*, pp 69-116 (https://doi.org/10.1016/S0016-7061(97)00039-6).
5. de Santana, F. B.; de Souza, A. M.; Poppi, R. J. *Sci. Total Environ.*, **2019**, *658*, pp 895-900 (https://doi.org/10.1016/j.scitotenv.2018.12.263).
6. Teixeira, P. C.; Donagemma, G. K.; Fontana, A. I.; Teixeira, W. G. *Manual de métodos de analise de solo*. Embrapa, Brasília, **2017**. Available at: https://ainfo.cnptia.embrapa.br/digital/bitstream/item/181717/1/Manual-de-Metodos-de-Analise-de-Solo-2017.pdf [Accessed 16 January 2021].
7. Rossel, R. A. V.; Hicks, W. S. *Eur. J. Soil Sci.*, **2015**, *66*, pp 438-450 (https://doi.org/10.1111/ejss.12237).
8. Lefèvre, C.; Rekik, F.; Alcantara, V.; Wiese, L. *Soil Organic Carbon: the hidden potential*. Food and Agriculture Organization (FAO). Rome, **2017**. Available at: http://www.fao.org/3/i6937e/i6937e.pdf [Accessed 16 January 2021].
9. de Santana, F. B.; de Souza, A. M.; Poppi, R. J. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, **2018**, *191*, pp 454-462 (https://doi.org/10.1016/j.saa.2017.10.052).
10. Stenberg, B.; Rossel, R. A. V.; Mouazen, A. M.; Wetterlind, J.; Mouazen, M.; Wetterlind, J. *Adv. Agron.*, **2010**, *107*, pp 163-215 (https://doi.org/10.1016/S0065-2113(10)07005-7).
11. Rossel, R. A. V.; Behrens, T.; Ben-Dor, E.; Brown, D. J.; Demattê, J. A. M.; Shepherd, K. D.; Shi, Z.; Stenberg, B.; Stevens, A.; Adamchuk, V.; et al. *Earth-Science Rev.*, **2016**, *155*, pp 198-230 (https://doi.org/10.1016/j.earscirev.2016.01.012).
12. Jaconi, A.; Poeplau, C.; Ramirez-Lopez, L.; Van Wesemael, B.; Don, A. *Eur. J. Soil Sci.*, **2019**, *70*, pp 127-139 (https://doi.org/10.1111/ejss.12761).
13. Alvares, C. A.; Stape, J. L.; Sentelhas, P. C.; De Moraes Gonçalves, J. L.; Sparovek, G. *Meteorol. Zeitschrift*, **2013**, *22*, pp 711-728 (https://doi.org/10.1127/0941-2948/2013/0507).
14. EMBRAPA *Sistema brasileiro de classificação de solos*; Solos, E., Ed.; 5th ed.; Rio de Janeiro - Brasil, **2018**.
15. United States Department of Agriculture (USDA), Natural Resources Conservation Service. *Keys to soil taxonomy*, 12th Ed., Washington, **2014**. Available at: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/class/taxonomy/?cid=nrcs142p2_053580 [Accessed 16 January 2021].
16. Ab’Sáber, A. N. *Domínios morfoclimáticos e províncias fitogeográficas do Brasil*. “Departamento de Geografia, Universidade de São Paulo”, **1967**.
17. Bongiovanni, M. D.; Lobartini, J. C. *Geoderma*, **2006**, *136*, pp 660-665 (https://doi.org/10.1016/j.geoderma.2006.05.002).
18. Savitzky, A.; Golay, M. J. E. *Anal. Chem.*, **1964**, *36*, pp 1627-1639 (https://doi.org/10.1021/ac60214a047).
19. Bro, R.; Smilde, A. K. *Anal. Methods*, **2014**, *6*, pp 2812-2831 (https://doi.org/10.1039/c3ay41907j).
20. Pasquini, C. *Anal. Chim. Acta*, **2018**, *1026*, pp 8-36 (https://doi.org/10.1016/j.aca.2018.04.004).

21. Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta*, **1986**, *185*, pp 1-17 (https://doi.org/10.1016/0003-2670(86)80028-9).
22. Haaland, D. M.; Thomas, E. V. *Anal. Chem.*, **1988**, *60*, pp 1193-1202 (https://doi.org/10.1021/ac00162a020).
23. Westad, F.; Marini, F. *Anal. Chim. Acta*, **2015**, *893*, pp 14-24 (https://doi.org/10.1016/j.aca.2015.06.056).
24. American Society for Testing Materials. ASTM E1655-17. *Standard Practices for Infrared Multivariate Quantitative Analysis*. West Conshohocken, PA, US, **2017**.
25. Roger, J. M.; Chauchard, F.; Bellon-Maurel, V. *Chemom. Intell. Lab. Syst.*, **2003**, *66*, pp 191-204 (https://doi.org/10.1016/S0169-7439(03)00051-0).
26. de Santana, F. B.; de Giuseppe, L. O.; de Souza, A. M.; Poppi, R. J. *Microchem. J.*, **2019**, *145*, pp 1094-1101 (https://doi.org/10.1016/j.microc.2018.12.027).
27. Minasny, B.; McBratney, A. B.; Bellon-Maurel, V.; Roger, J. M.; Gobrecht, A.; Ferrand, L.; Joalland, S. *Geoderma*, **2011**, *167-168*, pp 118-124 (https://doi.org/10.1016/j.geoderma.2011.09.008).
28. Stenberg, B.; Rossel, R. A. V.; Mouazen, M.; Wetterlind, J. *Adv. Agron.*, **2010**, *107*, pp 163-215 (https://doi.org/10.1016/s0065-2113(10)07005-7).
29. van der Voet, H. *Chemom. Intell. Lab. Syst.*, **1994**, *25*, pp 313-323 (https://doi.org/10.1016/0169-7439(94)85050-X).
30. Olivieri, A. C. *Anal. Chim. Acta*, **2015**, *868*, pp 10-22 (https://doi.org/10.1016/j.aca.2015.01.017).

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 91–106
doi: 10.30744/brjac.2179-3425.AR-11-2021

BrJAC

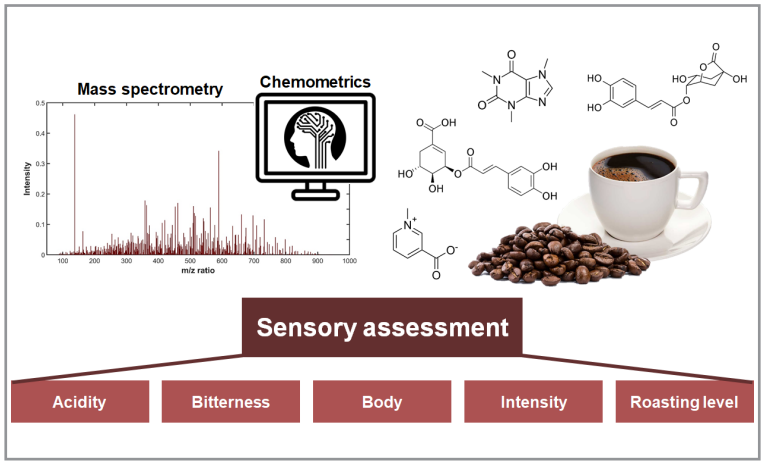
ARTICLE

Fast High–Resolution Mass Spectrometry and Chemometrics for Evaluation of Sensory Parameters of Commercial Coffee Blends

Victor Gustavo Kelis Cardoso[1], Guilherme Post Sabin[1,2], Leandro Wang Hantao[1]*

[1]Instituto de Química, Universidade Estadual de Campinas, Rua Monteiro Lobato, 270, 13083-862, Campinas, SP, Brazil

[2]OpenScience, Rua Conceição, 233, Sala 916, Centro, Campinas, SP, Brazil



This study presents a fast method to estimate sensory parameters of commercial capsules of roasted coffee using flow injection analysis coupled to high–resolution mass spectrometry (FIA–HRMS) as an alternative to traditional sensory analysis, which is a laborious and subjective method. Over 25 types of coffee capsules were studied. The samples were partitioned into an aqueous and organic extract, which were analyzed by FIA–HRMS in the positive and negative ionization modes. Data fusion of such mass spectra was performed to explore the complementary information of sample preparation and ionization conditions. Orthogonalized partial least square discriminant analysis (OPLS–DA) models were built and trained to determine the type of capsule and to estimate important coffee parameters (e.g., acidity, bitterness, body, intensity, and roasting level), achieving accuracy values higher than 91.1%. In addition, variable importance in projection (VIP) scores enabled assignment of the elemental composition and, in some cases, putative identification of compounds in coffee (e.g., caffeine, caffearine, and quinides) that exhibited an important role in class discrimination.

Keywords: chemical fingerprint, chemometrics, food and beverage analysis, high throughput, sensory assessment.

INTRODUCTION

Coffee is one of the most consumed beverages worldwide, attracting a large number of consumers due to its enjoyable flavor [1-3]. Coffee has played an important role in the global economy with a total trade of US$30 bi in 2019 [4]. In the same year, Brazil was the top exporter of coffee responsible for almost half

Cite: Cardoso, V. G. K.; Sabin, G. P.; Hantao, L. W. Fast High–Resolution Mass Spectrometry and Chemometrics for Evaluation of Sensory Parameters of Commercial Coffee Blends. *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 91–106. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-11-2021

Submitted 20 January 2021, Resubmitted 12 April 2021, Accepted 20 April 2021, Available online 03 May 2021.

of the global trade together with Vietnam, Colombia, Germany, and Switzerland, in this order of export volume [4]. In this context, the coffee industry requires extensive and oriented product development to meet current consumer demands in an ever increasing and diversified coffee market. Numerous types of coffee can be prepared by blending different types of coffee and/or modifying process parameters to increase the quality and improve the sensory coverage of the product.

The evaluation of coffee relies on sensory analysis by estimating their acidity, bitterness, and body, for instance. Sensory assessment is the characterization of products based on their organoleptic properties using the five human senses [5]. This process is based on a range of methods to obtain responses about taste, smell, texture, appearance, and other properties [5]. However, sensory analysis is a complex and time-consuming task that depends on professional tasters [6]. Despite the rigorous training of the sensory panel, sensory analysis remains a very subjective process and it is affected by physiological and physical factors, which may result in low accuracy and poor repeatability [6,7]. In the coffee industry, sensory analysis is even more challenging due to the overwhelming quantity of samples [7].

Alternatives to sensory analysis are instrumental techniques that have been applied to estimate the sensory parameters of coffee, like vibrational spectroscopy [8,9] and chromatography coupled to mass spectrometry (MS) [10-12]. The former approach may be limited by the low selectivity of the analytical signal, while chromatographic methods may struggle to provide compatible sample throughput required for large studies, such as over 200 samples per day. An interesting alternative to contemporary techniques is direct analysis using high-resolution mass spectrometry (HRMS) for fingerprinting purposes. After suitable sample preparation, HRMS analyses can provide molecular-level information in just a few seconds [13-16], which is an important feature for coffee analysis. Liquid samples may be introduced into the HRMS by flow injection analysis (FIA–HRMS), providing fast and reliable results [15,17]. FIA–HRMS–based methods have been successfully applied in the analysis of glyphosate in fruits, berries, and cereals [18], milk adulteration [19], discrimination of peppermints [20], and aerosols assessment in tobacco products [21].

Despite the obvious advantages of HRMS, the direct analysis may be unsuitable for the identification of sensory active compounds, providing instead only putative identification. In order to fully explore the potential of HRMS for fingerprinting, multivariate data analysis is mandatory to establish meaningful and statistically valid correlations between mass spectra and sensory properties, especially if very large data comprising hundreds of mass spectra are available for chemometric modeling [15-17].

In this study, we report an analytical method for coffee classification using FIA–HRMS and chemometric modelling. The FIA–HRMS provided the ideal platform for high sample throughput analyses compatible with the logistics of a typical coffee cooperative. OPLS–DA [22] was used to train the statistical model and predict the sensory properties of coffee samples using the MS–based coffee fingerprinting [23]. Important properties were successfully assigned using the OPLS–DA model, like the type of capsule and many sensory parameters (e.g., acidity, bitterness, body, intensity, and roasting level). We believe that this solution represents an important development in the coffee industry, providing rigorous quality control and intelligence for coffee production, moving towards the industry 4.0 [24].

MATERIALS AND METHODS

Reagents

Water, methanol, chloroform, formic acid, and acetonitrile (Chromasolv – MS grade) were purchased from Honeywell Riedel–de–Haën™ (Germany). Ultrapure water (resistivity below 18.2 MΩ.cm and TOC below 5 ppb) was obtained from a Milli–Q® purifier (Millipore®, USA). All the materials used in the analytical procedures were carefully washed using MS grade solvents and/or ultrapure water.

Coffee samples

The study used 25 capsules of espresso coffee produced by the same manufacturer, containing different blends and sensory properties. The samples were purchased in the local market. The manufacturer

provided coffee blend information about acidity, bitterness, body, intensity, and roasting on its website, which is also available in Table S1 (Supplementary Material). The samples were analyzed in triplicates on the same day over three different days, totalizing 9 analyses per sample. For 25 samples, 225 analyses were performed.

Sample preparation

Approximately 100 mg of coffee from the capsule was weighed in analytical balance ATX–224 (Shimadzu Corporation, Japan) and placed in 2 mL Eppendorf Tubes. The compounds were extracted using a modified Folch method through a ultrasound assisted dispersive liquid-liquid microextraction (UA-DLLME) [25,26], which was performed by adding a mixture of 400 µL of water and 400 µL of methanol, and then 800 µL of chloroform. The extractions were assisted by ultrasound for 15 min in an Q5.9L ultrasound bath (Eco–Sonics, Brazil), followed by vortexing at 3200 rpm during 15 min in a Vortex–Genie 2 (Scientific Industries, United States), and centrifuged at 13400 rpm for 15 min in a MiniSpin® (Eppendorf, Germany). After phase separation, the organic extract was composed of chloroform and the aqueous extract was a mixture of water and methanol. Each extract was filtered through Millex® syringe filters (Merck, United States) with a diameter of 13 mm and 0.22 µm pore size membranes, being polyvinylidene fluoride (PVDF) and polytetrafluoroethylene (PTFE) membranes to filtrate aqueous (hydrophilic) and organic (hydrophobic) extracts, respectively. Aliquots containing 80 µm of extract were transferred to 2 mL vials and diluted 1–20 with acetonitrile for the organic extract, and 1:1 (v/v) acetonitrile:water solution for the aqueous extract. After those steps of sample preparation, the samples were analyzed by FIA–HRMS. Blank samples were prepared using the extraction procedure without powdered coffee.

Instrumentation

The analyses were performed using a Xevo G2–XS hybrid quadrupole time–of–flight sequential mass spectrometer (QTOFMS) (Waters Corporation, United States). The FIA system employed an ultra–high performance liquid chromatography (UHPLC) to introduce the sample into the QTOF. The UHPLC autosampler was directly connected to the electrospray ionization (ESI) source, bypassing the use of a chromatographic column. An injection volume of 0.3 µL and a constant volumetric flow rate of 100 µL min^{-1} were used for the analyses. The mobile phase was water acidified with 0.1% (v/v) formic acid. Each sample extract (organic and aqueous) was analyzed separately using ESI in the positive (ESI+) and negative (ESI–) modes. The mass spectra were recorded from 100 Da to 1000 Da in the high–resolution mode. The mass resolving power was approximately 30,000 at *m/z* 400. The interested reader is directed elsewhere for a complete description of the MS conditions [25,26]. A total of four mass spectra per sample were available for modelling. Before the analysis, a system suitability check was performed, i.e., detector check and mass calibration. A solution of leucine enkephalin (1 µg mL^{-1}) was used for mass correction.

Data analysis

By combinatorics, two extraction phases and two ionization modes provided four blocks of mass spectra, which were imported to MATLAB R2016b (MathWorks, United States). These blocks were normalized by the length and concatenated into a unique data matrix by data fusion of the individual MS analyses, containing 5253 *m/z* signals for each one of 225 fused mass spectra. The *m/z* peaks were filtered by removing signals with relative intensity lower than 0.001%. Next, the signals were scaled by Pareto scaling to reduce signal magnitude (mask effect) without amplifying deviations and spectral noise. The 225 mass spectra were randomly organized and divided into training (60%) and validation (40%) data, being 135 and 90 spectra, respectively. Also, we have kept the same-day replicates together.

PLS_Toolbox 8.6 (Eigenvector, United States) was used to train the OPLS–DA [22] models for each property modelled (i.e., acidity, bitterness, body, intensity, and roasting level, and type of capsule). To train discriminant analysis models through OPLS, we have used OPLS2–DA, which models the classes through a single model instead of the OPLS1–DA, which models the classes singly. The selection of the number

of latent variables (LV) was according to the lowest average error of cross–validation (CV), which was performed by 10–fold venetian blinds. To ensure no overfitting in the trained models, the permutation test was performed by random permutation of sample labels for 50 times, aiming to find significant differences between permuted and unpermuted models [15,27]. The VIP score algorithm was applied to identify the most important variables for class discrimination, providing a suitable variable selection [28,29]. Thus, the *m/z* peaks that presented VIP scores below 1 were excluded from the data of each model, since they were not relevant for class discrimination [28,29]. Finally, the models were trained again, following the same criterion to select the number of LV.

The results were evaluated using the accuracy values (Equation 1), instead of the sensitivity and specificity values, due to the high number of classes. The accuracy takes into account the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) by predicting the test sets [30]. Sensitivity (Equation 2) and specificity (Equation 3) were also used to evaluate the trained models [30].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad \text{Equation 1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad \text{Equation 2}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \qquad \text{Equation 3}$$

RESULTS AND DISCUSSION

FIA-HRMS analysis

An important concern about sample preparation was the quality of the extraction to obtain a reliable measurement by preserving the original characteristics of the sample matrix [31,32]. Hence, mild extraction conditions were employed to avoid sample deterioration and/or contamination. A modified Folch method [25,26] was used to obtain the hydrophilic and hydrophobic extracts from coffee to increase the molecular coverage of the mass spectra (i.e., a broad range of polarity and molecular weights). Hydrophilic extraction promoted the extraction of polar compounds, such as sugars, alkaloids, acids, quinides, and polyphenols [11]. Conversely, hydrophobic extraction promoted the extraction of nonpolar compounds, such as fatty acids, terpenes, and carotenoids [11]. Figure 1 shows the mass spectra of a COL sample highlighting the diversity of *m/z* peaks according to the ionization mode and extraction steps. Each block of mass spectra contained important and complementary information for class discrimination. Therefore, the individual blocks were normalized and concatenated for chemometric modelling using data fusion.

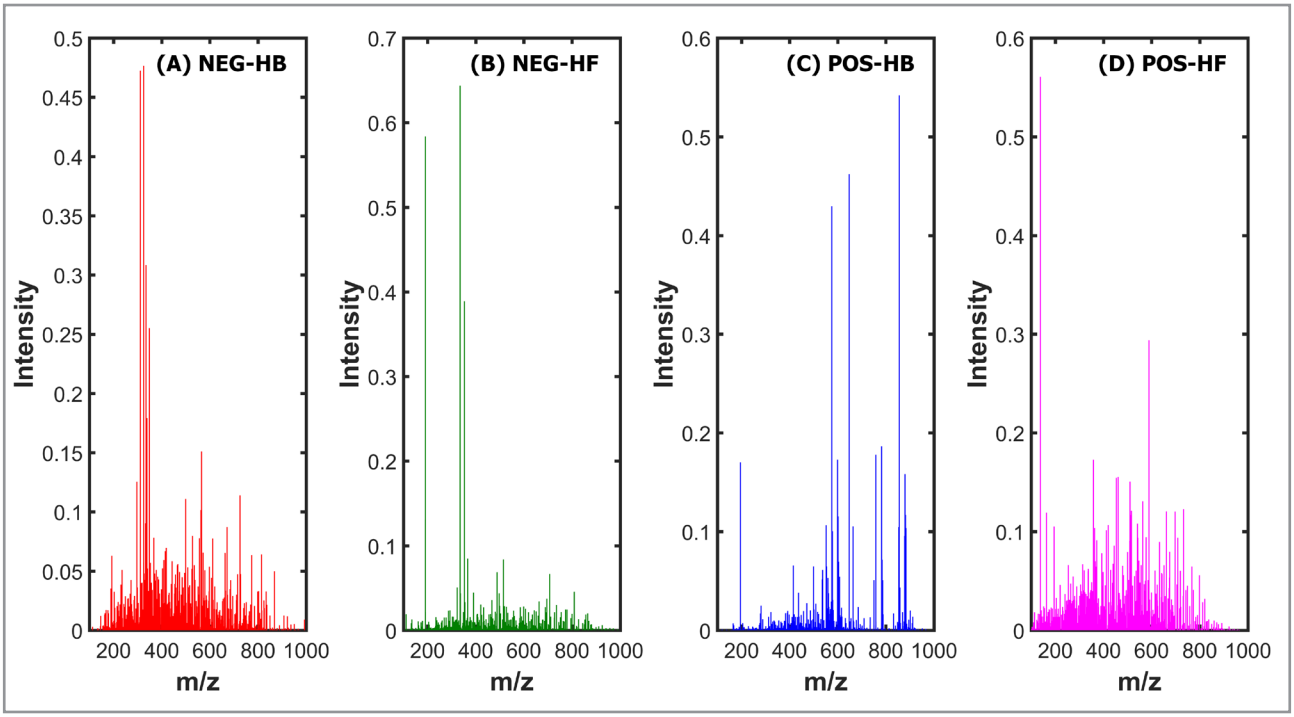


Figure 1. Mass spectra of a COL coffee obtained by the FIA–HRMS analysis and (a) hydrophobic extraction and negative ionization (NEG–HB), (b) hydrophilic extraction and negative ionization (NEG–HF), (c) hydrophobic extraction and positive ionization (POS–HB), and (d) hydrophilic extraction and positive ionization (POS–HF).

OPLS–DA models for sensory parameter evaluation

OPLS–DA models were trained for each sensory property evaluated, namely, acidity, bitterness, body, intensity, and roasting level, and type of capsule. The number of LV was estimated using CV, as shown in Figure 2. The high number of LV was explained due to the challenging heterogeneity of the sample composition and the extremely broad set of sensory properties. For example, the data input consisted of over 25 types of coffee blends with different species and origins, in addition to the four blocks of mass spectra (both extraction methods and ionization modes). The permutation test alongside the training sets ascertained the lack of model overfitting. Noteworthy, the number of LV depended on the sensory property, ranging from 33 to 57 LVs, which is related to many non-correlated *m/z* values (Table I), to the variability of samples (25 types of capsules), and to the number of classes (from 5 to 25 modelled classes). Thus, a good analytical system must capture various variation sources as cited above, increasing the number of LV as in these OPLS–DA models. The blocks X and y presented at least 95% and 98% of the cumulative explained variance, respectively, as shown in Table I.

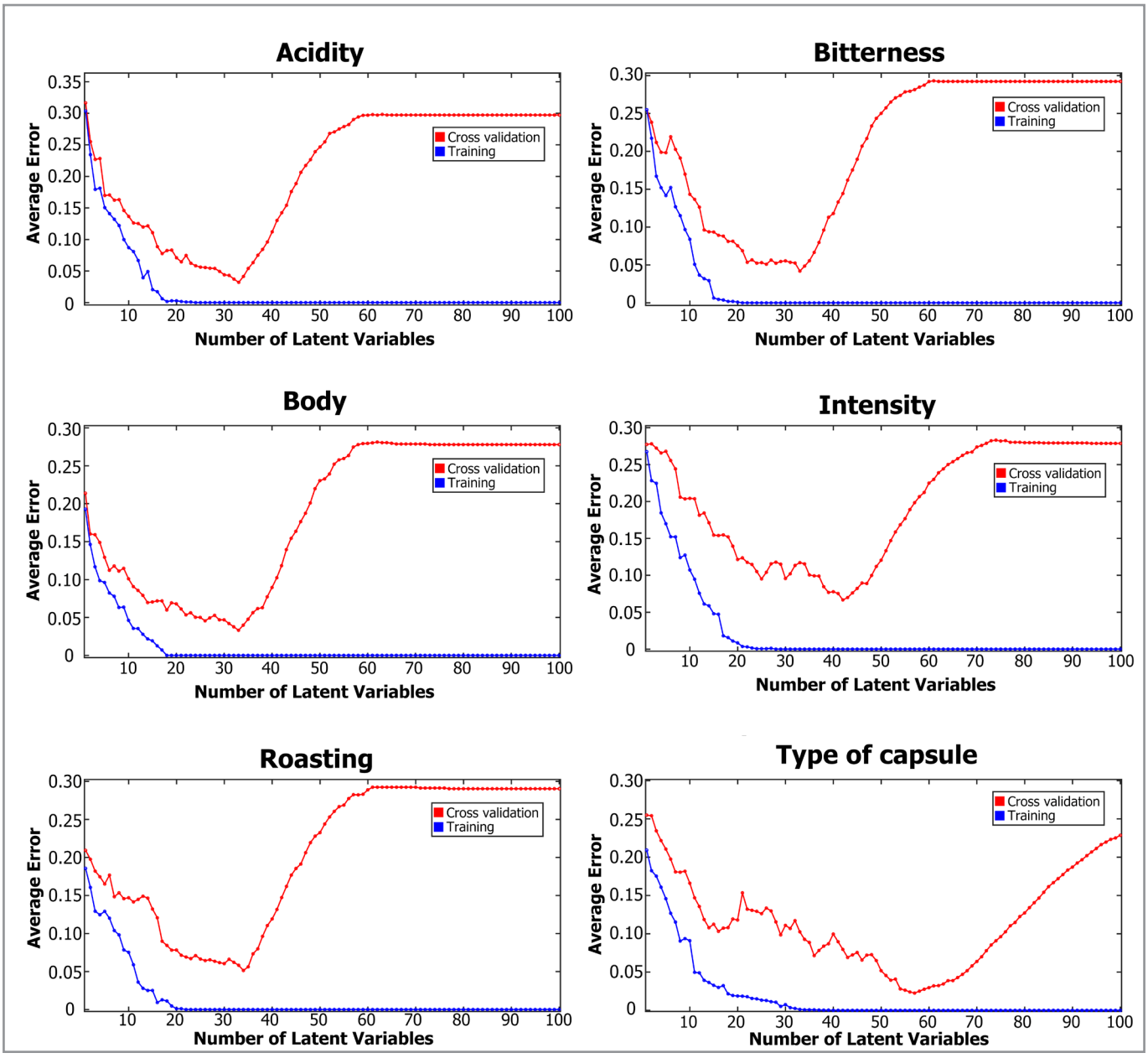


Figure 2. The average error of training and CV of OPLS–DA models according to the number of latent variables for each sensory assessment modelled.

Overall, the models presented an excellent performance to discriminate samples of the validation set, providing high accuracy values (Table I). The worst models evaluated the bitterness and intensity of coffees, achieving 91.1% of accuracy due to the misclassification of 8 amongst 90 samples (Figure 3). Conversely, evaluation of acidity provided the best results with an accuracy of 97.8% with only 2 mismatched from 90 samples (Figure 3). Also, Tables S2–S7 (Supplementary Material) provide the sensitivity and specificity values for each class and sensory property. Overall, the minimum values were 66.7% for sensitivity and 93.1% for specificity, but more than 75% of the sensitivity values were higher than 90%, indicating excellent performance. Such results indicate that the predictive models are well–fitted, i.e., without under– or overfitting.

Table I. Information about OPLS–DA models and the sensory properties

Parameter	Cumulative variance on X-block (%)	Cumulative variance on y-block (%)	Number of Variables	Number of LV	Accuracy of validation set (%)
Acidity	95.22	99.02	1638	33	97.8
Bitterness	95.49	99.03	1500	33	91.1
Body	95.85	99.11	1534	33	93.3
Intensity	97.17	99.15	1778	42	91.1
Roasting Level	95.76	99.23	1625	34	96.7
Type of Capsule	97.83	98.43	2164	57	93.3

The results provided by the models are very promising in the field of coffee analysis due to the high accuracy in the prediction of non-modelled and external samples (validation sets). This is an advantage over sensory analysis, which even performed by well–trained tasters are prone to large deviations between tasters [33,34]. Similarly, mid– [35,36] and near-infrared [8,37] spectroscopy have been used to estimate parameters of coffee, but achieving high accuracy is still a limitation due to the low selectivity of the spectral bands. This fact ascertains a benefit of using rich–information MS–based methods to obtain reliable models. For instance, our acidity model exhibited 97.8% of accuracy, whereas a previous model featured results with sensibility as low as 75% [35]. In addition, the body model presented in this study provided 93.3% of accuracy, compared to a previous model with a correlation coefficient lower than 0.8 [8].

The high accuracies provided by FIA–HRMS indicate that this method has high scalability, i.e., it can be easily automatized, being ideal for industrial applications to analyze coffee. Furthermore, the high sample throughput is important to significantly expand the number of samples analyzed per day when compared to conventional sensory analysis [15,21]. Although FIA–HRMS might exhibit a higher capital expenditure (CAPEX) is substantially higher due to the instrument price, compared to spectroscopy instruments, its operational expenditure (OPEX) is much more attractive, costing less than US$2 per analysis.

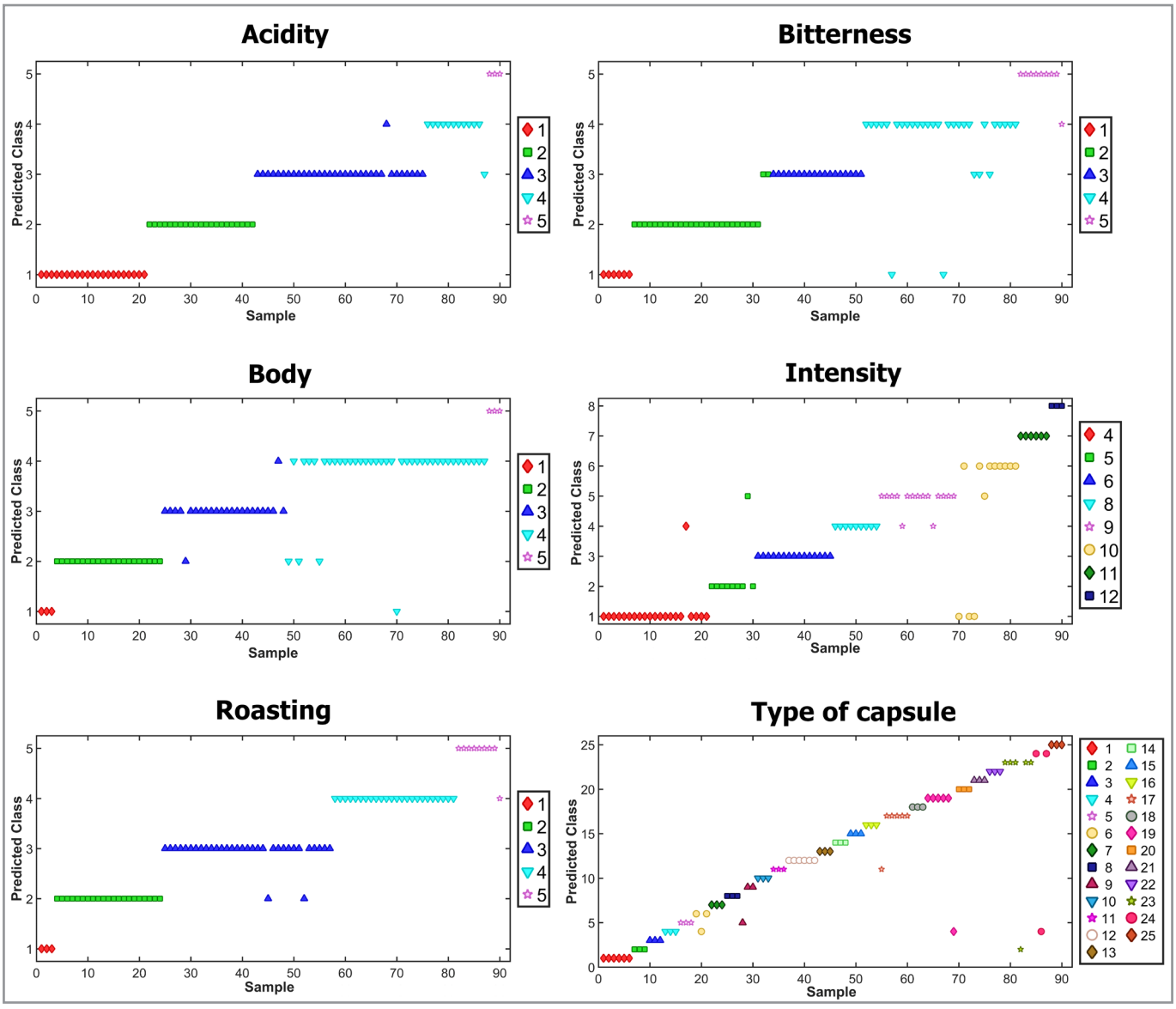


Figure 3. Predicted classes of validation sets of OPLS–DA models for each sensory assessment.

Evaluation of important variables

VIP scores algorithm was used to evaluate the most important variables to discriminate classes in all trained models. Table II presents the *m/z* peaks of the signals with the highest VIP score values alongside the respective elemental composition. A putative identification was supplied for common analytes found in coffee. The VIP scores threshold chosen was 4 for acidity, bitterness, body, roasting, and type of capsule, and 5 for the intensity model due to a larger number of important compounds in the latter.

Table II. VIP assignment of the OPLS-DA models and the sensory properties

	Adduct type	Measured *m/z*	Accurate mass	Empirical formula	DBE	Error (ppm)	Putative identification	Property modelled
NEG – HB	$[M-H]^-$	265.1465	266.1538	$C_{16}H_{18}N_4$	10	4.5	-	4,6
	$[M-H]^-$	293.1768	294.1841	$C_{18}H_{22}N_4$	10	0.7	-	4,6
	$[M-H]^-$	335.0769	336.0842	$C_{16}H_{16}O_8$	9	0.7	Caffeoyl-quinides [38,39] and caffeoyl-shikimic acids [39,40]	1,2,3,4,5,6
	$[M-H]^-$	349.0926	350.0999	$C_{17}H_{18}O_8$	9	0.8	Feruloyl-quinides [38-40]	2,3,4,6
	$[M-H]^-$	353.0912	354.0985	-	-	-	-	1,2,3,5,6
	$[M-H]^-$	502.1079	503.1152	-	-	-	-	1,2,4,5
	$[M-H]^-$	565.3018	566.3091	$C_{25}H_{43}O_5N_8P$	9	0.4	Unknown monophosphate metabolite	1
	$[M-H]^-$	569.2942	570.3015	-	-	-	-	4
	$[M-H]^-$	725.4456	726.4529	-	-	-	-	6
NEG – HF	$[M-H]^-$	335.0766	336.0839	$C_{16}H_{16}O_8$	9	0.2	Caffeoyl-quinides [38,39] and caffeoylshikimic acids [39,40]	1,2,3,4,5,6
	$[M-H]^-$	341.8730	342.8803	-	-	-	-	1,2,3,4,5,6
	$[M-H]^-$	353.0876	354.0949	$C_{17}H_{18}O_8$	9	0.2	Caffeoylquinic acids [38-41]	1,2,3,4,5,6
POS – HB	$[M+H]^+$	195.0883	194.0810	$C_8H_{10}O_2N_4$	6	6	Caffeine [41]	1,2,3,4,5,6
	$[M+H]^+$	647.4593	646.4520	-	-	-	-	1,2,4,5,6
	$[M+H]^+$	758.5700	757.5627	-	-	-	-	3
	$[M+H]^+$	782.5699	781.5626	-	-	-	-	3
POS – HF	$[M+H]^+$	138.0557	137.0484	$C_7H_7O_2N$	5	9.2	Caffearine [41]	1,2,4,5,6
	$[M+H]^+$	138.0767	137.0694	$C_5H_7N_5$	5	1.4	-	4,6
	$[M+H]^+$	195.0882	194.0809	$C_8H_{12}O_2N_4$	1	5.5	Caffeine [41]	1,2,4,5,6
	$[M+H]^+$	229.1416	228.1343	$C_{12}H_{20}O_4$	3	5.7	-	6
	$[M+H]^+$	453.1675	452.1602	-	-	-	-	6
	$[M+H]^+$	589.2253	588.2180	-	-	-	-	1,2,4,5,6
	$[M+H]^+$	732.2697	731.2624	-	-	-	-	3,5,6

Properties: 1 – acidity, 2 – bitterness, 3 – body, 4 – intensity, 5 – roasting, 6 – type of capsule. DBE – double bond equivalent.

Some important peaks assigned by the VIP algorithm (Table II) were tentatively identified using previous reports of coffee analysis [38-41]. The positive blocks provided information based on nitrogen-containing compounds in coffee. Caffeine [41], the most famous compound in coffee, presented important relevance to discriminate classes in the hydrophobic and hydrophilic positive blocks. In turn, caffearine [41], an alkaloid, also known as trigonelline, presented relevance only in the latter block. The negative block provided information based on oxygen–containing compounds. Caffeoyl–quinides [38,39] and caffeoyl shikimic acid [39,40] information presented relevance in both negative blocks, feruloyl–quinides [38-40] only in the hydrophobic negative block, and caffeoylquinic acids [38-41] only in the hydrophilic negative block.

CONCLUSIONS

This proof–of–concept study showed that FIA–HRMS combined with OPLS–DA can successfully estimate parameters of coffee blends, providing high values of accuracy. The models could estimate the acidity, bitterness, body, intensity, roasting level, and type of capsule with accuracy values higher than 91.1%, indicating excellent performance. The acidity estimation was the best model, achieving 97.8% of accuracy. VIP scores from OPLS–DA models indicated that various compounds in coffee already described in the literature, e.g., caffeine, quinides, and caffearine, were responsible for the class discrimination. We believe that this method has great potential for routine analysis in the coffee industry due to its high–throughput, scalability, and low OPEX, meeting the requirements of smart production and industry 4.0.

Conflicts of interest

The authors declare there are no conflicts of interest.

Acknowledgements

The São Paulo Research Foundation (FAPESP) [Grant N. 2020/01064-6], National Council for Scientific and Technological Development (CNPq) [Grant N. 302748/2018-0], Coordination for the Improvement of Higher Education Personnel (CAPES) (Finance code 001), Waters Corporation are acknowledged for supporting our research.

REFERENCES

1. Bona, E.; Marquetti, I.; Link, J. V.; Makimori, G. Y. F.; Arca, V. C.; Lemes, A. L.; Ferreira, J. M. G.; Scholz, M. B. S.; Valderrama, P.; Poppi, R. J. *LWT - Food Sci. Technol.*, **2017**, *76*, pp 330–336 (https://doi.org/10.1016/j.lwt.2016.04.048).
2. Forchetti, D. A. P.; Poppi, R. J. *Food Anal. Methods*, **2020**, *13*, pp 44–49 (https://doi.org/10.1007/s12161-019-01502-x).
3. Ayseli, M. T.; Kelebek, H.; Selli, S. *Food Chem.*, **2021**, *338*, 127821 (https://doi.org/10.1016/j.foodchem.2020.127821).
4. https://oec.world/en/profile/hs92/coffee?growthSelector=value2&yearSelector2=tradeYear24 [Accessed March 17, 2021].
5. Yu, P.; Low, M. Y.; Zhou, W. *Trends Food Sci. Technol.*, **2018**, *71*, pp 202–215 (https://doi.org/10.1016/j.tifs.2017.11.013).
6. Hernández-Ramos, P.; Vivar-Quintana, A. M.; Revilla, I.; González-Martín, M. I.; Hernández-Jiménez, M.; Martínez-Martín, I. *Sensors (Switzerland)*, **2020**, *20*, 5624 (https://doi.org/10.3390/s20195624).
7. Liu, P.; Zhu, X.; Hu, X.; Xiong, A.; Wen, J.; Li, H.; Ai, S.; Wu, R. *Vib. Spectrosc.*, **2019**, *103*, 102923 (https://doi.org/10.1016/j.vibspec.2019.05.005).
8. Baqueta, M. R.; Coqueiro, A.; Valderrama, P. *J. Food Sci.*, **2019**, *84*, 1247 (https://doi.org/10.1111/1750-3841.14617).
9. Ribeiro, J. S.; Ferreira, M. M. C.; Salva, T. J. G. *Talanta*, **2011**, *83*, 1352 (https://doi.org/10.1016/j.talanta.2010.11.001).

10. Bressanello, D.; Liberto, E.; Cordero, C.; Rubiolo, P.; Pellegrino, G.; Ruosi, M. R.; Bicchi, C. *Food Chem.*, **2017**, *214*, pp 218–226 (https://doi.org/10.1016/j.foodchem.2016.07.088).
11. Rocchetti, G.; Braceschi, P. G.; Odello, L.; Bertuzzi, T.; Trevisan, M.; Lucini, L. *Metabolomics*, **2020**, *16*, 127 (https://doi.org/10.1007/s11306-020-01751-6).
12. Pua, A.; Lau, H.; Liu, S. Q.; Tan, L. P.; Goh, R. M. V.; Lassabliere, B.; Leong, K. C.; Sun, J.; Cornuz, M.; Yu, B. *Food Chem.*, **2020**, *302*, 125370 (https://doi.org/10.1016/j.foodchem.2019.125370).
13. Quintanilla-Casas, B.; Dulsat-Serra, N.; Cortés-Francisco, N.; Caixach, J.; Vichi, S. *LWT - Food Sci. Technol.*, **2015**, *64*, 1085 (https://doi.org/10.1016/j.lwt.2015.07.010).
14. Sittipod, S.; Schwartz, E.; Paravisini, L.; Tello, E.; Peterson, D. G. *J. Agric. Food Chem.*, **2020**, *68*, 10424 (https://doi.org/10.1021/acs.jafc.0c01479).
15. Kaiser, S.; Dias, J. C.; Ardila, J. A.; Soares, F. L. F.; Marcelo, M. C. A.; Porte, L. M. F.; Gonçalves, C.; Canova, L. dos S.; Pontes, O. F. S.; Sabin, G. P. *Talanta*, **2018**, *190*, pp 363–374 (https://doi.org/10.1016/j.talanta.2018.08.007).
16. Schwanz, T. G.; Bokowski, L. V. V.; Marcelo, M. C. A.; Jandrey, A. C.; Dias, J. C.; Maximiano, D. H.; Canova, L. S.; Pontes, O. F. S.; Sabin, G. P.; Kaiser, S. *Talanta*, **2019**, *202*, pp 74–89 (https://doi.org/10.1016/j.talanta.2019.04.060).
17. Kaiser, S.; Soares, F. L. F.; Ardila, J. A.; Marcelo, M. C. A.; Dias, J. C.; Porte, L. M. F.; Gonçalves, C.; Pontes, O. F. S.; Sabin, G. P. *Chem. Res. Toxicol.*, **2018**, *31*, pp 964–973 (https://doi.org/10.1021/acs.chemrestox.8b00154).
18. Ciasca, B.; Pecorelli, I.; Lepore, L.; Paoloni, A.; Catucci, L.; Pascale, M.; Lattanzio, V. M. T. *Food Chem.*, **2020**, *310*, 125813 (https://doi.org/10.1016/j.foodchem.2019.125813).
19. Du, L.; Lu, W.; Cai, Z.; Bao, L.; Hartmann, C.; Gao, B.; Yu, L. *Food Chem.*, **2018**, *240*, pp 573–578 (https://doi.org/10.1016/j.foodchem.2017.07.107).
20. Gao, B.; Lu, Y.; Qin, F.; Chen, P.; Shi, H.; Charles, D.; Yu, L. *J. Agric. Food Chem.*, **2012**, *60*, pp 11987–11994 (https://doi.org/10.1021/jf303415d).
21. Soares, F. L. F.; Marcelo, M. C. A.; Dias, J. C.; Juliano, L. C.; Porte, L. M. F.; Canova, L. dos S.; Ardila, J. A.; Pontes, O. F. S.; Sabin, G. P.; Kaiser, S. *J. Chemom.*, **2020**, *34*, e3297 (https://doi.org/10.1002/cem.3297).
22. Ergon, R. *J. Chemom.*, **2005,** 19, pp 1–4 (https://doi.org/10.1002/cem.899).
23. Burgués, J.; Marco, S. *Anal. Chim. Acta,* **2018**, *1019*, pp 49–64 (https://doi.org/10.1016/j.aca.2018.03.005).
24. Lee, J.; Bagheri, B.; Kao, H. A. *Manuf. Lett.*, **2015**, *3*, pp 18–23 (https://doi.org/10.1016/j.mfglet.2014.12.001).
25. Draper, J.; Lloyd, A. J.; Goodacre, R.; Beckmann, M. *Metabolomics*, **2013**, *9*, pp 4–29 (https://doi.org/10.1007/s11306-012-0449-x).
26. Fuhrer, T.; Zamboni, N. *Curr. Opin. Biotechnol.*, **2015**, *31*, pp 73–78 (https://doi.org/10.1016/j.copbio.2014.08.006).
27. Szymańska, E.; Saccenti, E.; Smilde, A. K.; Westerhuis, J. A. *Metabolomics*, **2012**, 8, pp 3–16 (https://doi.org/10.1007/s11306-011-0330-3).
28. Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S. *Chemom. Intell. Lab. Syst*., **2012**, 118, pp 62–69 (https://doi.org/10.1016/j.chemolab.2012.07.010).
29. Cardoso, V. G. K.; Poppi, R. J. *Microchem. J*., **2021**, *164*, 106052 (https://doi.org/10.1016/j.microc.2021.106052).
30. Cardoso, V. G. K.; Poppi, R. J. *Food Control*, **2021**, 125, 107917 (https://doi.org/10.1016/j.foodcont.2021.107917).
31. Kim, H. K.; Verpoorte, R. *Phytochem. Anal.*, **2010**, *21*, pp 4–13 (https://doi.org/10.1002/pca.1188).
32. De Vos, R. C. H.; Moco, S.; Lommen, A.; Keurentjes, J. J. B.; Bino, R. J.; Hall, R. D. *Nat. Protoc.*, **2007**, *2*, pp 778–791 (https://doi.org/10.1038/nprot.2007.95).

33. Pereira, L. L.; Cardoso, W. S.; Guarçoni, R. C.; da Fonseca, A. F. A.; Moreira, T. R.; Caten, C. S. *Eur. Food Res. Technol.*, **2017**, *243*, pp 1545–1554 (https://doi.org/10.1007/s00217-017-2863-9).
34. Nebesny, E.; Budryn, G. *Eur. Food Res. Technol.*, **2006**, *224*, pp 159–165 (https://doi.org/10.1007/s00217-006-0308-y).
35. Belchior, V.; Botelho, B. G.; Oliveira, L. S.; Franca, A. S. *Food Chem.*, **2019**, *273*, pp 178–185 (https://doi.org/10.1016/j.foodchem.2017.12.026).
36. Craig, A. P.; Botelho, B. G.; Oliveira, L. S.; Franca, A. S. *Food Chem.*, **2018**, *245*, pp 1052–1061 (https://doi.org/10.1016/j.foodchem.2017.11.066).
37. Baqueta, M. R.; Coqueiro, A.; Março, P. H.; Valderrama, P. *Food Anal. Methods*, **2020**, *13*, pp 50–60 (https://doi.org/10.1007/s12161-019-01503-w).
38. Frank, O.; Zehentbauer, G.; Hofmann, T. *Dev. Food Sci.*, **2006**, *43*, pp 165–168 (https://doi.org/10.1016/S0167-4501(06)80039-7).
39. Jaiswal, R.; Matei, M. F.; Golon, A.; Witt, M.; Kuhnert, N. *Food Funct.*, **2012**, *3*, pp 976–984 (https://doi.org/10.1039/c2fo10260a).
40. Correia, R. M.; Loureiro, L. B.; Rodrigues, R. R. T.; Costa, H. B.; Oliveira, B. G.; Filgueiras, P. R.; Thompson, C. J.; Lacerda Jr, V.; Romão, W. *Anal. Methods*, **2016**, *8*, pp 7678–7688 (https://doi.org/10.1039/c6ay02501c).
41. Jeszka-Skowron, M.; Zgoła-Grześkowiak, A.; Grześkowiak, T. *Eur. Food Res. Technol.*, **2015**, *240*, pp 19–31 (https://doi.org/10.1007/s00217-014-2356-z).

SUPPLEMENTARY MATERIAL

Table S1. Codes and sensory assessment of each class used in to train the models

Code	Acidity	Bitterness	Body	Intensity	Roasting level	Type of capsule
ARP	2	4	4	9	4	1
ARP_DEC*	2	4	4	9	4	2
CAP	3	3	2	5	2	3
CAR	3	3	3	6	3	4
CIO	3	3	3	6	3	5
COL	4	3	3	5	3	6
COS	5	1	1	4	1	7
DHA	1	5	4	11	5	8
ENV	1	4	4	9	4	9
ETH	4	2	2	4	2	10
FOR	2	3	3	8	4	11
IND	1	5	4	11	5	12
INN	2	3	4	8	4	13
KAZ	1	5	5	12	5	14
LIN	1	1	2	4	2	15
LIV	3	3	3	6	3	16
NIC	2	2	2	5	2	17
RIS	3	4	4	10	4	18
RIS_DEC*	3	4	4	10	4	19
ROM	4	4	3	8	3	20
VAN	3	3	3	6	3	21
VIV	2	2	2	4	3	22
VIV_DEC*	2	2	2	4	3	23
VOL	3	2	2	4	2	24
VOL_DEC*	3	2	2	4	2	25

*Decaffeinated capsules

Table S2. Sensitivity and specificity of validation set of acidity model according to each class

Acidity		
Class	**Sensitivity**	**Specificity**
1	100%	100%
2	100%	100%
3	97.0%	98.2%
4	91.7%	98.7%
5	100%	100%

Table S3. Sensitivity and specificity of validation set of bitterness model according to each class

Bitterness		
Class	**Sensitivity**	**Specificity**
1	100%	97.6%
2	92.6%	100%
3	100%	93.1%
4	83.3%	98.3%
5	88.9%	100%

Table S4. Sensitivity and specificity of validation set of body model according to each class

Body		
Class	**Sensitivity**	**Specificity**
1	100%	99.0%
2	100%	94.2%
3	91.7%	100%
4	89.7%	98.0%
5	100%	100%

Table S5. Sensitivity and specificity of validation set of intensity model according to each class

Intensity		
Class	**Sensitivity**	**Specificity**
4	95.2%	95.7%
5	88.9%	100%
6	100%	100%
8	100%	96.3%
9	86.7%	97.3%
10	66.7%	100%
11	100%	100%
12	100%	100%

Table S6. Sensitivity and specificity of validation set of roasting-level model according to each class

Roasting level		
Class	**Sensitivity**	**Specificity**
1	100%	100%
2	100%	97.1%
3	93.9%	100%
4	100%	98.5%
5	88.9%	100%

Table S7. Sensitivity and specificity of validation set of type-of-capsule model according to each class

Type of Capsule		
Class	**Sensitivity**	**Specificity**
1	100%	100%
2	100%	98.9%
3	100%	100%
4	100%	96.6%
5	100%	98.9%

Table S7. Sensitivity and specificity of validation set of type-of-capsule model according to each class (Continuation)

Type of Capsule		
Class	**Sensitivity**	**Specificity**
6	66.7%	100%
7	100%	100%
8	100%	100%
9	66.7%	100%
10	100%	100%
11	100%	98.9%
12	100%	100%
13	100%	100%
14	100%	100%
15	100%	100%
16	100%	100%
17	83.3%	100%
18	100%	100%
19	83.3%	100%
20	100%	100%
21	100%	100%
22	100%	100%
23	83.3%	100%
24	66.7%	100%
25	100%	100%

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 107–115
doi: 10.30744/brjac.2179-3425.AR-12-2021

BrJAC

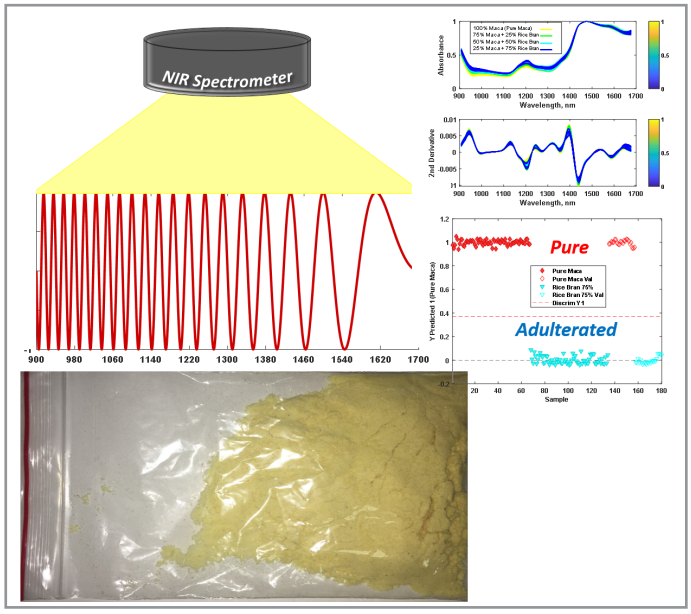
ARTICLE

The use of Chemometrics to Discriminate Sample Adulteration in Different Levels: the case of Peruvian Maca

Heloísa de Carvalho Rodrigues[1], Hellen Fernanda da Silva Paulino[2], Patrícia Valderrama[1], Paulo Henrique Março[1]*

[1]Programa de Pós-Graduação em Tecnologia de Alimentos, Universidade Tecnológica Federal do Paraná – Campus Campo Mourão, Via Rosalina Maria dos Santos, 1233, Área Urbanizada I, 87301-899, Campo Mourão, PR, Brazil

[2]Departamento de Engenharia de Alimentos, Universidade Tecnológica Federal do Paraná – Campus Campo Mourão, Via Rosalina Maria dos Santos, 1233, Área Urbanizada I, 87301-899, Campo Mourão, PR, Brazil



This study aimed to use chemometrics as a tool to allow for near-infrared (NIR) spectroscopy to detect adulteration in Peruvian maca powder. To adulterate the samples, pure maca powder was mixed with rice flour and rice bran, which are the most employed adulterants, using proportions of 25%, 50%, and 75% of each. After adulteration, the mixtures were submitted to NIR spectroscopy. The spectral data were used to build discriminant models through Partial Least Squares Discriminant Analysis (PLS-DA), and each model was evaluated through permutation tests to check for reliability. It was observed that all samples were discriminated against the pure ones, providing unitary sensitivities and specificities. All the models built were tested by the pairwise Wilcoxon signed-rank test, pairwise signed-rank test, and a randomization t-test, and the results indicated that the permuted and un-permuted models were not similar. The results indicated that NIR spectroscopy can be used to discriminate between pure maca powder samples and maca powder adulterated with different levels of rice flour and rice bran, achieving unitary sensitivity and specificity in the PLS-DA models.

Keywords: food analysis, fraud, Peruvian maca, chemometrics, PLS-DA.

Cite: Rodrigues, H. C.; Paulino, H. F. S.; Valderrama, P.; Março, P. H. The use of Chemometrics to Discriminate Sample Adulteration in Different Levels: the case of Peruvian Maca. *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 107–115. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-12-2021

Submitted 21 January 2021, Resubmitted 7 April 2021, 2nd time Resubmitted 16 April 2021, Accepted 20 April 2021, Available online 04 May 2021.

INTRODUCTION

As a consequence of the modernization of analytical techniques, in the food industry, the interpretation of the results obtained by chemical analysis referring to the composition of food increasingly demand chemometric strategies, to allow more assertive interventions. Food fraud and adulteration generally aim to increase profit, distorting the desired or required characteristics of the original product. These infractions are recurrent, often due to deficiency or lack of knowledge and or an inspection by the competent authorities. One of the reasons for that is due to the inspection for fraud detection which requires, in most cases, time-consuming analysis and/or expensive techniques, such as the ones derived from chromatography, often demanding sample preparation and trained operator. In this sense, food analysis is a field in which analytical chemistry is mandatory to allow for food authentication.

According to the Food and Agriculture Organization of the United Nations, the term “superfoods” was introduced in 2005 by the food and beverage industry to describe a range of fruit and vegetables thought to confer significant and essential components to human health and nutrition [1]. Nonetheless, there is a need to check not only the active substances which are present in it but also to verify the possibility of adulteration. One of the considered super-foods is the Peruvian maca (*Lepidium meyenii*), which is a species of tuber grown mainly in the high Andes region of Peru. The consumption of this cultivar is widespread among inhabitants of that region and is usually related to benefits for human health, with the effects attributed mainly to the content of the hypocotyl, the region of the plant where many of the most desired nutrients are concentrated [2]. Interest in maca has been increasing in recent years due to popular appeal related mainly to fertility. However, considering the form in which this plant is legally marketed (flour), makes it susceptible to adulteration by the addition of other flours of similar appearance, which is mainly due to the high market value [3].

Spectroscopic techniques such as in the ultraviolet (UV), near-infrared (NIR), mid-infrared (MIR) regions appear as possible alternatives for industrial processes that can lead to the detection of fraud, bringing advantages such as the analysis speed, little or no sample preparation, non-invasive and non-destructive characteristics, resulting in reduced costs compared to other consolidated techniques, that are used in standard analytical protocols, as shown in earlier studies as in the case of oil capsules [4], that uses to be evaluated by gas-chromatography [5]. However, the use of spectroscopic techniques is challenging when dealing with samples that produce signals with a high degree of similarity, being necessary to use multivariate statistical methods associated with the chemical knowledge of the samples to allow for a feasible interpretation [6].

Among these multivariate options, the partial least squares discriminant analysis (PLS-DA) [7] is a supervised pattern recognition method used for discrimination based on the assignment of classes to known data and has been successfully used for food analysis. In a previous study, we did apply PLS-DA to discriminate lettuce cultivated with and without pesticides evaluated through mass spectrometry [8], to discriminate roasted arabica coffee evaluated by NMR spectroscopy [9], to check for milk adulteration [10] beside other applications. Therefore, this study aimed to discriminate samples of Peruvian maca powder adulterated by adding different proportions of rice flour and rice bran. As far as we know, in Brazil, the main adulteration in this food is done by adding rice bran due to its organoleptic similarity, even though it has never been formally reported, which encourage us to offer this alternative methodology to be used for authentication purposes.

MATERIALS AND METHODS

Samples

Nine different samples of Peruvian maca were purchased in the trade of Campo Mourão (Paraná State, south of Brazil) and region. The samples were found in the form of finely divided powder (in bulk) and in capsules, totaling 9 different brands. Each sample was divided into 10 different portions containing 10 g of maca powder, totalizing 90 samples. The acquisition was done during the period from November 2018 to February 2019. Besides the Peruvian maca samples, though for adulteration purposes.

Sample preparation

The Peruvian maca was adulterated separately with rice flour and rice bran, choosing proportions considered as low (25%), half (50%), and high (75%) adulteration levels (percentage indicating w/w). These proportions were chosen considering the discriminant capability and the amount of sample available to the study. The weighing was performed using an analytical scale model 2204 Bioscale balance, being samples stored in individual plastic bags, duly identified, and stored under refrigeration to be measured the next day. Considering 9 different pure maca, the option was using 10 replicates of each sample, resulting in 90 pure maca samples, 10 pure rice flour samples, 10 pure rice bran samples, 90 samples adulterated with 25% rice flour, 90 samples adulterated with 25% rice bran, 90 samples adulterated with 50% rice flour, 90 samples adulterated with 50% rice bran, 90 samples adulterated with 75% rice flour, and 90 samples adulterated with 75% rice bran. The granulometry was not standardized to simulate a real “in field” application, in which the models would be able to be used for sample screening regarding adulteration.

Near-Infrared Spectroscopy

The NIR spectra were acquired using a JDSU MicroNIR 1700® spectrophotometer, located in the CAMulti laboratory at the UTFPR - Campus Campo Mourão. The measurements were performed using 20 scans for each spectrum. Samples were removed from the refrigerator two hours before starting the measurements to reach room temperature. From each sample, 10 replicates were produced, with the average spectrum of the 10 measurements used as representative of each sample for all data processing. The measurements were taken with the samples inside plastic packages, being the same procedure used for the calibration of 100% reflectance (with the reflectance sphere also packed in plastic) and turning off the light source to calibrate the zero reflectance.

Chemometrics

The chemometric procedures adopted were mainly related to the elimination of noise and irrelevant information. Multiplicative Scatter Correction (MSC), standard normal variation (SNV), and second derivatives were used, and the models obtained from these different treatments were compared in order to allow for the choice of the best ones in terms of sensibility/specificity [7]. To build the models, the Kennard-Stone [11] approach was used to select 67 samples for calibration and 23 samples for validation. The best models were built after mean-centering the spectra and performing MSC to the samples adulterated with rice bran. To the samples adulterated with rice flour, the best models also demanded performing second derivatives. All the models were built through cross-validation using the “venetian blinds” method with 10 splits and 1 sample per split.

After pre-processing, PLS-DA was applied to discriminate pure maca samples from 25% adulterated with rice flour (model 1). Then, pure maca versus 50% adulteration with rice flour (model 2) and so pure maca vs 75% adulterated with rice flour (model 3). The same procedure was done to the adulterant rice bran, providing model 4 (pure maca vs 25% adulteration with rice bran), model 5 (pure maca vs 50% adulteration with rice bran), and model 6 (pure maca vs 75% adulteration with rice bran). The spectral data were processed using Matlab R2017® and PLS Toolbox 7.8® provided by EMBRAPA Solos do Rio de Janeiro.

PLS-DA performance

To check for PLS-DA models’ performance, the diagnostic tests used were sensitivity and specificity, which details can be found somewhere else [7]. Moreover, the models were evaluated through permutation tests, which are tests in which the “y-block” is randomly reordered. The permutation tests consisted of verifying the Pairwise Wilcoxon signed rank test, Pairwise signed rank test, and a Randomization t-test according to Souza et al [8].

RESULTS AND DISCUSSION

The models built to discriminate between maca powder from rice bran demanded mean-centering and multiplicative scatter correction, while the models for the discrimination between maca and rice flour demanded, besides mean-centering and multiplicative scatter correction, the use of second derivatives. Otherwise, the models are not able to discriminate between these classes (pure vs contaminated). Table I presents the sensitivity, specificity, Pairwise Wilcoxon signed rank test (Wilcoxon), Pairwise signed rank test (Sign Test), and the Randomization *t*-test (Rand *t*-test) to each achieved model. *Model 1* refers to the discrimination between pure maca and maca adulterated with 25% rice bran; *model 2* to the discrimination between pure maca and adulteration with 50% rice bran, and *model 3* to the discrimination between pure maca and the adulterated with 75% rice bran. Models *4*, *5*, and *6* refer to the adulteration of 25%, 50%, and 75% with rice flour, respectively.

Table I. Parameters of merit of the discrimination models

Model	Latent Variables	Sensitivity	Specificity	Wilcoxon	Sign Test	Rand *t*-test
1	2	1.00	1.00	0.000	0.000	0.005
2	2	1.00	1.00	0.000	0.000	0.005
3	2	1.00	1.00	0.000	0.000	0.005
4	4	1.00	1.00	0.000	0.000	0.005
5	4	1.00	1.00	0.000	0.000	0.005
6	4	1.00	1.00	0.000	0.000	0.005

Figure 1 (A) presents the spectra of pure maca adulterated with 25%, 50%, and 75% of rice bran, and (B) its respective derivatives, while Figure 1 (C) presents the spectra of pure maca adulterated with 25%, 50% and 75% of rice flour and (D) its respective derivatives. It can be observed that the spectra are influenced by the adulteration, increasing vibration intensities below 1400 nm when adulterated with rice bran and decreasing it when adulteration is made by adding rice flour. Moreover, when the adulteration is done by adding rice bran, it is not possible to see any trend in wavelengths higher than 1500 nm, while for rice flour adulteration the absorbance to this region trends to go down. The rice bran is made from the rice skin (obtained from the polishing of the pealed rice) and rice flour is produced by grinding the rice grains. Therefore, besides the difference in the composition, they differ in the particle size, which is bigger in rice flour than in rice bran.

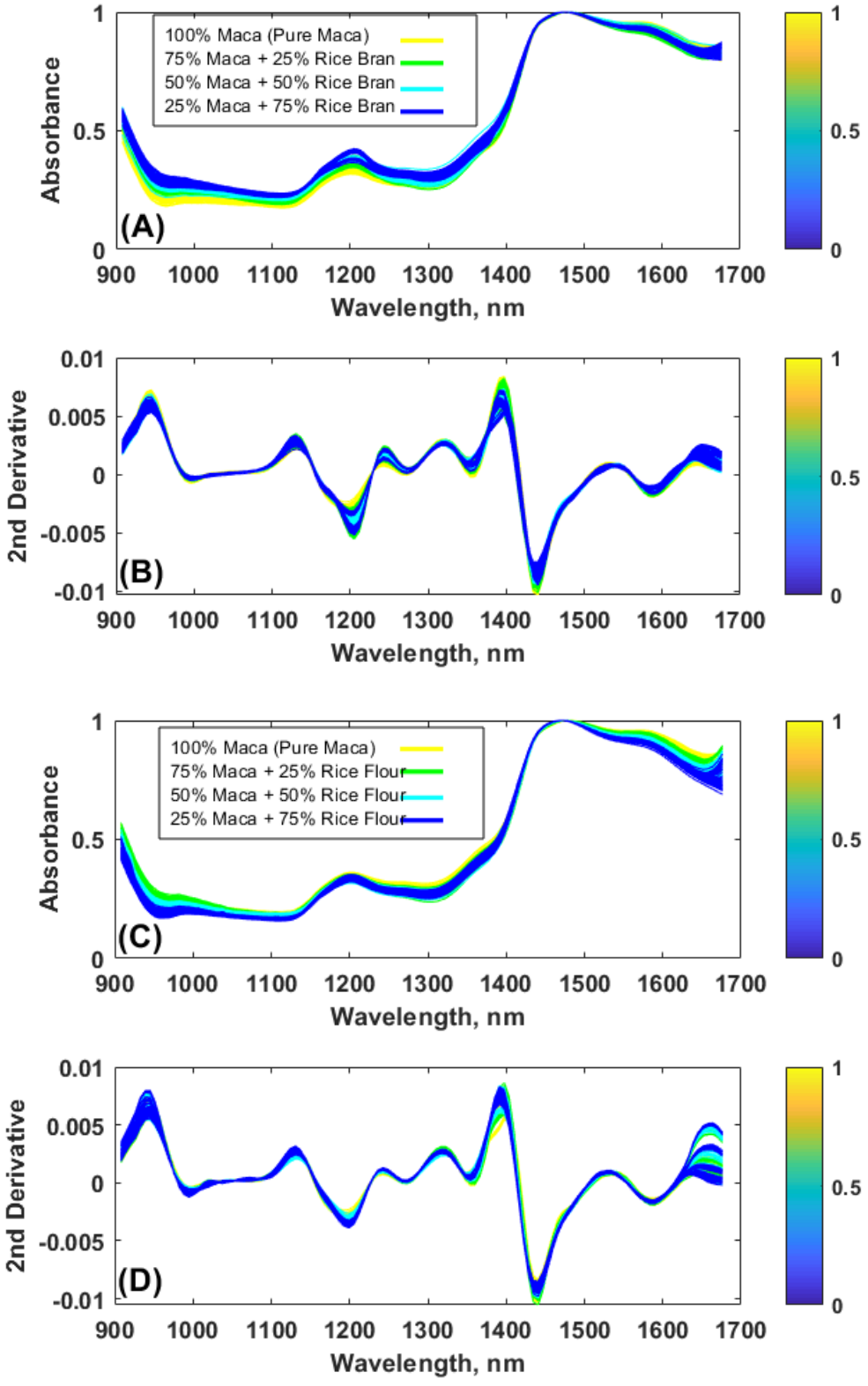


Figure 1. (**A**) Spectra of the pure maca and respective adulteration with rice bran and (**B**) its respective derivatives; (**C**) Spectra of the pure maca and respective adulteration with rice flour and (**D**) its respective derivatives.

Figure 2 presents the model's discrimination considering only class 1 (pure maca), to differentiate pure from the adulterated with different bran proportions and the respective VIP scores. The VIP scores indicate that the most important region to be verified is located around 1100–1200 nm, which is assigned to second overtones of methylenic stretches [12]. The regions of 960 nm, 1140 nm and 1200 nm correspond to OH, $R\text{-}NH_2$, and CH, and are highly prevalent in the protein molecule and can be related to the protein content in the samples [13]. Even though the region around 1400 nm, which is assigned to second combination bands of OH and CH bonds, is down the importance limit in the VIP scores (below 1), it is remarkable that this region increases together with the addition of rice bran.

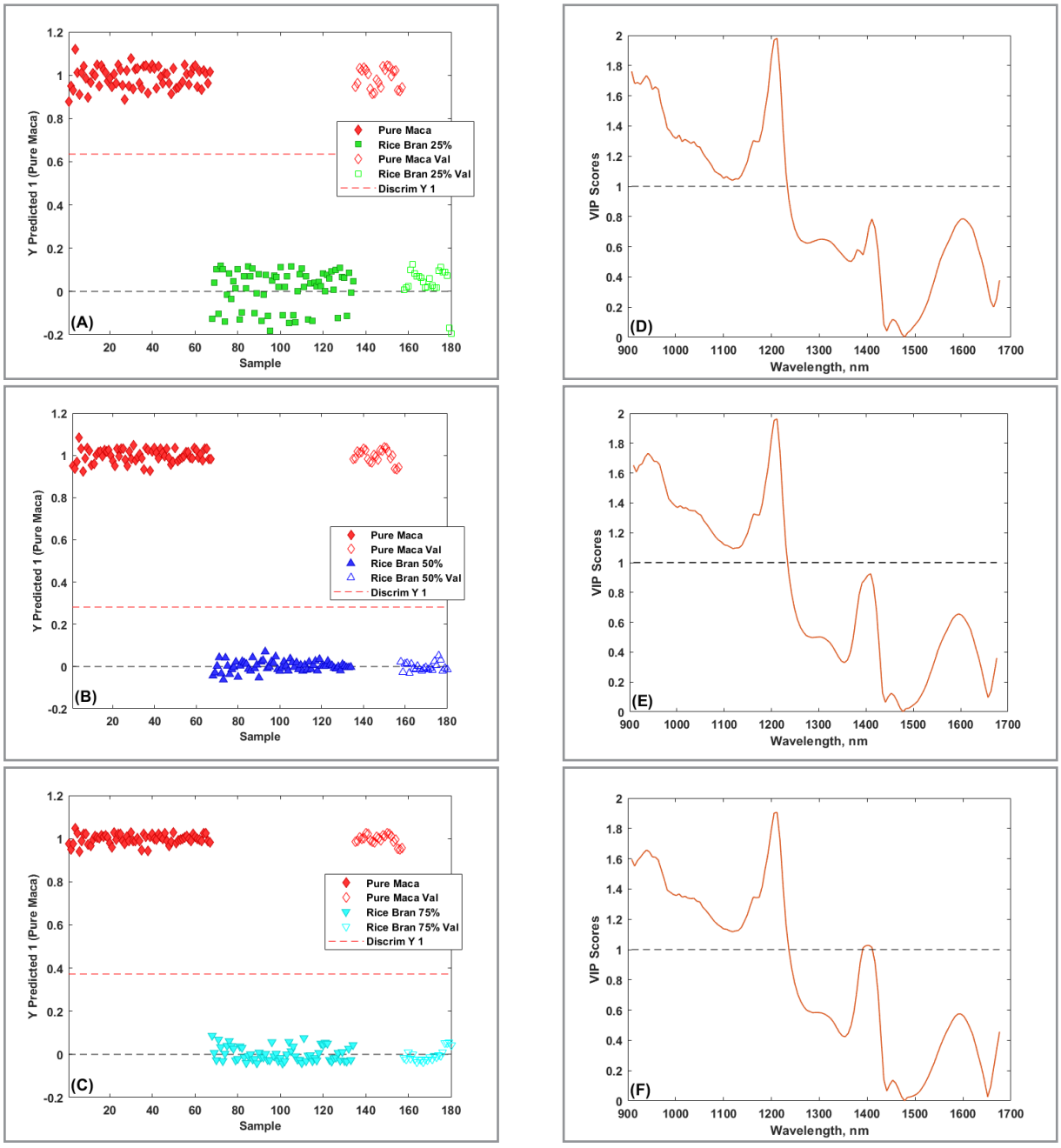


Figure 2. Estimated class values for training and prediction sets for discrimination between (**A**) pure maca and maca adulterated with 25% rice bran; (**B**) pure and adulterated with 50% rice bran; (**C**) pure and adulterated with 75% rice bran. (**D**), (**E**) and (**F**) are the respective VIP Scores.

The same strategy was used to discriminate between samples adulterated with rice flour, as shown in Figure 3. Looking at the VIP scores, it can be observed that, besides 1200 nm, the region between 1400 -1500 nm is above the threshold (value 1), being also important for this discrimination. This region is assigned to the first overtones of OH stretches [12].

Comparing these effects, the VIP scores suggest that the adulteration made by rice flour is more susceptible to effects attributed to OH stretches, being possibly related to water absorption, while rice bran is richer in proteins compared to rice flour. Considering the rice bran composition, the protein content varies from 13-18% [14,15], while to rice flour, this quantity falls to about 4-7% [16].

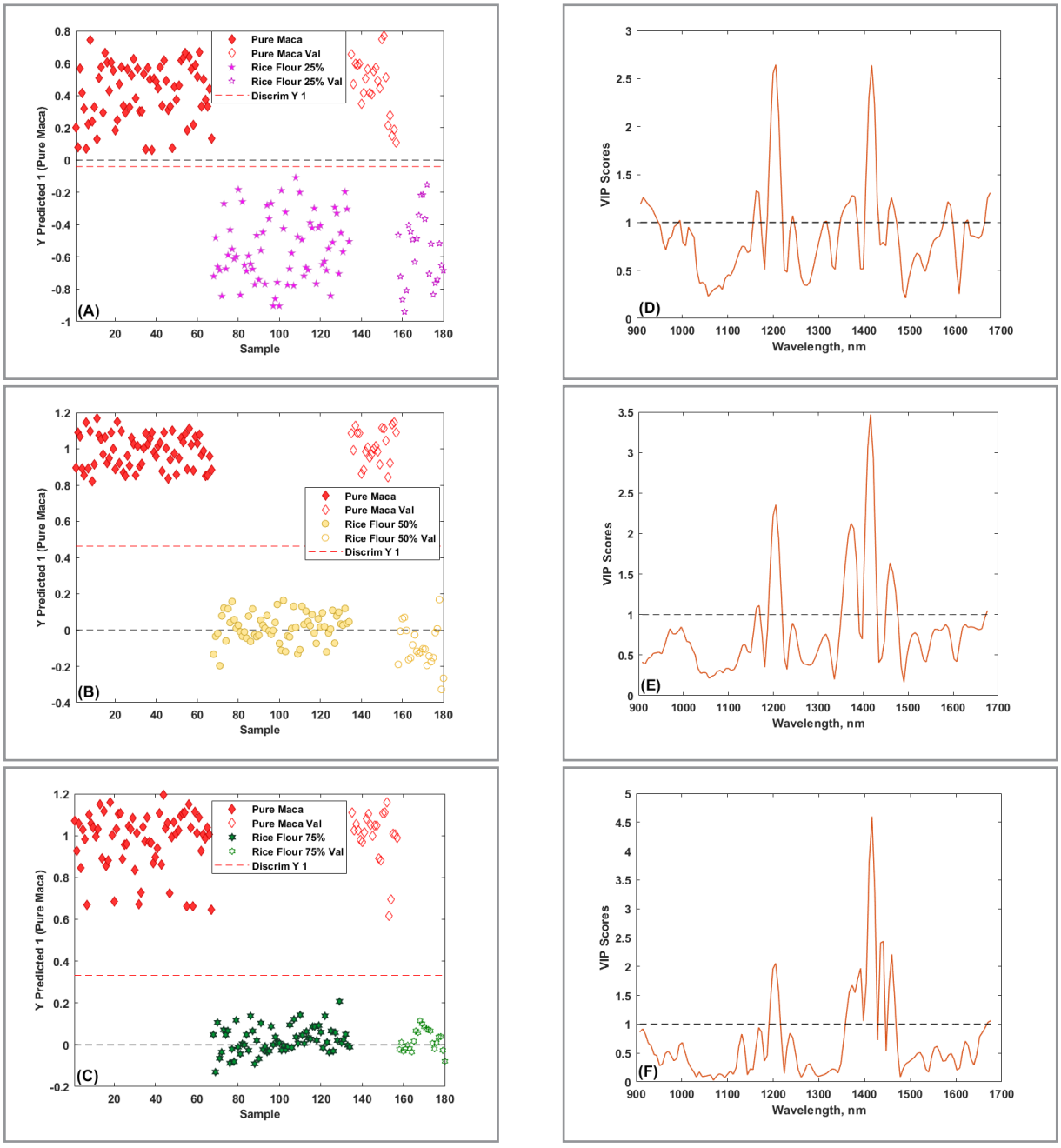


Figure 3. Estimated class values for training and prediction sets for discrimination between (A) pure maca and maca adulterated with 25% rice flour; (B) pure and adulterated with 50% rice flour; (C) pure and adulterated with 75% rice flour. (D), (E) and (F) are the respective VIP Scores.

The main drawback of using NIR spectroscopy to discriminate samples regarding adulteration is the difficulty to characterize the substances responsible for the differentiation. Nonetheless, it seems to be a feasible tool for sample discrimination, which could allow for a significant increase in the number of evaluated samples, improving the quality control on these emerging foods.

CONCLUSIONS

NIR spectroscopy can be used as a technique to discriminate Peruvian maca adulteration with rice bran and rice four since it could be associated with chemometrics tools, showing potential to be used as a feasible method for sample discrimination regarding adulteration. The use of PLS-DA on the NIR spectral data allowed for the exploration of different levels of adulteration and the permutation tests confirmed that the models achieved discrimination of maca samples. Moreover, the adulterants did impose different influences on the spectra, increasing the absorbance below 1400 nm when maca was adulterated with rice bran and decreasing it when adulteration was made by adding rice flour. Also, even though all the models achieved unitary sensibility and specificity, it would be required to increase the variability of the samples to reach more robustness.

Acknowledgements

This work was supported by the Brazilian National Council for Scientific Research (CNPq), [grant number 485640/2013-9 and Scientific Beginner], and CAPES (001), besides the CAMult-CM laboratories.

Conflicts of interest

The authors declare that this study has no relationships which may constitute a conflict of interest at any level.

REFERENCES

1. Hewett, W. E. *Post-harvest requirements: farm to market and is there a case for certifying superfruit?* In: Proceedings of the International Symposium on Superfruits: Myth or Truth? Ho Chi Minh, Vietnam, on 1-3 July **2013**, p 140.
2. Lozano-Canales, A.; Janampa-Santome, M.; Clark, D.; Gonzáles, W. L. *Sci. Hortic*., **2019**, *253*, pp 341–348 (https://doi.org/https://doi.org/10.1016/j.scienta.2019.04.059).
3. Beharry, S.; Heinrich, M. *J. Ethnopharmacol*., **2018**, *211*, pp 126-170 (https://doi.org/10.1016/j.jep.2017.08.003)
4. Leme, L. M.; Nakamura, F.; Tanamati, A. A. C.; Valderrama, P.; Março, P. H. *LWT*, **2019**, *109*, pp 179–185 (https://doi.org/10.1016/j.lwt.2019.03.088).
5. Martinez, N.; McDonald, B. *Veterinary Dermatology*, **2021** (https://doi.org/https://doi.org/10.1111/vde.12950)
6. Ferreira, M. M. C. *Quimiometria: Conceitos, Métodos e Aplicações* (ISBN: 9788526810631) Editora da UNICAMP, Campinas, SP, **2015**.
7. Brereton, R. G.; Lloyd, G. R. *J. Chemom*., **2014**, *28*, pp 213–225 (https://doi.org/10.1002/cem.2609).
8. Souza, S. J.; Valderrama, P.; Consolin Filho, N.; Pilau, E. J.; Tanamati, A. A. C.; Wentzell, P. D.; Março, P. H. *J. Chemom*., **2020**, *34* (12), e3299 (https://doi.org/10.1002/cem.3299).
9. Happyana, N.; Hermawati, E.; Syah, Y. M.; Hakim, E. H. H. *Current Research in Nutrition Food Science*, **2020**, *8* (2), pp 479–488 (https://doi.org/10.12944/CRNFSJ.8.2.13).
10. Mabood, F.; Jabeen, F.; Ahmed, M.; Hussain, J.; Al Mashaykhi, S. A. A.; Al Rubaiey, Z. M. A.; Farooq, S.; Boqué, R.; Ali, L.; Hussain, Z.; et al. *Food Chem*., **2017**, *221*, pp 746–750 (https://doi.org/10.1016/j.foodchem.2016.11.109).
11. Kennard, R. W.; Stone, L. A. *Technometrics*, **1969**, *11*, pp 137–148 (https://doi.org/10.2307/1266770).
12. Burns, D. A.; Ciurczak, E. W. (Editors) *Handbook of near-infrared analysis*, 3rd Edition. (ISBN 9780849373930) CRC Press, **2007**.

13. Workman Jr, J.; Weyer, L. *Practical Guide and Spectral Atlas for Interpretive Near-infrared Spectroscopy*, 2nd Edition. CRC Press, Boca Raton, FL, **2012**.
14. Faria, S. A. S. C.; Bassinello, P. Z.; Penteado, M. V. C. *Braz. J. Pharm. Sci*., **2012**, *48* (4), pp 651–657 (https://doi.org/10.1590/S1984-82502012000400008).
15. Liu, Y.; Zhang, H.; Yi, C.; Quan, K.; Lin, B. *Food Chem*., **2021**, *342*, 128352 (https://doi.org/10.1016/j.foodchem.2020.128352).
16. He, Y.; Chen, F.; Shi, Y.; Guan, Z.; Zhang, N.; Campanella, O. H. *Food Science and Human Wellness*, **2021**, *10* (1), pp 45–53 (https://doi.org/10.1016/j.fshw.2020.05.010).

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 116–130
doi: 10.30744/brjac.2179-3425.AR-13-2021

BrJAC

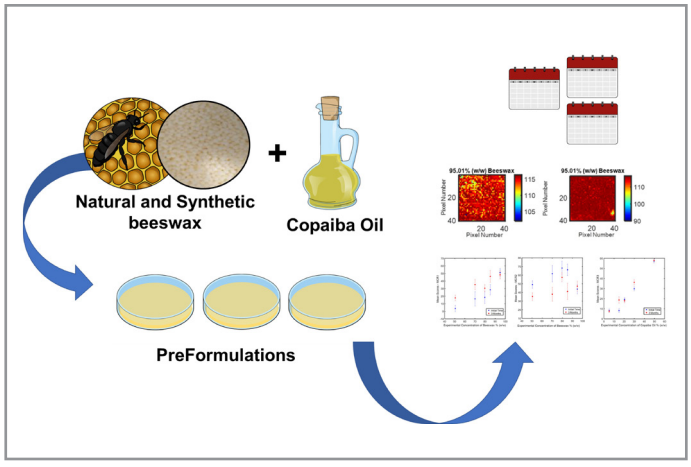
ARTICLE

Raman Imaging and Chemometrics Evaluation of Natural and Synthetic Beeswaxes as Matrices for Nanostructured Lipid Carriers Development

Hery Mitsutake[1,2*], Gustavo Henrique Rodrigues da Silva[2], Lígia Nunes de Morais Ribeiro[2], Eneida de Paula[2], Ronei J. Poppi[1]†, Douglas N. Rutledge[3,4], Márcia C. Breitkreitz[1]

[1]*Department of Analytical Chemistry, Institute of Chemistry, University of Campinas (Unicamp), Rua Josué de Castro, s/n Cidade Universitária Zeferino Vaz, 13084-970, Campinas, SP, Brazil*
[2]*Department of Biochemistry and Tissue Biology, Institute of Biology, University of Campinas (Unicamp), Rua Monteiro Lobato, 255, 13083-862, Cidade Universitária Zeferino Vaz, Campinas, SP, Brazil*
[3]*Université Paris-Saclay, INRAE, AgroParisTech, UMR SayFood, 75005, Paris, France*
[4]*National Wine and Grape Industry Centre, Charles Sturt University, Wagga Wagga, Australia*

† In memoriam (1961 – 2020)



Beeswaxes are interesting solid lipids for the development of nanostructured lipid carriers (NLC), and their origin can be either natural or synthetic. Due to this difference, their performance should be distinct and unstable formulations can be generated. The objective of this work was to investigate miscibility and structural changes (polymorphism) in pre-formulations (blends of solid and liquid lipids) using synthetic and natural beeswaxes in combination with copaiba oil (a natural liquid lipid), in the concentration range of 5.0 to 50.0% (w/w). Raman spectra were acquired over a region of 4 mm^2 (mapping mode), dead pixels were removed using Independent Components Analysis (ICA) and Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) was then used to generate the images. Samples were analyzed at the initial time and after 3 months, using the Distributional Homogeneity Index (DHI) and standard deviation of the histograms. The pre-formulation containing synthetic beeswax showed different structural forms before and after melting, and structural changes over time, depending on the amount of the liquid lipid incorporated. These results demonstrate how spectroscopic imaging techniques can be valuable in pharmaceutical development, as well as the importance of choosing the type and proportion of solid lipid to achieve stable NLC formulations.

Cite: Mitsutake, H.; da Silva, G. H. R.; Ribeiro, L. N. M.; de Paula, E.; Poppi, R. J.; Rutledge, D. N.; Breitkreitz, M. C. Raman Imaging and Chemometrics Evaluation of Natural and Synthetic Beeswaxes as Matrices for Nanostructured Lipid Carriers Development. *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 116–130. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-13-2021

Submitted 21 January 2021, Resubmitted 16 April 2021, Accepted 16 April 2021, Available online 28 April 2021.

Keywords: beeswax, copaiba oil, DHI, MCR-ALS, ICA

INTRODUCTION

Beeswaxes (BW) are a source of lipids that can be used in the development of nanostructured lipid carriers (NLC) [1,2] and they can be of either natural or synthetic sources. Natural beeswaxes (nBW) are produced by the species *Apis mellifera* and *Apis cerana*. The composition of this product is variable, depending both on the bee species and the region of production. It is a complex mixture, composed of more than 300 components, divided into: hydrocarbons – mainly C27-C33 chains (12-16%), free fatty acids with C24-C32 (12-14%), free fatty alcohols – C28-C35 (~ 1%), monoesters and hydroxymonoesters – C40-C48 derived mainly from palmitic, 15-hydroxpalmitic and oleic acids (35-45%) and complex esters (15-27%) such as 15-hydroxy-palmitic acid or diols [3]. Synthetic BW (sBW) (CAS number 71243-51-1) are defined by the United States Environmental Protection Agency (EPA) as "the product of the complex reaction of mixtures of acids and alcohols that simulate the composition of natural BW. It consists mainly of alkyl esters – C18-C30 of fatty acids C16-C32, fatty acids – C16-C32, and alkanes – C22-C34" [4]. Copaiba (*Copaifera officinalis*) is a tree native to South America and the oil-resin produced by it is known for its medicinal properties (e.g., anti-inflammatory, antiseptic and antitetanic). Copaiba oil is rich in sesquiterpene hydrocarbons, mainly *β*-caryophyllene which is responsible for its biological activites [5,6]. Other compounds found in this oil are diterpene acids, copalic acid, hardwickiic acid, *allo*-aromadendrene, germacrene B, *β*-bisabolene, *α*-cadinene, *γ*-cadinene, trans-*α*-bergamotene and *α*-humulene [7].

Attama and coworkers characterized mixtures of nBW and cocoa oil for NLC production and demonstrated that formulations with 50 and 75% (w/w) of oil could be used for the development of lipid nanoparticles [8]. Ribeiro *et al.* developed formulations of nBW with copaiba oil, which showed good physicochemical stability for 12 months, with prolonged *in vitro* release and *in vivo* effect of lidocaine [1]. Lima *et al.* developed solid lipid nanoparticles (SLN) and NLC formulations with beeswax and stearic acid for the delivery of tacrolimus. BW allowed formulations with high drug upload without stability issues [9]. The polymorphism of solid excipients used in the development of SLN and NLC is one of the major problems that can promote shelf-instability in these formulations, probably due to the expulsion of the Active Pharmaceutical Ingredient (API) from the nanoparticles. However, this characteristic does not hamper the use of a given excipient, but rather indicates that attention with miscibility issues should be taken at the very early stages of formulation. One example is the cocoa and shea butter used on NLC by Ribeiro *et al.* [1].

Beeswax is a solid lipid that can show polymorphism. Therefore, the characterization of this raw material is essential prior to its use in pharmaceutical formulations [10]. In this sense, the use of Raman imaging in the early stages of pharmaceutical development can give much information, since this technique is sensitive to both chemical and physical changes in the material. Using this technique, mixtures of solid and lipid excipients (which will here be called 'pre-formulations') can be evaluated prior to their mixing with surfactants and other excipients, with the aim of forecasting stability issues. This approach allows a rational screening of excipients during the NLC development. Thus, it is possible to avoid instabilities and waste of money during this step. We have recently demonstrated how this tool can be useful for evaluating the miscibility of excipients such as Capryol® 90, Precirol ATO5®, Dhaykol 6040® and cetyl palmitate by combining Raman imaging with classical least squares (CLS). In this previous work, the standard deviation of the histograms of pixel was used to evaluate the miscibility. However, this parameter can also be assessed with the help of the distributional homogeneity index (DHI), a technique based on macropixel analysis developed by Sacré *et al.* [11]. DHI does not require a calibration model and may be used to compare the homogeneity of different formulations. Also, it avoids the subjective evaluation through the use of the kurtosis of the histograms.

It should be noted that defects may occur in vibrational spectroscopy images, due to instrumental, sample presentation, or radiation issues. These problems generate defects: the affected pixels being called dead/bad pixels, non-informative background or outliers, which can all affect the performance of models for identification, classification and quantification [12]. Therefore, it is necessary to eliminate these defects before applying a chemometric method. In some cases, these pixels can be purged with adequate

preprocessing or replaced by the mean or median of the neighbors. Since many of these defects are physical and independent phenomena, independent components analysis (ICA) can be used to correct these defective pixels [13,14]. ICA is a Blind Source Separation (BSS) method based on Central Limit Theorem that states that mixtures of several independent source signals should give intensity distributions which are more gaussian than those of the individual source signals.

Therefore, the objective of this work was to evaluate the structural transformations and miscibility between synthetic and natural beeswaxes and copaiba oil - in the range of 5.0-50.0% (w/w) - using Raman mapping and chemometrics. In a first step, the two different beeswaxes were analyzed by X-ray diffraction, differential scanning calorimetry (DSC) and Raman spectroscopy. In the second step, ICA was used to eliminate artefacts in the spectral images. Afterwards, chemical maps were built using MCR-ALS scores and their homogeneity was evaluated using the standard deviation of histograms (STD) and the DHI criterion.

MATERIALS AND METHODS

Materials

Natural beeswax was donated from GM Ceras (Brazil) while synthetic beeswax (Meghwax SEW 200) was provided by Megh Industria e Comércio (Brazil). Copaiba oil (*Copaifera officinalis*) was purchased from Phytoterápica Cosméticos (Brazil).

DSC Curves

A TA Instruments equipment (model Q100) was used in the temperature range of 20 to 250 ºC, with heating rate of 10 ºC/min and inert atmosphere (argon 50 mL/min). Beeswax samples (3-14 mg) were weighed and placed in hermetic aluminum pans for the analysis.

X-Ray Diffraction

The beewaxes used in pre-formulations were analyzed by X-ray diffraction before and after melting. A Shimadzu X-ray diffractometer (model XRD7000) was used under the following conditions: Cu radiation (1.54060 A), voltage of 40 kV, current of 30.0 mA, 2 degrees 2θ / minute, in the range of 5.0 to 50.0 degrees 2θ.

Sample Preparation and Raman Mapping

Five pre-formulations were developed, varying copaiba oil from 5.0-50.0 (% w/w). The samples were prepared by heating (10 ºC above the melting point of the beeswax) and copaiba oil was added with stirring until a visually homogeneous mixture was obtained. The samples were cooled to room temperature (25 ± 2 ºC) in an aluminum cell and an area of 2.0 × 2.0 mm (4 mm^2) was mapped using a Raman Station 400 (Perkin Elmer, CT, USA). A 785 nm laser of was used as an excitation light and power of 100 mW.

The exposure time was 3s/pixel and each spectrum was the average of 2 exposures. The step size was 50 µm and the spectral range 600-3200 cm^{-1} with a resolution of 4 cm^{-1}. Each sample map (4 mm^2) generated a cube of data with dimensions of 40 × 40 × 651, where 40 was the number of pixels at *x* and *y* axis and 651 the number of spectral variables. The total mapping time for each image acquisition was 4 hours. The flasks were sealed and stored in the dark. Raman images of all samples were obtained at the initial time and 3 months after preparation. The period of 3 months was chosen based on recommendations of ICH Q1A(R2) guideline. Three months is the time that for preformulation stability studies are normally carried out by pharmaceutical companies [15].

Data Processing

Spikes from Raman spectra were excluded using an algorithm developed by Sabin and co-workers [16]. The spectral range of 1804 - 1044 cm^{-1} was selected. The data cube was unfolded to a 2D (NM × **λ**) matrix, where M is number of pixels on the *x* axis, N is the number of pixels on the *y* axis and **λ** is the number of spectral variables. The spectral baseline was corrected by asymmetric least squares [17] and unit vector normalization.

Chemometric Analysis

Independent Components Analysis

ICA is a BSS method proposed by Jutten and Herault [18] to extract pure underlying signals (source signals) from a set of mixed signals (Equation 1):

$$\boldsymbol{X} = \boldsymbol{AS} \quad (1)$$

where **X** $(n \times \lambda)$ is the data matrix, **S** $(k \times \lambda)$ is a matrix of independent source signals, **A** $(n \times k)$ is a matrix of mixing coefficients or proportions of the pure signals, *k* is the number of independent components (IC), λ is the number of variables (wavenumbers) and *n* is the number of samples/pixels.

ICA assumes that the source signals and their proportions in the analyzed mixtures are unknown, and it aims to extract them by using the criterion of maximum independence among the source signals. In this work the Joint Approximate Diagonalization of Eigenmatrices (JADE) algorithm [19] was used. This algorithm has the advantage of being deterministic, i.e., not subject to convergence problems since initial random estimates are not necessary [20]. An important parameter during the construction of the ICA model is the choice of the number of components since for a model with *A* factors and one with *A+1* factors, some ICs do not have the same index in both models, and/or do not have the same signal or are simply different. In this work the Durbin-Watson criterion was used [21]. After the application of ICA with the optimal number of ICs, one IC was identified as containing the information concerning the physical defect, which was removed from the data matrix using Equations 2 and 3:

$$\boldsymbol{IC}_{physical\ defect} = \boldsymbol{a}_{physical\ defect} * \boldsymbol{s}_{physical\ defect}{}^{T} \quad (2)$$

$$\boldsymbol{X}_{corrected} = \boldsymbol{X}_{preprocessed} - \boldsymbol{IC}_{physical\ defect} \quad (3)$$

where $\boldsymbol{a}_{physical\ defect}$ refer to the ICA proportions, $\boldsymbol{s}_{physical\ defect}$ is the source signal of the IC with physical information and $\boldsymbol{X}_{corrected}$ represents the matrix after subtraction of this information.

Multivariate Curve Resolution – Alternating Least Squares

MCR-ALS is the most popular algorithm for multivariate curve resolution, based on Equation 4:

$$\boldsymbol{X} = \boldsymbol{CS}_1^{\boldsymbol{T}} + \boldsymbol{E} \quad (4)$$

where **C** is the concentration profile (MN x G), $\mathbf{S}_1$ the spectral profile (λ x G) and **E** is residual matrix (MN x λ) for G components. **C** and $\mathbf{S}_1$ are iteratively calculated until convergence. This iterative aspect of the algorithm presents advantages as it is possible to model spectral profiles with small changes that could be caused by interactions. The number of components was chosen based on singular value decomposition (SVD). The constraints of non-negativity of both spectral and concentration profiles were used in this work. Equality constraint was also used for copaiba oil spectrum in the model for natural BW. The pure excipient spectra were used as initial estimates of **S**. For the MCR analysis the MCR_gui version 2 toolbox [22] was used.

Distributional Homogeneity Index

DHI is a subsampling technique based on analysis of macropixels. A macropixel is defined as “square cluster of neighboring pixels with an intensity value equal to the average value of the included pixels”. In the first step, the chemical map was sampled for all possible macropixels of size 22. In the next step, all macropixels of size 33 were evaluated. This continues until the macropixel size was equal to whole chemical map size (continuous-level moving block). At each step, the standard deviation of the macropixels was

calculated and plotted vs. the macropixel size to generate the homogeneity curve [11,23]. Afterwards the chemical map was randomized several times (usually 50 to 100) and a homogeneity curve of each map was calculated. DHI was obtained from the ratio of the area under the homogeneity curve (AUC) of the original map and the AUC of the randomized map. So, the greater the DHI values, the more heterogenous the sample [11]. All calculations were done using Matlab version 8.3 (Mathworks Inc., Natick MA, USA).

RESULTS AND DISCUSSION

Structural Characterization of the Beeswaxes

Before the construction of the models, Raman spectra, DSC curves and X-ray diffractograms of the beeswaxes were acquired, before and after the fusion (Figure 1). Synthetic wax displayed structural changes after melting, which was not observed for nBW. X-ray diffraction analysis was performed on natural and synthetic BW in order to verify the alteration in the crystalline structure, (Figure 1a and 1b). While there was a clear loss of structure in sBW (due to peak disappearance), no significant changes (except for a slight shift to higher 2θ, from 21.0 to 24.0, after melting) were observed in the natural sample. A very weak peak observed at 5.7º in the sBW diffractogram before melting - corresponding to second or higher order of the long spacing reflection of the diester fraction (monoclinic structure) - disappeared after the melting process. Higher intensities localized at 21.3 to 24.2 in both samples were related to the orthorhombic structure of hydrocarbon/monoester fractions [10] and these peaks seem to be characteristic of beeswax [2,9]. In the DSC curves the polymorphism of sBW was identified by the decrease in the melting point (63.76 to 62.06 ºC) and the appearance of a new peak at 54.5 ºC (Figure 1c). In nBW, the curve does not change significantly.

This change in sBW was also observed in the Raman spectra due to the disappearance of the peak around 1410 cm^{-1} (as highlighted in Figure 1e). Therefore, for the construction of the sBW model, the Raman spectra of BW before and after melting were considered as two distinct excipients, while in case of nBW only one form of solid excipient was considered.

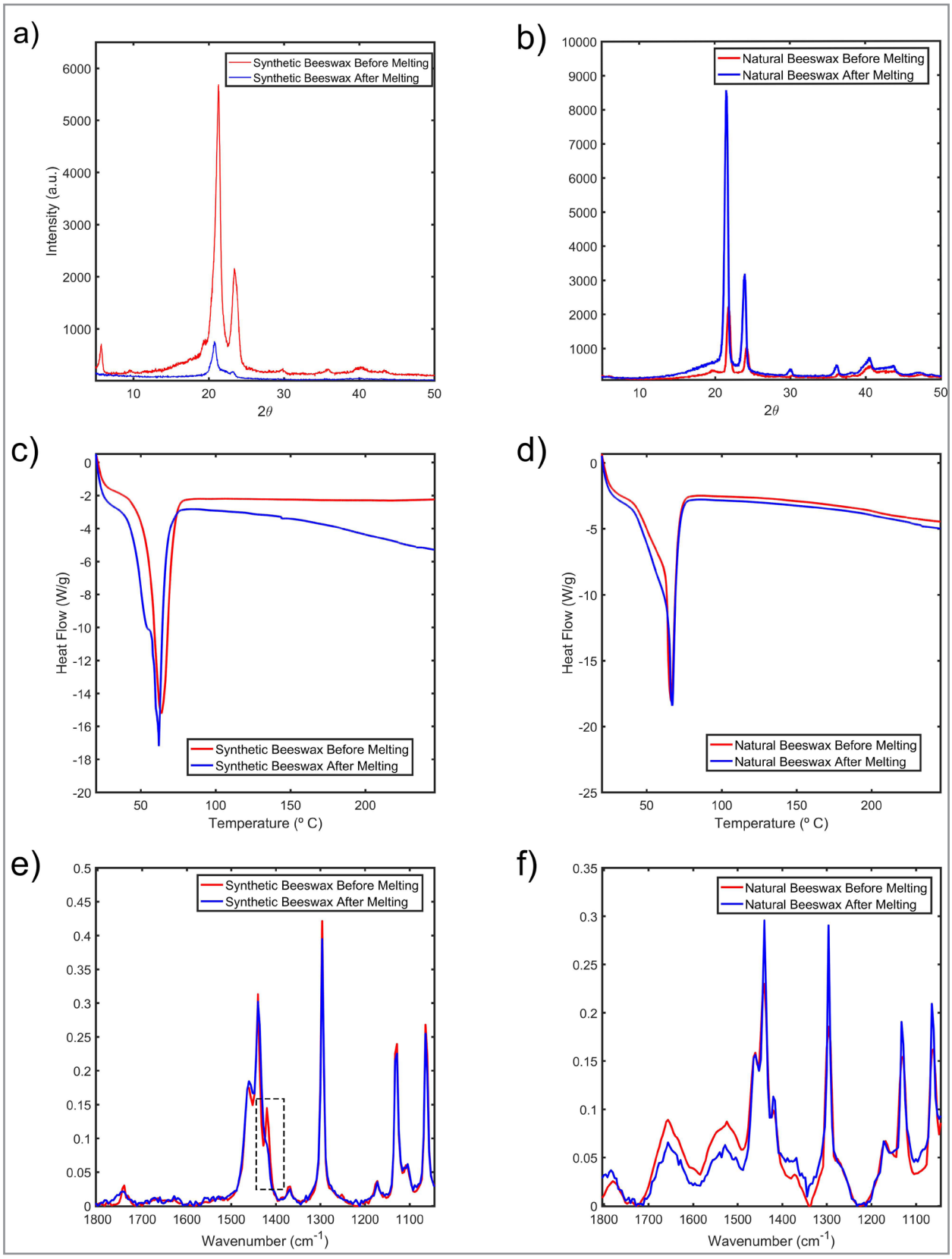


Figure 1. X-ray diffractograms of: (a) synthetic and (b) natural BW. DSC curves of: (c) synthetic and (d) natural BW. Raman spectra of: (e) synthetic and (f) natural BW.

Removal of Dead Pixels

Before the use of MCR-ALS to build chemical maps, an initial analysis was made using CLS to detect defective pixels in the Raman images. Three examples are shown in Figure S1 for samples with 5.0% (w/w) copaiba oil at time zero (freshly prepared) and after 3 months, and for the 20.0% (w/w) copaiba oil sample, after 3 months.

In the first case (Figure S1a), some lines showed unusually high scores (red color) in the original dataset and because of this, the histogram displayed a tail on the left-hand side. In some situations these defective pixels could be replaced by the median of the neighbors [12]. In this dataset, however, this approach was not possible because some neighboring pixels were also defective. ICA could be a good tool to solve this problem. In this dataset there are only a few defective pixels, thus, 10 lines containing only these problematic pixels were created (entitled 'augmented defective spectra' in Figure S1a), with the objective to extract more easily these physical defects, as an independent component. Figure S1d shows the source signal of IC2 related to this phenomenon.

In the second and third cases (Figures S1b-S1c), it is clear that several pixels have defects, and thus, it is not necessary to use the strategy of augmenting with defective pixels, since ICA could be applied directly to these samples to isolate the problem of dead pixels. Figures S1d to S1f show that IC2 and IC4 are related to this problem, in the second and third cases, respectively. In other words, these ICs refer to physical problems in the images and are not due to a specific chemical compound.

After identification of the ICs containing information about dead pixels, Equations 2 and 3 were applied in Raman spectra. The chemical maps from CLS after this correction (Figures S1g-S1i) indicated that ICA was efficient in the correction.

MCR-ALS Results of Natural Beeswax

After ICA correction of these three samples, the chemical maps were created using MCR-ALS with an augmented matrix with all concentrations and times. In the case of nBW, the model converged in 15 iterations, the lack of fit (LOF) of PCA was 7.51% and the percent of explained variance was 96.96%. Figure 2 shows the **S** matrix recovered (pure spectra recovered) and the original Raman spectra of excipients in these pre-formulations. Correlation coefficients were higher than 0.9302, without rotational ambiguity in any of the recovered spectral profiles. It should be emphasized that, as the equality constraint was used for the copaiba oil spectrum, this profile did not suffer alterations and, therefore, its correlation coefficient was equal to 1. Without the use of this constraint, rotational ambiguity occurred in the spectral region of 1320 to 1270 cm^{-1} and 1150 to 1050 cm^{-1} (Figure S2) resulting in a decrease of the correlation coefficient for the copaiba oil.

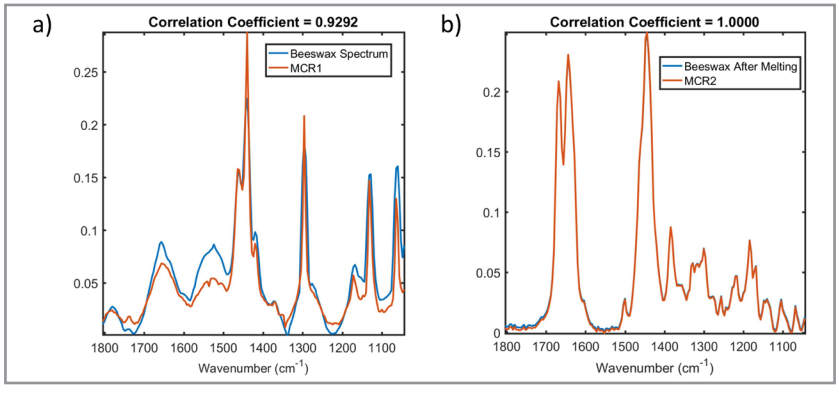


Figure 2. Original Raman spectra and spectral profile recovered by MCR-ALS, for nBW and copaiba oil pre-formulations.

Figure 3 shows the solid lipid maps and histograms obtained for these pre-formulations using nBW. However, the total number of chemical maps generated was high: for just one excipient, 10 maps and

histograms were obtained (five concentrations at two different times) giving a total of 20 Raman maps for the two lipids. This large volume of data makes the interpretation difficult, and for this reason, the experimental concentrations were plotted against the mean values of the scores, along with the standard deviations of the histograms (Figure 4).

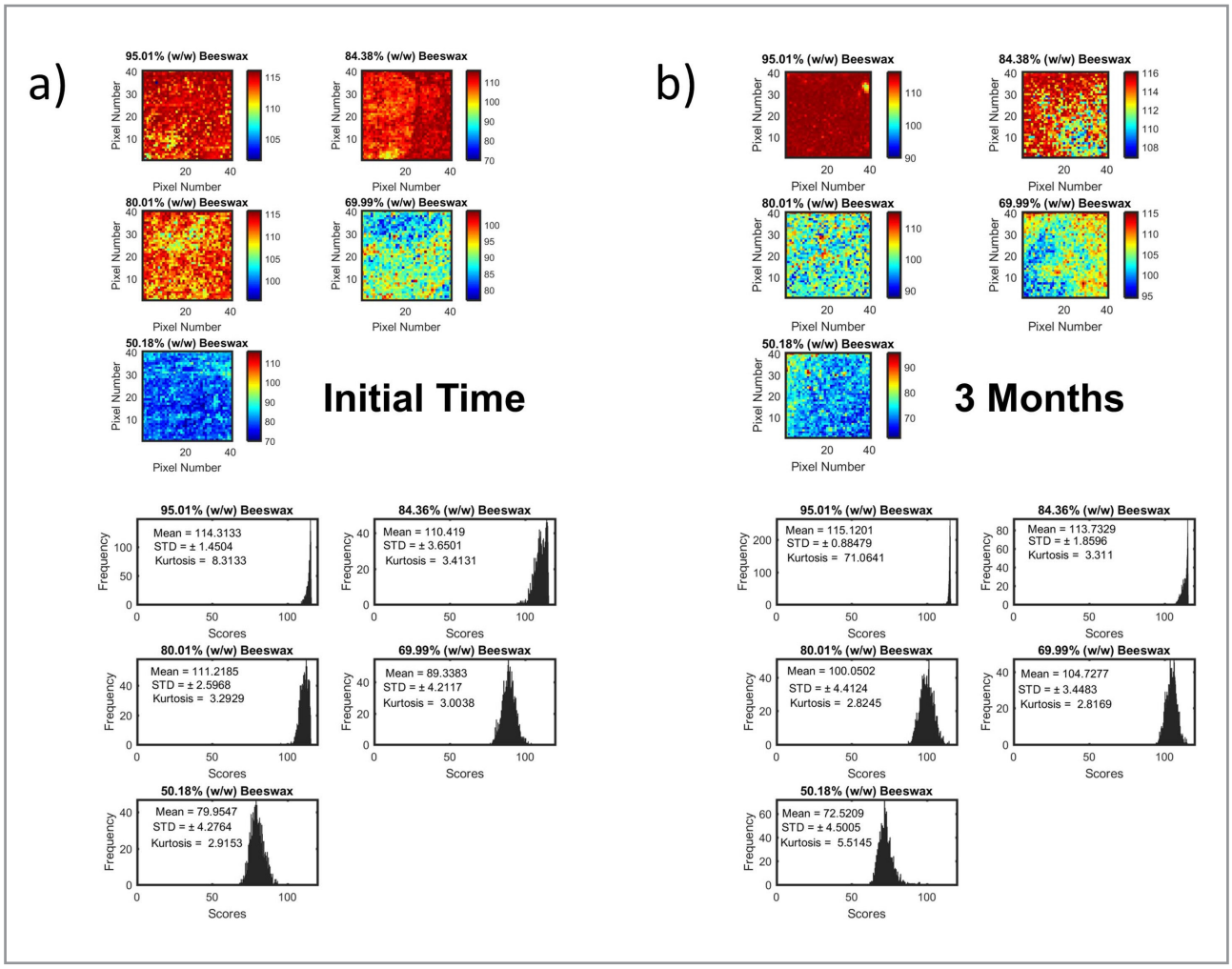


Figure 3. MCR-ALS score maps for nBW excipient a) at initial time and b) after 3 months.

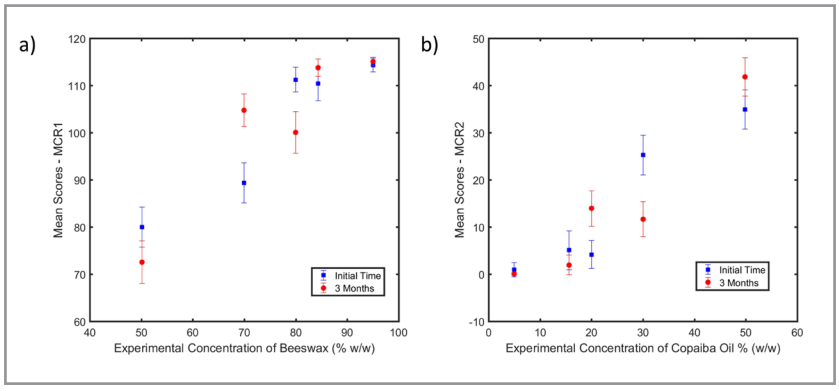


Figure 4. STD and mean of MCR-ALS scores for: a) nBW and b) copaiba oil.

As can be seen in Figure 4, there is no tendency to differentiate the samples in relation to time, i.e., samples of the initial time have sometimes higher and sometimes lower mean values than samples after 3 months. However, it is interesting to note that the differences between the samples (initial *vs.* 3 months) generally tend to be higher as the amount of the liquid excipient increases, indicating that more stable samples are produced with lower amounts of the liquid lipid. In addition to visual inspection of maps, the DHI and standard deviation of the histograms (STD) were used for evaluation of the results, as shown in Table I. As reported by Ma *et al*. [24] and Farkas, Nagy and Marosi [23], DHI is a suitable tool to evaluate the miscibility of components in chemical images. However, its use alone is not recommended because sometimes it cannot represent the distributional homogeneity and, therefore, it is also interesting to evaluate the images associated with the STD of the corresponding histograms. Farkas, Nagy and Marosi [23] also suggested one modification, achieved by weighting the scores and relative standard deviations.

Table I. STD and DHI obtained for samples developed using nBW

		nBW		Copaiba Oil	
Time	**Concentration of copaiba oil (% w/w)**	**STD**	**DHI**	**STD**	**DHI**
Initial	5.0	1.4	2.8	1.5	3.5
	15.6	3.6	4.7	4.1	4.7
	20.0	2.6	2.7	2.9	2.9
	30.0	4.2	4.0	4.2	3.8
	49.8	4.3	3.2	4.2	3.2
After 3 Months	5.0	0.9	2.2	0.3	1.8
	15.6	1.9	4.5	2.1	4.7
	20.0	4.4	1.8	3.8	4.3
	30.0	3.5	4.3	3.7	4.3
	49.8	4.5	3.2	4.0	3.2

The results shown in Figure 4 and Table I indicate that copaiba oil does not have miscibility problems with the nBW, as low STD and DHI were observed. The standard deviations (STD) varied between 0.3 to 4.5 and DHI from 1.8 to 4.7.

Another factor to be evaluated is that there is no difference between the STD and DHI values in relation to the time of the samples. This implies no alterations of formulations/phase separation in natural BW pre-formulations over the evaluated time period (3 months).

MCR-ALS Results of Synthetic Beeswax

The model for formulations using sBW converged after 63 iterations, with lack of fit (LOF) in Principal Component Analysis (PCA) of 0.39% and 99.31% of the explained variance. Figure 5 shows $\mathbf{S}_1$ (spectral profile recovered by MCR-ALS) and original Raman spectra of excipients. There were no problems of rotational ambiguity, that is, the peak due to one compound was not recovered in the spectrum of another. In addition, the correlation coefficients were above 0.9743. The spectral profile recovered for copaiba oil (Figure 5c) had a signal intensity lower than the those of the beeswaxes, therefore the *y*-axis was changed. In this case, three profiles were recovered, since sBW showed spectral differences before and after melting, especially in the region of 1400-1500 cm^{-1}.

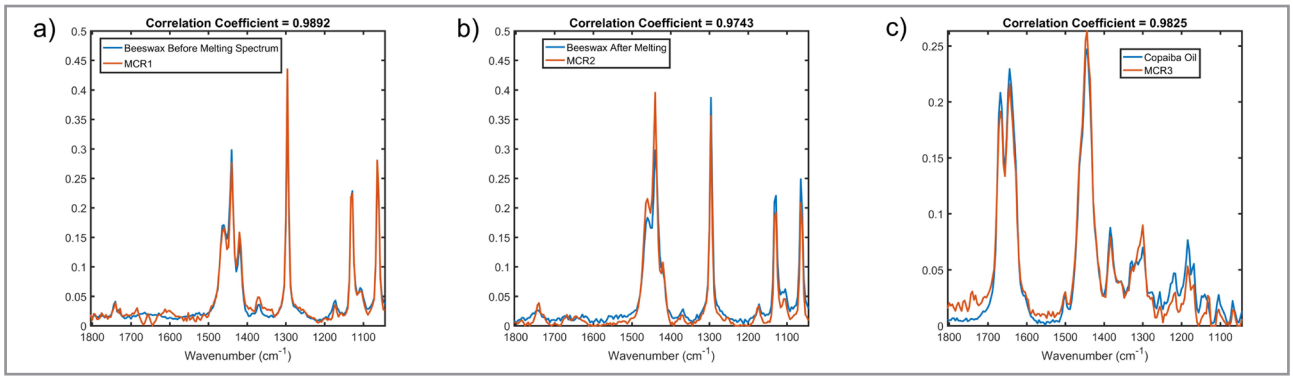


Figure 5. Original Raman spectra and spectral profiles recovered by MCR-ALS (see text) for sBW and copaiba oil formulations.

The histograms and chemical maps built using MCR-ALS scores for sBW before melting are shown in Figure S3. Table II shows the STD and DHI values obtained for each excipient. As can be seen, the standard deviations (STD) and DHI values tend to be slightly higher than those observed for nBW. Copaiba oil did not have miscibility problems with two forms of synthetic beeswax, with the low STD and DHI values.

Figure 6 displays the plot of experimental concentrations against the mean values of the scores, along with the standard deviations of the histograms. As observed in Figure 6a, the mean values of the images at the initial times are lower than in the images acquired after 3 months. On the other hand, in the case of the sBW form found after melting (Figure 6b), the inverse occurs: the values at the initial time are higher than in the other times. This tendency did not appear in case of copaiba oil. Therefore, a polymorphic transformation is probably taking place in the sBW over time, in which the form that appeared after the melting process is converting back into the initial form.

Moreover, the sample with the highest amount of BW (90%) showed smaller difference between the initial time and after 3 months, similar to the results obtained for the nBW. Thus, one assumption is that the presence of liquid excipient caused changes in the mobility of the structures of the solid excipient, with a greater impact at higher concentrations. This can also be one of the explanations for the better performance of NLC formulations at lower concentrations of the liquid lipid [1,25].

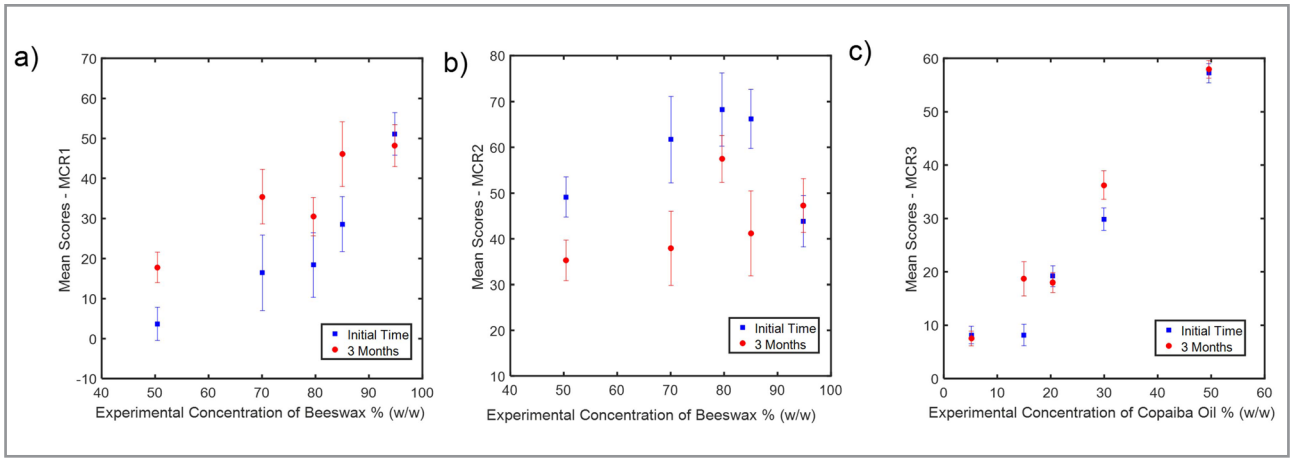


Figure 6. STD and mean of MCR-ALS scores for: a) sBW before melting, b) sBW after melting and c) copaiba oil for formulations developed using synthetic wax.

Table II. STD and DHI obtained for samples developed using sBW

		sBW Before Melting		sBW After Melting		Copaiba Oil	
Time	**Concentration of copaiba oil (% w/w)**	**STD**	**DHI**	**STD**	**DHI**	**STD**	**DHI**
Initial	5.1	5.3	2.6	5.6	2.6	1.6	3.4
	15.0	6.9	4.4	6.5	4.1	2.0	4.1
	20.4	8.0	5.7	8.0	5.1	1.9	2.5
	29.9	9.4	5.1	9.4	4.8	2.1	2.1
	49.6	4.1	3.2	4.3	1.9	1.8	4.9
After 3 Months	5.1	5.2	4.2	5.9	4.1	1.4	2.9
	15.0	8.1	5.2	9.3	5.0	3.2	3.7
	20.4	4.8	3.8	5.1	3.3	1.8	4.1
	29.9	6.8	4.3	8.1	4.2	2.6	2.9
	49.6	3.8	2.1	4.4	2.3	1.6	2.9

Comparing the STD and DHI values of sBW with those of nBW, sBW have higher values than nBW. However, copaiba oil STD and DHI are similar in both cases, indicating this difference is due to the two polymorphic forms. Moreover, the homogeneity of the formulations does not seem to change significantly over the time (i.e., neither the STD and DHI increased significantly nor did two distributions appear in the histograms).

CONCLUSIONS

A structural change was identified in the synthetic BW after melting, which was not observed for the natural BW using DSC, DRX and Raman techniques. But the miscibility of natural and synthetic BW with copaiba oil was not significantly different, based on the STD and DHI values. Nevertheless, the form that appeared after melting of sBW, seems to transform back to the initial form (Figures 6a,b).

It was also observed that the differences between freshly prepared and 3 months samples tended to be higher as the amount of the liquid lipid increased in the pre-formulation. This influence of the liquid excipient in solid materials has also been found in other works [26,27] which makes it even more important to analyze the various properties of these mixtures of solid and liquid excipients, as a function of their concentrations, miscibility and physical changes.

Conflicts of interest

The authors declare they have no conflict of interest.

Acknowledgements

The authors thank Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP [2016/05636-9], Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq [H.M. scholarship - 142441/2015-5], Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES [Finance Code 001] and Instituto Nacional de Ciência e Tecnologia em Bioanalítica - INCT-Bioanalitica [FAPESP/INCT 14/50867-3; CNPq/INCT 465389/2014-7] for financial support; GM Ceras and Megh Indústria e Comércio for the donation of beeswaxes.

REFERENCES

1. Ribeiro, L. N. M.; Breitkreitz, M. C.; Guilherme, V. A.; da Silva, G. H. R.; Couto, V. M.; Castro, S. R.; de Paula, B. O.; Machado, D.; de Paula, E. *Eur. J. Pharm. Sci.*, **2017**, *106*, pp 102–112 (http://dx.doi.org/10.1016/j.ejps.2017.05.060).
2. Attama, A. A.; Schicke, B. C.; Müller-Goymann, C. C. *J. Drug Deliv. Sci. Technol.*, **2007**, *17*, pp 103–112 (http://dx.doi.org/10.1016/S1773-2247(07)50016-X).
3. Fratini, F.; Cilia, G.; Turchi, B.; Felicioli, A. *Asian Pac. J. Trop. Med.*, **2016**, *9*, pp 839–843 (https://doi.org/10.1016/j.apjtm.2016.07.003).
4. Environmental Protection Agency. Beeswax, synthetic. **2020**. Available from: https://ofmpub.epa.gov/sor_internet/registry/substreg/searchandretrieve/substancesearch/search.do?details=displayDetails&selectedSubstanceId=6927 [Accessed 18 January 2021].
5. Dias, D. D. O.; Colombo, M.; Kelmann, R. G.; Kaiser, S.; Lucca, L. G.; Teixeira, H. F.; Limberger, R. P.; Veiga, V. F.; Koester, L. S. *Ind. Crop. Prod.*, **2014**, *59*, pp 154–162 (https://doi.org/10.1016/j.indcrop.2014.05.007).
6. Reátegui, J. L. P.; Fernandes, F. P.; dos Santos, P.; Rezende, C. A.; Sartoratto, A.; Queiroga, C. L.; Martínez, J. *J. Supercrit. Fluid*, **2018**, *140*, pp 364–371 (https://doi.org/10.1016/j.supflu.2018.07.021).
7. Da Trindade, R.; da Silva, J. K.; Setzer, W. N. *Int. J. Mol. Sci.* **2018***, 19* (5), 1511 (https://doi.org/10.3390/ijms19051511).
8. Attama, A. A.; Schicke, B. C.; Müller-Goymannb, C. C. *Eur. J. Pharm. Biopharm.*, **2006**, *64* (3), pp 294–306 (https://doi.org/10.1016/j.ejpb.2006.06.010).
9. Dantas, I. L.; Bastos, K. T. S.; Machado, M.; Galvão, J. G.; Lima, A. D.; Gonsalves, J. K. M. C.; Almeida, E. D. P.; Araújo, A. A. S.; de Meneses, C. T.; Sarmento, V. H. V.; et al. *J. Therm. Anal. Calorim.*, **2018**, *132*, pp 1557–1566 (https://doi.org/10.1007/s10973-018-7072-7).
10. Gaillard, Y.; Mija, A.; Burr, A.; Darque-Ceretti, E.; Felder, E.; Sbirrazzuoli, N. *Thermochim. Acta*, **2011**, *521*, pp 90–97 (https://doi.org/10.1016/j.tca.2011.04.010).
11. Sacré, P.-Y. Y.; Lebrun, P.; Chavez, P.-F. F.; Bleye, C. De; Netchacovitch, L.; Rozet, E.; Klinkenberg, R.; Streel, B.; Hubert, P.; Ziemons, E. *Anal. Chim. Acta*, **2014**, *818*, pp 7–14 (https://doi.org/10.1016/j.aca.2014.02.014).
12. Vidal, M.; Amigo, J. M. *Chemom. Intell. Lab. Syst.*, **2012**, *117*, pp 138–148 (https://doi.org/10.1016/j.chemolab.2012.05.009).
13. Rutledge, D. N.; Bouveresse, D. J-R. *TrAC Trends Anal. Chem.*, **2013**, *50*, pp 22–32 (https://doi.org/10.1016/j.trac.2013.03.013).
14. Rutledge, D. N.; Bouveresse, D. J-R. *TrAC Trends Anal. Chem.*, **2015**, *67*, pp 220 (https://doi.org/10.1016/j.trac.2015.02.001).
15. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use - ICH; Q1A(R2). *Stability testing of new drug substances and products.* Brussels, ICH, **2003**.
16. Sabin, G. P.; de Souza, A. M.; Breitkreitz, M. C.; Poppi, R. J. *Quim. Nova*, **2012**, *35*, pp 612–615 (https://doi.org/10.1590/S0100-40422012000300030).
17. Mazet, V.; Carteret, C.; Brie, D.; Idier, J.; Humbert, B. *Chemom. Intell. Lab. Syst.*, **2005**, *76*, pp 121–133 (https://doi.org/10.1016/j.chemolab.2004.10.003).
18. Jutten, C.; Herault, J. *Signal Processing*, **1991**, *24*, pp 1–10 (https://doi.org/10.1016/0165-1684(91)90079-X).
19. Cardoso, J. F.; Souloumiac, A. *IEE Proceedings, Part F Radar Signal Process*, **1993**, *140*, pp 362–370 (https://doi.org/10.1049/ip-f-2.1993.0054).
20. Bouveresse, D. J-R.; Rutledge, D. N. Independent Components Analysis: Theory and Applications. In: Ruckebusch, C. (Ed.). *Resolving Spectral Mixtures with Applications from Ultrafast Time-Resolved Spectroscopy to Super-Resolution Imaging*, v. 30 of Data Handling in Science and Technology. Elsevier, [s.l.], **2016**, pp 225–277 (https://doi.org/10.1016/B978-0-444-63638-6.00007-3).

21. Bouveresse, D. J-R.; Moya-González, A.; Ammari, F.; Rutledge, D. N. *Chemom. Intell. Lab. Syst.*, **2012**, *112*, pp 24–32 (https://doi.org/10.1016/j.chemolab.2011.12.005).
22. Jaumot, J.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2015**, *140*, pp 1–12 (https://doi.org/10.1016/j.chemolab.2014.10.003).
23. Farkas, A.; Nagy, B.; Marosi, G. *Period. Polytech. Chem. Eng.*, **2017**, *62*, pp 1–7 (https://doi.org/10.3311/PPch.10143).
24. Ma, L.; Zhou, L.; Xu, M.; Huang, X.; Zhang, Q.; Dai, S.; Qiao, Y.; Wu, Z. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, **2018**, *204*, pp 783–790 (https://doi.org/10.1016/j.saa.2018.06.081).
25. da Silva, G. H. R.; Ribeiro, L. N. M.; Mitsutake, H.; Guilherme, V. A.; Castro, S. R.; Poppi, R. J.; Breitkreitz, M. C.; de Paula, E. *Int. J. Pharm.*, **2017**, *529*, pp 253–263 (https://doi.org/10.1016/j.ijpharm.2017.06.066).
26. Mitsutake, H.; Castro, S. R.; De Paula, E.; Poppi, R. J.; Rutledge, D. N. *Int. J. Pharm.*, **2018**, *552*, pp 119–129 (https://doi.org/10.1016/j.ijpharm.2018.09.058).
27. Unga, J.; Matsson, P.; Mahlin, D. *Int. J. Pharm.*, **2010**, *386*, pp 61–70 (https://doi.org/10.1016/j.ijpharm.2009.10.049).

SUPPLEMENTARY MATERIAL

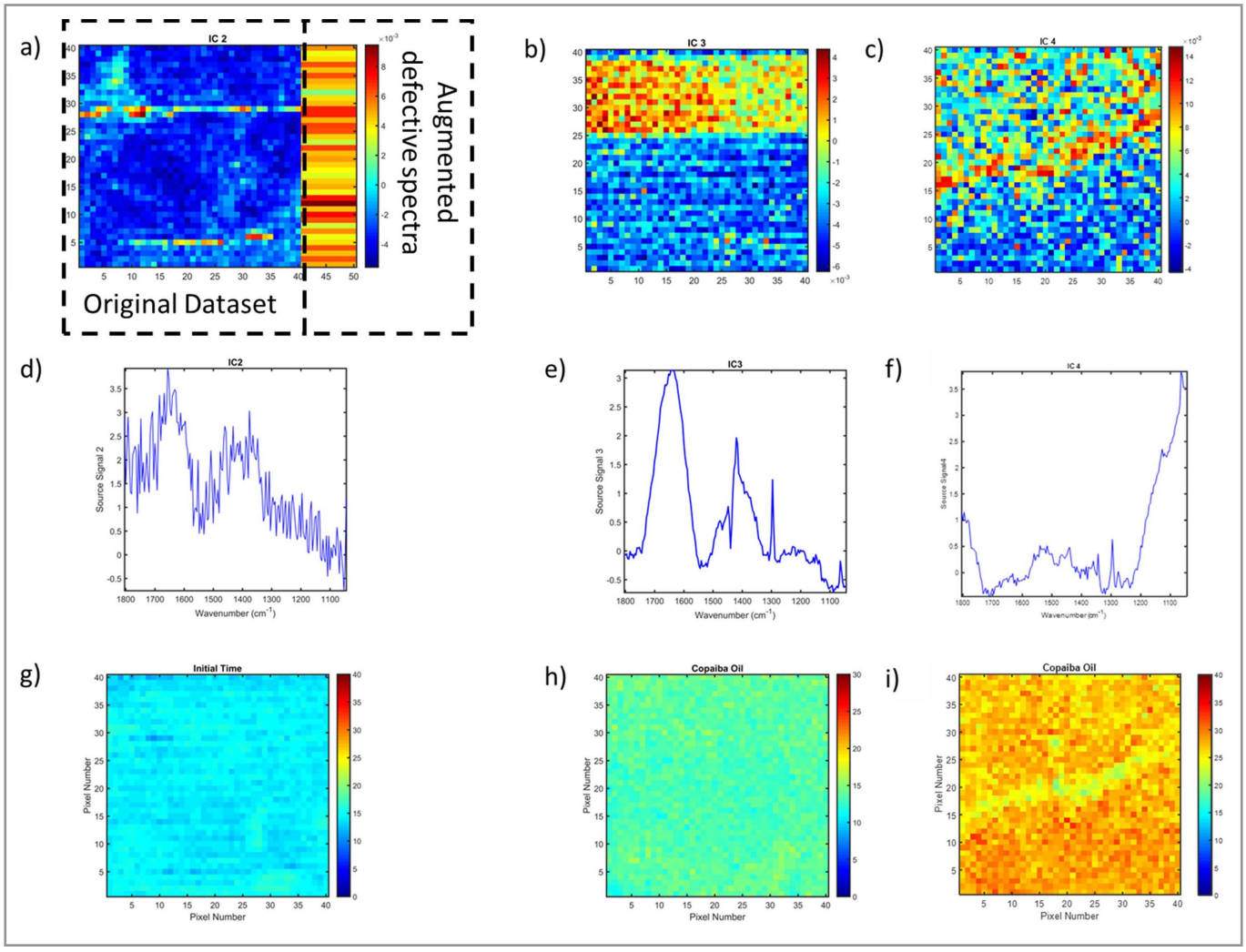


Figure S1. Chemical maps obtained by ICA before correction for samples: (a) 5.0% (w/w) copaiba oil at time zero (freshly prepared), (b) 5.0% (w/w) copaiba oil after 3 months, and (c) for the 20.0% (w/w) copaiba oil sample, after 3 months. Source signals: (d) IC2 for 5.0% (w/w) copaiba oil at time zero (freshly prepared), (e) IC3 for 5.0% (w/w) copaiba oil after 3 months, and (f) IC4 for the 20.0% (w/w) copaiba oil sample, after 3 months. Chemical maps obtained by CLS after correction for samples: (g) 5.0% (w/w) copaiba oil at time zero (freshly prepared), (h) 5.0% (w/w) copaiba oil after 3 months, and (i) for the 20.0% (w/w) copaiba oil sample, after 3 months.

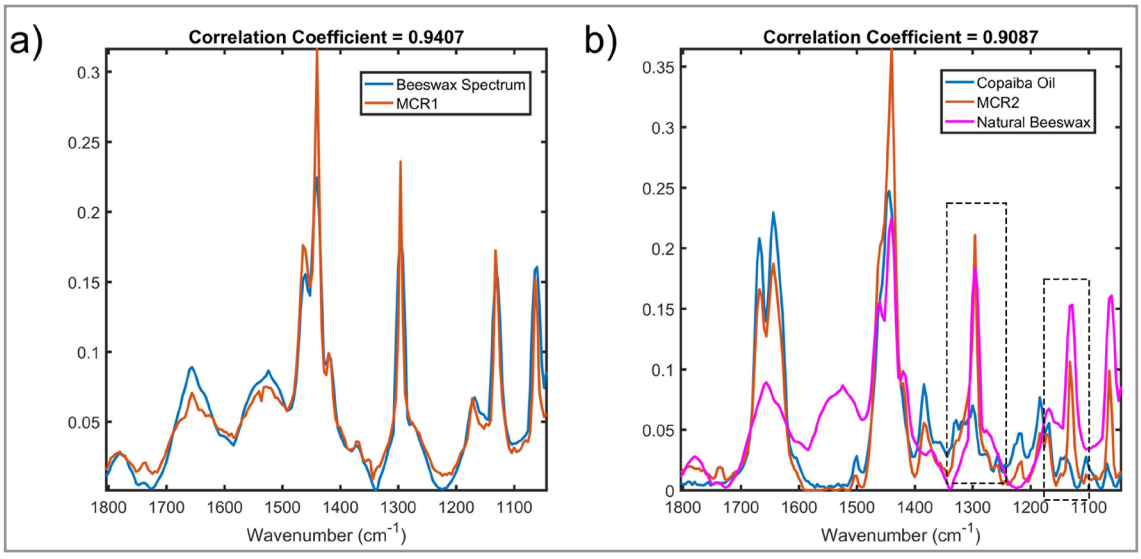


Figure S2. Original Raman spectra and spectral profile recovered by MCR-ALS, for nBW and copaiba oil pre-formulations without equality constraint.

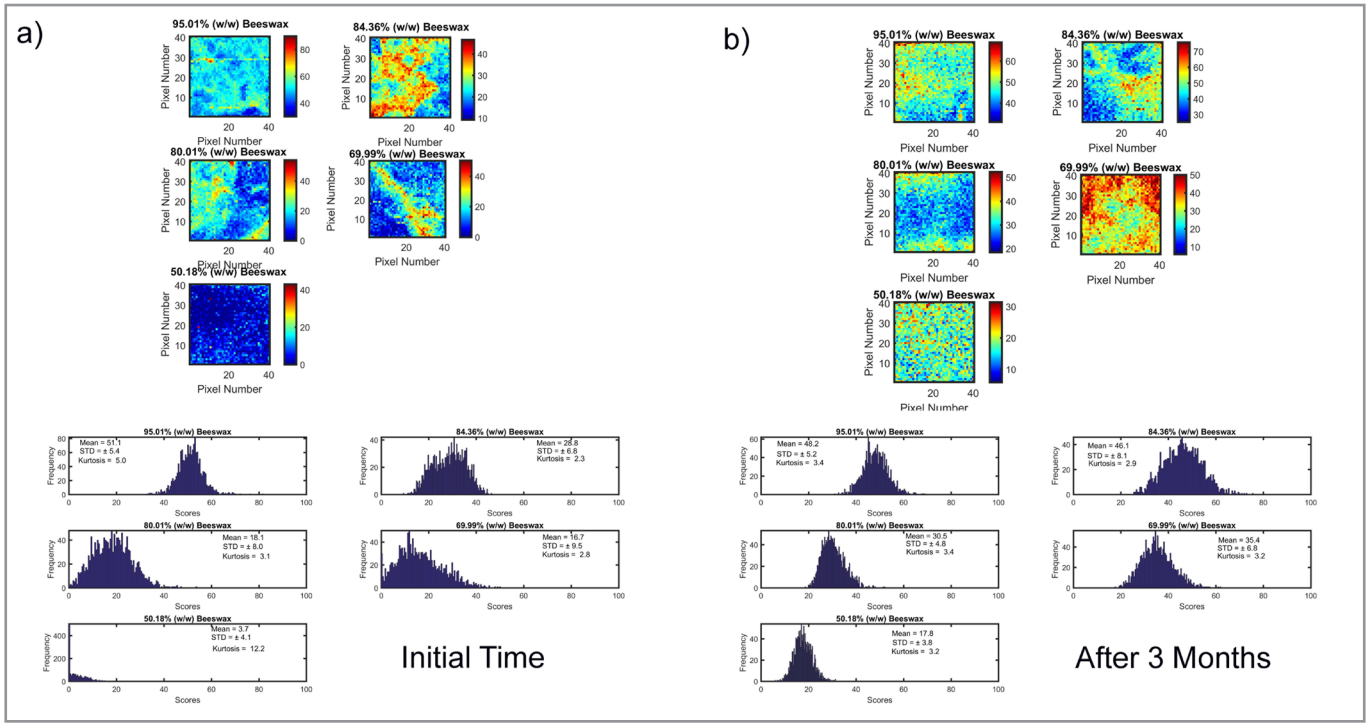


Figure S3. Histograms and chemical maps built using MCR-ALS scores for sBW before melting.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 131–140
doi: 10.30744/brjac.2179-3425.AR-23-2021

BrJAC

ARTICLE

Correlating Comprehensive Two-dimensional Gas Chromatography Volatile Profiles of Chocolate with Sensory Analysis

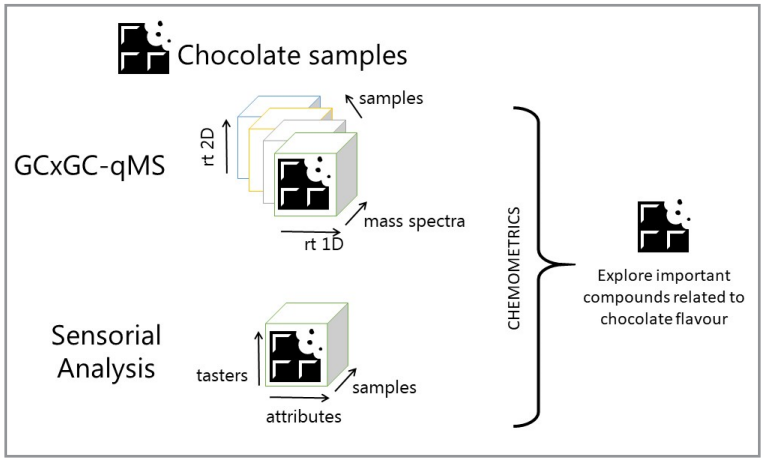
Luciana Fontes de Oliveira[1]*, Soraia Cristina Gonzaga Neves Braga[2], Fabio Augusto[3], Ronei Jesus Poppi[3]†

[1]*Signal and Information Processing for Sensing Systems Group - Institute for Bioengineering of Catalonia (IBEC), Baldiri Reixac 10-12, 08028, Barcelona, Spain*

[2]*Instituto Federal de Educação, Ciência e Tecnologia do Espírito Santo (IFES), Rodovia BR-262, Universal 29134400, Viana, Espírito Santo, Brazil*

[3]*Instituto de Química e Instituto Nacional de Ciência e Tecnologia de Bioanalítica (INCTBio) - Universidade Estadual de Campinas (Unicamp), CP 6154, 13083–970, Campinas, SP, Brazil*

† In memoriam (1961 – 2020)



The identification of key components relevant to sensory perception of quality from commercial chocolate samples was accomplished after chemometric processing of GC×GC-MS (Comprehensive Two-dimensional Gas Chromatography with Mass Spectrometric Detection) profiles corresponding to HS-SPME (Headspace Solid Phase Microextraction) extracts of the samples. Descriptive sensory evaluation of samples was carried out using Optimized Descriptive Profile (ODP) procedures, where sensory attributes of 24 commercial chocolate samples were used to classify them in two classes (low and high chocolate flavor). 2D Fisher Ratio analysis was applied to four-way chromatographic data tensors (1st dimension retention time $^{1}t_{R}$ × 2nd dimension retention time $^{2}t_{R}$ × m/z × sample), to identify the crucial areas on the chromatograms that resulted on ODP class separation on Principal Component Analysis (PCA) scores plot. Comparing the relevant sections of the chromatograms to the analysis of the corresponding mass spectra, it was possible to assess that most of the information regarding the sample main sensory attributes can be related to only 14 compounds (2,5-dimethylpyrazine, 2,6-dimethyl-4-heptanol, 1-octen-3-ol, trimethylpyrazine, β-pinene, o-cimene, 2-ethyl-3,5-dimethylpyrazine, tetramethylpyrazine, benzaldehyde, 1,3,5-trimethylbenzene, 6-methyl-5-hepten-2-one, limonene,

Cite: de Oliveira, L. F.; Braga, S. C. G. N.; Augusto, F.; Poppi, R. J. Correlating Comprehensive Two-dimensional Gas Chromatography Volatile Profiles of Chocolate with Sensory Analysis. *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 131–140. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-23-2021

Submitted 25 February 2021, Resubmitted 19 April 2021, Accepted 19 May 2021, Available online 25 May 2021.

benzeneethanol and 1,1-dimethylbutylbenzene) among the complex blend of volatiles found on these extremely complex samples.

Keywords: chocolate flavor; sensorial analysis; GC×GC-MS, Fisher ratio; principal component analysis

INTRODUCTION

Chocolate is a complex matrix, exhibiting a volatile fraction with more than 500 different chemical components [1]. Several analytical tools have been used for identification of chocolate chemical constituents, mainly based on gas chromatography with mass spectrometry identification [2–4]. However, due to the high complexity of this matrix, in last years the use of comprehensive two-dimensional gas chromatography (GC×GC) [5] is becoming popular: compounds that could co-eluate in the first dimension chromatographic column could be separated in the second dimension column [6]. Described in 1991 [7], GC×GC technique is being constantly improved: *e.g.* with the introduction of new modulation devices [8,9] and data processing strategies (due to the complexity and large size of the chromatographic data sets) [10–12]. Consequently, the number of successful applications of GC×GC has increased in the literature, specially for complex samples such as petroleum [13,14], food [15–17], biological fluidics [18], among others.

Many components from the volatile fraction of chocolates are relatable to sensorial quality parameters assessable by procedures such as Quantitative Descriptive Analysis (QDA) [19,20]. QDA is the most complete and popular tool in food science, but requires long-term training of the panelists, being relatively slow and expensive [22]. In terms of time and cost, there are some more appropriate alternatives for QDA [20] such as Optimized Descriptive Profile (ODP) analysis [21]. ODP is based on semi-trained panelists, who evaluate the products using an unstructured anchored scale corresponding to the extremes (weak and strong) of the range of the sensorial property being assessed. Samples representing the extremes of the attribute scale (weak and strong standard) and samples being tested should be offered to panelists in the same tasting session; consequently, people with limited training will be able to consistently evaluate the samples. In this approach, on each tasting session just a single sensorial attribute should be evaluated, since the samples and the references are presented simultaneously [21,23,24].

The combination of volatile profile characterization by GC×GC and sensory analysis can be a powerful tool to study chocolate. If sensory analysis results could be objectively correlated to GC×GC data, it would be possible to pinpoint volatile chemical components responsible for specific quality features of the samples. The modelling of quantitative sensory data correlation to GC×GC chromatograms should be made using multivariate chemometric tools [10,25,26] such as Principal Component Analysis (PCA), Partial Least Squares regression (PLS) [27] as well as neural network modelling [28] have been successfully applied in these studies. Among them, PLS [29] is by far the most useful when chromatographic data has to be correlated to some macroscopic parameter of the samples; however, it has some limitations for large and complex data structures [30] such as GC×GC chromatograms. In these situations, preliminary selection of sections of the chromatograms potentially related to the sensory property being assessed before multivariate modelling can be crucial. Fisher Ratio analysis (FR) has been successfully applied as pre-processing strategy to filter non-relevant section from large GC×GC data sets and improve the quality of subsequent multivariate analysis [31–33]. Essentially, FR analysis locates regions of the chromatograms were extra-class variations (pertinent to the distinction of samples of different sample classes) are significantly larger than intra-class variations (that does not contain useful information regarding sample categorization). Only the relevant regions of the chromatographic data sets are feed to multivariate data analysis algorithms, improving the quality of the chemometric processing and demanding less computational power and resources [34].

For each signal point in the $^{1}t_{R}\times{}^{2}t_{R}\times m/z$ chromatographic space, Fisher Ratio is defined as a scalar value calculated as:

$$FR = \frac{\sigma_{cl}^2}{\sigma_{err}^2}$$

where $\sigma_{cl}^2 = \sum \frac{(\bar{x}_i - \bar{x})^2 \cdot n_i}{k-1}$ and $\sigma_{err}^2 = \frac{\sum\left(\sum(\bar{x}_{ij} - \bar{x})^2\right) - \left(\sum((\bar{x}_i - \bar{x})^2 \cdot n_i)\right)}{(N-k)}$

In these equations, $\mathbf{n_i}$ is the number of measurements in the **i**th class, $\mathbf{x_i}$ is the mean of the **i**th class, **x** is the overall mean, **k** is the number of classes, $\mathbf{x_{ij}}$ is the **i**th measurement of the **j**th class and **N** is the total number of sample profiles. Sections of the chromatogram with values for **FR** significantly larger than 1 are relevant to the distinction of sample classes; points where **FR** ~ 1 correspond to sections of the chromatograms where intra- and extra-class variations of the signal are similar and are not relevant to sample qualification. To select regions of the chromatographic space for multivariate analysis, data points with **FR** smaller than a user-defined threshold $\mathbf{FR_{cut}}$ are excluded without loss of information.

In this work, the volatile fraction of commercial chocolate samples was isolated by headspace Solid Phase Microextraction (HS-SPME) an analyzed by GC×GC-MS; the so-called "chocolate flavor" sensorial attribute of these samples was previously determined by ODP, which were split in two classes according to the corresponding chocolate flavor values (low and high). After preliminary selection of potentially relevant areas of the chromatograms using 2D Fisher Ratio analysis, the data tensors were subject to exploratory inspection by Principal Component Analysis. The loadings from PCA were compared to the chromatograms, and the mass spectra and retention indexes of the peaks appearing on the relevant regions allowed identification of the key chemical compounds responsible for class separation.

MATERIALS AND METHODS

Samples and Sample Preparation

Twenty-four chocolate samples of different suppliers and with nominal cocoa content from 45 to 90% were obtained at the local market. The volatile fraction of the samples was isolated by Headspace Solid Phase Microextraction (HS-SPME) using a previously described procedure [35]. Aliquots of (1.000 ± 0.005) g of sample were exposed to N_2-frozen and then grounded manually and then the grounded samples were weighed in 15 mL septum-sealed extraction vials. After a 5 min period for sample/headspace equilibration at 60 °C a 50/30 µm Divinylbenzene / Carboxen / Polydimethylsiloxane (DVB/CAR/PDMS) SPME fiber (Supelco, Bellefont – PA, USA) was exposed for 50 min to the sample headspace and then immediately inserted into the GC×GC system injector for 5 min. Before the extractions, the fiber was conditioned for 2 h at 250 ºC according to the supplier instructions.

GC×GC-MS analysis

GC×GC-MS analysis was performed using a prototype based on a Shimadzu (Tokyo, Japan) QP2010+ GC-MS [36]. The column set consisted on a 30 m × 0.25 mm × 0.25 µm HP-5 column (Agilent Technologies, Wilmington, DE) connected to a 0.80 m × 0.1 mm × 0.1 µm Solgel Wax column (SGE Analytical Science, Ringwood - Victoria, Australia). The injector was operated on splitless mode at 260 ºC. The column oven temperature was programmed as: 40 ºC to 110 ºC at 3 ºC min^{-1} and 110 ºC to 240 ºC at 10 ºC min^{-1}; the modulation period was set to 6.0 s and a cryogenic modulator was used. The interface temperature was 260 ºC. MS detector photomultiplier high voltage was programmed from 0.8 kV up to 10 min and then increased to 0.9 kV until final of the chromatographic run. Data was collected at an acquisition rate of 25 scan s^{-1} with scanned mass range set from *m/z* = 40 D to 340 D. The determination 1st dimension linear temperature programming retention indexes (LTPRI) of relevant peaks was performed after analysis of selected samples spiked with n-alkane mixture (C_8 to C_{20}). Detection of chromatographic peaks on the chromatograms was performed using GCImage software (Zoex, Houston – TX, USA) and analyte identification achieved by matching their mass spectra and LTPRI with literature data and spectra on NIST 2010 library.

Optimized Descriptive Profile (ODP)

ODP of samples was performed according to the methodology described by Silva *et al.* [21,23]. Triangular tests were applied in 20 judges using two samples from the same supplier but with different cocoa content (50% and 75% cocoa). The triangular test was performed in triplicate and the judges that have at least two right answers were selected for the ODP tests. In this case, 75% of the judges were approved in the triangular test.

2D Fisher ratio for four-way data

In order to simplify further data processing, Fisher Ratio Analysis was used to select areas on the $^1t_R \times ^2t_R \times m/z$ chromatographic space generated by GC×GC-MS that really contained pertinent information related to the sample differentiation. The main algorithm used to calculate the variance inside one class, between two classes and the Fisher Ratios is described elsewhere [37]; this procedure was later extended to be applied directly in four-way data [34] such as the case here. Fisher Ratio analysis applied in four-way chromatograms allows both pixel-based (direct application to the raw **Signal = f(^1t_R, ^2t_R, m/z)** tensors without peak detection and integration) and peak-level (where input data are peak area tables obtained after conventional detection and integration) [31,38].

Figure 1 shows the general scheme of 2D Fisher ratio calculation in the four-way space. Samples with highest and lowest values for the chocolate flavor attribute obtained in ODP analyses were chosen to define the two classes and (**Sample × ^1t_R × ^2t_R × m/z**) data cubes for these samples were generated for each mass channel (Figure 1-A). These cubes were unfolded by concatenating the second and third dimensions (retentions times), resulting in matrices where each line corresponds to a sample and retention times are in columns for each mass channel (Figure 1-B). Fisher Ratios for all matrices were calculated, resulting in a vector with dimensions equal of the number of retention times for each mass channel (Figure 1-C). Finally, a new matrix was formed with Fisher Ratios for all mass channels (Figure 1-D); this matrix was folded back, resulting in a data cube with (**Fisher Ratios × ^1t_R × ^2t_R**) (Figure 1-E). Each signal point in this cube is obtained as the sum of all individual m/z signals (Figure 1-F) [31,34]. All operations here were performed in MATLAB R2016b (MathWorks, USA) using self-made routines for 2D Fisher ratio calculation.

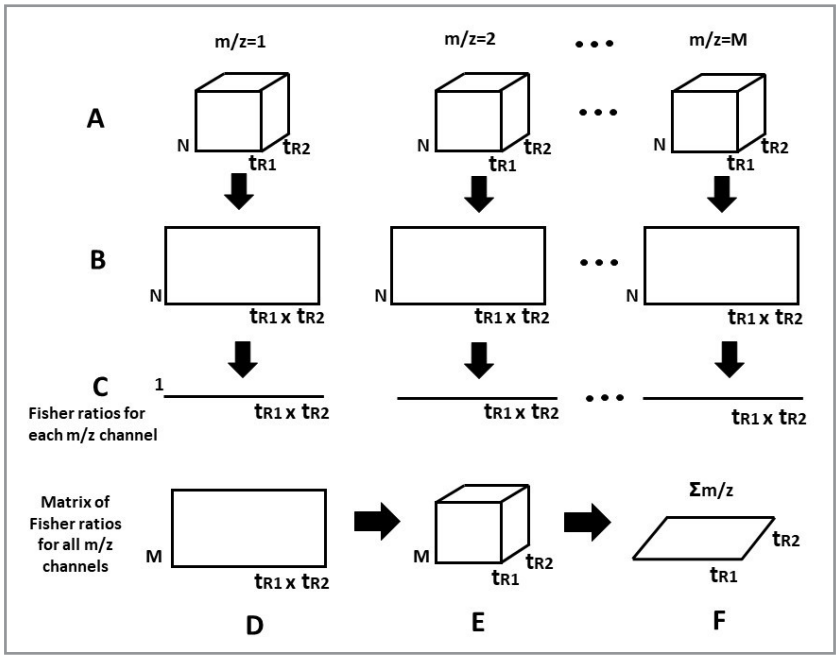


Figure 1. General scheme of 2D Fisher ratio calculation in GC×GC-MS. N is the number of samples, M is the number of mass (m/z) channels, t_{R1} is the retention time in the first column and t_{R2} is the retention time in the second column.

Identification of volatile components essential to predict chocolate flavor sensory attribute

The identification of the set of key volatile compounds that need to be assessed to model the evaluated sensory attribute (chocolate flavor) from GC×GC-MS chromatograms was performed through an iterative process. After calculation of the tensor of Fisher Ratio values, selection of potentially relevant sections of the chromatograms for differentiation between sample classes (high and low chocolate flavor) was performed using an initial tentative value for FR_{cut}. Data sets containing only the selected areas of the GC×GC-MS were unfolded, and the mass spectra channel was summed obtaining the TIC chromatograms and after that the data were submitted to Principal Component Analysis, using MatLab PLS Toolbox v. 7.3.1 (Eigenvector, USA) with pareto scaling as pre-processing method and the resulting scores plots inspected to check for distinction between sample classes. This process was repeated, at each time using larger FR_{cut} values (and therefore smaller input data sets on PCA) until clear separation of sample classes (low and high chocolate flavor) appears in the scores plots.

After definition of the maximum Fisher Ratio threshold for PCA class separation – and therefore isolating only the absolutely essential parts of the chromatograms that contain information regarding class differentiation, the correspondent loadings matrixes were matched to GC×GC chromatograms to point the location on the $^1t_R \times {}^2t_R$ plane of compounds relevant to characterization of these categories.

RESULTS AND DISCUSSION

Figure 2 shows the average values of panelist responses for chocolate flavor attribute for the studied samples studied. Samples with intermediary sensory parameter values between the classes were marked as “undefined”; Samples with chocolate flavor attribute defined as “low” and “high” are indicated in the figure.

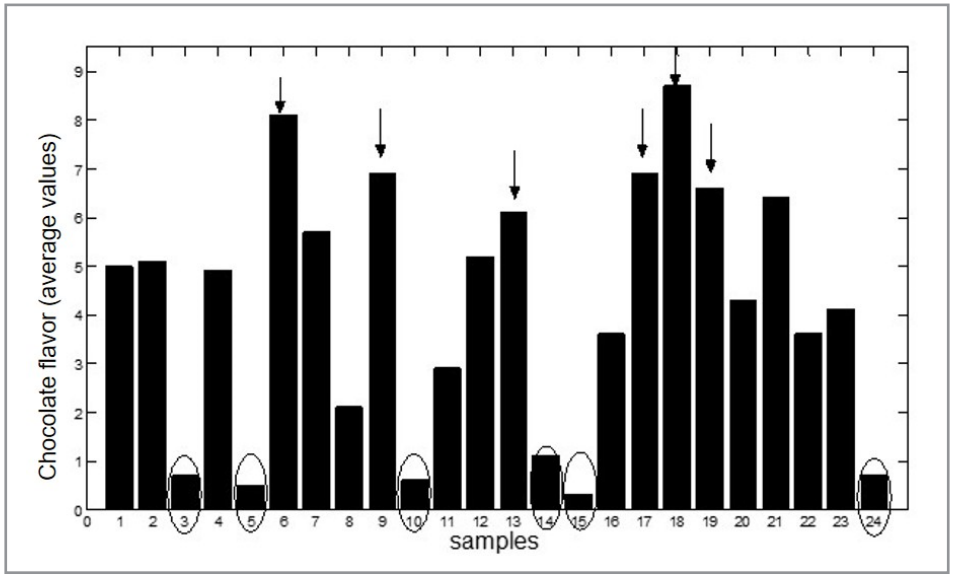


Figure 2. Chocolate flavor values of all commercial chocolates studied. Samples indicated with a circle were considering as ‘low chocolate flavor’ class and samples with an arrow were considered as ‘high chocolate flavor’ class.

Figure 3 shows the chromatogram parts selected as relevant to sensory classification of samples using the final threshold value of $\mathbf{FR_{cut}} = 2\times10^7$ overlapped with a typical chromatogram. The spots selected by Fisher Ratio analysis were fed as input data for PCA modelling.

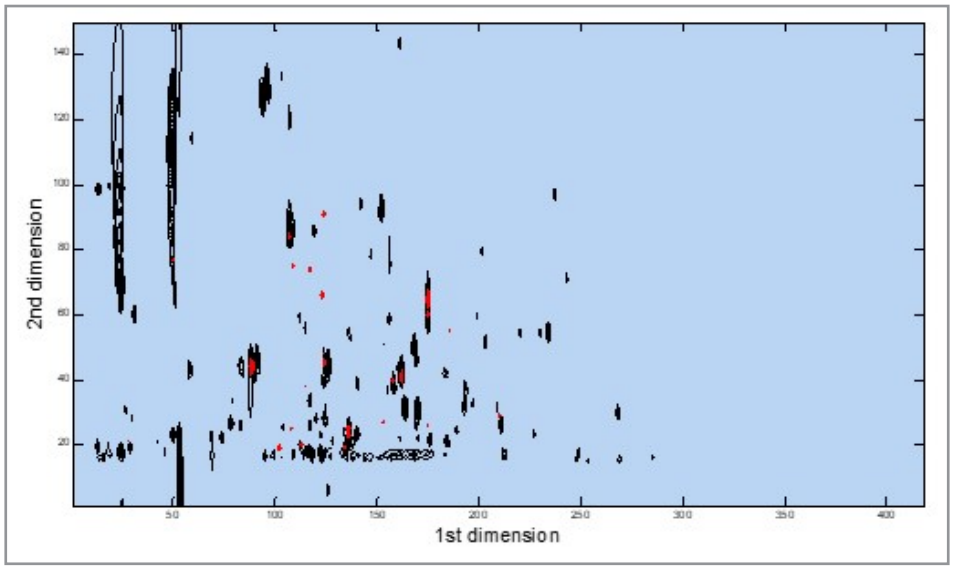


Figure 3. Spots selected by 2D Fisher ratio (in red) overlapped in a typical chromatogram (black color).

Figure 4 compares PCA scores plots obtained after modelling using the raw, whole chromatographic data sets without variable selection and scores plots of PCA using partial chromatographic data selected after FR preliminary analysis. In the former, it is quite clear that samples group in two distinct regions of the plane - confirming that this value of threshold was adequate for sample classification; as for the PCA scores plot resulting after using the whole, unfiltered data no class separation is observable.

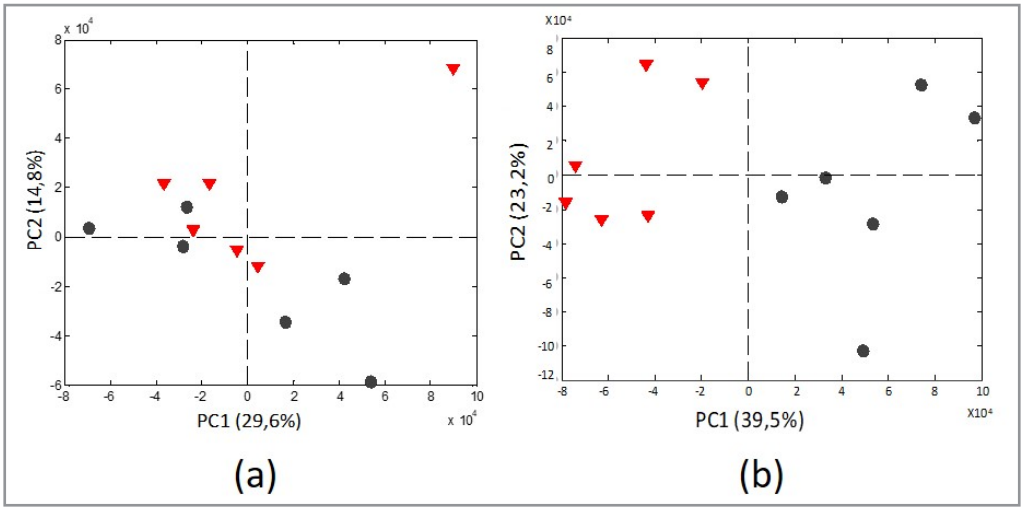


Figure 4. PCA scores plot. (a) all variables included in the model, (b) only variables with Fisher ratio above $2x10^7$. ▼ samples of "low chocolate flavor" class and ● samples of "high chocolate flavor" class.

By using the loadings of the first PC of the data set formed by the selected variables, it is possible to point the areas in the GC×GC chromatograms where compounds responsible for class separation elute. Those compounds were tentatively identified from their retention indexes and mass spectra. Figure 5 shows the positive (Figure 5a) and negative loadings (Figure 5b), where is possible to note a reduced number of spots with significant loading values. These results indicated that for the class separation (low and high chocolate flavor) a reduced number of variables were selected by the 2D Fisher ratio (only 5% of

the total of variables). Positive loadings were related to samples with high chocolate flavor and negative loadings with samples with low values for this attribute.

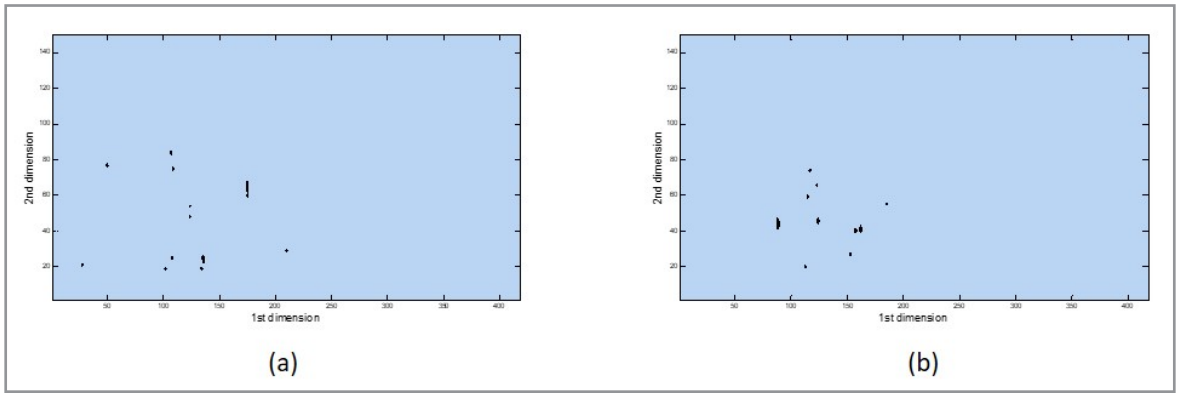


Figure 5. Plot of the loadings of the first PC. (a) positive loadings. (b) negative loadings.

As a reduced number of spots in GC×GC chromatograms were selected, the number of identified components was also low. The identified compounds are listed in Table I. The compounds with positive loadings were mostly pyrazines (tetramethylpyrazine, 2-ethyl-3,5-dimethylpyrazine, trimethylpyrazine and 2,5-dimethylpyrazine) which are known to be related to the chocolate flavor [39,40]. Also, Counet *et al*. [41] described in their work that the tetramethylpyrazine was the most abundant pyrazine found in dark chocolates evaluated and that this compound could have a reasonable influence in the final odor of chocolates. Besides the recognized influence of pyrazines in the chocolate odor 1-octen-3-ol was also observed as important compound by Nightingale *et al*. [2] when evaluating the changes in the headspace of dark chocolates during different storage times.

Table I. Compounds identified as significant for classes separation

Positive loadings	Negative loadings
2,5-dimethylpyrazine	benzaldehyde
2,6-dimethyl-4-heptanol	1,3,5-trimethylbenzene
1-octen-3-ol	6-methyl-5-hepten-2-one
trimethylpyrazine	limonene
β-pinene	benzeneethanol
o-cimene	1,1-dimethylbutylbenzene
2-ethyl-3,5-dimethylpyrazine	
tetramethylpyrazine	

The opposite can be understood for the negative loadings. Samples with low chocolate flavor values show this behavior due to the compounds indicated in the negative loadings. Among them, the compounds were benzaldehyde, 1,3,5-trimethylbenzene, 6-methyl-5-hepten-2-one and limonene.

Among the identified compounds limonene was already described for being a important compound in chocolates related with the crystallization process, physical properties and the aroma [42–44]. Furthermore,

benzaldehyde was previously related with the bitter sensorial attribute [45] and leads to confirm the obtained results in this work, since chocolates described as low chocolate flavor present more bitterness flavor (benzaldehyde). Here is important to know that Brazilian chocolate consumers are more used to consume milk chocolate (that present more sugar than dark chocolates used in this study) [46] and for this reason describe chocolates with less chocolate flavor as chocolates with more bitterness.

Finally, the threshold value stipulated by the Fisher ratio was applied for all 24 samples studied. A new data set was formed with all samples, and the PCA was calculated. The plot of the scores of the first three PCs for this new data is presented in Figure 6. We can verify that the samples were again separated into high and low chocolate flavor values (red and black samples), and between them, samples with intermediate values of this attribute (in blue). This is an additional indication of the correct variable selection by Fisher ratio, showing that samples with intermediate chocolate flavor attribute have also intermediate content of the chemical constituents.

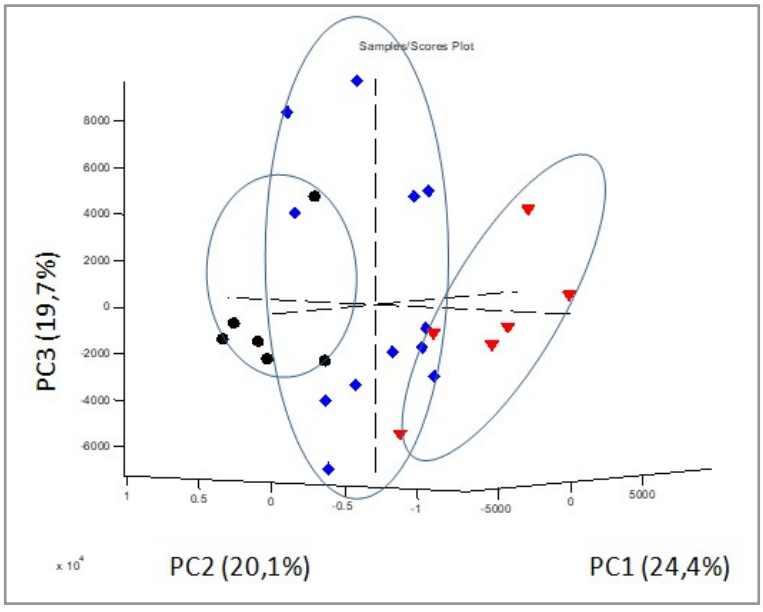


Figure 6. Plot of the scores of the PCA model using all samples studied. ▼ low values of chocolate flavor attribute ♦ intermediate values of chocolate flavor attribute and ● high values of the chocolate flavor attribute.

CONCLUSIONS

In this work we demonstrated a strategy for identification of the essential set of volatile components related to the sensory parameter “chocolate flavor” from GC×GC-MS data in tandem with sensory analysis PCA combined to variable selection using Fisher Ratio analysis. The 2D Fisher ratio was able to point the sections of chromatograms where key compounds that separate the predefined classes (low and high chocolate flavor) in the PCA scores plot were located. It is important to emphasize that the separation was accomplished using a reduced number of variables and consequently only 14 chemical components were identified. This result does not mean that only these compounds are related to the differentiation of the chocolate flavor – several other minor components can contribute with this attribute, but they certainly are associated with this feature; also, this approach does not account for synergistic effects that could appear on compounds which are on non-selected areas of the chromatographic space. The compounds identified for positive loadings, and with high chocolate flavor attribute, are mostly pyrazines that are recognized for the characterization of chocolate flavor.

Conflicts of interest

No conflicts of interest – financial or otherwise – existed on the composition of this manuscript.

Acknowledgements

This work was supported by grants provided by Fapesp - São Paulo Research Foundation [Grants 2012/14695-8 and 2014/50867-3], CNPq - National Council for Scientific and Technological Development [Grant 465389/2014-7] and CAPES - Brazilian Coordination for the Improvement of Higher Education Personnel [Finance Code 001].

This paper is the last paper resulting from of the last and fruitful cooperation between Prof. Ronei Poppi and our research group. He left this world prematurely in April 2020; saying that he is going by to be missed as scientist, associate and friend is a bitter understatement. It is in honor of his memory we dedicate this paper.

REFERENCES

1. Beckett, S. T. *Industrial chocolate manufacture and use*, 4th Edition. Blackwell Publishing, York, **2008**, Chapter 8, p 169.
2. Nightingale, L. M.; Cadwallader, K. R.; Engeseth, N. J. *J. Agric. Food Chem.*, **2012**, *60*, pp 4500–4507 (https://doi.org/10.1021/jf204718z).
3. Aculey, P. C.; Snitkjaer, P.; Owusu, M.; Bassompiere, M.; Takrama, J.; Nørgaard, L.; Petersen, M. A.; Nielsen, D. S. *J. Food Sci.*, **2010**, *75*, pp S300–S307 (https://doi.org/10.1111/j.1750-3841.2010.01710.x).
4. Cambrai, A.; Marcic, C.; Morville, S.; Sae Houer, P.; Bindler, F.; Marchioni, E. *J. Agric. Food Chem.*, **2010**, *58*, 1478–1483 (https://doi.org/10.1021/jf903471e).
5. Humston, E. M.; Knowles, J. D.; McShea, A.; Synovec, R. E. *J. Chromatogr.*, *A* **2010**, *1217*, pp 1963–1970 (https://doi.org/10.1016/j.chroma.2010.01.069).
6. Seeley, J. V.; Seeley, S. K. *Anal. Chem.*, **2013**, *85*, pp 557–578 (https://doi.org/10.1021/ac303195u).
7. Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.*, **1991**, *29*, pp 227–231.
8. Tranchida, P. Q.; Salivo, S.; Franchina, F. A.; Mondello, L. *Anal. Chem.*, **2015**, *87*, pp 2925–2930 (https://doi.org/10.1021/ac5044175).
9. Tranchida, P. Q.; Purcaro, G.; Dugo, P.; Mondello, L.; Purcaro, G. *TrAC Trends Anal. Chem.*, **2011**, *30*, pp 1437–1461 (https://doi.org/10.1016/j.trac.2011.06.010).
10. Matos, J. T. V.; Duarte, R. M. B. O.; Duarte, A. C. *J. Chromatogr. B*, **2012**, *910***,** pp 31–45 (https://doi.org/10.1016/j.jchromb.2012.06.039).
11. Pierce, K. M.; Hoggard, J. C. *Anal. Methods*, **2014**, *6*, pp 645–653 (https://doi.org/10.1039/C3AY40965A).
12. Eftekhari, A.; Parastar, H. *J. Chromatogr. A*, **2016**, *1466*, pp 155–165 (https://doi.org/10.1016/j.chroma.2016.09.016).
13. Alexandrino, G. L.; Prata, P. S.; Augusto, F. *Energy & Fuels*, **2017**, *31*, pp 170–178 (https://doi.org/10.1021/acs.energyfuels.6b01925).
14. Prata, P. S.; Alexandrino, G. L.; Mogollón, N. G. S.; Augusto, F. *J. Chromatogr. A*, **2016**, *1472*, pp 99–106 (https://doi.org/10.1016/j.chroma.2016.10.044).
15. Chin, S.-T.; Eyres, G. T.; Marriott, P. J. *Food Chem.*, **2015**, *185*, pp 355–361 (https://doi.org/10.1016/j.foodchem.2015.04.003).
16. Cordero, C.; Kiefl, J.; Schieberle, P.; Reichenbach, S. E.; Bicchi, C. *Anal. Bioanal. Chem.*, **2015**, *407*, pp 169–191 (https://doi.org/10.1007/s00216-014-8248-z).
17. Cordero, C.; Schmarr, H.-G.; Reichenbach, S. E.; Bicchi, C. *J. Agric. Food Chem.*, ***2018***, *66*, pp 2226–2236 (https://doi.org/10.1021/acs.jafc.6b04997).
18. Yu, Z.; Huang, H.; Reim, A.; Charles, P. D.; Northage, A.; Jackson, D.; Parry, I.; Kessler, B. M. *Talanta*, **2017**, *165*, pp 685–691 (https://doi.org/10.1016/j.talanta.2017.01.003).
19. Hootman, R., Ed. *Manual on Descriptive Analysis Testing for Sensory Evaluation*. West Conshohocken, PA: ASTM International**, 1992** (https://doi.org/10.1520/MNL13-EB).
20. Murray, J. M.; Delahunty, C. M.; Baxter, I. A. *Food Res. Int.*, **2001**, *34*, pp 461–471 (https://doi.org/10.1016/S0963-9969(01)00070-9).

21. Minim, V. P. R.; Simiqueli, A. A.; Moraes, L. E. S.; Gomide, A. I.; Minim, L. A. *Food Quality and Preference*, **2012**, *24*, pp 190–200 (https://doi.org/10.1016/j.foodqual.2011.10.014).
22. Minin, V. P. R. *Análise sensorial - estudos com consumidores*, 1th ed. Publisher: Universidade Federal de Viçosa, **2014**.
23. Silva, R. C. S. N.; Minim, V. P. R.; Carneiro, J. D. D. S.; Nascimento, M.; Della Lucia, S. M.; Minim, L. A. *Food Quality and Preference*, **2013**, *30*, pp 169–179 (https://doi.org/10.1016/j.foodqual.2013.05.011).
24. Silva, R. C. S. N.; Minim, V. P. R.; Silva, A. N.; Peternelli, L. A.; Minim, L. A. *Food Quality and Preference*, **2014**, *36*, pp 3–11 (https://doi.org/10.1016/j.foodqual.2014.02.011).
25. Arancibia, J. A.; Damiani, P. C.; Escandar, G. M.; Ibañez, G. A.; Olivieri, A. C. *J. Chromatogr. B*, **2012**, *910*, pp 22–30 (https://doi.org/10.1016/j.jchromb.2012.02.004).
26. Harvey, P. M.; Shellie, R. A. *Anal. Chem*., **2012**, *84*, pp 6501–6507 (https://doi.org/10.1021/ac300664h).
27. Ballabio, D.; Consonni, V. *Anal. Methods*, **2013**, *5*, pp 3790–3798 (https://doi.org/10.1039/c3ay40582f).
28. Carvalho, N. B.; Minim, V. P. R.; Silva, R. de C. dos S. N.; Della Lucia, S. M.; Minim, L. A. *Food Sci. Technol*., **2013**, *33*, pp 722–729 (https://doi.org/10.1590/S0101-20612013000400018).
29. Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst*., **2001**, *58*, pp 109–130 (https://doi.org/10.1016/S0169-7439(01)00155-1).
30. Zeng, Z.; Li, J.; Hugel, H. M.; Xu, G.; Marriott, P. J. *TrAC Trends Anal. Chem*., **2014**, *53*, pp 150–166 (https://doi.org/10.1016/j.trac.2013.08.009).
31. Parsons, B. A.; Marney, L. C.; Siegler, W. C.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. *Anal. Chem*., **2015**, *87*, pp 3812–3819 (https://doi.org/10.1021/ac504472s).
32. Beckstrom, A. C.; Humston, E. M.; Snyder, L. R.; Synovec, R. E.; Juul, S. E. *J. Chromatogr. A*, **2011**, *1218*, pp 1899–1906 (https://doi.org/10.1016/j.chroma.2011.01.086).
33. Hantao, L. W.; Toledo, B. R.; de Lima Ribeiro, F. A.; Pizetta, M.; Pierozzi, C. G.; Furtado, E. L.; Augusto, F. *Talanta*, **2013**, *116*, pp 1079–1084 (https://doi.org/10.1016/j.talanta.2013.08.033).
34. Pierce, K. M.; Hoggard, J. C.; Hope, J. L.; Rainey, P. M.; Hoofnagle, A. N.; Jack, R. M.; Wright, B. W.; Synovec, R. E. *Anal. Chem.*, **2006**, *78*, pp 5068–5075 (https://doi.org/10.1021/ac0602625).
35. Braga, S. C. G. N.; Oliveira, L. F.; Silva, A. R. A; Efraim, P.; Poppi, R. J.; Augusto, F. *Braz. J. Anal. Chem.*, **2019**, *6*, pp 16–26 (https://doi.org/10.30744/brjac.2179-3425.ar-02-2019).
36. Hantao, L. W.; Aleme, H. G.; Passador, M. M.; Furtado, E. L.; Ribeiro, F. A. de L.; Poppi, R. J.; Augusto, F. *J. Chromatogr. A*, **2013**, *1279*, pp 86–91 (https://doi.org/10.1016/j.chroma.2013.01.013).
37. Pierce, K. M.; Hope, J. L.; Johnson, K. J.; Wright, B. W.; Synovec, R. E. *J. Chromatogr. A*, **2005**, *1096*, pp 101–110 (https://doi.org/10.1016/j.chroma.2005.04.078).
38. Marney, L. C.; Christopher Siegler, W.; Parsons, B. A.; Hoggard, J. C.; Wright, B. W.; Synovec, R. E. *Talanta*, **2013**, *115*, pp 887–895 (https://doi.org/10.1016/j.talanta.2013.06.038).
39. Afoakwa, E. O.; Paterson, A.; Fowler, M.; Ryan, A. *Food Chem.*, **2009**, *113*, pp 208–215 (https://doi.org/10.1016/j.foodchem.2008.07.088).
40. Frauendorfer, F.; Schieberle, P. *J. Agric. Food Chem*., **2006**, *54*, pp 5521–5529 (https://doi.org/10.1021/jf060728k).
41. Counet, C.; Callemien, D.; Ouwerx, C.; Collin, S. *J. Agric. Food Chem*., **2002**, *50*, pp 2385–2391 (https://doi.org/10.1021/jf0114177).
42. Ray, J.; MacNaughtan, W.; Chong, P. S.; Vieira, J.; Wolf, B. *J. Am. Oil Chem. Soc*., **2012**, *89*, pp 437–445 (https://doi.org/10.1007/s11746-011-1934-5).
43. Do, T.-A. L.; Vieira, J.; Hargreaves, J. M.; Wolf, B.; Mitchell, J. R. *J. Am. Oil Chem*. Soc., **2008**, *85*, pp 911–920 (https://doi.org/10.1007/s11746-008-1281-3).
44. Maikhunthod, B.; Marriott, P. J. *Food Chem*., **2013**, *141*, pp 4324–4332 (https://doi.org/10.1016/j.foodchem.2013.05.156).
45. Bonvehí, J. S. *Eur. Food Res. Technol*., **2005**, *221*, pp 19–29 (https://doi.org/10.1007/s00217-005-1147-y).
46. Oliveira, D.; Reis, F.; Deliza, R.; Rosenthal, A.; Giménez, A.; Ares, G. *Food Res. Int*., **2016**, *89*, pp 448–453 (https://doi.org/10.1016/j.foodres.2016.08.019).

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 141–154
doi: 10.30744/brjac.2179-3425.AR-25-2021

BrJAC

ARTICLE

Fingermark Analysis by Fourier Transform Infrared Microscopy Using Chemometric Tools

Marina González[1]*, Kristiane de Cássia Mariotti[2,3], Adriano de Araújo Gomes[4], Marco Flôres Ferrão[4,5], Renata Pereira Limberger[1,3]

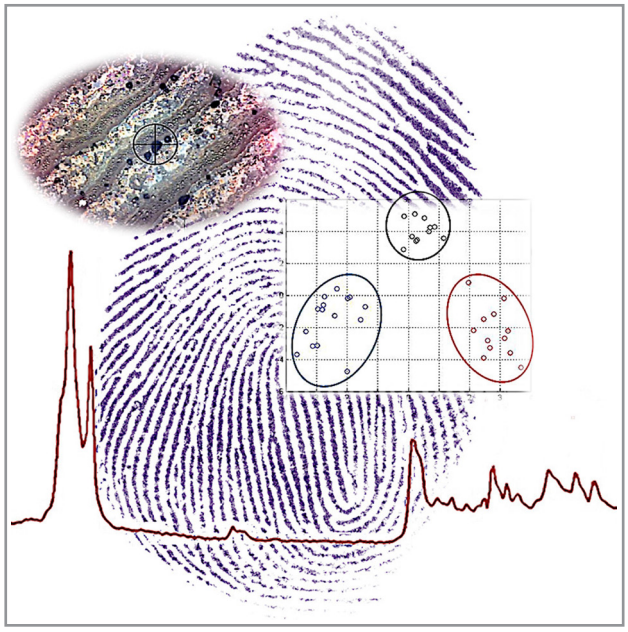
[1]*Departamento de Farmácia, Universidade Federal do Rio Grande do Sul, Av. Ipiranga, 2752, Azenha, 90610-000, Porto Alegre, RS, Brazil*

[2]*Grupo de Biometria, Polícia Federal, Av. Ipiranga, 1365, Azenha, 90610-093, Porto Alegre, RS, Brazil*

[3]*Instituto Nacional de Ciência e Tecnologia Forense (INCT Forense), Porto Alegre, RS, Brazil*

[4]*Departamento de Química, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Agronomia, 90650-001, Porto Alegre, RS, Brazil*

[5]*Instituto Nacional de Ciência e Tecnologia em Bioanalítica (INCT Bioanalítica), Campinas, SP, Brazil*



The temporal estimation of the fingermarks deposition at crime scenes is a recurring issue in forensic identification. To study this challenging topic, a preliminary study was proposed to develop a method of fingermark analysis by Fourier Transform Infrared Microscopy (μ-FTIR), using chemometric tools for time separation in a six-day aging study. The samples were collected and analyzed at hour zero, 3 days (72 hours), 4 days (96 hours), 5 days (120 hours) and 6 days (144 hours). The samples were separated into a calibration set and a test set, using Kennard Stone. Following, a comparison between variable selection tools was made of the Ant of Colony (AOC) and the Genetic Algorithm (GA) were used with subsequent application of the Linear Discriminant Analysis (LDA). The results showed that in the analyzed samples there was a predominance of sebaceous material because of the presence of saturated esters signals, with two regions of interest in the infrared spectra, the first being 1800 - 1100 cm^{-1} and the second region being 3000 - 2800 cm^{-1}. The statistical tools could group the fingermarks by donors and by age, emphasizing the separation within the tested period. More studies need to be carried out, but this work provide that μ-FTIR associated with chemometric analysis was able to separate fingermarks samples aged for up to a week.

Keywords: forensic chemistry, latent fingerprints, aging, FTIR, chemometrics.

Cite: González, M.; Mariotti, K. C.; Gomes, A. A.; Ferrão, M. F.; Limberger, R. P. Fingermark Analysis by Fourier Transform Infrared Microscopy Using Chemometric Tools. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 141–154. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-25-2021

Submitted 01 March 2021, Resubmitted 28 May 2021, Accepted 31 May 2021, Available online 08 June 2021.

INTRODUCTION

Fingermarks are the most common vestiges in crime scenes, bringing information about an individual present who may or may not be responsible for the commission of a crime [1]. For criminal investigation, the fingermark identification remains the most widely used [2]. One of the great challenges in the fingermark analysis is the aging establishment, that is, the period elapsed between the trace deposition by the donor and the analysis or development by forensic expertise. Questions about this subject have been the object of investigation by the academy and expert forces [1–7]. Placing events and evidence in time is an essential issue in investigations and, among the typical issues of forensic science, time is generally not explored. The reason for this can be attributed to the issue complexity [4]. The latent fingermarks aging is one of the most challenging problems in criminalistics. In many cases, it can be unbelievably valuable information, providing tools that will nourish the court's conviction regarding participation or elimination of potential suspects, reducing the time spent in the investigative process, minimizing errors during the prosecution and allowing the correct application of duties, rights and sanctions [5,6,8]. Therefore, the methods development that can estimate latent fingermark aging may represent an improvement in forensic procedures [9].

Specifically to fingermarks, which are physical evidence formed by biological materials, there are three main factors that influence the temporal analysis: (1) conditions inherent to the donor of the fingermark; (2) transfer conditions (deposition and the substrate or support); (3) environmental conditions in which the evidence is exposed. The time lapse determination between the deposition of the fingermark and its analysis will be the result of the complex and dynamic interaction of all these variables [7].

Different analytical methods have been employed to aging and chemical profile studies, such as gas chromatography, mass spectrometry and spectroscopy methods - Infrared and Raman [5-8]. Among these, spectroscopy methods are mostly non-destructive, low cost and quick analysis time (seconds) [10] and when it is combined with microscopy can provide good results in fingermark studies [9,10]. Spectroscopic methods, especially infrared, have been applied to characterize the composition of fingermarks, specifically the lipid components [11]. The Fourier Transform Infrared Microscopy (μ-FTIR) method is portrayed in the literature as the most common spectroscopic technique for fingermark analysis [12]. The operation of the μ-FTIR in atmospheric conditions is a significant attribute for the study of latent fingermarks, since compositional or morphological changes can be observed on different surfaces both as a function of the time elapsed after deposition and according to the environmental conditions [11,12].

The practical implementation in the forensic routine of aging fingermark studies is still put to the test. For forensic practices, the time estimation of fingermarks can change the course of an investigation. Weynermann et al., (2011), developed as the main strategy for the improvement of fingermark aging methods, the selection of a chemical target present in the sample that changes over time, but being reproducible and measurable [5]. Focusing on the analysis of lipids, one of the principal classes of compounds of interest in aging studies, the characterization of di- and triglycerides demonstrated rapid degradation over time and are frequently present in this type of sample [13,14,15].

Spectroscopic methods are intrinsic to the subsequent use of chemometrics, generating results that can increase the correlation between samples [11]. Multivariate data analysis (MDA) involves many tools, like variables selection using the ant of colony algorithm (AOC) genetic algorithm (GA) and one of the supervised methods of pattern recognition, like Linear Discriminant Analysis (LDA), that select the information on infrared spectra and describing relationships between samples and variables.

Therefore, this work developed a preliminary study for aged fingermarks analysis using μ-FTIR and chemometric tools for analysis optimization and pattern recognition between samples.

MATERIALS AND METHODS

Latent Fingermarks Deposition and Development

Latent fingermarks were collected from three Caucasian female donors, aged 30-34 years with a typical diet and without using cosmetics to minimizes the chance of anomalous results [16] Donors performed

their daily activities, the only condition where they did not wash their hands 45 minutes before collection. In moments before the collection, donors were asked to rub their fingers on the forehead and nose [7,16]. The deposition protocol comprised the collection of the fingermark of the right thumb on a reflective microscope slide provided by Agilent Technologies, previously cleaned with ethanol, exerting a force between 1.0 and 1.5 kg for 15 seconds. Two samples were collected, one from the right thumb and other from the right index finger.

To simulate real materials that are collected at crime scenes, different deposition surfaces was tested like: soda can; glass slide covered with self-adhesive film; candy packaging; silver tape; stiletto blade; aluminum foil; Tetra Pak® box.

Kinetic Conditions

The monitored times in the analysis were: hour zero, 3 days (72 hours), 4 days (96 hours), 5 days (120 hours) and 6 days (144 hours). It is described in the literature that the most significant chemical changes in fingermarks occur in the first week after deposition. The choice of short intervals was made to observe these changes [12,16]. The samples were kept at room temperature with daily monitoring and exposed to light. All analyzes were performed in the morning, five spectra were collected per donor and three spectra were selected, considering the best quality of the collection, with less external interferences and noise. Zero-time analyzes were performed up to one hour after the first collection. The analysis region was observed in the microscope connected to the equipment, choosing five random points on the fingermark ridges as shown at Figure 1.

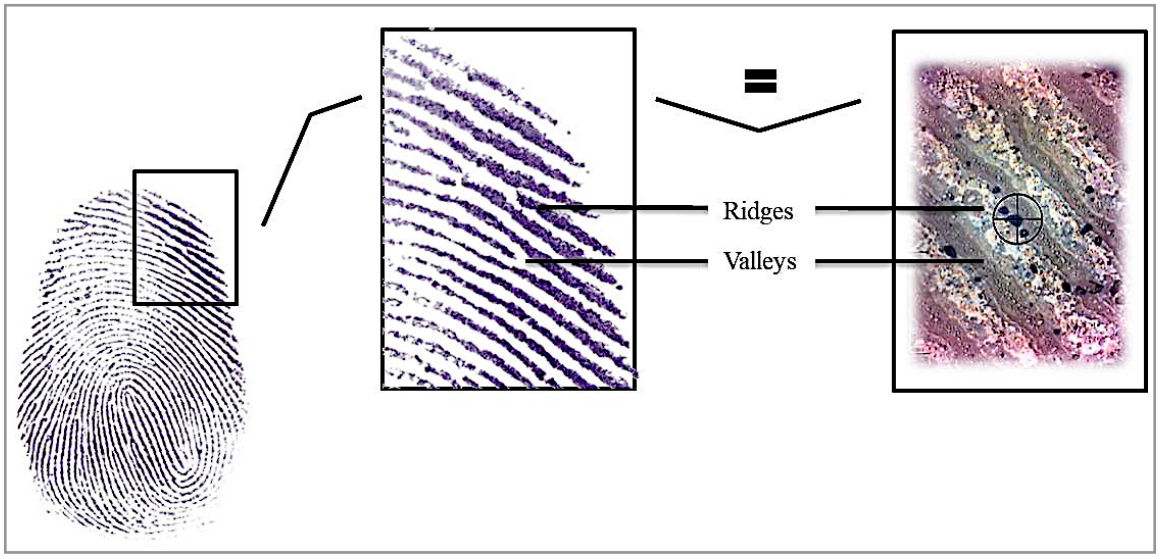


Figure 1. Visualization of fingerprint ridge details with naked eye and with microscopy (1:500 µm).

Equipment and Method Conditions

The analyzes were performed on the Cary 630 FTIR Spectrometer (Agilent Technologies - Santa Clara, CA - USA) with Survey IRTM Infrared Microspectroscopy Accessory (Czitek-Danbury, CT - USA). 320 scans were collected per sample at a resolution of 16 cm^{-1} with a spectral range of 4000 - 650 cm^{-1}. The aperture of the accessory lens was 200 µm and infrared mode in reflection and Y-Axis unit in absorbance. All samples were visualized, the analysis points were photographed, and the infrared spectra were collected using MicroLab FTIR Software (Agilent Technologies - Santa Clara, CA - USA). Background collection was carried out at each analysis to minimize external influences and level the conditions of each sample.

Chemometric Tools

The Matlab Version 7.10.0.499 software (R2010a - MathWorks Inc. Natick, MT, USA) [17] was used with the Kennard-Stone algorithm [18]; the routines of: PCA Toolbox 1.5 (Milano Chemometrics and QSAR research Group) [19]; Ant of Colony (AOC); Genetic Algorithm (GA) and Linear Discriminant Analysis (LDA) Variable Selection Toolbox [20].

The Kennard-Stone algorithm selects samples with a uniform distribution, starting with a selection of two samples with large Euclidean distance in a sample space. For each of the remaining samples, the lower distance is calculated with respect to the samples already selected. Then, the sample with the smaller distance is retained, and the procedure is repeated until a certain number of samples is selected [18].

It was used a supervised method to perform the samples classification. Due to the large number of spectral variables, it was necessary to apply a variable selection tool. Among the available resources in the software, the GA was chosen for being a classic tool while the AOC is an emerging tool, both consolidated in chemometrics studies [21].

The GA is an adaptation of the evolutionary systems of natural selection developed by Darwin applied as a resource in mathematical and computational tools. The principle of GA is to create a primordial set with several possible solutions to the problem. This set is called the initial population and each solution belonging to this population is called an individual. In this case, the genetic material is a chromosome, and this is an individual itself. This codification of individuals aims to enable the application of genetic operators and the concept of natural selection on existing solutions, thus being able to combine their genetic materials, searching for more and more adapted individuals over several generations, that is, solutions with a lower degree of error for the problem addressed [21]. The AOC is an algorithm inspired by the cooperative behavior of real ant of colonies, which search for the best path (shorter distances and fewer obstacles) between the colony and a food source, coordinated by pheromones. This tool has been successfully employed as a powerful resource for selecting variables from chemical data in multivariate calibration, with linear discriminant analysis (LDA) [21].

The LDA is a linear combination of original characteristics of the sample set which is characterized by producing the maximum separation between two populations. And its main objectives are to verify if the groups are correctly discriminated, to classify unknown observations and to verify which are the most important variables for the discrimination of these groups [19,21]. This tool takes a different approach in that it considers the existence of classes for the data; projecting the probability distribution of the data on the axes, and therefore not only maintains but highlights a linear separation of the data if it exists [19,21].

RESULTS AND DISCUSSION

Spectroscopy Findings

First, it was select the spectra region of analysis choosing the region between 1749 and 649 cm^{-1} totalizing 150 variables. This region was selected because it contains characteristic information about the sample, called fingerprint region, and other relevant signals for this study. The region selection is also necessary for limiting variables during the process of understanding the results, focusing on the recognition of patterns in the sample [22] The spectra were normalized between 0 and 1 (Figure 2). The normalization pre-processing step was used to transform the original data into an appropriate model for dataset processing, performed by scaling values in the indicated range [23,24].

It should be noted some limitations of the analytical tool, only fingermarks deposited on the company's microscopy slide showed spectra with adequate resolution for identifying bands. There was also a limitation regarding the image resolution, not being possible to observe the entire fingermark in a widefield microscopy. Therefore, it was not possible to view a complete fingermark image or make an automatic point selection for spectral analysis.

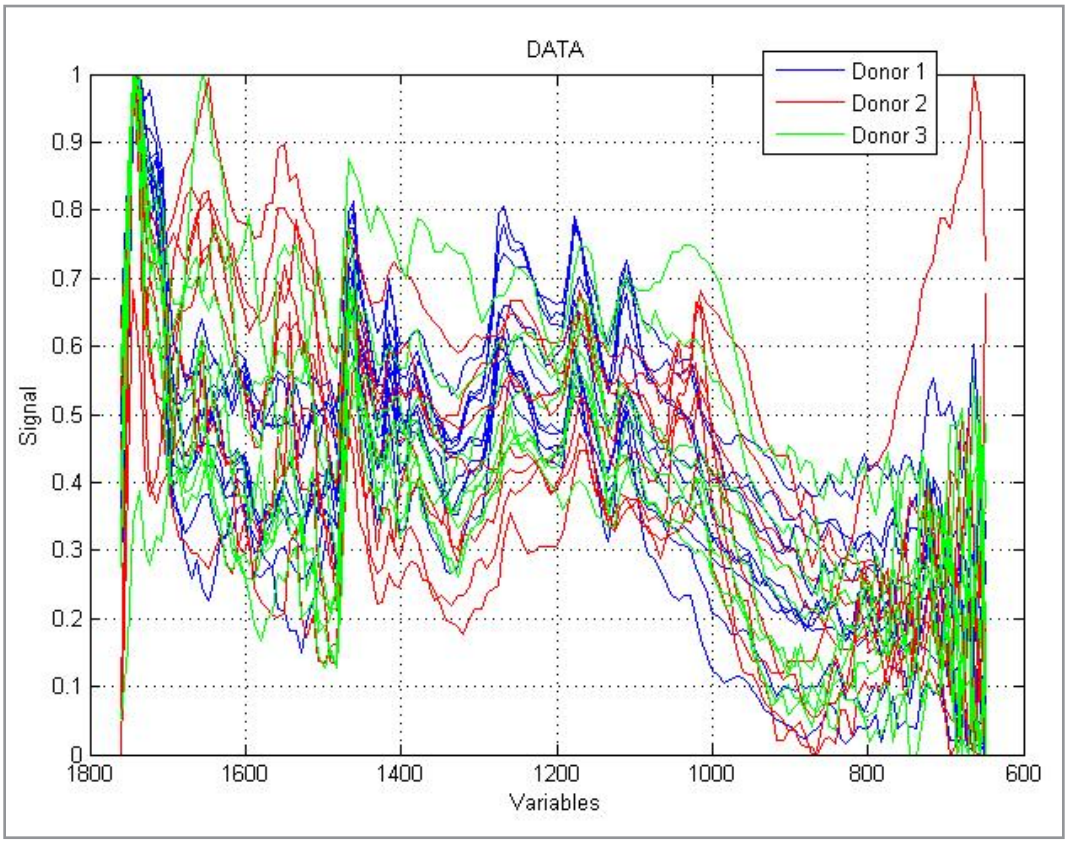


Figure 2. Normalized spectra with the selected fingerprint region.

Studies show that the infrared bands of fingermarks are similar between donors [5–7,25–28] as can be seen in Figure 3, corresponding to the overlay spectra of the three donors and their microscopic images.

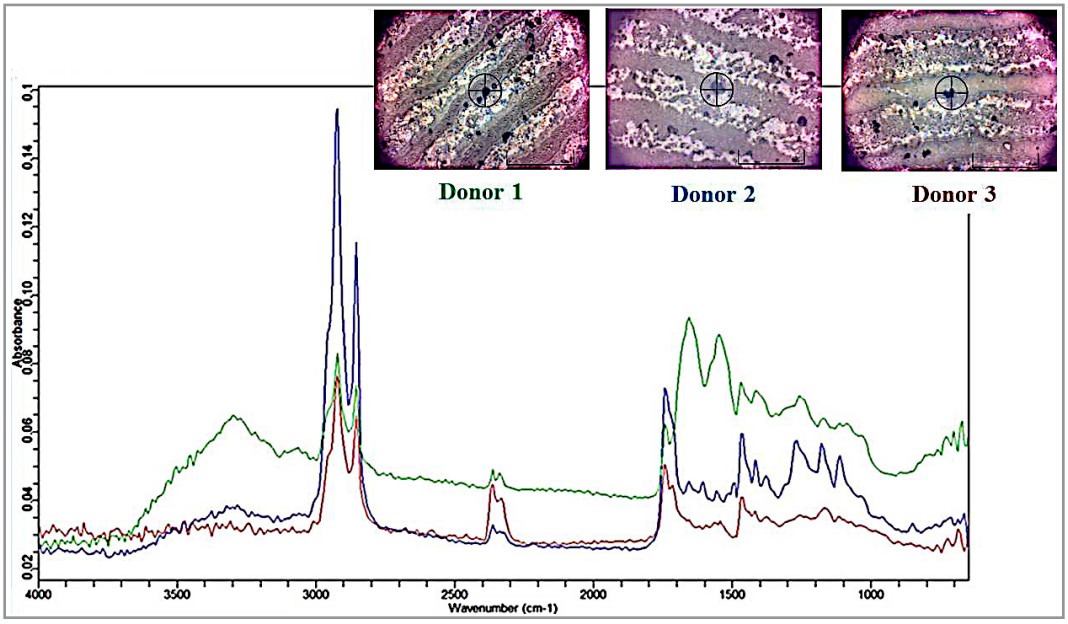
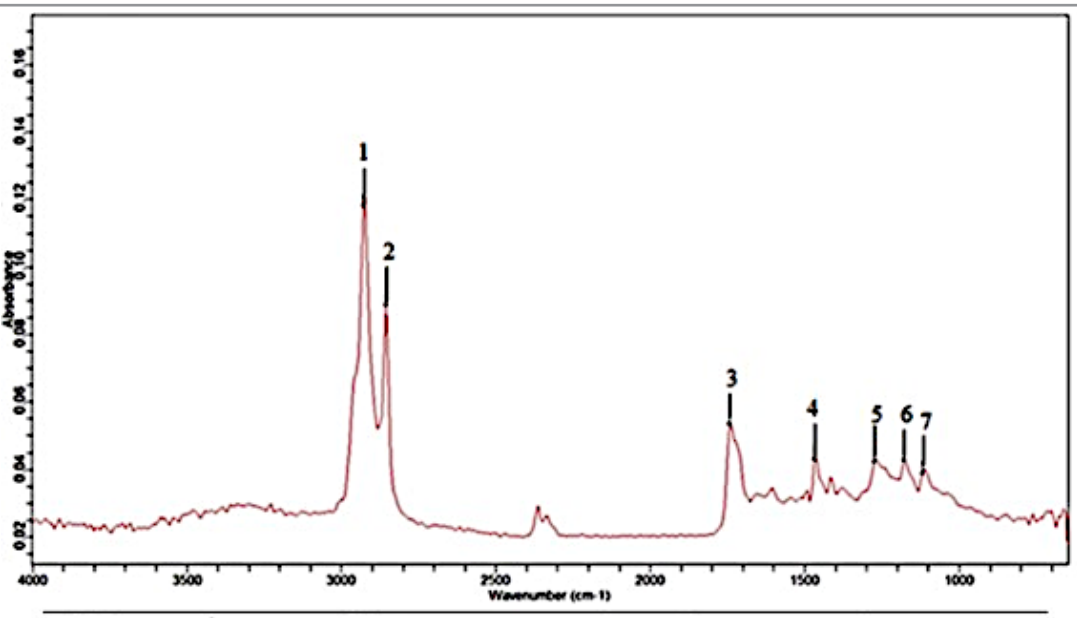


Figure 3. Comparison of spectra of the three donors with visualization of the microscopic region with an increase of 200 µm of selected analysis.

The signals correspond mainly to the major functional groups that form wax esters and fatty acids. Figure 4 summarizes the main signs observed in the standard spectrum. In the samples collected from the three donors, it was not possible to evidence eccrine secretions, as there are no signs that indicate vibrations of secondary amides, coming from proteins. According to Girod et al. (2015), eccrine secretions are represented by N-H stretches in the regions of 3200 and 1500 cm^{-1} [16], which is not reported in our study. In fact, it is more common to find patterns of spectra corresponding only to sebaceous secretions [12,16], as evidenced in our samples. In all samples collected, it was possible to see a pattern in the spectrum in which there are two more informative absorbance regions, with characteristic vibrations indicative of the presence of lipids because they have signs of saturated esters. The first region being 1800 - 1100 cm^{-1} and the second region being 3000 - 2800 cm^{-1}.



Peak	cm^{-1}	Vibrations	Compounds	Secretion
1	2929	C–H stretch (1st carbon)	Long aliphatic C chains	Sebaceous
2	2855	C–H stretch (2st carbon)	Long aliphatic C chains	Sebaceous
3	1744	C=O stretch	Saturated esters	Sebaceous
4	1468	C–H bend (asymmetric bend)	Aliphatic C chains	Sebaceous
5	1274	C–C–O (asymmetric bend)	Esters	Sebaceous
6	1177	C–C–O stretch	Saturated esters	Sebaceous
7	1118	O–C–C stretch	Saturated esters	Sebaceous

Figure 4. Fresh fingermark sample spectrum containing sebaceous material with main vibrational bands identified in the table [12,16].

The tests carried out on the materials to simulate different surfaces was inconclusive. All materials allowed the equipment to focus on the sample, making a clear image and allowing the colection of the background stage prior to each analysis. However, the materials had different roughness, making it difficult to locate the region of interest for analysis; thus, the infrared spectrum was full of noise, with baseline elevation and overlapping signals. Therefore, only the reflective slide provided by the equipment company resulted in a possible image to identify the fingermarks and the region of interest for analysis. Despite differences in the amount of material deposited by the 3 donors, it is possible to see characteristic signs pattern among the donor's samples.

Chemometric Findings

The Kennard-Stone algorithm was applied to divide the samples into the training and test sets in order to assess whether variable selection models would be applicable to the samples. Then, each variable selection model was tested five times in which the test set was tested against the model. Figures 5 show the regions of the spectrum selected by AOC and GA algorithms for donor and sample separation, respectively. It was acchieve better separation of the sample set in relation to the test set. Some regions chosen in the two models are similar, however, it was observed that the region with less intense signals is not chosen by the AOC model, which may suggest a performance slightly lower than the GA. The results were organized below according to the sample separation between time and donors.

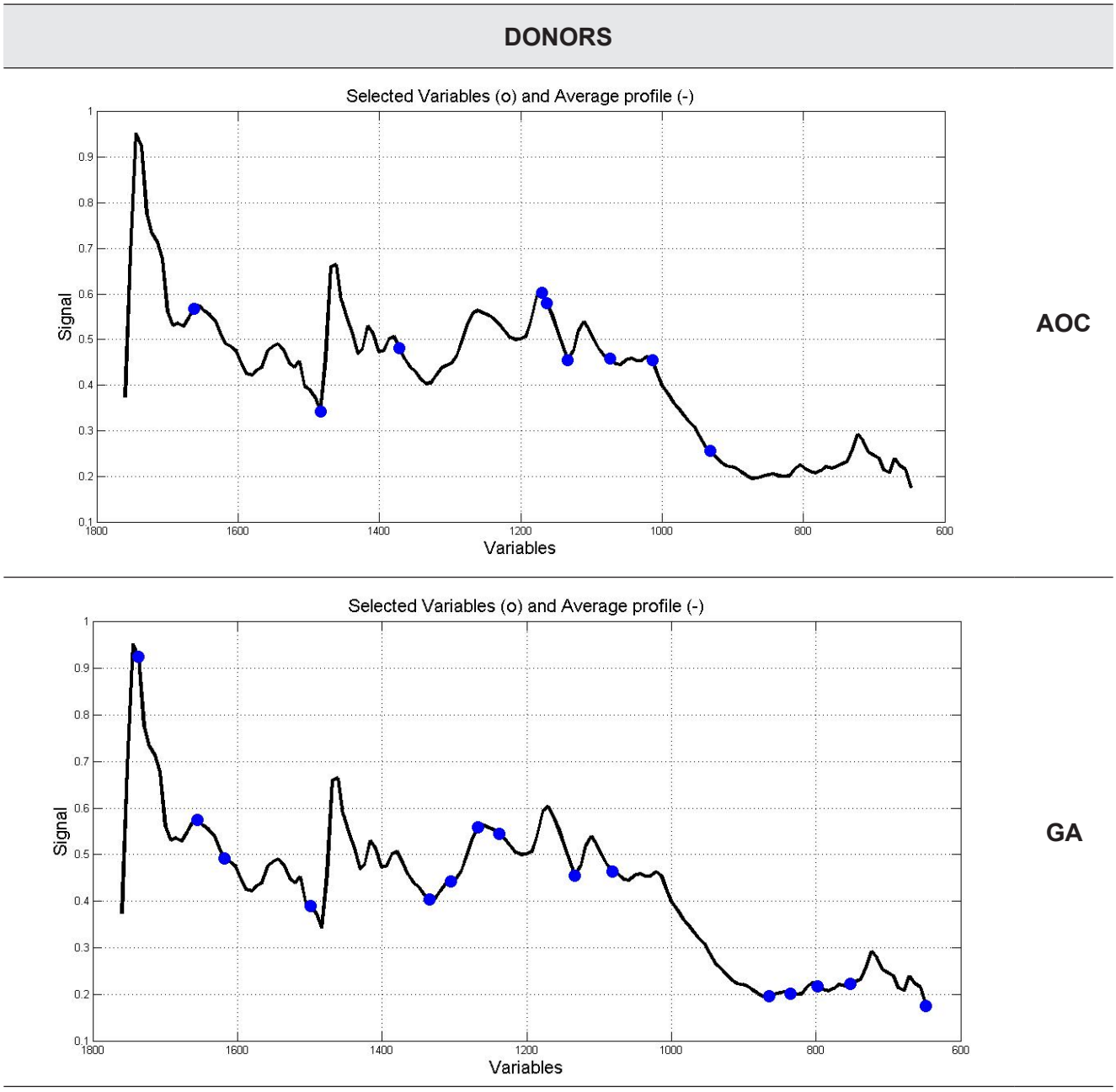


Figure 5. Regions of interest responsible for sample separation scores.

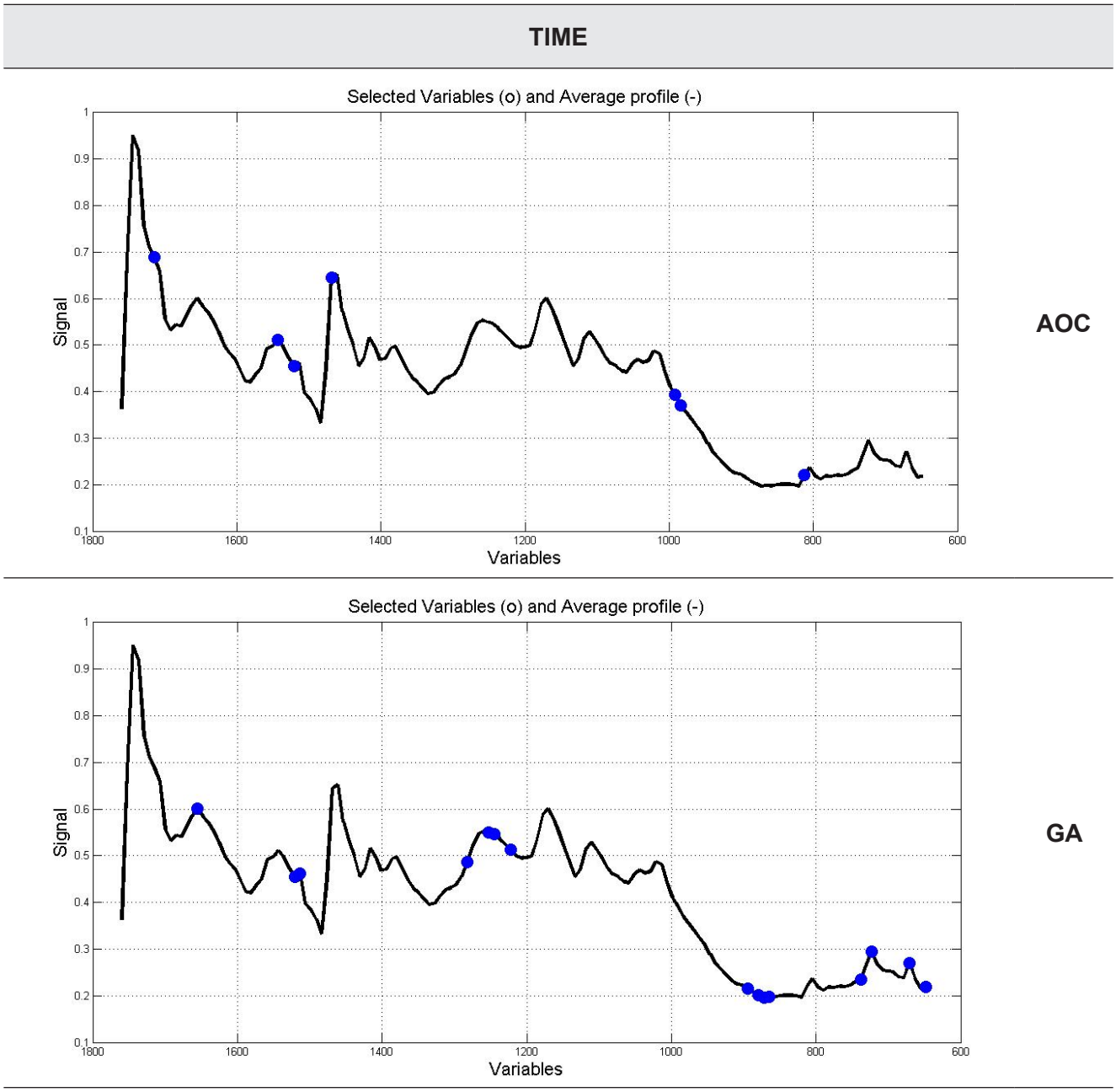


Figure 5. Regions of interest responsible for sample separation scores. (Continuation)

Donor Separarion Results

Comparing the results of AOC and GA with subsequent application of the LDA (Figure 6) it is possible to observe a clear sampling separation of both the training set and the test set for the three different donors. Thus, with these tools, it was possible to see that even with a small amount of samples the differentiation of classes (donors and time) could be performed. With regard specifically to fingermark samples, it is considered impossible to conduct an aging model that works for all types of fingerprints because of the great variability between different donors and the same donor [5,6,11,16]. Thus, Girod et al. (2015) suggested the construction of a new model in order to study from the initial composition to the aging of the sample from a single donor [16]. Based on this idea, we use only three female donors, because female samples tend to have greater intra and inter donor variation (make up and cosmetics uses and hormonal

differences). It can be seen in Figure 6 that the prediction samples (stars) are slightly apart compared to the training samples (balls). In both results only a sample of the test group is shifted in the graphs, this shows that in terms of responses the methods are equivalent. In addition, GA is a slightly better model, but there is no greater representativeness than the point of discarding the AOC model.

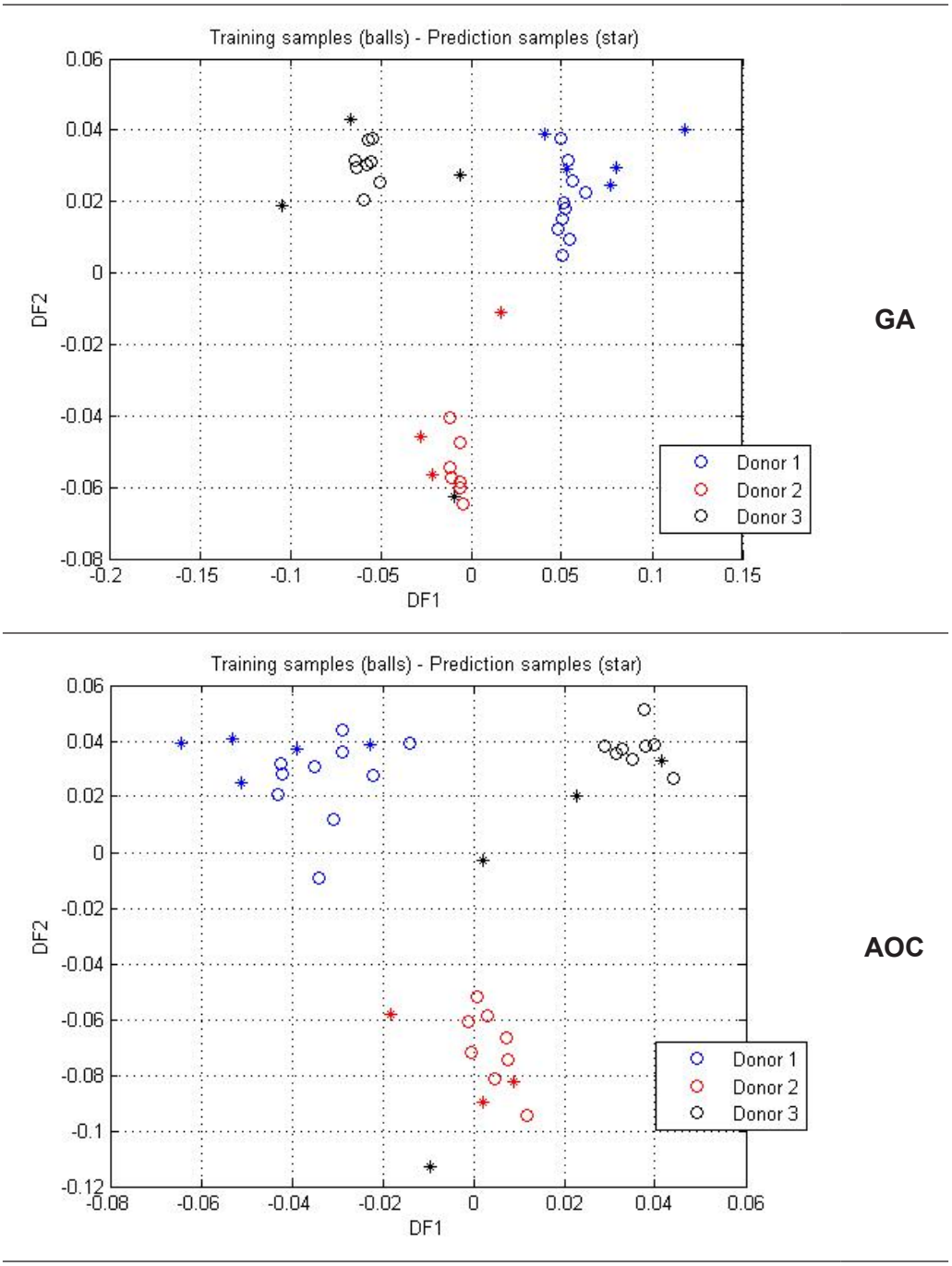


Figure 6. Results of sample separation by GA and AOC models.

Time Separation Results

One of the forms to see fingermarks aging is to follow the time evolution of the sample. According to the number of samples available, only four of the six times analyzed were used to create the test set and

training set. The result was similar for both GA and AOC, in which, even with a small number of samples, it was possible to observe a trend of separation, as shown in Figure 7. Thus the application of this model can provide consistent results if applied to a larger number of samples.

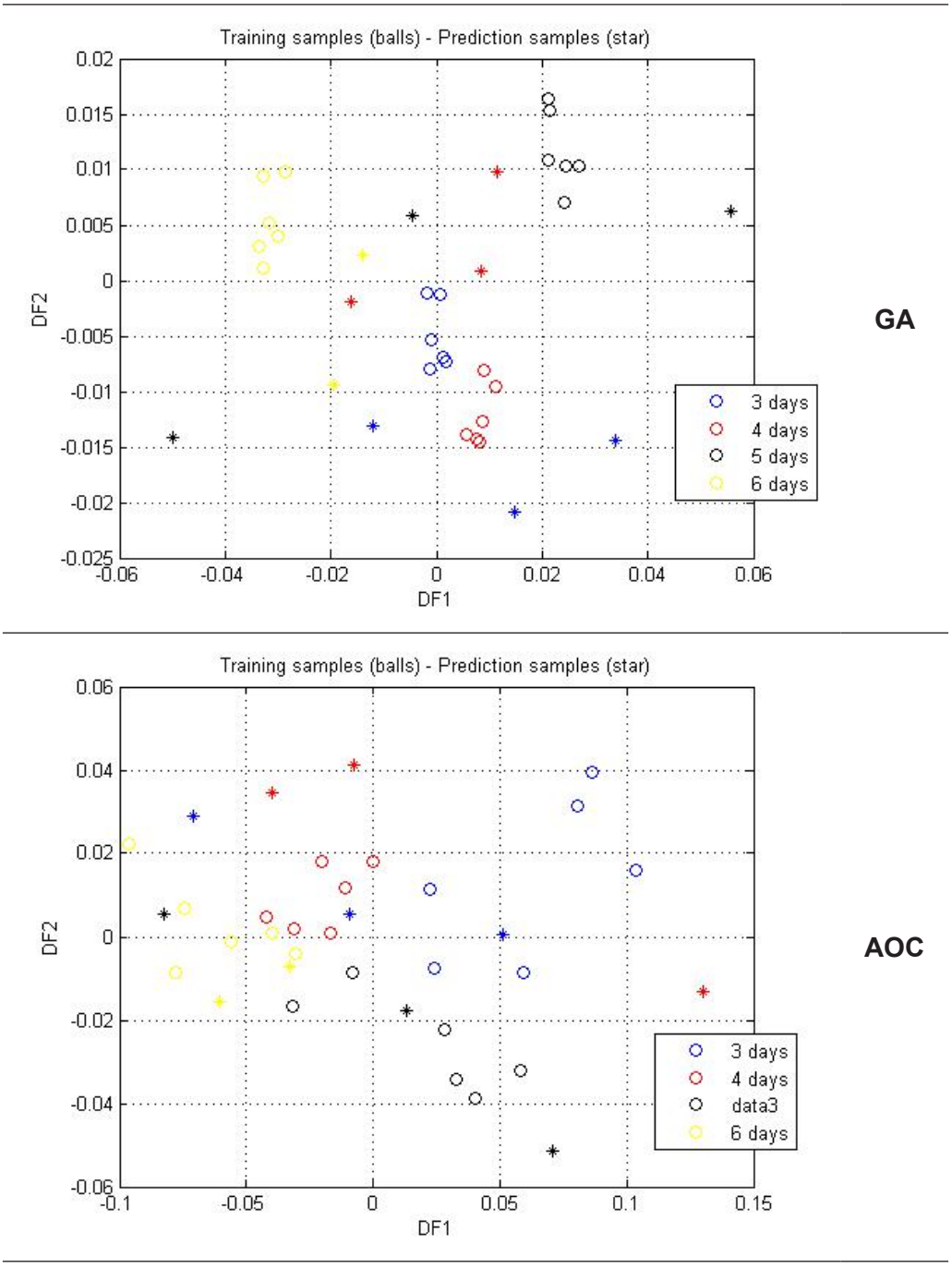


Figure 7. Samples separation in four days.

The spectra of aged fingermarks originally containing mainly lipid compounds (i.e., sebaceous secretions) showed a decrease in the intensity of all vibrational bands. According to Girod et al. (2015) study, fingermarks could be grouped by age, but the aging profile was significantly influenced by storage conditions and also by substrate when specimens were stored in the dark [16].

Table I shows the performance of the models for the separation between donors and the analysis times. The classes are the sample categories, being three donors and four times evaluated, excluding the test set. Precision expresses the agreement degree between the results of a series of measurements carried out for the same homogeneous sample under determined conditions [29–31] and it was calculated according the Equation 1, for each class. The sensitivity, samples belonging to the class and classified correctly in this class, were considered for the evaluation of the classification achieved with the multivariate methods and was calculated to Equation 2. The specificity, samples not belonging to the modeled class and correctly classified as not belonging, was calculated according to Equation 3:

$$Precision = \frac{TP}{TP + FP}\ x\ 100 \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN}\ x\ 100 \quad (2)$$

$$Specificity = \frac{TN}{TN + FP}\ x\ 100 \quad (3)$$

where TP is the number of true positives; TN is true negatives; FP is false positives and FN is false negatives [29-31]. It is possible to observe the performance of the models proved to be superior for sample separation by donor, due to greater accuracy and less errors. However, although there is no clear division as to the analysis times, it is possible to see a homogeneity of the classes in the samples for time analysis.

Table I. Performance of selected models for separating samples into donors and time

Samples	**Donors**						**Time**							
Variable Selection	**GA**			**AOC**			**GA**				**AOC**			
Accuracy rate (%)	93.98			91.67			65.97				77.43			
Misclassification rate (%)	6.01			6.02			34.02				22.56			
Total errors	1			1			6				4			
Class	1	2	3	1	2	3	1	2	3	4	1	2	3	4
Precision	1	0.75	1	1	0.75	1	0.5	0.5	0.33	0.5	0.5	0.66	1	0.5
Sensitivity	1	1	0.75	1	1	0.5	0.33	0.33	0.33	1	0.33	0.66	0.66	1
Specificity	1	0.88	1	1	0.88	1	0.87	0.87	0.75	0.77	0.87	0.87	1	0.77

The fingermarks samples have great variability between donors [4-7], so, it was decided to carry out the pilot study with a small number of donors in order to establish a method that could be reproduced in the forensic routine. Thus, an preliminary aging model was built with 3 donors of the same sex to verify whether samples would be separated without directly considering gender as a variant. The use of the μ-FTIR method provided comparable spectra in a short period of analysis. The chemometric tools allowed to see that in a single infrared spectrum collected in the described parameters it contains thousands of variables that, although correlated, much information does not necessarily explain a significant variance related to differentiation by time, environmental exposure or effects of the sample itself. Therefore, the selection of variables performed by chemometrics was important [30].

The use of AOC and GA reduced the data set, selecting the relevant variables and the LDA allowed the separation of samples by time. Even for a period of six days, it was possible to observe that the most important transformations occurred in the samples from time zero to 72 h, 120 h and 144 h, as observed in Figure 8. This opens the possibility of sparing the analyzes for the next studies.

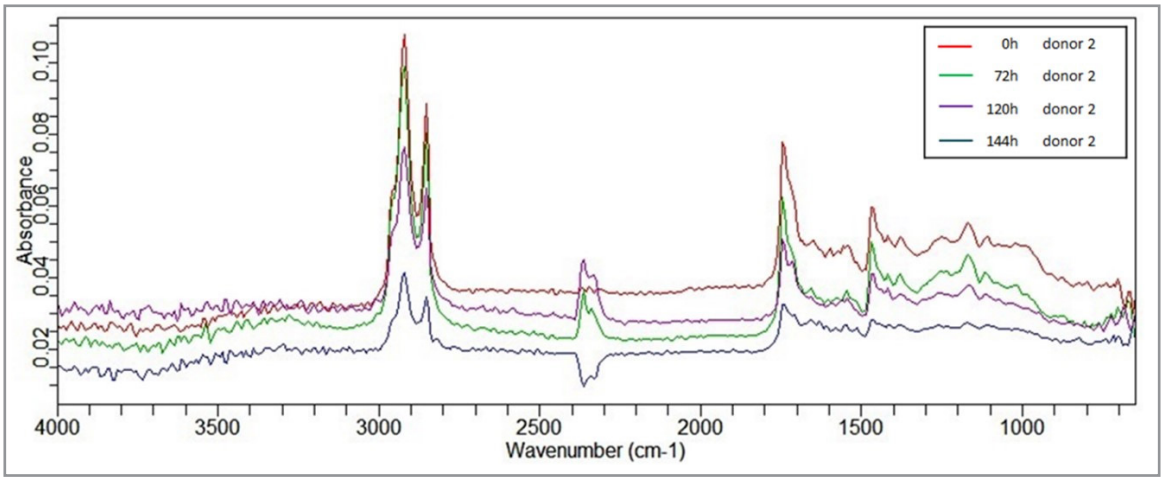


Figure 8. Overlaid samples of donor 2 in aging parameters. The inversion of bands shown in the middle of the spectrum, between 2300 - 200 cm $^{-1}$, corresponds to CO_2. These changes can occur in the spectrum because the sample is not subjected to vacuum during an analysis, thus, the spectrum can undergo changes according to the instability of the environment.

The limiting steps of the study are: (i) it was not possible adapted the method for other surfaces of sample collection; (ii) need to increase the number of the samples and expand to both genders to be more representative; (iii) need to adapted to be used in real cases.

In general, it was possible to observe a pattern in the fingerprint spectra analyzed by μ-FTIR corresponding to sebaceous components of the samples. For a period of six days it was possible to see an decreasing intensity of the compounds as the days passed, suggesting a characteristic degradation capable of being observed in studies related to fingermark aging analysis.

CONCLUSIONS

A preliminary fingermark study was carried out with three donors in order to establish a sample aging profile to suggest a temporal estimate of the forensic evidence. The μ-FTIR method enabled a fast and non-destructive analysis, but with some but with some limitations concerning about the surface were fingermark can be analysed. It was possible to verify the presence of sebaceous components that had degradation by decreasing the signal in the infrared spectrum in the period of six days and also evidenced by the use of the GA and LDA chemometric tools. In addition, it was possible to separate donors, even with a small number of samples and a short study interval. The μ-FTIR method linked to the use of

chemometrics is promising for the reality of the forensic expertise in Brazil and is already consolidated in other countries, but it needs more concrete studies to be placed in the forensic routine as an additional resource in the forensic identification protocols.

Conflicts of interest

The authors declare that there is no conflict of interest. This work was supported by "Fundo de Amparo à Pesquisa do Rio Grande do Sul" (FAPERGS) – fund number 17/2551-0000839-1 and by National Institute of Science and Technology (INCT Forensics – CNPq 465450/2014-8).

Acknowledgements

This work was supported by: "Fundo de Amparo à Pesquisa do Rio Grande do Sul" (FAPERGS) – grant nº 17/2551-0000839-1; National Institute of Science and Technology (INCT Forensics – CNPq 465450/2014-8; "Conselho Nacional de Desenvolvimento Científico e Tecnológico" (CNPq) and "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) – fund nº 88882.345923/2019-01.

REFERENCES

1. Mariotti, K. C. *Brazilian Journal of Forensic Sciences, Medical Law and Bioethics*, **2020**, *9* (2), pp 210–228 (http://dx.doi.org/10.17063/bjfs9(2)y2020210).
2. Peterson, J.; Sommers, I.; Baskin, D.; Johnson, D. *The Role and Impact of Forensic Evidence in the Criminal Justice Process*. US Department of Justice, Washington DC, **2010**, Chapter 2, p 14.
3. Girod, A; Spyratou, A; Holmes, D.; Weyermann, C. *Sci Justice*, **2016**, 56 (3), pp 165–180 (http://dx.doi.org/10.1016/j.scijus.2015.12.004).
4. Girod, A.; Ramotowski, R.; Lambrechts, S.; Misrielal, P.; Aalders, M.; Weyermann, C. *Forensic Sci. Int*., **2016**, *262*, pp 212–226 (http://dx.doi.org/10.1016/j.forsciint.2016.03.021)
5. Weyermann, C.; Roux, C.; Champod, C. *J. Forensic. Sci*., **2011**, *56* (1), pp 102–108 (http://dx.doi.org/10.1111/j.1556-4029.2010.01523.x).
6. Weyermann, C.; Ribaux, O. *Sci. Justice*, **2012,** *52* (2), pp 68–75 (http://dx.doi.org/10.1016/j.scijus.2011.09.003).
7. Girod, A.; Ramotowski, R.; Weyermann, C. *Forensic Sci. Int*., **2012**, *223* (1-3), pp 10-24 (https://doi.org/10.1016/j.forsciint.2012.05.018).
8. De Alcaraz-Fossoul, J.; Patris, C. M.; Freixat, C. B.; McGarr, L.; Brandelli, D.; Stow, K.; Badia, M. G. *J. Forensic. Sci*., **2016**, *61* (2), pp 322–333 (http://dx.doi.org/10.1111/1556-4029.13007).
9. Barros, R. M.; Faria, B. E. F.; Kuckelhaus, S. A. S. *Sci. Justice*, **2013**, *53* (4), pp 402–408 (http://dx.doi.org/10.1016/j.scijus.2013.08.002).
10. De Alcaraz-Fossoul, J.; Feixat, C. B.; Tasker, J.; McGarr, L.; Stow, K.; Carreras-Marin, C.; Oset, J. T.; Badia, M. G. *J. Forensic. Sci*., **2016**, *61* (4), pp 947-958 (http://dx.doi.org/10.1111/1556-4029.13099).
11. Girod, A.; Ramotowski, R.; Lambrechts, S.; Misrielal, P.; Aalders, M.; Weyermann, C. *Forensic Sci Int*., **2016**, *262*, p 212.
12. Johnston, A.; Rogers, K. *Sci. Justice*, **2018**, *58* (2), pp 121–127 (https://doi.org/10.1016/j.scijus.2017.11.004).
13. Cadd, S.; Islam, M.; Manson, P.; Bleay, S. *Sci. Justice*, **2015**, *55* (4), pp 219–238 (http://dx.doi.org/10.1016/j.scijus.2015.02.004).
14. Frick, A. A.; Weyermann, C. *Analyst*, **2019**, *144* (11), pp 3590–3600 (http://dx.doi.org/10.1039/c9an00521h).
15. Pleik, S.; Spengler, B.; Bhandari, D. R.; Luhn, S.; Schäfer, T.; Urbach, D.; Kirsch, D. *Analyst*, **2018**, *143* (5), pp 1197–1209 (http://dx.doi.org/10.1039/c7an01506b).
16. Girod, A.; Xiao, L.; Reedy, B.; Roux, C. *Forensic Sci. Int*., **2015**, *254*, pp 185–196 (http://doi.org/10.1016/j.forsciint.2015.07.022).
17. https://www.mathworks.com/products/matlab.html [Accessed 21 June 2020].

18. Kennard, R. W.; Stone, L. A. *Technometrics*, **1969**, *11* (1), pp 137–148 (https://doi.org/10.1080/00401706.1969.10490666).
19. Ballabio, D. *Chemom. Intell. Lab. Syst*., **2015**, *149* (Part B), pp 1–9 (https://doi.org/10.1016/j.chemolab.2015.10.003).
20. Homemade routine available under request to araujo.gomes@ufrgs.br
21. Consonni, V.; Baccolo, G.; Gosetti, F.; Todeschini, R.; Ballabio, D. *Chemom. Intell. Lab. Syst*., **2021**, *213* (https://doi.org/10.1016/j.chemolab.2021.104313).
22. Smith, B. C. *Fundamentals of Fourier transform infrared spectroscopy*, 2nd ed., CRC Press, **2011**, Chapter 1, p 10.
23. Ferreira, M. M. C. *Quimiometria: Conceitos, Métodos e Aplicações*. Editora UNICAMP, Campinas, **2016**, Capítulo 5, p 409.
24. Pontes, A. S.; Araújo, A.; Marinho, W.; Diniz, P. H. G. D.; Gomes, A. A.; Goicoechea, H. C.; Silva, E. C.; Araújo, M. C. U. *J. Chemom*., **2020**, *34* (12) (https://doi.org/10.1002/cem.3292).
25. Kumar, R.; Sharma, V. *TrAC, Trends Anal. Chem*., **2018**, *105*, pp 191-201 (https://dx.doi.org/10.1016/j.trac.2018.05.010).
26. Champod, C. *Fingerprints and Other Ridge skin Impressions.* 2nd Ed. CRC Press, Boca Raton, **2016**, Chapter 4, p 119.
27. Frick, A. A.; Fritz, P.; Lewis, S. W. Chemistry of Print Residue. In: Siegel, J. A.; Saukko, P. J.; Houck, M. M. (Eds). *Encyclopedia of Forensic Sciences* (2nd Edition). Academic Press, Waltham, **2013**, pp 92–97 (https://doi.org/10.1016/B978-0-12-382165-2.00262-2).
28. Francese, S.; Bradshaw, R.; Ferguson, L. S.; Wolstenholme, R.; Clench, M. R.; Bleay, S. *Analyst*, **2013**, *138*, pp 4215–4228 (https://doi.org/10.1039/C3AN36896C).
29. Grasel, F. S.; Marcelo, M. C. A.; Ferrão, M. F. *RSC Adv*., **2016**, *6*, pp 32358-32364 (http://dx.doi.org/10.1039/c6ra00900j).
30. Grasel, F. S.; Ferrão, M. F. *Anal. Methods*, **2016**, *8*, 644 (http://dx.doi.org/10.1039/c5ay02526e)
31. Valderrama, P.; Braga, J. W. B.; Poppi, R. J. *Quim. Nova,* **2009***, 32* (5), pp 1278-1287 (http://dx.doi.org/10.1590/S0100-40422009000500034).

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 155–164
doi: 10.30744/brjac.2179-3425.AR-26-2021

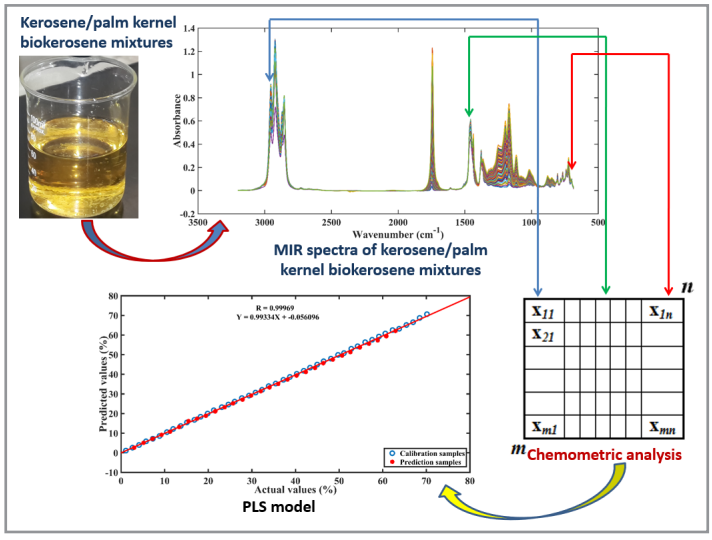
BrJAC

ARTICLE

Rapid Quantification of the Palm Kernel Biokerosene Content in Mixtures with Aviation Kerosene using MIR Spectroscopy and Multivariate Regression by PLS

Ademar Domingos Viagem Máquina[1,2*], Maria Teresa Carvalho Ferreira[1], Edvando Souza Teles[1], Douglas Queiroz Santos[3], Waldomiro Borges Neto[1]

[1]*Instituto de Química, Universidade Federal de Uberlândia, Campus Santa Mônica, 38408-100, Uberlândia, MG, Brazil*
[2]*Universidade de Púnguè, Faculdade de Ciências Exactas e Tecnológicas, Rua Nacional 6, CEP 323, Chimoio, Manica, Moçambique*
[3]*Escola Técnica de Saúde, Universidade Federal de Uberlândia, Campus Umuarama, 38400-902, Uberlândia, MG, Brazil*



An alternative methodology was developed to monitor the biokerosene content of palm kernel in blend with kerosene using medium infrared spectroscopy associated with partial least squares (PLS). The efficiency of this methodology was analyzed based on the parameters of accuracy and figures of merit. The values of root-mean-square error of cross-validation (RMSECV), root-mean-square error of calibration (RMSEC) and root-mean-square error of prediction (RMSEP) were in agreement because the RMSEP was higher than RMSECV and RMSEC. In addition, the RMSEP value is considered acceptable according to the Brazilian standard ABNT NBR 15568 because it is less than 1%. The figures of merit were performed in agreement with the requirements established in the standard ASTM E1655-05. The linearity of the model was assessed based on the analysis of the model fit through the correlation of the actual and predicted values of the calibration and prediction sets, where a high correlation between the values was evidenced, with a correlation coefficient (R) exceeding 0.99. The good results of the application of MIR spectroscopy combined with multivariate regression by PLS suggest that this analytical methodology is feasible, efficient and suitable for use by inspection agencies to control the biokerosene content of palm kernel in mixture with diesel.

Keywords: Chemometric method, Monitoring, Biofuels, Fuel.

Cite: Máquina, A. D. V; Ferreira, M. T. C.; Teles, E. S.; Santos, D. Q.; Borges Neto, W. Rapid Quantification of the Palm Kernel Biokerosene Content in Mixtures with Aviation Kerosene using MIR Spectroscopy and Multivariate Regression by PLS. *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 155–164. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-26-2021

Submitted 28 February 2021, Resubmitted 05 May 2021, Accepted 25 May 2021, Available online 30 June 2021.

INTRODUCTION

Aviation kerosene, commercially known as QAV, is a fuel derived from non-renewable energy sources used mainly in the airline sector. According to data provided by the International Air Transport Association - IATA, the use of this fuel by this sector generates approximately 2% of carbon dioxide emissions into the atmosphere, reaching a level of 3% by 2030 [1,2].

Several countries have sought to impose limits on current and future emissions to mitigate these polluting gases in the environment. These emissions represent a global concern in the 21st century and may increase with the expansion of the aviation industry. Thus, it is necessary to use environmentally sustainable fuel as an alternative to fossil fuel [3].

Aviation biokerosene (BioQAV) is a fuel derived from renewable sources with the potential to replace fossil kerosene. Some oilseeds such as jatropha (*Jatropha curcas L.*), camelina (*Camelina sativa*), babassu (*Attalea speciosa*), macaúba (*Acrocomia aculeata*) and palm kernel (*Elaeisguineensis*) stand out in the production of BioQAV because they have high levels of fatty oils in their composition whose carbon chain is similar to fossil kerosene [4].

The use of biokerosene in the aviation sector also produces carbon dioxide, but its production cycle reduces this environmental damage with the absorption of CO_2 by plants in the process of photosynthesis. This process allows the biofuel to be carbon neutral during its life cycle. Another advantage of biofuel is that it is sulfur-free, which is one of the elements responsible for acid rain [5,6].

In Brazil, there is the National Biokerosene Program created in 2009 by law No. 3213/2009, which establishes the development of research on renewable fuels from biomass without changes in the technologies established in the turbine engines. That is, the composition of these fuels must guarantee the safety of the aviation system [7,8]. In this context, it is necessary to develop analytical methods that provide fast and reliable responses to assess the biokerosene content in the kerosene/biokerosene mixture.

For quantification purposes, several studies report the use of spectroscopic techniques associated with multivariate regression by partial least squares (PLS). Some of these works present the quantification of adulterants in biodiesel / diesel mixtures, as is the case of the authors De Souza et al. [9], proposes a new method for the quantitative analysis of soybean oil and sunflower oil as adulterants in oil of extra virgin flaxseed, using MIR spectroscopy associated with PLS. The models obtained were built according to the standard ASTM E1655-05, having obtained acceptable error values and good correlation between the measured and predicted values of the calibration and prediction sets. Máquina et al. [10], developed two methodologies to quantify and classify the cotton biodiesel content in mixtures with diesel, using MIR spectroscopy associated with PLS and Discriminant Analysis by Partial Minimum Squares (PLS-DA) methods. The PLS model developed to determine the biodiesel content showed a good fit, with a correlation coefficient of the measured and predicted values exceeding 0.99.

However, there are no reports of published studies on the quantification of the biokerosene content of palm kernel in the kerosene/biokerosene mixture, using rapid and non-destructive analytical techniques associated with PLS regression methods. Thus, the present work aims to develop an analytical methodology that can be used by inspection bodies to quantify the biokerosene content of palm kernel in mixtures with kerosene, using MIR spectroscopy associated with the PLS chemometric method.

MATERIALS AND METHODS

Sample preparation

In this study, palm kernel biokerosene provided by LABIO (Biofuels Laboratory of the Chemistry Institute) of the Federal University of Uberlândia was used. Transpetro S/A (Brazil) supplied the pure kerosene used to prepare the kerosene/palm kernel biokerosene mixture, adding biokerosene to the kerosene in a concentration ranging from 1.00 to 70.00% (v/v). For the construction of the PLS model, 45 samples were used in the calibration set and 30 samples in the forecast set.

Acquisition of spectral data

The MIR spectra were acquired using a PerkinElmer Spectrum Two spectrometer (PerkinElmer, Waltham, MA, USA) equipped with a HATR accessory and ZnSe crystal (Pike Techonologies). The spectra were recorded in the range of 4000–600 cm^{-1} with a 4 cm^{-1} resolution and were acquired using 16 scans for each of the quintuplicates.

Chemometrics analysis

MATLAB software, version R2018b (Mathworks, Inc.) and PLS_Toolbox, version 8.9.1 (Eigenvector Research) were used to pre-process the data and develop the PLS model. To perform the multivariate procedures, the MIR spectra data were organized in an ordered array of rows and columns, constituting a matrix **X**, where each row corresponds one sample *"m"* and each column corresponds to one variable *"n"*, in the which *m* = 1,2,3, …, 75 and *n* = 1,2,3, …, 1306.

The spectral baseline was corrected in the bands of 4000–3100 and 2500–1850 cm^{-1} to minimize undesirable systematic variations in the data and then they were centered on the mean. Finally, a **Y** matrix was created containing concentration values (from 1.00 to 70.00% (v / v)) of the samples.

In the construction of the PLS model, matrix **X** is correlated with matrix **Y** through mathematical operations to obtain Latent Variables (LVs) and Regression Coefficients, used to achieve the maximum covariance between the spectra and concentrations of the species of interest and determine the concentration value of each spectral profile, respectively [9]. The purpose of this process is to find a small number of relevant factors that are predictive of **Y** and that use **X** efficiently. To do this, matrix **X** is decomposed into a set of orthogonal factors that are used to adjust **Y**, according to Equation 1. Matrix **Y** is decomposed according to Equation 2: [10]

$$\mathbf{X} = \mathbf{TP^T} + \mathbf{E}_X = \sum \boldsymbol{t}_h \boldsymbol{p}_h^T + \boldsymbol{E}_X \quad (1)$$

$$\mathbf{Y} = \mathbf{UQ^T} + \mathbf{E}_Y = \sum \boldsymbol{u}_h \boldsymbol{q}_h^T + \boldsymbol{E}_Y \quad (2)$$

where, **X** and **Y** are matrices that contain instrumental measurement and response (concentration) data, respectively; **T** and **U** are the (*m x A*) scores for the two data matrix; **P** and **Q** are the (*n x A*) loadings respective, ***h*** is the latent variable number (LV), $\mathbf{E}_X$ and $\mathbf{E}_Y$ are the respective residues. However, if block **Y** is univariate, weight **Q** is set to 1 [11].

The linear relationship between the two matrices is established by the correlation of the **X** and **Y** scores for each LV (***h***), according to Equation 3: [12]

$$\mathbf{U_h} = \mathbf{b_h T_h} + \mathbf{E} \quad (3)$$

where, $\mathbf{U}_h$ is a matrix that contains the properties of all samples (concentration), $\mathbf{b}_h$ is a vector that contains the model parameters, $\mathbf{T}_h$ is a matrix that contains instrumental measurement data (spectrum) for the calibration samples and **E** is a matrix that represents noise.

The PLS model was constructed by minimizing the waste matrices $\mathbf{E}_X$ and $\mathbf{E}_Y$, at the same time, thereby obtaining a linear relationship between **t** and **u**, through the method of cross-validation by the venetian blind criterion, with 14 blocks of data division and with two samples per block. The number of LVs that provide the lowest RMSECV was selected, following the requirements of ASTM E1655-55 (2012) [13].

Once completed, the model was validated based on the determination of the following figures of merit: selectivity, sensitivity, analytical sensitivity, limit of detection, limit of quantification and test for systematic error (bias and t_{bias}), according to the equations presented in Table I [14–17]. The fit of the PLS model

was analyzed based on the correlation between the current and predicted concentration values of the calibration and forecast sets.

Table I. Equations used to calculate the merit figures of the PLS model*

Figure of merit	Equation
Selectivity	$\hat{\mathrm{SEL}}_i = \frac{\hat{\mathrm{nas}}_i}{\|\mathrm{x}_i\|}$
Sensitivity	$\hat{\mathrm{SEN}} = \frac{1}{\|\mathrm{b}_k\|}$
Analytical sensitivity	$\gamma = \frac{\hat{\mathrm{SEN}}}{\|\delta_x\|}$
Limit of detection	$\mathrm{LD} = 3.3\delta_x \frac{1}{\hat{\mathrm{SEN}}}$
Limit of quantification	$\mathrm{LQ} = 10\delta_x \frac{1}{\hat{\mathrm{SEN}}}$
bias	$\mathrm{bias} = \frac{\sum_{i=1}^{n_{val}}(y_i - \hat{y}_i)}{n_{val}}$
SDV	$\mathrm{SDV} = \sqrt{\frac{\sum[(y_i - \hat{y}_i) - \mathrm{bias}]^2}{n_{val} - 1}}$
t_{bias}	$t_{bias} = \frac{\lvert\mathrm{bias}\rvert\sqrt{n_{val}}}{\mathrm{SDV}}$

*where, $\hat{y}_i$ is the predicted value using the PLS model; y_i is the reference value for the sample I; $\boldsymbol{n_{val}}$ is the number of validation samples; $\mathbf{nâs}_i$ is the norm of the NAS vector and $\|\mathbf{x}_i\|$ is the norm for each spectrum; $\mathbf{b}_k$ is the vector of the final regression coefficients, δ_x is the standard deviation value of 9 kerosene samples.

RESULTS AND DISCUSSION

Spectrum MIR

Figure 1 shows the MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v) before being pre-processed, where noise is observed in the region from 600 cm^{-1} to 700 cm^{-1} and low signal variation in the region from 3100 cm^{-1} to 4000 cm^{-1}.

Figure 2 shows the MIR spectra of (a) kerosene (b) palm kernel biokerosene and Figure 3 shows the MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v) pre-processed. In these spectra, similarities are noted in the following significant absorption bands: at 2950 cm^{-1} – attributed to the asymmetric stretching vibrations of the C−H bond of methyl groups (−CH_3); at 2923 cm^{-1} – attributed to the asymmetric stretching vibration of the C−H bond of methylene group (−CH_2); at 2851 cm^{-1} – attributed to the symmetrical stretching vibration of the C−H bond of methylene group (−CH_2); at 1451 cm^{-1} – attributed to the asymmetric angular deformations of the C−H bond of methyl groups (−CH_3) and at 1379 cm^{-1} – attributed to the symmetric angular deformations of the C−H bond of methyl groups (−CH_3) [18,19].

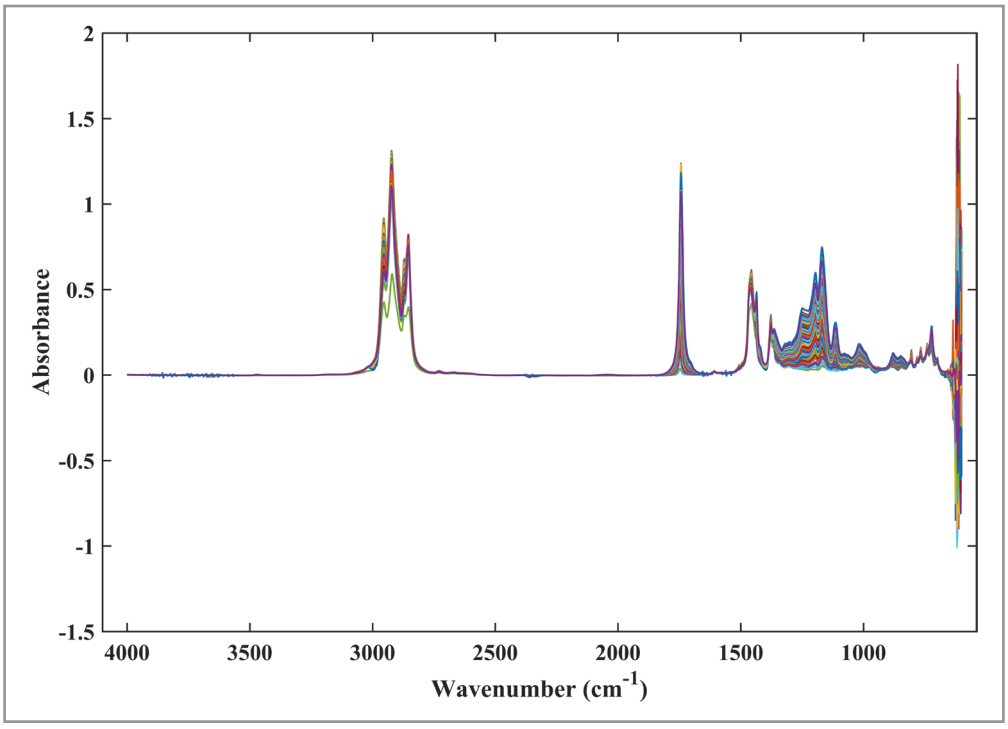


Figure 1. Unpreprocessed MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v).

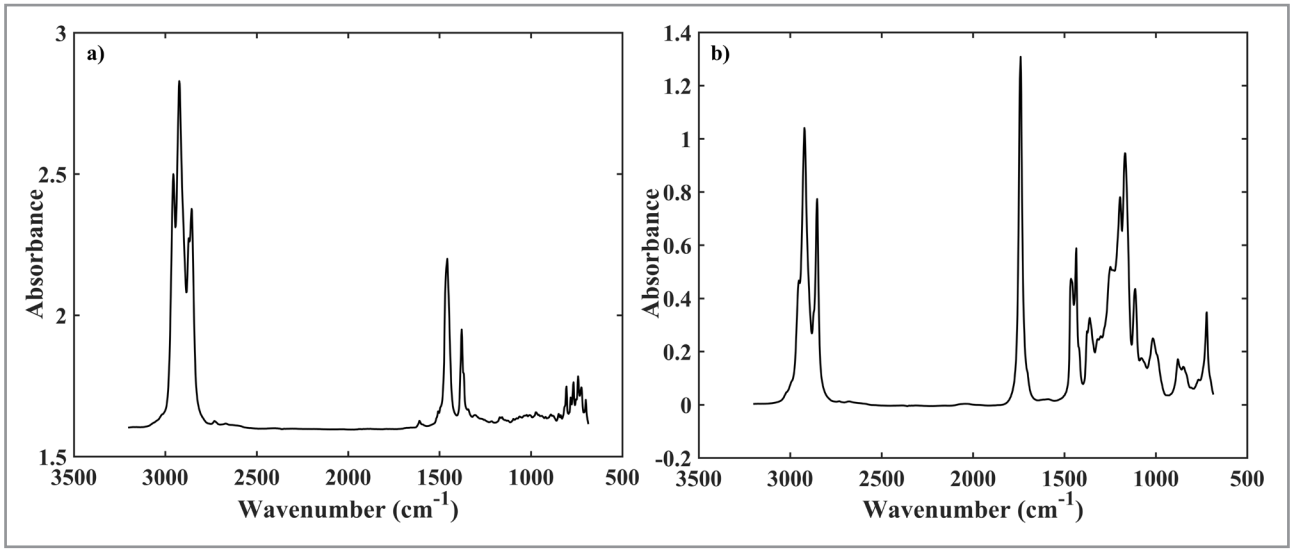


Figure 2. MIR spectra of (a) kerosene (b) palm kernel biokerosene.

However, the spectrum of kerosene differs from the spectrum of bio-kerosene, mainly in the characteristic bands of bio-kerosene, found in the 1744 cm^{-1} spectral regions, attributed to vC=O of acylglycerols; from 1200 to 1119 cm^{-1}, attributed to the stretching vibrations of the C−O group bond in esters of type O=C−(OR) [18,19]. That is why, when mixed (Figure 3), these characteristic bands are evident. However, it is difficult to attribute each spectral profile to a specific concentration based on visual analysis. To this end, the use of chemometric tools is essential.

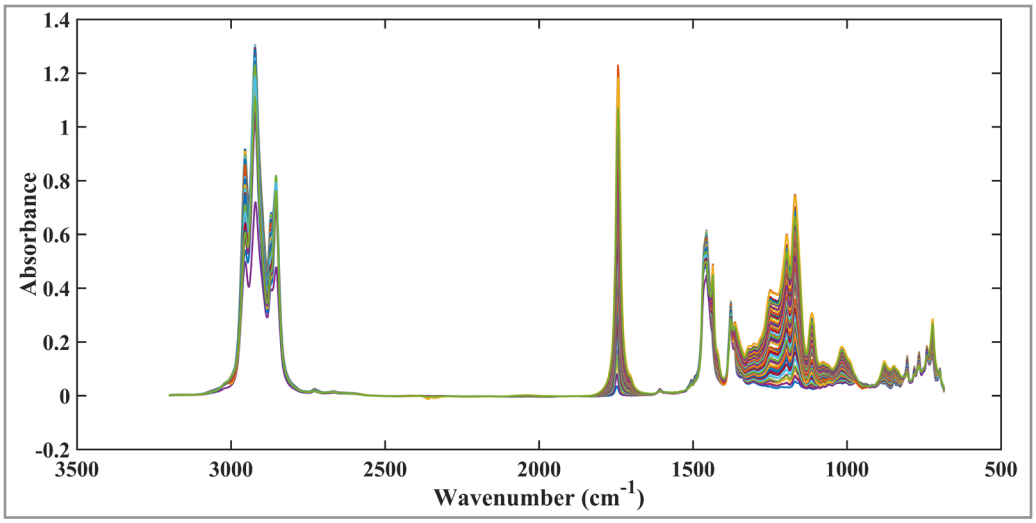


Figure 3. Pre-processed MIR spectra of kerosene/palm kernel biokerosene mixtures in the concentration range from 1.00 to 70.00% (v/v).

PLS Model

Figure 4 shows: a) the root-mean-square error of cross-validation (RMSECV) and root-mean-square error of Calibration (RMSEC); b) the accumulated variance captured in blocks **X** and **Y** obtained as a function of the amount of the Latent Variable (LV) chosen for construction of the PLS model. In this figure, it can be seen that the 5 LVs chosen to build the PLS model provide less errors and greater captured variance, that is, they represent 99.99% and 99.98% of the variance explained in blocks **X** and **Y**, respectively.

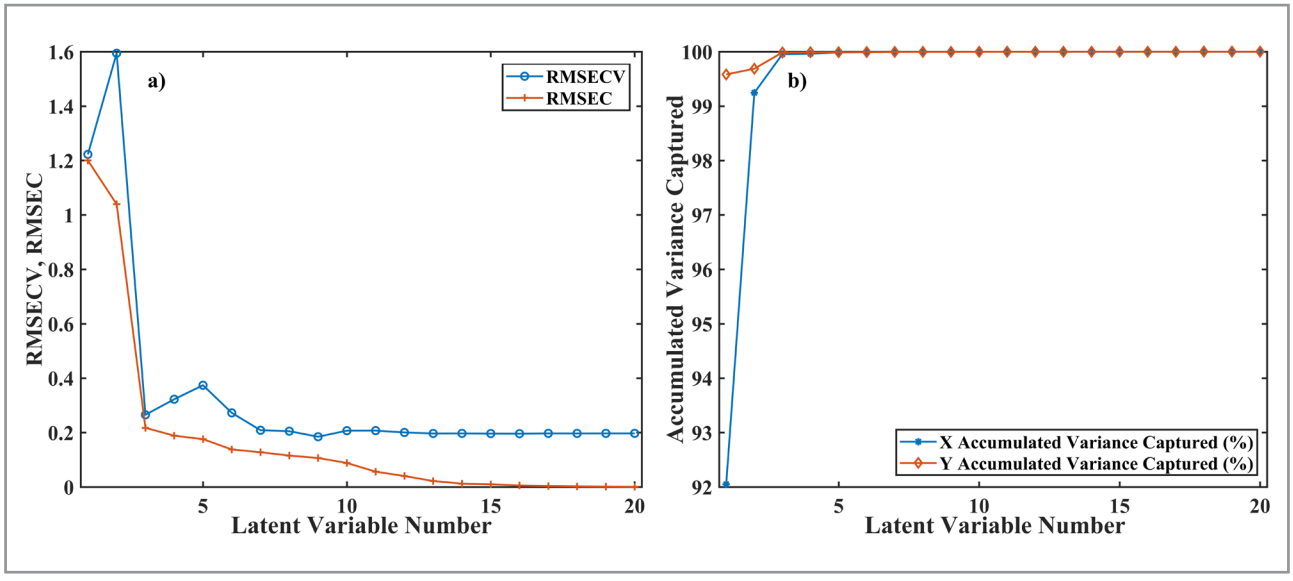


Figure 4. (a) Root-mean-square error of cross-validation (RMSECV) and root-mean-square error of Calibration (RMSEC), (b) Accumulated variance captured in blocks **X** and **Y**.

Table II presents results of the parameters and figures of merit, calculated for the PLS model, where it is observed that the values of root-mean-square error of cross-validation (RMSECV), root-mean-square error of calibration (RMSEC) and root-mean-square error of prediction (RMSEP) were in agreement because the RMSEP was higher than RMSECV and RMSEC. In addition, the RMSEP value is considered acceptable

according to the Brazilian standard ABNT NBR 15568 because it is less than 1% [20] and are close to the values obtained in some quantification models published in the literature [21,22]. It is also observed that a good agreement between the parameters is evident because RMSEC was lower than RMSEP. The calculated t_{bias} value was lower than $t_{critical}$, which means that there are no systematic errors in the model.

The value of the LQ obtained in the model is higher than the LD, demonstrating a good agreement between the parameters because the lowest concentration of the substance of interest that is measured with a maximum uncertainty of 10% is within what can be detected. The sensitivity expresses an increase in the signal fraction when the concentration of the analyte of interest has a high value for one unit, however, its value was estimated at 0.871. The analytical sensitivity of the model is relatively high, showing that the influence of residues in the prediction of unknown samples is low. The inverse of the analytical sensitivity value shows that the PLS model is able to distinguish differences between samples with a variation of 0.052. The selectivity was good, which means that the PLS model did not present significant overlap of the interference signal with the analyte.

Table II. Results of the accuracy parameters and figures of merit

Figure of Merit		Value
Accuracy	RMSEC (% v/v)	0.175
	RMSECV (% v/v)	0.373
	RMSEP (% v/v)	0.708
Analytical sensitivity/ % (v/v)		19.20
Inverse of analytical sensitivity /$(v/v)^{-1}$		0.052
Sensitivity/ % $(v/v)^{-1}$		0.871
Selectivity		0.375
Limit of detection / % (v/v)		0.171
Limit of quantification / % (v/v)		0.521
bias		0.118
SDV		0.364
t_{bias}		1.782
$t_{crítico}$		2.051

The evaluation of the fit of the PLS model by correlating the measured and predicted values of the calibration and prediction sets is shown in Figure 5. In this figure, it can be seen that the PLS model showed a good fit because a high correlation between the two values was evident, with the correlation coefficient (R) exceeding 0.99. Figure 6, illustrates the plot of the real concentration versus absolute errors, showing uniform distribution in a horizontal range. Ideally, the residuals should show random behavior and constant variance in the concentration range, as seen in the calibration set samples. Although the residuals from the prediction set show a trend, the predicted concentrations are not severely affected.

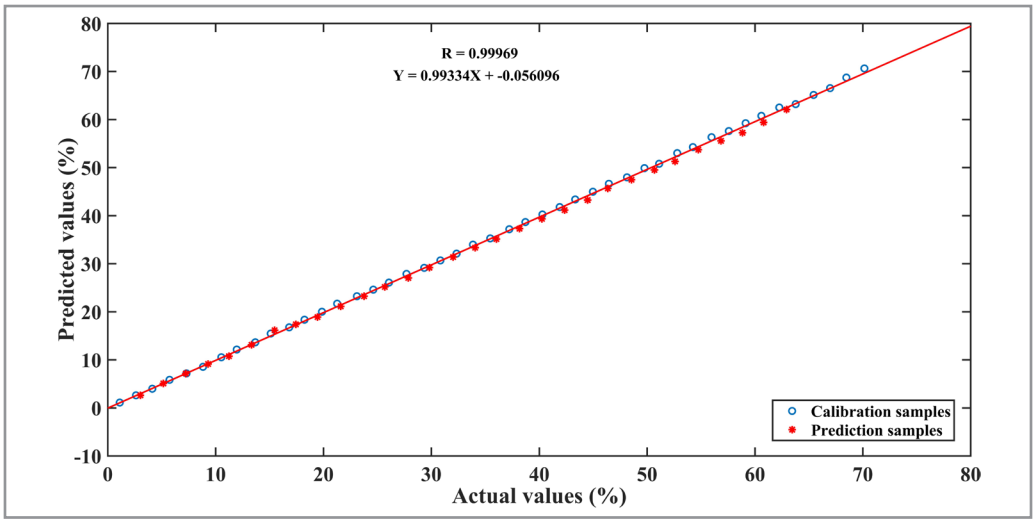


Figure 5. Fit of the PLS model through the actual values (experimental concentrations) versus predicted values of the calibration and forecast sets.

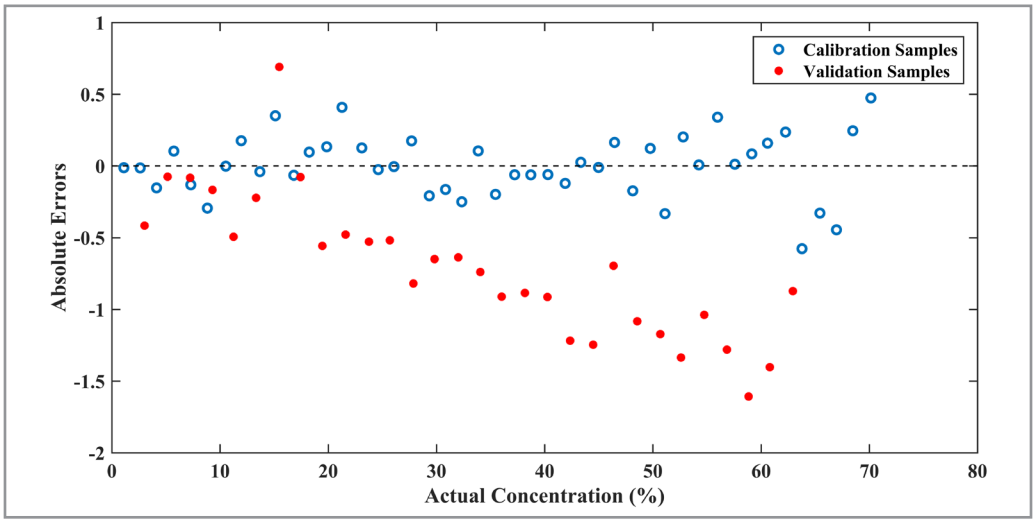


Figure 6. Residues of the PLS model for the calibration and prediction set

Figure 7 shows the weights of the latent variables of the PLS model, where it can be seen that the LV1 describes the variables of the characteristic bands of the biokerosene, with 92.05% of the explained variance. The remaining latent variables describe the variables of the common bands, with 7.94% of the explained variance.

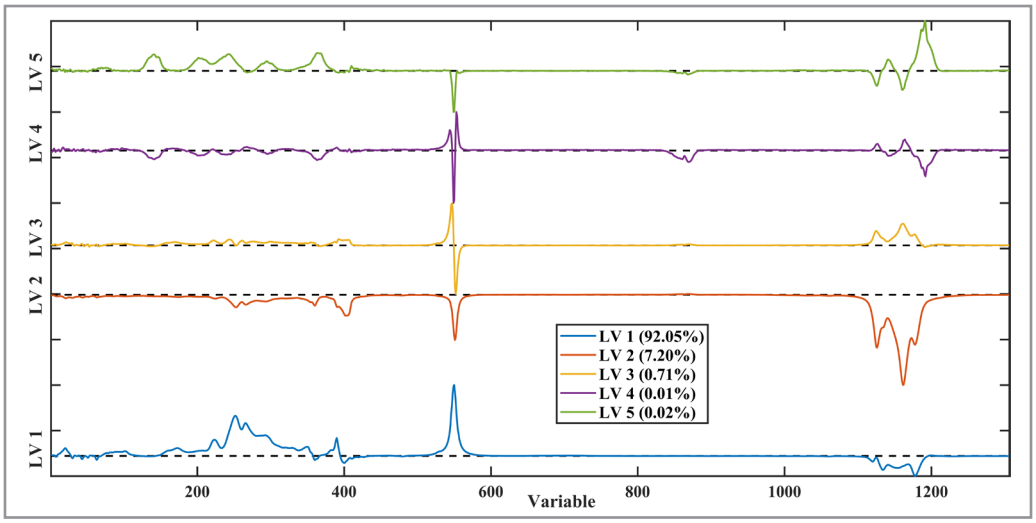


Figure 7. Plot of the loadings of LV1, LV2, LV3, LV4 and LV5 versus variables for the application of the PLS model.

CONCLUSIONS

The application of multivariate regression by PLS in the MIR spectra of kerosene/palm kernel biokerosene mixtures allowed the development of a methodology to quantify the content of this biokerosene. The efficiency of this methodology was analyzed based on the figures of merit and the fit of the model. The results of the figures of merit were in agreement with the requirements established in the standard ASTM E1655-05. The model fit showed a high correlation between actual and predicted concentration values of the calibration and prediction sets, with a correlation coefficient exceeding 0.99 and with relatively low errors.

These results demonstrate that the developed methodology has the potential to be explored by regulatory agencies to monitor the content of biokerosene in a mixture with kerosene, because it is relatively low cost and allows for quick, direct and in situ analyzes.

Conflicts of interest

Regarding conflicts of interest and on behalf of all authors, I declare that for this manuscript there are no conflicts of interest.

Acknowledgements

The authors acknowledge the partnership program PEC-PG (Process 190279/2014-1) for financial support; Transpetro, S/A. for the supply of pure kerosene and Biofuels Laboratory of the Federal University of Uberlândia for the supply of palm kernel biokerosene.

REFERENCES

1. Grampella, M.; Martini, G.; Scotti, D.; Zambon, G. *Transp. Res. Part A Policy Pract.*, **2016**, *92*, pp 310–325 (https://doi.org/10.1016/j.tra.2016.06.013).
2. Hofer, C.; Dresner, M. E.; Windle, R. J. *Transp. Res. Part D Transp. Environ.*, **2010**, *15* (1), pp 37–45 (https://doi.org/10.1016/j.trd.2009.07.001).
3. Geller, H. S. *Revolução Energética: Políticas Para Um Futuro Sustentável*; Relume Dumará, Rio de Janeiro, **2003**.

4. Biocombustíveis Aeronáuticos: Progressos e Desafios. Série Documentos Técnicos do Centro de Gestão e Estudos Estratégicos (CGEE). **2010**, Nº 08. Available at: https://www.cgee.org.br/documents/10195/734063/biocombustiveis_aeronauticos_24012011_9559.pdf/378f8f90-fa5c-4e0c-aad7-7adcbf607063?version=1.5 [Accessed January 11, 2021].
5. Llamas, A.; Al-Lal, A. M.; Hernandez, M.; Lapuerta, M.; Canoira, L. *Energy and Fuels*, **2012**, *26* (9), pp 5968–5976 (https://doi.org/10.1021/ef300927q).
6. Reynol, F. *Inovação Uniemp*, **2007**, *3* (3). Available at: http://inovacao.scielo.br/scielo.php?script=sci_arttext&pid=S1808-23942007000300011&lng=en&nrm=iso&tlng=pt [Accessed January 11, 2021].
7. Bonassa, G.; Schneider, L. T.; Damaris, K.; Frigo, D. A.; Severo, F.; Lins, M. A.; Frigo, E. P. *Rev. Bras. Energias Renov*., **2014**, *3*, pp 97–106.
8. Baroutian, S.; Aroua, M. K.; Raman, A. A. A.; Shafie, A.; Ismail, R. A.; Hamdan, H. *J. Taiwan Inst. Chem. Eng*., **2013**, *44* (6), pp 911–916 (https://doi.org/10.1016/j.jtice.2013.02.007).
9. Máquina, A. D. V.; Sitoe, B. V.; Buiatte, J. E.; Santos, D. Q.; Neto, W. B. *Fuel*, **2019**, *237*, pp 373–379 (https://doi.org/10.1016/j.fuel.2018.10.011).
10. Máquina, A. D. V.; Souza, L. M.; Buiatte, J. E.; Santos, D. Q.; Neto, W. B. *Energy and Fuels*, **2017**, *31* (1), pp 571–577 (https://doi.org/10.1021/acs.energyfuels.6b02079).
11. Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; Jong, S. De; Lewi, P. J.; Smeyers, J. *Handbook of Chemometrics and Qualimetrics, Volume 20B*. Elsevier Science, Amsterdam, **1998**.
12. Ferreira, M. M. C. *Quimiometria: Conceitos, Métodos e Aplicações*;.Editora da Unicamp, Campinas, SP, **2015** (https://doi.org/10.7476/9788526814714).
13. American Society for Testing and Materials (ASTM). Standard E1655-05. *Standard Practices for Infrared Multivariate Quantitative Analysis*. United States, **2012**, pp 1–29 (https://doi.org/10.1520/E1655-05R12).
14. Ferreira, M. H.; Braga, J. W. B.; Sena, M. M. *Microchem. J*., **2013**, *109*, pp 158–164 (https://doi.org/10.1016/j.microc.2012.03.008).
15. Valderrama, P.; Braga, J. W. B.; Poppi, R. J. *Quim. Nova*, **2009**, *32* (5), pp 1278–1287 (https://doi.org/10.1590/S0100-40422009000500034).
16. Silva, M. A. M.; Ferreira, M. H.; Braga, J. W. B.; Sena, M. M. *Talanta*, **2012**, *89*, pp 342–351 (https://doi.org/10.1016/j.talanta.2011.12.039).
17. De Carvalho Rocha, W. F.; Nogueira, R.; Vaz, B. G. V. *J. Chemom*., **2012**, *26* (8–9), pp 456–461 (https://doi.org/10.1002/cem.2420).
18. Barbosa, L. C. A. *Espectroscopia no Infravermelho*, 1th Ed. Editora da Universidade Federal de Viçosa (UFV), Viçosa, MG, Brazil, **2007**.
19. Silverstein, R. M.; Webster, F. X.; Kiemle, D. J.; Bryce, D. L. *Spectrometric Identification of Organic Compounds*, 8th Ed. John Wiley & Sons, Nova Jersey, **2014**.
20. Associação Brasileira de Normas Técnicas. ABNT NBR-15568. *Biodiesel – Determinação do Teor de Biodiesel em Óleo Diesel por Espectroscopia na região do Infravermelho Médio*. Brazil, **2008**.
21. Shi, T.; Zhu, M. T.; Zhou, X. Y.; Huo, X.; Long, Y.; Zeng, X. Z.; Chen, Y. *Food Chem*., **2019**, *287*, pp 46–54 (https://doi.org/10.1016/j.foodchem.2019.02.072).
22. Singh, K.; Kumar, S. P.; Blümich, B. *Fuel*, **2019**, *243*, pp 192–201 (https://doi.org/10.1016/j.fuel.2019.01.084).

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 165–177
doi: 10.30744/brjac.2179-3425.AR-27-2021

BrJAC

ARTICLE

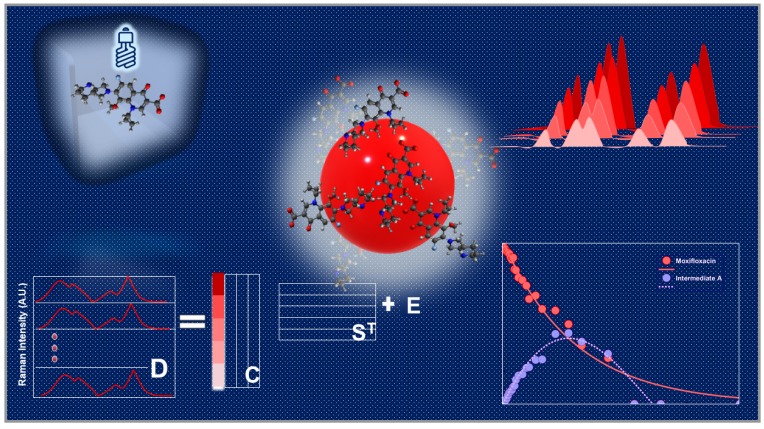
Soft Modelling of the Photolytic Degradation of Moxifloxacin Combining Surface Enhanced Raman Spectroscopy and Multivariate Curve Resolution

Mónica Benicia Mamián-López[1]*, Ronei J. Poppi[2]†

[1]*Universidade Federal do ABC, Av. dos Estados, 5001, Prédio B, Sala 1052. 09210-580, Santo André, SP, Brazil*

[2]*Universidade Estadual de Campinas, Rua Josué de Castro, s/n, Cidade Universitária, 13083-970, Campinas, SP, Brazil*

† In memoriam (1961 – 2020)



The antibiotic moxifloxacin had a recent surge in its use due to its broad spectrum of activity. However, due to the low metabolization inside the organism, it became an environmental concern. Here, the photolytic degradation of moxifloxacin antibiotic in alkaline medium was carried out and monitored through Surface enhanced Raman spectroscopy (SERS). Multivariate curve resolution method was applied to extract quantitative and kinetic information about the whole process, using correlation constraint to simultaneously quantify the variation of moxifloxacin concentration. The results showed that the photolysis follows an apparent first order kinetics with half-life of 47.5 min. Also, SERS spectrum along with the calculated Raman spectra suggest that cleavage of the diazabicyclonyl substituent is the preferred photodegradation pathway, in agreement with previous reports.

Keywords: Chemometrics, Multivariate Curve Resolution, SERS spectroscopy, Gold Nanoparticles

INTRODUCTION

Moxifloxacin is a fourth-generation synthetic antibiotic belonging to the fluoroquinolone family. Its use has expanded recently due to its broad spectrum of activity, mainly towards gram-positive bacteria compared to other synthetic quinolones. It is used principally to treat pneumonia, bacterial sinusitis, and bronchitis. Furthermore, it can be administrated via intravenous, oral, and topical [1]. One of its characteristics is that it is not fully metabolized when used in medical treatments, being more of 50% excreted in its pharmaceutically active form. Due to its limited biodegradability and its increasing use, the research for efficient degradation

Cite: Mamián-López, M. B.; Poppi, R. J. Soft Modelling of the Photolytic Degradation of Moxifloxacin Combining Surface enhanced Raman Spectroscopy and Multivariate Curve Resolution. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 165–177. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-27-2021

Submitted 01 March 2021, Resubmitted 27 April 2021, Accepted 21 May 2021, Available online 02 June 2021.

methodologies of this kind of compounds has become a chief concern. Hence, studying its degradation pathways, products, and intermediate species is critical for searching for new methods.

Advanced oxidation processes (AOPs) include many techniques such as sonication, photocatalysis, ozonation, photolysis, among others, which are perhaps the most popular methods applied for environmental remediation [2-4]. Degradation of moxifloxacin has been studied [5] through photocatalytic and photolytic processes in aqueous medium [6] and in organic solvent [7]. In both cases the degradation was monitored via chromatographic techniques and mass spectrometry.

The introduction of spectroscopic techniques to these studies provides a very complete structural landscape of the process, associated with high analytical sensitivity, and more straightforward experimental procedures, not requiring sample treatment and time-consuming stages before instrumental analysis. One of the more suitable techniques to reach these conditions is surface-enhanced Raman spectroscopy (SERS), which drastically intensifies the Raman scattering signal, providing vibrational fingerprint information. Its performance in aqueous (even biological) medium is optimal and requires, in most of cases, minimal experimental procedures before spectroscopic analysis. Although reproducibility has been a major issue in its application for quantitative purposes and researchers have been reluctant for several years about applying it in analytical studies, the scientific literature has already shown that quantifications are successful, and today, SERS spectroscopy is considered an analytical technique itself.

In a standard SERS analysis, the analyte of interest should be adsorbed onto a noble metal nanostructured surface, generally synthesized following a controlled reduction reaction of a salt of gold, silver, or copper. This way, a colloidal suspension containing nanoparticles with a given morphology is obtained. Then the next step includes the addition of an aliquot of the sample before to the spectral acquisition. Several variations of this procedure are widely available and allow analysis accordingly to the sample-surface system's chemical nature.

SERS spectra are typically information-rich eventually, even with overlapped signals. Since high volume of spectral data can be obtained quickly, computational methods for processing this data are mandatory. Still, nowadays, data processing and information extraction have an increasing role in research design and execution [8]. Usually, a first approach to the data sets is through exploratory methods such as Principal Component Analysis (PCA), aimed to reduce dimensionality of a data matrix by calculating new matrices of scores and loadings, projected on a new set of coordinates or principal components (PCs). Thus, trends or patterns can be easily visualized and then associated with their corresponding variables, responsible for a given behavior.

Decomposition methods are handy when the problem under study represents a mixture of components or changing processes [9]. The Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) method has become an exciting option when studying this kind of data, and the number of applications to truly diverse mixture problems has increased strikingly in recent years. Here, the original dataset, **D**, is decomposed into a bilinear equation containing the profiles of the "purest" components in a matrix $\mathbf{S}^T$ and its respective relative concentrations in a matrix **C**, following Equation 1. The **E** object is a residual matrix.

$$\boldsymbol{D} = \boldsymbol{C}\boldsymbol{S}^{\boldsymbol{T}} + \boldsymbol{E} \qquad (1)$$

In MCR-ALS, $\mathbf{S}^T$ and **C** matrix are calculated iteratively using an initial estimation of pure signals (either calculated or acquired experimentally). Constraints related to the system being studied, such as non-negativity of concentrations or signals, closure, or correlation constraints, can be applied, to provide chemical meaning to the calculated profiles and to minimize rotational ambiguity in the **C** and $\mathbf{S}^T$ calculation. This chemometric tool has been successfully used for modelling kinetic processes [10,11], and calculating kinetic parameters. In this work, we have combined the high amount of structural information stored in SERS spectra and its ability to reach extremely low limit of detection values (values in the femtomolar regime has been reported [12]), with powerful chemometric tools to model and study the possible intermediates compounds in the photolytic degradation of moxifloxacin. In addition, the calculated Raman spectrum of

moxifloxacin along with its observed SERS spectra allowed us to perform a band assignment which, to the authors knowledge, is being reported for the first time in this work.

MATERIALS AND METHODS

Reagents and Solutions

Chloroauric acid ($HAuCl_4$) 30% wt was purchased from Sigma-Aldrich, and Moxifloxacin hydrochloride was acquired from EMS-Brazil. Other reagents used, such as sodium citrate, sodium phosphate, sodium acetate, and sodium chloride were purchased from Synth. All solutions were prepared in Milli-Q water.

Synthesis of the SERS substrate

Colloidal Gold nanoparticles in suspension (AuNPs) was used as enhancing surface. The synthesis procedure was a modification of the Lee-Meisel method [13,14], where a $HAuCl_4$ solution was initially heated with constant stirring until boiling. Then, a preheated solution of sodium citrate was added to initiate the gold reduction to produce semi-spheric AuNPs (≈ 60 nm).

Photolysis

Stock solutions of moxifloxacin (200.00 μg mL^{-1}) were prepared in pH values below, higher and between the two pKa values associated with its protonated and deprotonated forms (pKa = 5.7 – 9.2) to evaluate optimal condition favoring the degradation process. Thus, photolytic reactions were run at pH= 4.5 (acetate buffer), pH = 6.5 (not buffered) e pH 10.20 (phosphate buffer). Four different initial concentrations were also tested (20.00, 15.00, 10.00, and 7.00 μg mL^{-1}).

The photolysis experiments were carried out in a closed chamber, temperature-controlled bath, as, where 200 mL of moxifloxacin solution (at the given pH) was put in a Petri dish (of diameter equal to 12 cm) during a total time of 2 h, taking aliquots at reaction times of 0.60 and 120 minutes. A UV-C (16 W, 100–280 nm with main peak at 254 nm) was used as a light source.

The optimal condition for the photolysis reaction (pH=10.20 and 10 μg mL^{-1} of initial concentration) was monitored for 8 hours, taking 26 aliquots of 400 μL each, successively extracted according to Table I, and stored away from light. The process was run in triplicate and all samples were analyzed on the same day. Additionally, a control experiment was run simultaneously without exposition to UV-C light, and SERS spectra were registered before and at the end of the experiments. AuNPs were not part of the photolytic degradation, instead, they were used as enhancing surface for acquiring SERS spectra.

Table I. Aliquots acquisition time from photolytic process

Aliquot (number)	1	2	3	4	5	6	7	8	9	10	11	12	13	…
Time (min)	0	1	2	3	4	5	6	7	8	9	10	12	15	…

Aliquot (number)	…	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Time (min)	…	18	20	25	30	40	50	60	80	100	120	180	240	360	480

Standard solutions for the analytical curve

Seven moxifloxacin standard solutions were prepared in triplicate between 0 – 12 μg mL^{-1} (in pH 10.20), and analytical curves were built from its SERS spectra.

Spectroscopic analysis

Volumes of 200 µL of each aliquot were added to 1.00 mL of AuNPs in a quartz cuvette and spectra were collected with six exposures of 8 s each and 250 mW of laser power (100%). The spectral region used was between 200 and 2000 cm^{-1}, with 4.0 cm^{-1} of spectral resolution.

The spectrometer was a Perkin Elmer Raman Station 400F, equipped with a 785 nm laser source, an Echelle grating, and a temperature-controlled (-50 ºC) charged coupled device detector. The same conditions were used for all SERS spectra shown in this work.

Band assignment and Raman spectrum calculations

SERS bands attribution was based mainly on quantum chemistry calculations of the Raman spectrum. Furthermore, we compared our SERS spectra to those previously reported of similar fluoroquinolones (ofloxacin, enrofloxacin and ciprofloxacin) [15,16] from the literature.

Quantum chemistry calculations were done using the Orca 4.2.1 package for a single moxifloxacin molecule using a conductor-like polarizable continuum model to mimic water solvation. The calculations were done using the M062X functional with the D3 Grimme dispersion correction and basis set def2-TZVP(-f), following previous results from the literature for other fluoroquinolones [17–20].

Data processing and multivariate analysis

Data processing was performed in Matlab 7.8. Pre-processing was applied in PLS Toolbox 8.0 from *Eigenvector*. MCR-ALS analysis were run on MCR Toolbox [21] available at http://www.mcrals.info/.

Experimental data were arranged in matrix form with samples (SERS spectra from photolysis process) in the rows and variables (wavenumber, cm^{-1}) in the columns. For the PCA analysis, spectra were normalized, the baseline was corrected with WLS (Automatic Weighted Least Squares) filter and mean centered. For MCR-ALS, the data was neither normalized nor mean centered. For quantification, the data was arranged as augmented matrix, correlation constraint was applied and a closure constraint for concentrations was also included (see *Multivariate Curve Resolution: modelling and quantification* section below).

Quantification and assessment of Figures of Merit

A pseudo-univariate model was built by plotting the real concentration values from each standard against relative concentrations recovered from MCR-ALS calculation (values in **C** matrix). Accordingly, figures of merit were estimated as proposed in literature [22,23]. The Limit of Detection (LOD) was calculated using the Equation 2:

$$LOD = +t\frac{S_R}{b}\sqrt{\frac{1}{m}+\frac{1}{n}+\frac{(n-\bar{y})^2}{b^2(c_i-\bar{c})^2}} \qquad (2)$$

where, b is the slope of the curve, m and n are the number of replicates and standards, respectively.

The concentration for the i_{th} component and the mean concentration standard are c_i and $\bar{c}$, respectively, and $\bar{y}$ is the mean predicted response (estimated by the model). The value *t* corresponds to the *Student* value for n-2 degrees of freedom. The precision, S_R was calculated through the Equation 3:

$$S_R = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}-y_i)^2}{n-1}} \qquad (3)$$

where $\hat{y}$ and y_i, are the predicted and experimental response for the standard *i*, respectively.

The Limit of Quantification (LOQ) value was calculated following the Equation 4:

$$LOQ = 3.3\ LOD \qquad (4)$$

Finally, a test of significance for lack-of-fit (F_{LOF}) of the model was run using the unexplained variance and pure error. The F_{LOF} calculated was compared to F critical (F_c) to estimate the lack of fit to the linear behavior.

RESULTS AND DISCUSSION

Spectroscopic profile and band assignments

Structurally, the moxifloxacin ($C_{21}H_{24}FN_3O_4$) has a quinolone skeleton (drawn in Figure 1 with its atoms in blue), a fluorine atom at position 6, a methoxy group at position 8 and a diazabicyclonyl group at position 7, as shown in Figure 1 along with its SERS spectrum.

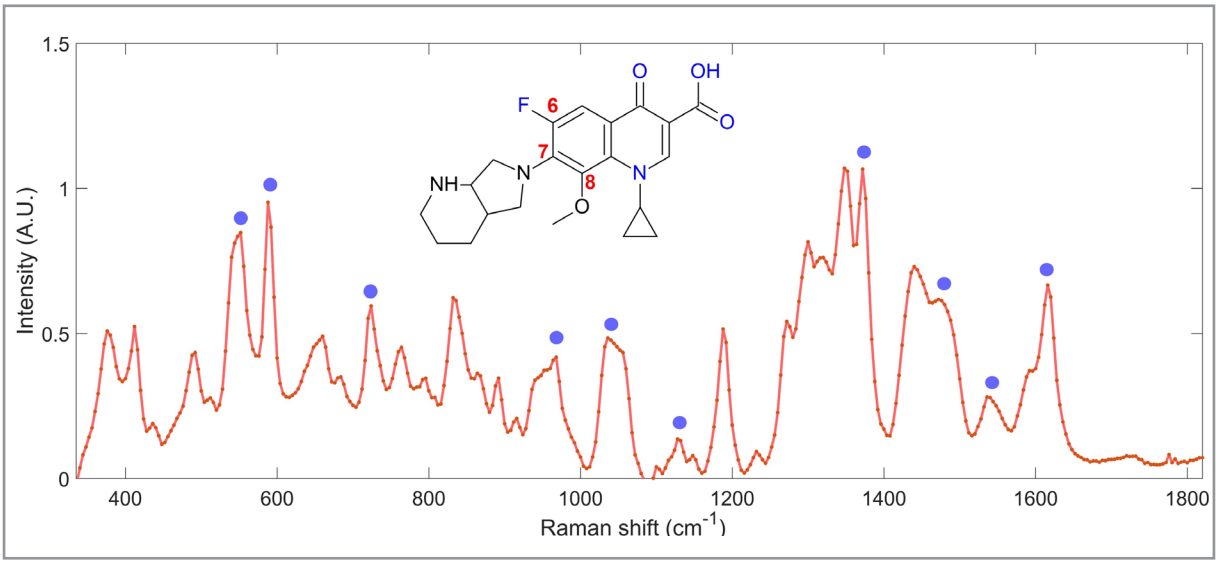


Figure 1. SERS spectrum of moxifloxacin (10 µg mL^{-1}) in aqueous solution, on AuNPs and main band signals. Inset plot: molecular structure of moxifloxacin.

Band assignments

The main SERS features shows that diazabicyclonyl substituent has an important contribution to the whole spectra. Carbonyl and carboxyl modes are located at 724 and 1536 cm^{-1}, the -C-O bond of methoxy group contributes at 1100 cm^{-1} and the -C-F bond has a small contribution at 1128 cm^{-1}. A more complete description of other vibrational modes is presented in Table II.

Table II. SERS bands observed and calculated (Raman) for deprotonated moxifloxacin. Contributions of diazabicyclonyl group are marked with an asterisk (*).

SERS (cm^{-1})	Calculated (cm^{-1})	Vibrational modes description
552	551.84	-CH_2 twisting in diazabicyclonyl ring Cyclopropyl ring deformation
588	596	-CH_2 rocking in diazabicyclonyl ring* Quinolone ring deformation

Table II. SERS bands observed and calculated (Raman) for deprotonated moxifloxacin. Contributions of diazabicyclonyl group are marked with an asterisk (*). (Continuation)

SERS (cm^{-1})	Calculated (cm^{-1})	Vibrational modes description
724	718.76	-C=O and -O-H bending -CH_2 rocking of diazabicyclonyl ring and cyclopropyl Quinolone ring deformation
944	945.96	-CH_2 stretch and twisting of diazabicyclonyl ring*
968	973.30	-CH_2 rocking of diazabicyclonyl ring*
1100	1100	-CH_2 wagging diazabicyclonyl ring and cyclopropyl Stretch -C-O of methoxy group
1128	1125.43	-CH_2 rocking and wagging diazabicyclonyl ring* Quinolone ring deformation -C-F stretching
1232	1236.52	-CH_2 wagging of diazabicyclonyl ring * and cyclopropyl Quinolone ring deformation
1372		-CH_2 wagging of diazabicyclonyl ring * Quinolone ring deformation
1436	1431	-CH_2 bending Quinolone ring deformation -CH_2 wagging of diazabicyclonyl ring*
1472	1471.77	-CH_2 bending of diazabicyclonyl ring*
1536	1529.4	Carbonyl and carboxyl stretch In plane ring deformation

Optimization of photolysis conditions

Since pH plays an essential role in photolytic and photocatalytic reactions of fluoroquinolones [7] preliminary experiments at three different pH values were run (data not shown), to set the best condition to execute a rapid degradation, suitable for practical applications. Moxifloxacin showed to be highly stable at natural pH (less than 10% of the initial concentration was degraded in 2 h of process). About 25% of the initial moxifloxacin concentration reacted at pH 4.5, with changes in the SERS spectrum obtained after 2 h of irradiation being evidenced by the decrease in correlation with the initial spectrum. Finally, alkaline condition (pH=10.20) showed to favor the degradation of ≈ 50% of the compound after 60 min of reaction, and this condition was chosen as optimal to study the kinetic process.

Exploratory analysis

SERS responses from the entire photolysis process were first explored using PCA analysis. The results allowed a more detailed visualization of how the compound(s) in the solution were changing with time as observed in Figure 2. Three main groups were observed along the reaction, with variables 756, 852, 828 and 1612 on PC1 dominating the first 80 min. On PC2, the first 8 min are dominated by the variables at 904 and 1128 cm^{-1} (Loadings are included in Figure1SI). After 100 min of reaction, the spectra's intensity strongly decreases, only showing some SERS features of very low intensity (Figure 2SI).

The patterns observed in the scores plot show that SERS spectroscopy is a technique capable of observing and monitoring the entire chemical transformation (the photolysis degradation). It can give us insights into the initial reactant's structural changes, allowing us to infer possible intermediates. Furthermore, as we will argue ahead, it also allows assessing variations of their concentration.

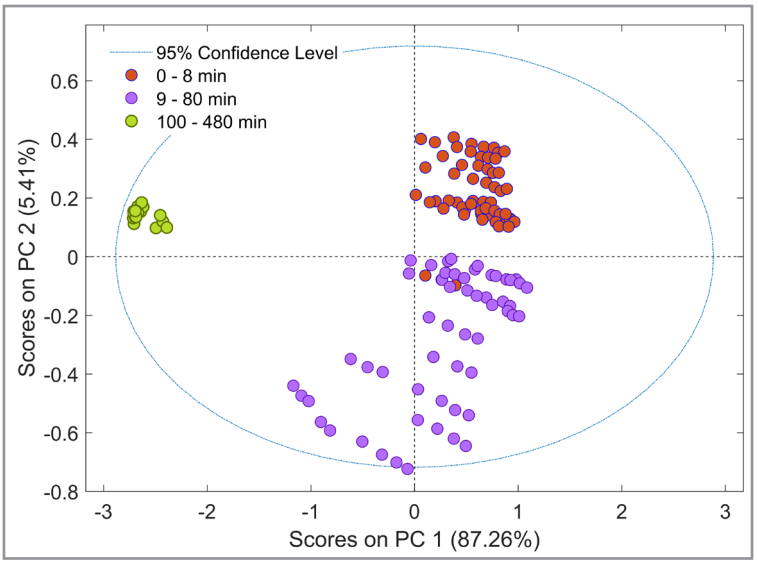


Figure 2. PCA scores plot for the SERS spectra of photolytic degradation of moxifloxacin in alkaline medium. Data were baseline-corrected and mean centered.

Multivariate Curve Resolution: modelling and quantification

As the objective of using MCR-ALS allied to SERS was both modelling the kinetic process and quantify the moxifloxacin degradation, the correlation constraint was chosen. This strategy has been successfully used [24] to address complex systems with overlapped signals and allows to reduce the range of feasible solutions.

To accomplish this, the SERS spectra acquired during the photolysis ($\mathbf{D}_t$) were arranged as a row-wise augmented matrix with the SERS spectra acquired for the analytical curve ($\mathbf{D}_s$). A constraint for concentrations was also introduced to run the MCR calculations, as depicted in Figure 3. Briefly, a $\mathbf{C}_{sel}$ matrix containing the concentration values of the moxifloxacin standard solutions (first column of $\mathbf{C}_{sel}$) was included to force the final values in **C** (concentration profiles) matrix to be equal to the real values. The two remaining columns in $\mathbf{C}_{sel}$ are filled with zeros, since for their corresponding spectra (**Ds**), there is a unique contributor (moxifloxacin). The second calculation was performed by running the MCR-ALS method for the $\mathbf{D}_t$ and $\mathbf{D}_s$ matrix, separately. An analytical curve was tested but the results were not satisfactory. The arrangement for MCR-ALS calculation with constraint is shown in Figure 3.

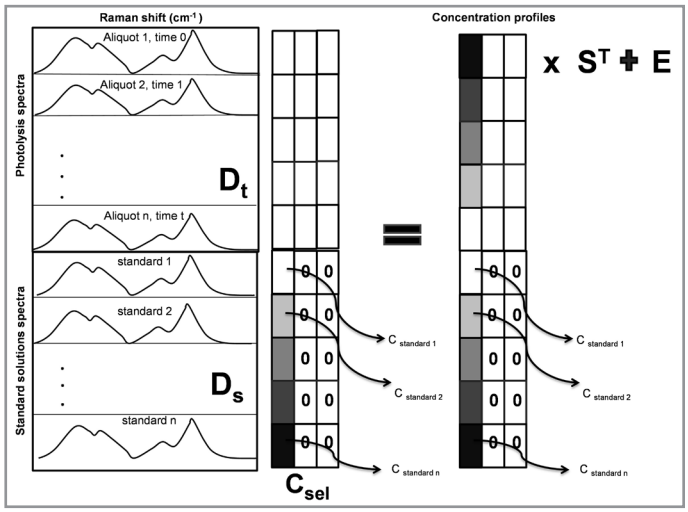


Figure 3. Data arrangement for MCR-ALS calculation with concentration constraint, including a $\mathbf{C}_{sel}$ matrix.

The number of purest components (as calculated by the Singular Value Decomposition algorithm) was three, and, since where it is not possible to have the SERS responses of all species for the studied system an initial estimate was calculated by SIMPLISMA algorithm [25].

The relative concentrations in the **C** matrix, along with the concentrations from standard solutions, were used to build an analytical curve (Figure 3SI), for the quantitative assessment of changes in the moxifloxacin concentration. Plots for determining kinetic reaction order showed that the photolysis in alkaline medium follows an apparent first-order degradation model, as shown in Figure 4. Table II shows the reaction rate (for both approaches) estimated from the regression and the half-life calculation, following the Equation 5:

$$t_{1/2} = \ln(2/k) \qquad (5)$$

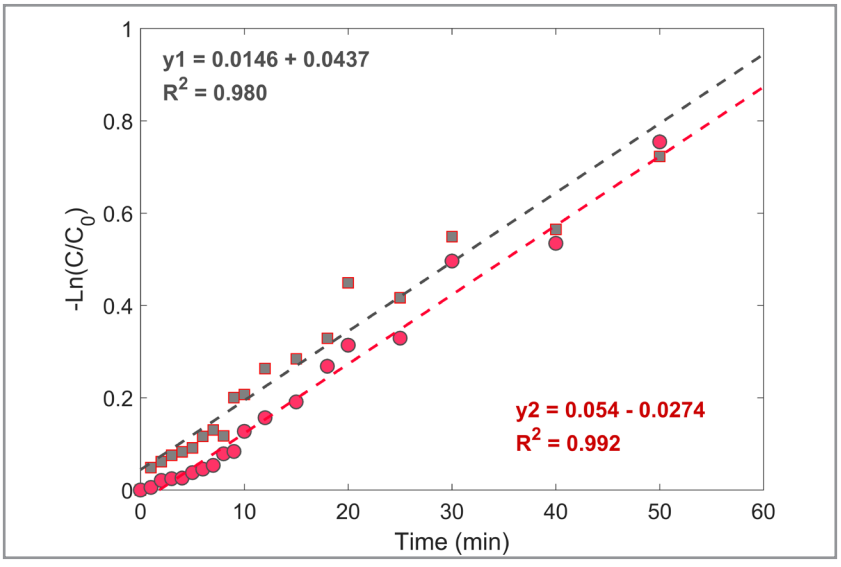


Figure 4. -Ln(C/C_0) *vs*. time (min) for determining kinetic parameters, *k* and $t_{1/2}$. Gray and red markers correspond to MCR-ALS without and with correlation constraint, respectively.

Table II. Reaction rate (*k*) and half-life time ($t_{1/2}$) for the studied compound with and without the augmented matrix correlation constraint

MCR approach	*k* (min^{-1})	$t_{1/2}$ (min)
Augmented matrix and correlation constraint	0.0146	47.5
Without correlation constraint	0.0154	46.2

These results follow the same trend as in previously reported studies for this reaction in similar conditions [6,7] where both strong pH dependence and apparent first-order kinetics were found. Relative concentrations in the **C** matrix also allowed to have an insight into the behavior of possible intermediates through the entire process. Plotting these values from MCR-ALS against time reaction provided a better landscape of the entire process, as shown in Figure 5. Here, the moxifloxacin loses about half its initial

concentration at *ca.* 50 min while an intermediate A is being formed. A third possible compound, or intermediate B (showed in Figure 2SI) could be detected after 30 min of reaction, reaching a maximum value at around two hours, and becoming undetectable after this time of photolysis. After 100 min of reaction, SERS features were low and not quantifiable (Figure 3SI).

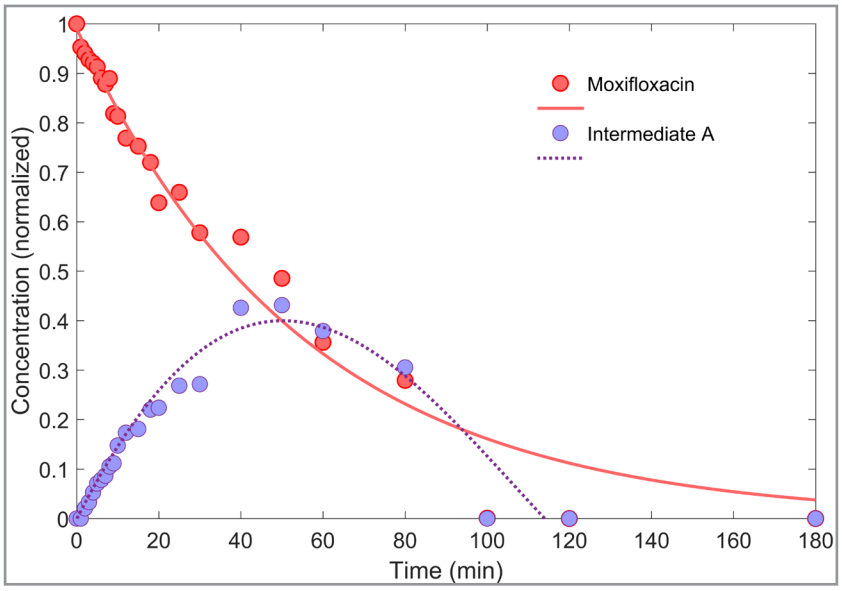


Figure 5. Photolytic degradation of moxifloxacin in pH=10.20 and formation of a byproduct. Plot built with relative concentration values from **C** matrix, with augmented matrix and correlation constraint.

Although the spectral profiles recovered from the photolysis process have several common features, the behavior of the relative concentrations of moxifloxacin, intermediates A and B throughout the reaction (Figures 5 and 2SI) evolve very differently, thus suggesting that structural changes are occurring. Key regions, or bands, identified by comparing the different spectra (Figure 6), show an interesting correlation with the route proposed by Ahmad *et al.* [6] where three species could be identified following hydroxylation at the pyrrole ring, hydroxylation/oxidation of the side chain of piperidine ring, and finishing with a cleavage of the entire diazabicyclononane moiety. The vibrational modes changing through the photolysis in this work, *i.e.* 968, 1128, 1232, 1472 cm^{-1} are related to the diazabicyclonyl substituent, pointing to changes happening before cleavage. On the other hand, the band at 1536 cm^{-1}, from carbonyl and carboxyl groups is almost disappears from moxifloxacin to intermediate B, which could be associated to a starting decarboxylation process. Also, subtle changes at 1128 cm^{-1}, could be related to a possible cleavage of the -C-F bond (although this band also has contribution of the diazabicyclonyl group), but further experiments would be necessary to reach a conclusion.

It is interesting to notice that the bands related to the quinolone backbone remain strong through at least two hours, then there is a drastic lowering, becoming barely detectable through SERS. The absence of SERS features could indicate that, upon a possible cleavage of the substituent diazabicyclonyl, the remaining products lose their affinity towards AuNPs' surface. This implies, from the perspective of the interaction molecule-metallic surface, that the substituent moiety would be governing the whole molecule anchoring to the AuNPs, due to its more flexible nature, causing a favorable orientation to be observable in SERS. This dependence would also be associated to the marked separation of the groups observed in the scores plot discussed before (Figure 2). Evidence of further degradation was neither observed nor expected under our conditions.

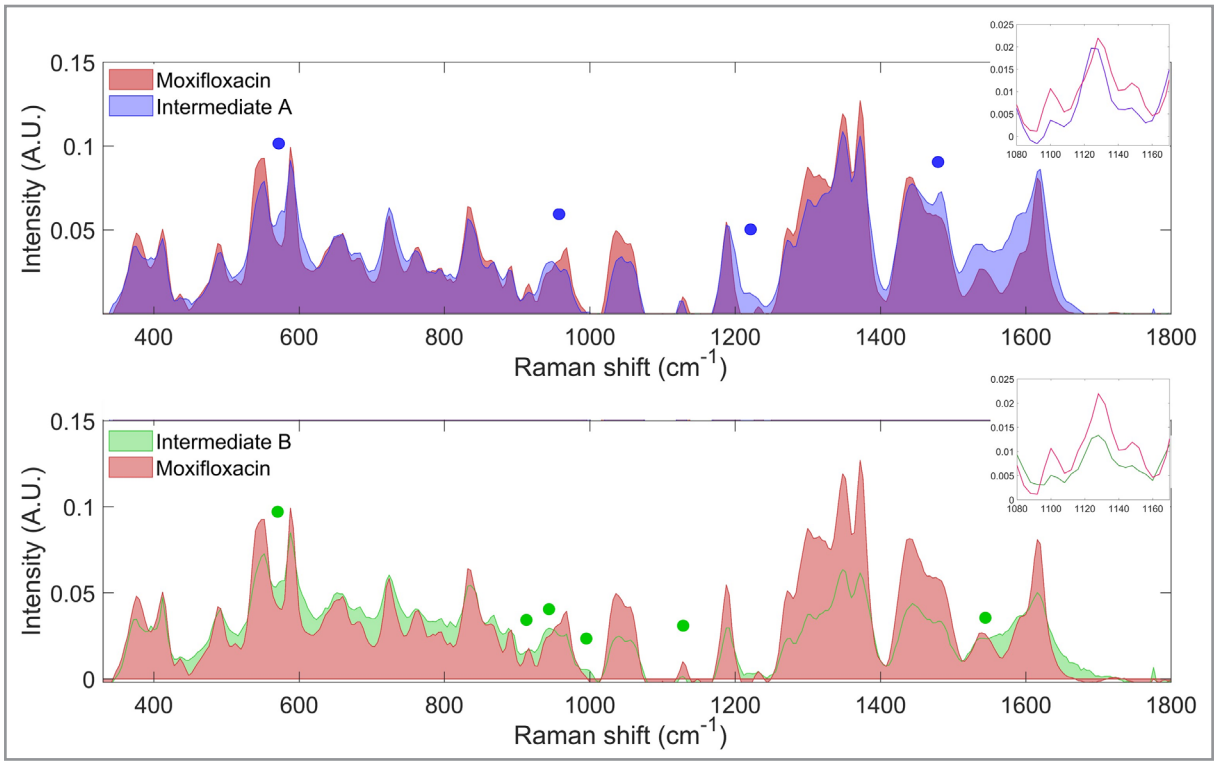


Figure 6. Spectral profiles calculated by MCR-ALS for the purest components of photolytic reaction of moxifloxacin and intermediates A and B). Inset plots: detail of the region between 1080-1170 cm^{-1}.

CONCLUSIONS

In this work, a typical SERS experimental procedure was applied in the monitoring of a photocatalytic process, which offered information-rich spectroscopic information. Soft-modelling through MCR-ALS allowed to obtain quantitative information, and calculate kinetic parameters, despite that the system under study presents intermediates of the degradation process that are difficult to isolate, or even measure them using SERS, so that their profiles could be used as initial estimates for the MCR-ALS routine. The relative concentrations calculated by MCR-ALS method constrained with known information about the photolytic reaction allowed to simultaneously accompany the changes in the starting species and to model an intermediate product. Although there is few information available in the literature about the band assignments of moxifloxacin in SERS, the analysis of the calculated Raman spectrum allowed to suggest a degradation pathway that agreed with those previously reported, showing the photolytic cleavage of diazabicyclonyl group under alkaline conditions. Both the kinetic modelling and exploratory analysis suggested that the presence of the substituent played a role in the anchoring of moxifloxacin (or the intermediate products), towards the metallic surface, but additional experiments would be necessary to confirm this hypothesis.

Acknowledgements

This paper is dedicated to the memory of professor Ronei J. Poppi whom we lost in 2020. The authors acknowledge the “Instituto Nacional de Ciência e Tecnologia de Bioanalítica” (INCTBio) and the “Coordenação de Aperfeiçoamento de Pessoal de Nível Superior” (CAPES, Brasilia, Brazil) for fellowships and financial support. Also, the authors are indebted to Dr. Vitor Hugo Paschoal from the Institute of Chemistry at the University of São Paulo for the Raman spectrum calculations.

Conflicts of interest

The authors declare that there is no conflict of interest.

REFERENCES

1. Scholar, E. “Moxifloxacin”. In: *xPharm: The Comprehensive Pharmacology Reference*, Elsevier, **2007**, pp 1–6.
2. Kanakaraju, D.; Glass, B. D.; Oelgemöller, M. *J. Environ. Manage.*, **2018**, *219*, pp 189–207 (http://dx.doi.org/10.1016/j.jenvman.2018.04.103).
3. Anjali, R.; Shanthakumar, S. *J. Environ. Manage.*, **2019**, *246*, pp 51–62 (http://dx.doi.org/10.1016/j.jenvman.2019.05.090).
4. Khare, P.; Patel, R. K.; Sharan, S.; Shankar, R. Recent trends in advanced oxidation process for treatment of recalcitrant industrial effluents. In: Shah, M. (Ed.). *Advanced Oxidation Processes for Effluent Treatment Plants*, Elsevier, **2020**, pp 137–160.
5. Yahya, M. S.; Beqqal, N.; Guessous, A.; Arhoutane, M. R.; El Kacemi, K. *Cogent. Chem.*, **2017**, *3* (1), 1290021 (http://dx.doi.org/10.1080/23312009.2017.1290021).
6. Ahmad, I.; Bano, R.; Musharraf, S. G.; Ahmed, S.; Sheraz, M. A.; ul Arfeen, Q.; Bhatti, M. S.; Shad, Z. *AAPS PharmSciTech*, **2014**, *15* (6), pp 1588–1597 (http://dx.doi.org/10.1208/s12249-014-0184-x).
7. Van Doorslaer, X.; Demeestere, K.; Heynderickx, P. M.; Van Langenhove, H.; Dewulf, J. *Appl. Catal., B.*, **2011**, *101* (3–4), pp 540–547 (http://dx.doi.org/10.1016/j.apcatb.2010.10.027).
8. Langer, J.; de Aberasturi, D. J.; Aizpurua, J.; Alvarez-Puebla, R. A.; Auguié, B.; Baumberg, J. J.; Bazan, G. C.; Bell, S. E. J.; Boisen, A.; Brolo, A. G.; et al. *ACS Nano*, **2020**, *14* (1), pp 28–117 (http://dx.doi.org/10.1021/acsnano.9b04224).
9. Bedia, C.; Sierra, À.; Tauler, R. *Anal. Bioanal. Chem.*, **2020**, *412* (21), pp 5179–5190 (http://dx.doi.org/10.1007/s00216-020-02595-8).
10. Gómez-Canela, C.; Bolivar-Subirats, G.; Tauler, R.; Lacorte, S. *J. Pharm. Biomed. Anal.*, **2017**, *137*, pp 33–41 (http://dx.doi.org/10.1016/j.jpba.2017.01.019).
11. Kaur, P.; Imteaz, M. A.; Sillanpää, M.; Sangal, V. K.; Kushwaha, J. P. *Chemom. Intell. Lab. Syst.*, **2020**, *203*, 104027 (http://dx.doi.org/10.1016/j.chemolab.2020.104027).
12. Indrasekara, A. S. D. S.; Meyers, S.; Shubeita, S.; Feldman, L. C.; Gustafsson, T.; Fabris, L. *Nanoscale*, **2014**, *6* (15), pp 8891–8899 (http://dx.doi.org/10.1039/C4NR02513J).
13. Lee, P. C.; Meisel, D. *J. Phys. Chem.*, **1982**, *86* (17), pp 3391–3395 (http://dx.doi.org/10.1021/j100214a025).
14. Mamián-López, M.; Temperini, M. *Quim. Nova*, **2019** (http://dx.doi.org/10.21577/0100-4042.20170424).
15. Hong, K. Y.; de Albuquerque, C. D. L.; Poppi, R. J.; Brolo, A. G. *Anal. Chim. Acta*, **2017**, *982*, pp 148–155 (http://dx.doi.org/10.1016/j.aca.2017.05.025).
16. El-Zahry, M. R.; Lendl, B. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, **2018**, *193*, pp 63–70 (http://dx.doi.org/10.1016/j.saa.2017.12.007).
17. Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. *J. Chem. Phys.*, **2020**, *152* (22), 224108 (http://dx.doi.org/10.1063/5.0004608).
18. Weigend, F. *Phys. Chem. Chem. Phys.*, **2006**, *8* (9), pp 1057-1065 (http://dx.doi.org/10.1039/b515623h).
19. Weigend, F. *J. Comput. Chem.*, **2008**, *29* (2), pp 167–175 (http://dx.doi.org/10.1002/jcc.20702).
20. Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.*, **2005**, *7* (18), pp 3297-3305 (http://dx.doi.org/10.1039/b508541a).
21. Jaumot, J.; de Juan, A.; Tauler, R. *Chemom. Intell. Lab. Syst.*, **2015**, *140*, pp 1–12 (http://dx.doi.org/10.1016/j.chemolab.2014.10.003).
22. Saurina, J.; Leal, C.; Compañó, R.; Granados, M.; Prat, M. D.; Tauler, R. *Anal. Chim. Acta*, **2001**, *432* (2), pp 241–251 (http://dx.doi.org/10.1016/S0003-2670(00)01378-7).
23. Olivieri, A. C. *Chem. Rev.*, **2014**, *114* (10), pp 5358–5378 (http://dx.doi.org/10.1021/cr400455s).
24. Bayat, M.; Marín-García, M.; Ghasemi, J. B.; Tauler, R. *Anal. Chim. Acta*, **2020**, *1113*, pp 52–65 (http://dx.doi.org/10.1016/j.aca.2020.03.057).
25. Windig, W.; Guilment, J. *Anal. Chem.*, **1991**, *63* (14), pp 1425–1432 (http://dx.doi.org/10.1021/ac00014a016).

SUPPORTING INFORMATION

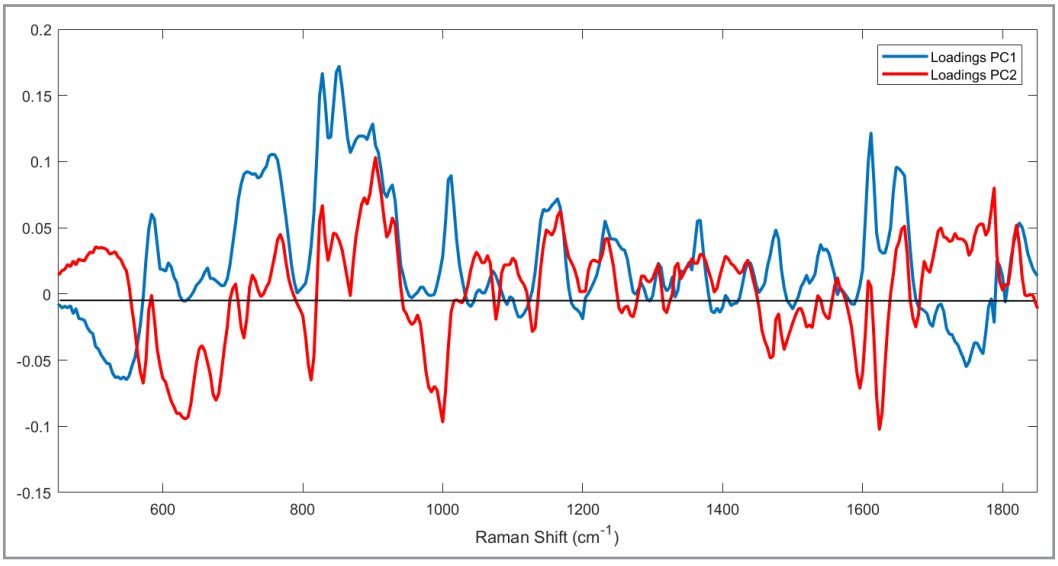


Figure 1SI. Loadings plot for PCA analysis of Figure 2 in main text.

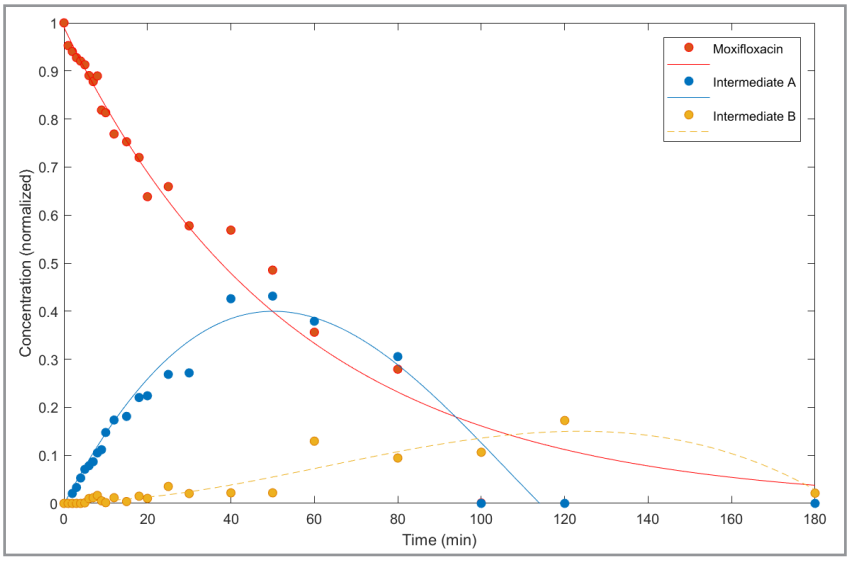


Figure 2SI. Photolytic degradation of moxifloxacin in pH=10.20 and formation of a byproduct. Plot built with relative concentration values from **C** matrix, with augmented matrix and correlation constraint.

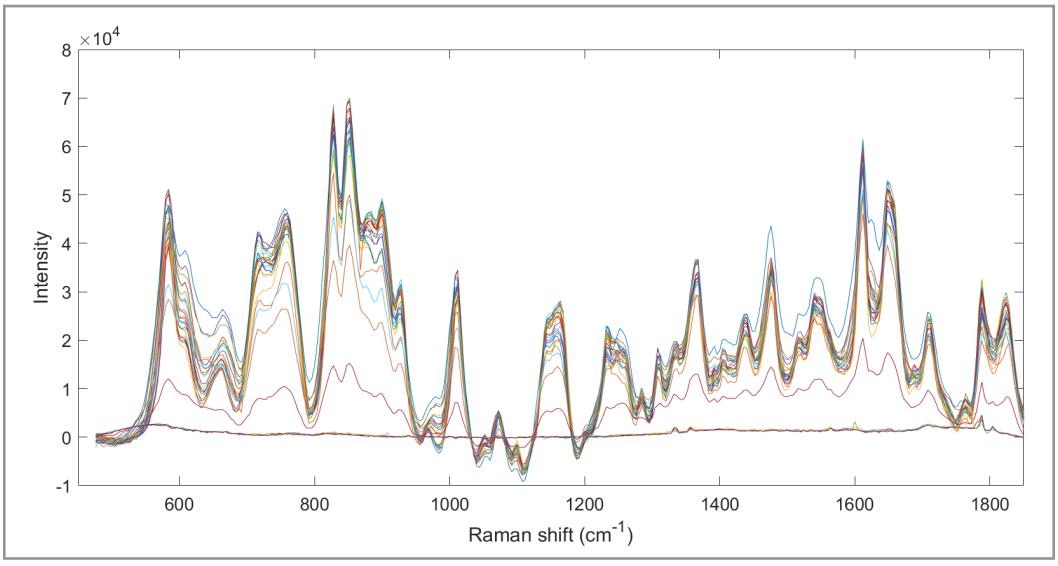


Figure 3SI. SERS Spectra through an entire photolytic process of moxifloxacin.

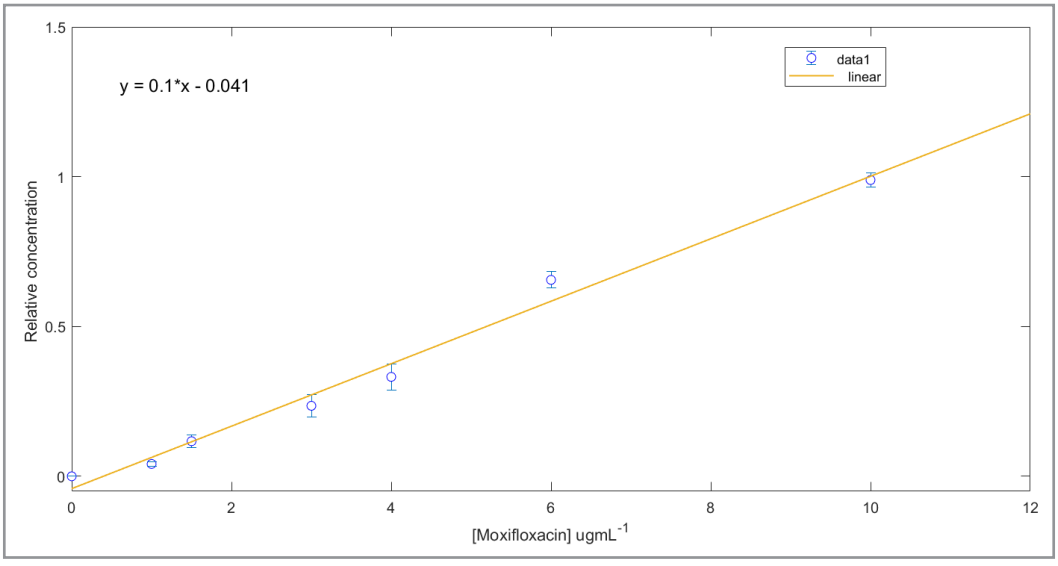


Figure 4SI. Analytical curve for the moxifloxacin quantification. Real concentration values of moxifloxacin standards *vs.* relative concentration values from MCR-ALS calculations.

Figures of Merit for Moxifloxacin quantification (Analytical Curve of Figure 3SI)

R^2	LOD (μg mL^{-1})	LOQ (μg mL^{-1})	F_{LOF}/F_C
0.989	0.12	0.36	2.06/3.03

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp178–197
doi: 10.30744/brjac.2179-3425.AR-31-2021

BrJAC

ARTICLE

Experimental Design to Enhance Dopamine Electrochemical Detection Using Carbon Paste Electrodes

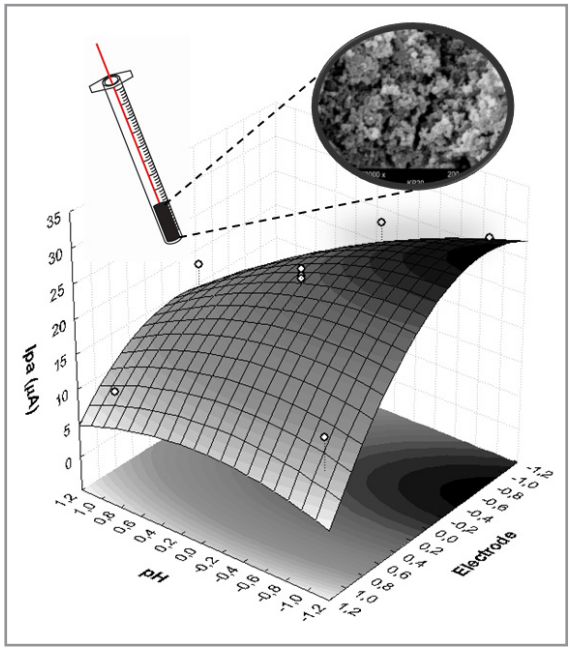
Soraya Adriane Blum[1], Felipe Zahrebelnei[1], Noemi Nagata[2], Valtencir Zucolotto[3], Luiz Henrique Capparelli Mattoso[4], Christiana Andrade Pessoa[1], Jarem Raul Garcia[1], Karen Wohnrath[1]*

[1]*Departamento de Química, Universidade Estadual de Ponta Grossa, Av. Carlos Cavalcanti, 4748, ZC 84030-900, Ponta Grossa, PR, Brazil*

[2]*Departamento de Química, Universidade Federal do Paraná, Centro Politécnico, ZC 81531-980, Curitiba, PR, Brazil*

[3]*Instituto de Física de São Carlos, Universidade de São Paulo, Av. Trabalhador São-carlense, 400, Parque Arnold Schimidt, ZC 13566-590, São Carlos, SP, Brazil*

[4]*Laboratório Nacional de Nanotecnologia Aplicada ao Agronegócio (LNNA), Embrapa Instrumentação, ZC 13560-970, São Carlos, SP, Brazil*



Efforts have been made on the development of new modified electrodes to be used in the fast determination of neurotransmitters, either in commercial drugs or in biological samples. Determination of dopamine (DA), for example, is of great importance since the lack of this neurotransmitter is related to many neurological disorders, including Parkinson's and Alzheimer's diseases. In this paper, we present a detailed electrochemical characterization, as well as DA detection studies of paste electrodes incorporating carbon materials in different allotropic forms, including carbon black modified with intrinsically conducting polymers (Eeonomers®), pristine carbon black, graphite, and carbon nanotubes. Emphasis is put on the smaller particle size and larger specific surface area of CB Eeonomers® materials, which led to an improved electroanalytical performance for the developed devices. The electrodes fabricated with Eeonomers® modified with polyaniline exhibited the highest current response towards DA detection, in addition to the ability of distinguishing DA from its natural interferent, ascorbic acid. Furthermore, a central composite design was used to investigate the influence of pH and electrode composition (proportion of Eeonomers®) on the electrochemical sensing of DA. A greater

Cite: Blum, S. A.; Zahrebelnei, F.; Nagata, N.; Zucolotto, V.; Mattoso, L. H. C.; Pessoa, C. A.; Garcia, J. R.; Wohnrath, K. Experimental Design to Enhance Dopamine Electrochemical Detection Using Carbon Paste Electrodes. *Braz. J. Anal. Chem.*, 2021, *8* (32), pp 178–197. doi: http://dx.doi.org/10.30744/brjac.2179-3425.AR-31-2021

Submitted 01 March 2021, Resubmitted 08 June 2021, Accepted 10 June 2021, Available online August 2021.

sensitivity was achieved for 50:50 (w/w) KP20/KPy20 electrode at pH 7.0. The optimized devices showed to be promising tools to perform quick, cheap and sensitive detection of this neurotransmitter in bioanalytical systems.

Keywords: carbon paste electrode, carbon black, intrinsically conducting polymers, electroanalysis, electrocatalysis, response surface, factorial design.

INTRODUCTION

Research on the development of new carbon paste electrodes (CPEs) have experienced a fast growth since the first reports by Adams [1]. CPEs have been widely used in electroanalysis due to their low background current, wide work potential window, chemical stability and low cost. CPEs can also be chemically modified, which can significantly improve their sensibility, selectivity and overall analytical performance towards the detection of a number of analytes [2-4].

The use of carbonaceous materials for constructing CPEs has been enabling the production of excellent detection devices with diversified applications [5]. Several carbon allotropes and nanomaterials have been investigated in this regard, including glassy carbon (GC) [6], graphite (GP) [7], carbon nanotubes (CNT) [8-9], carbon nanofibers [10] and graphene [11-12].

Carbon nanotubes (CNT) have also attracted attention as promising materials for modified electrodes, allowing the fabrication of highly sensitive and selective devices. Indeed, the electrochemical applications of CNTs have been greatly explored, particularly on battery technology [13], supercapacitors [14], and sensors and actuators [15-18]. CNTs are excellent electrode materials because of their unique electronic and mechanical properties, as well as their low chemical reactivity exhibited in most electrolyte solutions, retaining a high surface activity and a wide operational potential window [19]. Single-walled carbon nanotubes (SWCNTs) and multi-walled carbon nanotubes (MWCNTs) made into paste electrodes, using mineral oil as agglutinant, are also valuable tools for detection of several molecules [20-22]. Valentini et al [23] described for the first time the electrochemical behavior and the electroanalytical performance of a SWCNT paste electrode, which exhibited a sensitivity enhancement toward dopamine oxidation, in comparison to graphite-based CPEs.

Another important carbon material is carbon black (CB), which is a paracrystalline form of carbon, presenting small particle size and large specific surface area. CB particles grown together to form aggregates of different sizes and shapes. This material has been employed as an alternative carbon allotropic form for applications as electroconductive additive in hydrogen storage devices, and as cathode materials in lithium-ion micro-batteries [18]. The use of CB as active materials in CPEs has also been intensively investigated due to their excellent electrical conductivity and presence of several defect sites. Those properties contribute to faster electron transfer kinetics on the electrodic material, providing interesting abilities to mediate electron transfer reactions for some analytes [25-27]. Additionally, CB-based paste electrodes have been produced in a variety of configurations by exploiting the versatility of carbon paste and CB nanomaterial.

DA is a biological amine that has excitatory and inhibitory activity in postsynaptic membranes, performing essential roles in the endocrine, cardiovascular, excretory and central nervous systems [28]. In physiological samples, DA concentration ranges from 10^{-6} to 10^{-9} mol L^{-1} [29], and changes in these levels are related to several neurological disorders, such as Parkinson's disease, Alzheimer's disease and schizophrenia, among others. This neurotransmitter is also used as an intravenous resuscitation medication in hospitals, which demands a rigorous quality control to ensure the effectiveness and safety of this drug [30]. In these contexts, the development of fast and low-cost methods for detecting DA in biological fluids and pharmaceutical formulations is fundamental for the clinical field and pharmaceutical industry.

The method commonly used for DA analysis is high-performance liquid chromatography (HPLC). This method exhibits high sensitivity and excellent selectivity; however, it has important drawbacks, such as the need for expensive equipment, demand for trained technicians, complex sample preparation

procedures and the impossibility of in situ analysis. In contrast, electrochemical methods can overcome these disadvantages, in addition to present high sensitivity, good selectivity, quick response, and lower cost. Furthermore, the instrumentation required for electroanalytical determinations is simple and can be conveniently miniaturized, which enables its use for *in situ* and automated detections [31-33]. Due to the mentioned advantages, the development of new low-cost, selective, sensitive and simple devices for DA detection is justified.

When working with the advatangeous electrochemical sensors, it is also extremely important that these devices allow analyzes to be carried out on complex samples, such as biological samples. DA is easily detected due to its easy oxidation, however, in physiological samples, several substances coexist that can interfere with its electrochemical detection [34]. There is a large amount of work describing methodologies for the detection of DA with high sensitivity, but its selective detection remains a major challenge for researchers. The literature presents us two important interfering substances in DA electrochemical detection process: uric acid (UA) and ascorbic acid (AA), with AA as the most important because it is found naturally as a vital antioxidant against biological oxidation process [35]. These two compounds stand out due to their oxidation potential being close to the DA one and to the fact that their concentrations being hundreds of thousands of times higher than the concentration of DA in physiological environment. Additionally, the oxidation products of these compounds can be potentially absorbed or electropolymerized on the surface of the electrodes, reducing sensors reuse and reproducibility. Therefore, for a DA sensor to be actually used in real samples, selectivity is crucial [35].

The use of multivariate optimization tools, such as factorial design and response surface methodology (RSM), has showed to be advantageous for optimizing experimental parameters and conditions in a number of systems in both research laboratories and industry. In the electroanalytical field, the usage of such approaches has been leading to enhanced detection performances for several sensing devices. For instance, recent literature shows various studies in which the use of factorial designs has enabled significant improvements in important analytical parameters, such as selectivity [36-37] and sensitivity [38-39], for different electrochemical sensing systems.

In this study, we describe the preparation of CPEs for the electrochemical detection of DA, with emphasis on the use of CB modified with conducting polymers, namely polyaniline (Pani) and polypyrrole (PPy). Conducting polymer-modified carbon black (Eeonomers®) [29] is a thermally stable conductive material made *via in situ* deposition of intrinsically conducting polymers. These materials have found applications mainly as electroactive fillers in polymeric composites with improved processing properties [30-32]. However, the presence of Pani and PPy on the CB surface provides a significant increase in the electrical conductivity of the material and in the number of surface active sites due to the presence of alternating double and single bonds [46] and nitrogen atoms on the polymeric structure [47,48]. The incorporation of such polymers onto CB particles can enhance their conductive properties and electroactivity, which significantly impacts their sensing capability. These characteristics were explored in this work to develop sensitive, selective and feasible electrochemical sensors to perform DA determination in real samples.

MATERIALS AND METHODS

Chemicals

All chemicals were of analytical grade and used as received. Dopamine hydrochloride and multiwalled carbon nanotubes were purchased from Sigma-Aldrich. Ascorbic acid was purchased from Merck. Graphite powder was acquired from Fluka. Three different types of Eeonomers® (Eeonyx Co) were used, including pure CB (KP0 - specific surface area of *ca.* 1400 m^2/g), CB modified with 20 wt% polyaniline, KP20 (particle size 40 nm, specific surface area of *ca.* 570 m^2/g and density of 0.062 g/cm^3) and CB modified with 20 wt% of polypirrole, KPy20 (particle size 40 nm, specific surface area of *ca.* 390 m^2/g and density of 0.062 g/cm^3). Details on synthesis, characterization and applications of Eeonomers® can be found elsewhere [40-43]. Double distilled water was used to prepare all solutions.

Preparation of the carbon paste electrodes

The unmodified graphite CPE (GRPE) was prepared by mixing graphite powder and mineral oil (12% w/w). The unmodified MWCNT CPE (CNTPE) was prepared by mixing nanotube powder with 45% (w/w) of mineral oil. The unmodified pristine CB CPE (KP0PE) was prepared by mixing carbon black (KP0) powder with 14% (w/w) of mineral oil. The electrodes containing KP20 (KPPE) and KPy20 (KPyPE) were prepared by mixing the powders with 28% w/w and 26% w/w mineral oil, respectively.

0.1 g of each paste was inserted into a plastic needle-type capillary tube measuring 6 mm diameter and 5 cm length, and a 1 mm diameter Ni-Cr wire connected to a 5 mm diameter graphite cylinder was used as electrical contact. Electrode surfaces were smoothed using ordinary vegetal paper and rinsed carefully with double distilled water.

Electrochemical measurements

Electrochemical measurements were conducted with a conventional three-electrode cell using in a PalmSens potentiostat (Palm Instruments BV). An Ag/AgCl electrode (saturated with KCl) and a platinum wire (fixed area at 0.8 mm^2) served as the reference and auxiliary electrodes, respectively. The unmodified GRPE, CNTPE and KP0PE and the modified KPPE and KPyPE were used as working electrodes.

The electrochemical characterization of each CPE as well as DA and AA quantification were performed by using cyclic voltammetry, since this technique is the most suitable for a profound characterization in relation to the kinetic aspects of electrochemical reactions occurring in the electrode/solution interface. Cyclic voltammograms were recorded in the potential range from 0.0 to 0.8 V *vs* Ag/AgCl, at scan rates of 10 mV s^{-1} and 90 mV s^{-1}. NaCl 0.5 mol L^{-1} at pH 4.0 (adjusted with HCl 0.1 mol L^{-1}) was used as supporting electrolyte. Calibration curves were obtained for DA and AA in the range between 2 and 50 µmol L^{-1}. All experiments were performed at room temperature.

Experimental Design and Response Surface Methodology

A systematic study was carried out to define the optimal conditions to obtain a higher anodic peak current (Ipa) in DA determination upon varying the proportion of the two different Eonomers® and pH of the supporting electrolyte. Using these parameters, optimization based on a 3^2 factorial design and response surface methodology was performed. A set of 11 experiments was carried out.

The conducting polymer-modified Eeonomers® used were KP20 and KPy20. A mixture of these Eeonomers® (KP20/KPy20) at 50:50 (w:w) was also used. Each electrode was prepared by mixing Eeonomers® with mineral oil (28, 26 and 27%, respectively). The pH supporting electrolyte was controlled using acetate buffer solution (pH 3.0, 5.0 and 7.0). The choice of pH and the proportion between the two conducting polymer-modified Eeonomers® as parameters on this experimental design is justified due to: *(i)* pH of the electrolyte solution plays an important role on the electrode response, since it can directly influence the conductivity of the system by polymer doping; *(ii)* the conducting polymer significantly influences the electrochemical response; hence, it is believed that the effect of the proportions of the different materials can reveal the individual role of each conducting polymer towards DA detection.

The stabilization of electrodes was performed with 25 cycles in the presence of supporting electrolyte. Subsequently, randomized electrochemical analyses in presence of 50 µmol L^{-1} DA were carried out in the potential range from 0 to 800 mV *vs* Ag/AgCl, at 30 mV s^{-1}, according to the conditions shown in Table V. The Ipa data were processed with Statistica software (version 5.0).

DA Determination in Commercial Drugs

Standard addition methodology was applied in the analysis of commercial DA injections containing 200 mg of Dopamine Hydrochloride. Aliquots of the commercial sample (20 µL) were fortified with five additions of DA standard stock solution (1 x 10^{-2} mol L^{-1}). Finally, the solutions were analyzed by cyclic voltammetry using a previously prepared electrode containing the Eeonomer® KP20 (KPPE) and NaCl as supporting electrolyte.

RESULTS AND DISCUSSION

CPE preparation with different carbon materials

As usual for CPE preparation, a paste consisting of carbonaceous powder and mineral oil as binder composes all electrodes. The amount of binder depends on the granulation of the carbon-based material and this proportion can cause increment of background current. According to Malha et al., 2013, as CB possesses high surface area, carbon pastes prepared with this material do require high amounts of binder (25% w/w) [44]. Ideal amount of binding agent was verified using cyclic voltammetry, and it was considered optimal when the resistive contribution on the voltammetric profile was not observed with the reduction on the amount of mineral oil (data not shown).

As Eeonomers® present larger specific surface area (see morphological analysis in Figure S1), a proper amount of CB was mixed with a suitable amount of mineral oil. As shown in the experimental section, KP0 has the largest surface area and a lower percolation limit, thus it is expected a larger liquid surface accessible to the matrix and KP0PE was prepared by hand-mixing KP0 powder with 14% w/w mineral oil. This amount of binding agent is smaller than the one used for KP20 and KPy20, 38% w/w and 41% w/w, respectively. Higher amount of mineral oil was needed to disperse the modified Eeonomers® due to the presence of conducting polymers that present polar structure causing a decrease in the wettability of modified Eeonomers® related to the binder. Trying to avoid the use of higher amounts of mineral oil, KP20 and KPy20 materials were ground, and CPEs were prepared again and this new condition have resulted in a decreased mineral oil amount of 26% w/w and 28% w/w for KP20 and KPy20, respectively, which were employed as optimal electrode compositions.

GR and CNT-based paste electrodes were fabricated using the following proportions of mineral oil: 12% w/w and 45% w/w, respectively.

Electrochemical behavior of different carbon materials

The electrochemical behavior of all studied CPE was first investigated using the supporting electrolyte solution, NaCl 0.5 mol L^{-1} (pH = 4.0) at 30 mV s^{-1}. Results are shown in Figure 1.

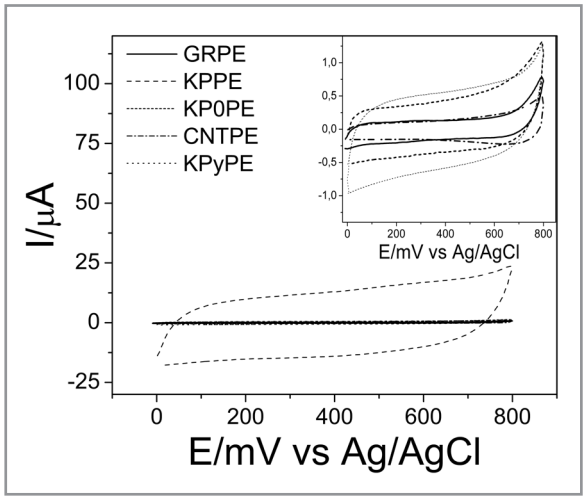


Figure 1. Cyclic voltammograms for the KPPE, KPyPE, KP0PE, GRPE and CNTPE, at ν= 30 mV s^{-1}. Supporting electrolyte: NaCl 0.5 mol L^{-1} at pH 4.0. For better visualization, the voltammograms for the KPyPE, KP0PE, GRPE and CNTPE electrodes are highlighted in an inset in the Figure.

The cyclic voltammograms for all paste electrodes exhibited different capacitive currents, according to the material employed for electrode fabrication. It is interesting to observe that neither the KPPE nor the KPyPE presented the characteristic voltammetric profile of the respective conducting polymer, since all of them did not present current peaks and only capacitive current was observed. This was assigned to the presence of polymeric films around the carbon black particles. In acidic solution, with the doping of the polymers, there is an increase on the conductivity of the materials. However, since the films are not thick enough to enable ion diffusion, only fast faradaic processes on the material surface occurs, creating a process known as pseudocapacitance [45].

It is interesting to notice that the capacitive current for the KPPE is 23 times higher than for its counterpart KPyPE, even considering that the specific area of the first one in only 1.5 times higher when compared to the former material. This can be understood considering that under acid doping PANI conductivity is higher than PPy [46]. In addition, the voltammograms obtained with the other CPE shows only capacitive current, probably due to Double Layer charging. It is also worth mentioning that the capacitive current for the unmodified CB is 30 times lower than the current observed for the KPPE, even though its specific area is 2.5 higher than the area for the KP20, fact that corroborate the occurrence of pseudocapacitive process in the modified Eeonomers®.

Eeonomers® materials are produced with different proportions of conducting polymers. However, our studies have indicated that the material with 20% of polymers have demonstrated the most promising electrochemical behavior aiming the construction of electrochemical sensors (*S2*).

Investigation of DA voltammetric behavior using different CPEs Electrodes

The response of the different CPEs in presence of DA were investigated toward the oxidation process of the analyte. A well-defined anodic current peak was observed for all electrodes, which is attributed to the oxidation of dopamine to dopaminequinone, with the involvement of 2 electrons and 2 protons. This redox process has been extensively studied in the literature [47-50]. Figure 2.a displays the voltammmetric responses of the modified and unmodified Eeonomers® KP0PE, KPPE and KPyPE in the presence of DA. As discussed above, the most noticeable characteristic of the modified Eeonomers® is related to the presence of intense capacitive current, however both KPPE and KPyPE presented peak currents with absolute values (capacitive + faradaic current) higher than the ones exhibited by the unmodified Eeonomers® KP0PE. Figure 2.b shows the voltammetric behavior for the electrodes prepared with different carbon allotropes used as reference. As expected, CNTPE (carbon nanotubes) presented a quasi-reversible voltammetric profile related to DA redox process. It is possible to observe a peak potential separation (ΔEp) *ca.* 83 mV and an anodic and cathodic peak current ratio for the process (*Ipa/Ipc)* of *ca.* 1.2. The other two CPE showed on Figure 2.b also presented pronounced anodic peaks compared to the modified Eeonomers® (Figure 2.a). The parameters obtained from the measured cyclic voltammetry for the different carbon-based electrodes in presence of DA are showed in Table I.

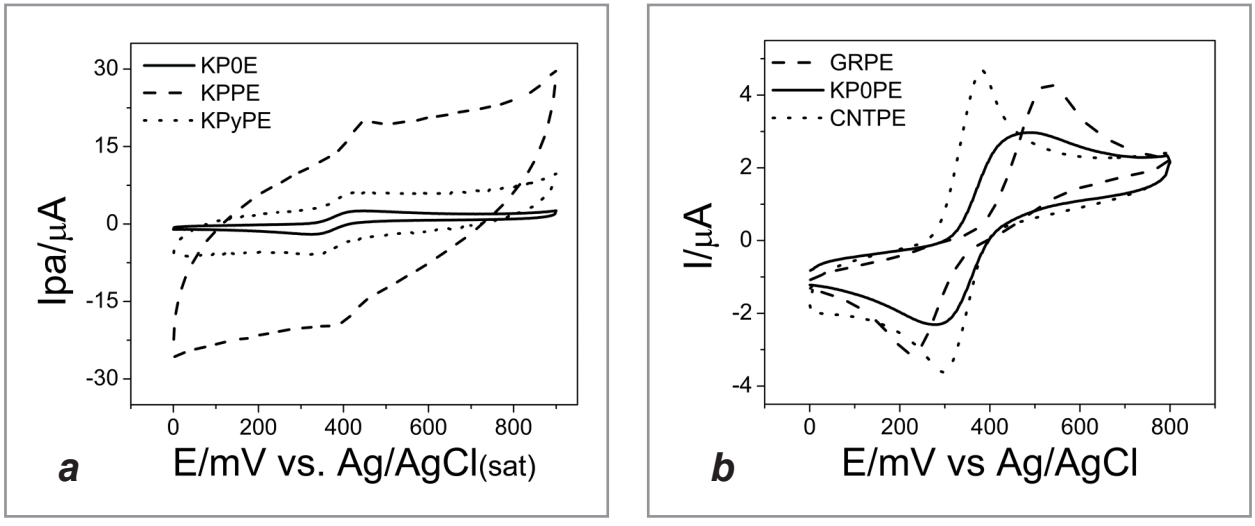


Figure 2. ***a)*** Cyclic voltammograms for the electrodes KP0PE, KPPE and KPyPE, and ***b)*** Cyclic voltammograms for the electrodes GRPE, KP0PE e CNTPE, in 50 μmol L^{-1} of DA. Supporting electrolyte: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 30 mV s^{-1}.

Table I. Electrochemical parameters obtained from the cyclic voltammograms of GRPE, KP0PE, CNTPE, KPyPE and KPPE electrodes in the presence of 50 μmol L^{-1} DA

CPE	Ep_a (mV)	Ep_c (mV)	$\Delta E_p = Ep_a - Ep_c$ (mV)	Ip_a (μA)	Ip_c (μA)	Ip_a/Ip_c
GRPE	530.0	226.0	304.0	3.88	-3.50	-1.11
CNTPE	380.0	297.0	83.0	4.25	-3.60	-1.20
KP0PE	484.0	280.0	199.0	2.69	-2.70	-0.99
KPyPE	433.0	329.0	102.0	2.82	-2.70	-1.04
KPPE	454.0	387.0	67.0	5.98	-6.01	-0.99

From Table I it is evidenced that, in addition to the CNTPE the two other electrodes that presented fast kinect for electron transfer for DA redox process are KPPE and KPyPE, even considering the presence of a high capacitive current. This behavior can be attributed to the smaller values of ΔE_p, 67 and 102 mV, respectively (criteria for reversible process: $\mathbf{\Delta E_p =} 57/\boldsymbol{n}\ mV$ at 25 ^0C - where ***n*** is the number of electrons involved in the process and **Ipa/Ipc =** 1.0) [51]. The faster kinetics for DA redox process on KP20 and KPy20 can be assigned to the presence of the pseudocapacitance phenomena that is caused by fast faradaic processes occurring on the material surface as a consequence of the presence of polymeric conducting film at the modified Eeonomers®. The Epa and Epc values shown in Table I were compared with those obtained with a conventional Pt electrode (Figure *S3*), in the same experimental conditions. We observed that oxidation of DA on a Pt electrode occurred at higher potentials (Epa= 504 mV). Another disadvantage on the use of Pt electrode is that DA adsorbs on the surface of this metal, which may lead to a fouling effect.

The effect of scan rate on the oxidation response of DA was examined in the range of 10-100 mV s^{-1} (Figure *S4*). The oxidation currents linearly increased with the square root of scan rate for all electrodes, indicating that the processes were diffusion-controlled, which is expected for catalytic systems and advantageous for quantitative determinations.

The ability of the electrodes to detect DA was investigated using CV, in the range of 2 to 50 μmol L^{-1}. For all electrodes employed, a linear relationship between the oxidation current and DA concentration was observed, allowing the determination of analytical parameters displayed in Table II.

Table II. Analytical parameters of DA determination using different carbon paste electrodes

CPE	Linear Regression Equation	Linear correlation factor	Sensitivity (A mol^{-1} L)	Linear Range of Concentration (mol L^{-1})
GRPE	Ipa = 0.06 + 0.051 [DA]	0.9930	0.051	2 – 50
CNTPE	Ipa = 0.02 + 0.065 [DA]	0.9936	0.065 ± 0.04	2 – 50
KP0PE	Ipa = 0.02 + 0.043 [DA]	0.9997	0.043	2 – 50
KPyPE	Ipa = 0.02 + 0.021 [DA]	0.9983	0.021	2 – 50
KPPE	Ipa = 0.15 + 0.110 [DA]	0.9958	0.110 ± 0.04	2 – 50

It is worth noticing that a wide linear range was obtained. A good sensitivity was observed for electrodes containing Eeonomer® KP20 (KPPE), followed by the CNT-containing electrodes (CNTPE).

Figure 3 shows the analytical curves obtained for the KPPE and CNTPE. The quantification of DA for these two electrodes were obtained in triplicate and the standard deviation is indicated for each point. It is also important to mention that the higher sensibility of KPPE electrode is clearly indicated in Figure 3. It is also interesting to notice that the error on the measurement of the peak current for each of the studied DA concentrations is slightly smaller for the CNTPE electrode, probably due to the lower definition of the peaks on the KPPE.

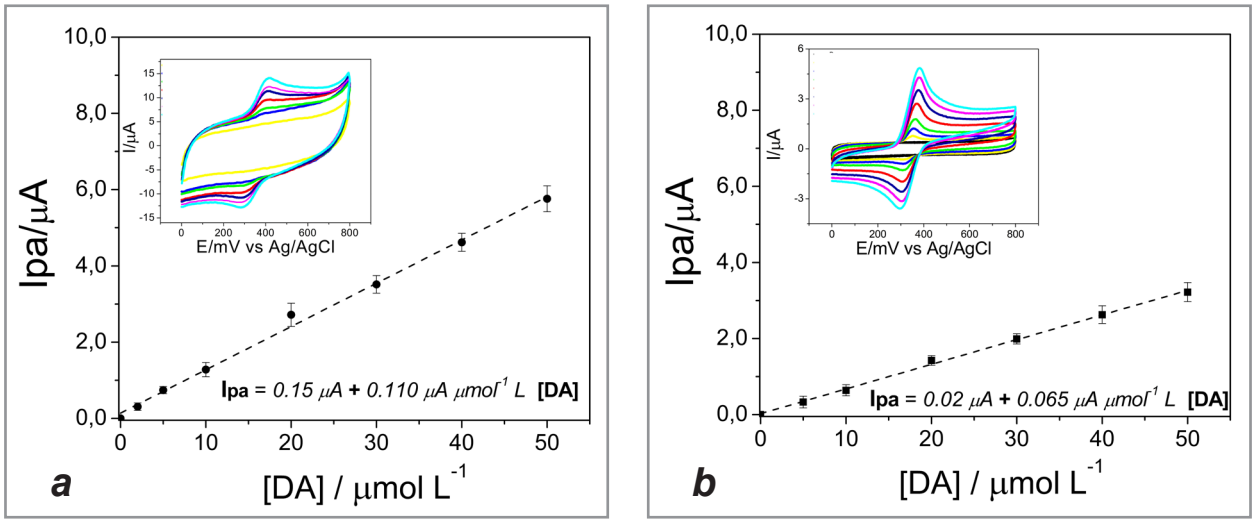


Figure 3. Analytical curves obtained by cyclic voltammetry in presence of DA for ***a)*** KPPE electrode and ***b)*** CNTPE electrode, both in the concentration range of 2 to 50 μmol L^{-1} of DA. Supporting electrolyte: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 30 mV s^{-1}.

In addition to the good sensitivity, a suitable reproducibility and response stability is required for an electrochemical sensing device. In this regard, the electrodes were submitted to consecutive redox cycles in the presence of DA (Figure *S5*). It was observed that the voltammetric profile remained nearly constant (the faradaic currents and potential values did not change significantly). It is also important to note that no fouling effect (caused by DA adsorption) was observed for the carbon paste electrodes, even after 100 cycles. This contrasts with the behavior observed for Pt electrodes, in which a blocking of the signal due to DA adsorption on the electrodic surface was observed.

To confirm that DA oxidation occurred via an electrocatalytical process when the modified electrodes were employed, the dependence of $Ipa/n^{1/2}$ on scan rate was plotted, as shown in Figure 4. According to Nicholson [41], a non-linear relationship of the $Ipa/n^{1/2}$ *vs* ν plot exhibited the typical behavior of a typical electrochemical-chemical catalytic process, as observed for the electrodes containing KP20 and KPy20.

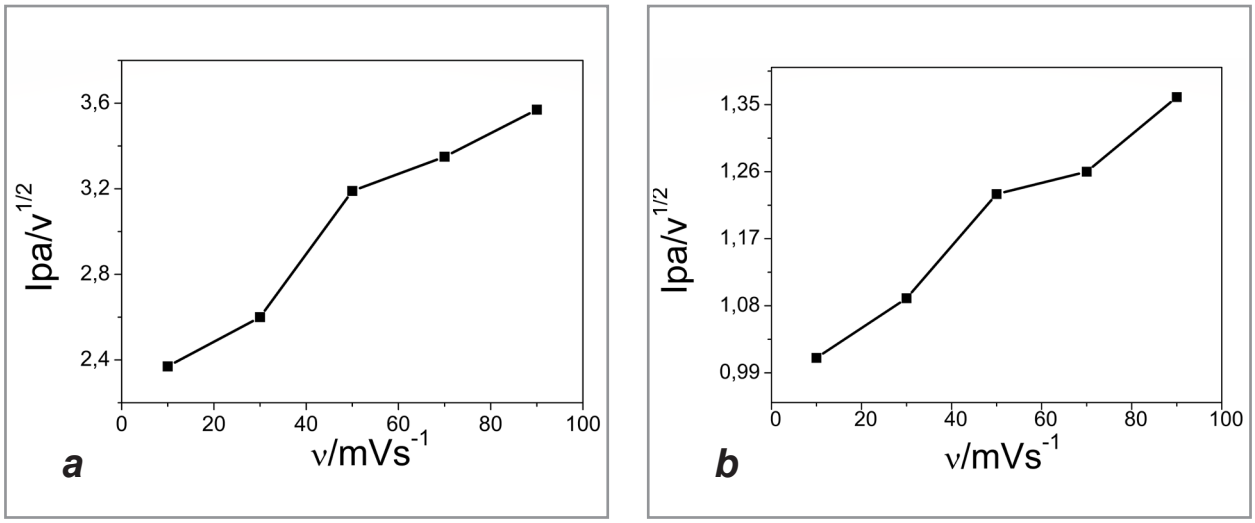


Figure 4. Plot of $Ip/\nu^{1/2}$ against scan rate (ν) for DA (50 μmol L^{-1}) at the modified electrodes ***a)*** KPPE and ***b)*** KPyPE. Supporting electrolyte: NaCl 0.5 mol L^{-1} and pH 4.0.

Electrochemical oxidation of ascorbic acid as a possible interferent for DA detection

The electrochemical response of the electrodes in the presence of ascorbic acid (AA) was evaluated due to its possible interference for physiological DA detection. AA is present along with DA in biological fluids, and its presence considerably decreases the sensitivity of most of the materials used in electrochemical sensors for DA determination [2,52-53]. This is due to the fact that the redox process of AA occurs at potentials that are close to those observed for DA oxidation [50,54].

Figure 5.a displays the cyclic voltammograms of the modified and unmodified Eeonomer® (KPPE, KPyPE and KP0PE) in the presence of a fixed concentration of 50 μmol L^{-1} AA. An irreversible oxidation peak was observed for KP0PE at Epa of *ca.* 450 mV, however, the voltammetric response for the KPPE and KPyPE showed no oxidation processes or just a small anodic peak. This is a particularly important finding for the production of selective sensors for DA quantification, due to the possibility of identification of this analyte in different complex samples. The cyclic voltammogram for the KPPE in the presence of AA is highlighted as an insert in Figure 5.a due to the high capacitive current presented by this material. For comparison, Figure 5.b shows the electrochemical behavior of the electrodes CNTPE, GRPE and KP0PE in the presence of AA. For those carbonaceous materials, it is possible to observe the same oxidation process as for the KP0PE, with a higher current peak for the GRPE, similarly to the response obtained in presence of DA.

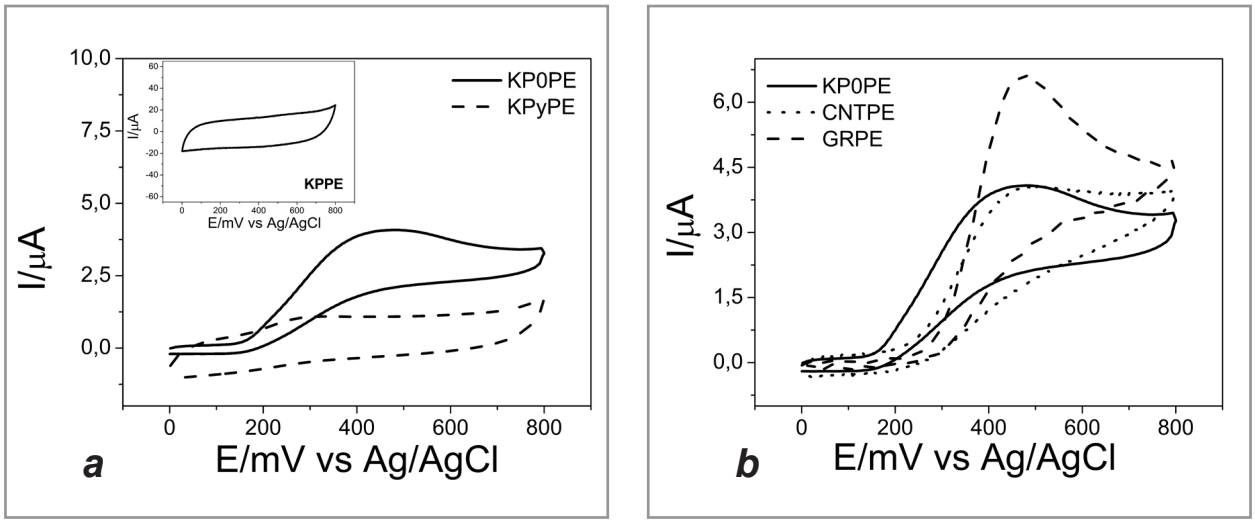


Figure 5. Cyclic voltammograms of ***a)*** KP0PE, KPPE and KPyPE and ***b)*** KP0PE, GRPE and CNTPE in presence of AA in a concentration of 50 μmol L^{-1}. Conditions: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 30 mV s^{-1}.

The obtained results indicate that the use of KP0PE, GRPE and mainly CNTPE electrodes are not suitable for the use in DA quantification in samples where AA is also present (S6). On the other hand, KPPE and KPyPE electrodes are both selective to DA determination, which is a desirable and important feature for applications in biological samples.

The information described above has showed that KP20 materials have the best performance regarding the application in the preparation of electrochemical sensors. Aiming to produce a proof of concept for the statement presented above, the quantification of DA in a commercial sample was performed by using cyclic voltammetry and standard addition methodology. The KPPE was chosen for this study due to its better sensitivity (0.11 A mol^{-1} L) for DA determination. Figure 6 shows the standard addition curve used for DA quantification. The relationship obtained between Ipa (μA) and DA concentration (μmol L^{-1}) was obtained by linear regression (Ipa = 10.23 + 0.21 [DA], R = 0.99773) and the determination of DA concentration in the sample was obtained by extrapolation to Ipa = 0. The obtained value, 50.8 μmol L^{-1}, agrees with the labeled concentration of DA in the tested commercial injection solution sample (50.4 μmol L^{-1}).

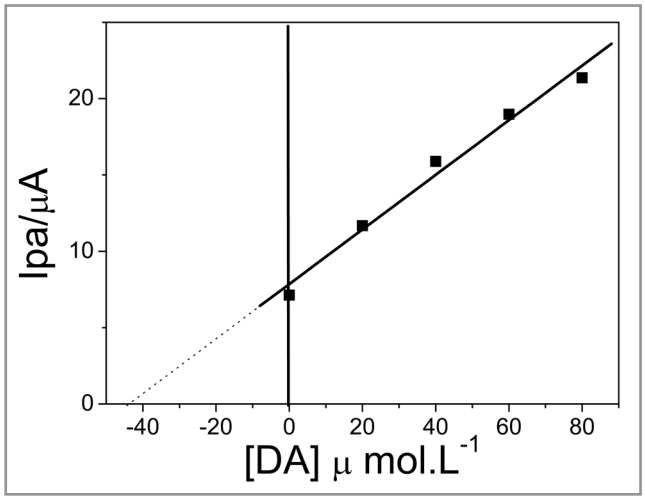


Figure 6. Ipa *vs* DA concentration plot obtained for the standard addition methodology aiming DA quantification in a commercial injection solution. Ipa values were obtained by using cyclic voltammetry and KPPE electrode in the following conditions: supporting electrolyte: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 30 mV s^{-1}.

Factorial Design

Intending to verify whether the combination of the two modified Eeonomers® could enhance the electrochemical response for DA detection, since both materials presented were able to electrocatalyze this process, a 3^2 full factorial design was performed including the mixture of KP20 and KPy20 materials for CPE construction, in different proportions, as one of the variables. The chosen second variable was pH due to the importance of this parameter in doping the conducting polymers present in the system.

Cyclic voltammograms acquired with the electrodes containing the Eeonomers® KP20 (KPPE), KPy20 (KPyPE), and a mixture of both (KP20/KPy20 50:50 w/w) in acetate buffer at pHs 3.0, 5.0 and 7.0 showed only capacitive currents. The electrochemical response of the electrodes toward AA oxidation revealed the great potential of the modified KPPE and KPyPE electrodes for DA detection, since the KPPE electrode did not show current peaks in the presence of AA, and the KPyPE could detect AA only at concentrations higher than 20 μmol L^{-1}. Thus, the Eeonomers® (KP20 and KPy20) are highly selective to DA even in the presence of its principal interferent (AA) in physiological medium.

The current results obtained in each assay of the 3^2 full factorial are displayed in the Table III. A central point assayed in triplicate was included in the factorial design for standard deviation estimate.

Table III. Current results obtained in each assay of the 3^2 full factorial design to investigate the influence of electrode composition and pH on the sensing capabilities of the developed devices

Assay	pH	CPE composition	Ipa (μA)
1	- (3.0)	- (KP20)	28.3
2	+ (7.0)	- (KP20)	15.2
3	- (3.0)	+ (KPy20)	6.64
4	+ (7.0)	+ (KPy20)	9.31
5	0 (5.0)	0 (KP20/KPy20)	24.3
6	0 (5.0)	0 (KP20/KPy20)	24.0
7	0 (5.0)	0 (KP20/KPy20)	25.4
8	0 (5.0)	- (KP20)	25.1
9	+ (7.0)	0 (KP20/KPy20)	20.4
10	0 (5.0)	+ (KPy20)	7.82
11	- (3.0)	0 (KP20/KPy20)	22.0

The main effects (pH = -5.18 and CPE composition = -13.8) indicated that both variables significantly affect the evaluated response. However, a significant interaction effect (pH *vs* CPE composition = +7.85) is also observed, which implies in a non-homogeneous effect of pH for both electrodes. For KPPE electrode, the pH effect reduces the Ipa 13.1 times, while for the KPyPE a 2.67-μA increase in the Ipa was observed. This interaction effect and the significant overall curvature of the response ($\bar{y}_f$ - $\bar{y}_c$ = -9.71 units, obtained by difference between the mean of the points of the 3^2 factorial and the mean at the center of the design) indicate that the first-degree equation is inadequate to represent the local response surface. The latter shows the need for construction of a quadratic response surface, a model that was evaluated by Analysis of Variance – ANOVA (Table IV).

Table IV. Analysis of Variance for the quadratic adjusted model **(Ipa = 24.2 – 2.00 pH – 7.47 EPC - 2.44 pH² – 7.18 EPC² + 3.94 pH x EPC)**

Source	Sum of Squares (SS)	Degrees of Freedom	Mean of Squares (MS)	F-test
Regression (R)	603.27	5	120.65	52.39
Residual (r)	11.51	5	2.31	
Lack of Fit (LF)	10.43	3	3.48	6.40
Pure Error (PE)	1.09	2	0.54	
Total	614.78	10		

% Total Explained Variance = 98.13

% Explained Variance Maximum = 99.82

ANOVA shows that the response surface satisfactorily describes the set of results in the studied range for the factors. Table IV showed that the model could explain 98.13% of the variability in the data. This value is very close to the maximum variance that could be explained (99.82%). The $F = MS_R/MS_r = 52.39 >>> F_{5,5,95\%} = 5.05$, indicating that the regression is significant. $F_{LF} = MS_{LF}/MS_{pe} = 6.40 < F_{3,2,95\%} = 19$, shows that there is no significant lack of fit in the quadratic model. Finally, the random distribution of the residues indicates the suitability of the model, as showed in the inset of Figure 7. Figure 7 shows the tridimensional plot of the model equation, which displays a plateau region where the Ipa maximum (28.3 µA) can be obtained using the pH close to 3.0 and KP20 as EPC. Regarding the determination of DA in biological samples (pH ~7.0), it can be performed with EPC/Kp20/KPy20 (Ipa = 20.4 µA). Therefore, this result shows the correct choice of factors, pH and electrode composition for the quantification of this neurotransmitter.

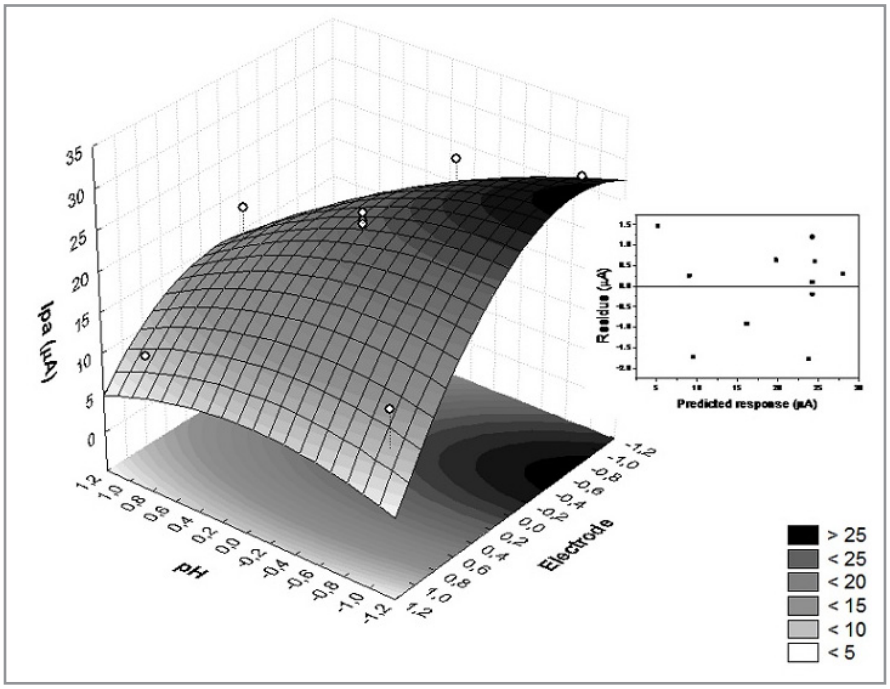


Figure 7. Response surface of the model (Ipa = 24.2 – 2.00 pH – 7.47 EPC - 2.44 pH² – 7.18 EPC² + 3.94 pH x EPC).

CONCLUSIONS

Eonomers® materials, prepared by incorporation of conducting polymers (Pani (KP20) and PPy (KPy20)) on CB, were used to prepare CPEs for the first time. These modified CPEs were investigated as candidates as electrochemical sensors for DA detection in commercial drugs. Other carbon allotropes such as GR, pristine CB (KP0) and MWCNT were also applied on CPE construction and those electrodes were evaluated toward DA detection for comparison.

The best DA detection performance was verified for CPE prepared with the KP20 Eonomer® with a sensitivity of 0.110 A mol^{-1} L at pH 4. The CPE prepared with KP20 doubled the sensitivity verified for the CPE produced with MWCNT. Besides the promising detection capacity, the prepared CPEs have presented good response stability in the presence of DA. The good analytical features of KP20 CPE enabled its successful application to detect DA in a commercial formulation.

The developed CPE were also tested in the presence of AA, since this compound is one of the most important interfering species for DA detection in physiological matrices. Both Eonomers®-based CPE showed to be selective, and no significant influence on DA determination was observed.

The response surface obtained after performing the 3^2 factorial design showed that DA voltammetric signal can be improved using the KP20 CPE at pH 3.0. However, even with lower sensitivity, the DA determination in commercial samples can also be carried out with the KP20/KPy20 (50:50 w/w) electrode at pH of 7.0, which represents a more suitable condition for the analysis of biological samples. These results showed that the electroactive modifiers (Pani and PPy) provide electrocatalytic properties for the electrodes, enhancing their electrochemical detection capabilities. Thus, our findings clearly suggest that the developed electrochemical devices can found promising applications to perform fast, cheap and sensitive determinations of DA in real samples.

Conflicts of interest

We hereby declare no financial conflicts of interest related with the information described in this text.

Acknowledgements

The authors would like to acknowledge the financial support from the following Brazilian Funding Agency: National Council for Scientific and Technological Development (CNPq), especially through the Program INCT-2014: Organic Electronic (INEO - acronym in Portuguese), [Grant n0 14/50869-6]; the financial support from Fundação Araucária (Paraná State) and Coordination for the Improvement of Higher Education Personnel (CAPES).

REFERENCES

1. Adams, R. N. *Anal. Chem*., **1958**, *30* (9), p 1576 (https://doi.org/10.1021/ac60141a600).
2. Gnahore, G. T.; Velasco-Torrijos, T.; Colleran, J. *Electrocatalysis*, **2017**, *8*, pp 459–471 (https://doi.org/10.1007/s12678-017-0402-x).
3. Matt, S. B.; Manjunatha, S.; Manjunatha, S.; Sidlingappa, D. M.; Sidlingappa, M. *ChemistrySelect*, **2019**, *4* (19), pp 5839-5844 (https://doi.org/10.1002/slct.201900642).
4. Naik, T. S. S. K.; Mwaurah, M. M.; Swamy, B. E. K. *J. Electroanal. Chem*., **2019**, *834* (1), pp 71-78 (https://doi.org/10.1016/j.jelechem.2018.12.054).
5. Tajik, S.; Beitollahi, H.; Nejad, F. G.; Safaei, M.; Zhang, K.; Le, Q. V.; Varma, R. S.; Jang, H. W.; Shokouhimehr, M. *RSC Adv*., **2020**, *10*, pp 21561–21581 (https://doi.org/10.1039/d0ra03672b).
6. Lopes, L. C.; Lima, D.; Mendes Hacke, A. C.; Schveigert, B. S.; Calaça, G. N.; Simas, F. F.; Pereira, R. P.; Iacomini, M.; Viana, A. G.; Pessôa, C. A. *Talanta,* **2020**, 121634 (https://doi.org/10.1016/j.talanta.2020.121634).
7. Švancara, I.; Kalcher, K.; Walcarius, A.; Vytras, K. *Electroanalysis with carbon paste electrodes*, CRC Pres, New York, **2012**.
8. Rajaei, M.; Foroughi, M. M.; Jahani, S.; Zandi, M. S.; Nadiki, H. H. *J. Mol. Liq*., **2019** (https://doi.org/10.1016/j.molliq.2019.03.135).

9. Pokhrel, L. R.; Ettore, N.; Jacobs, Z. L.; Zarr, A.; Weir, M. H.; Scheuerman, P. R.; Kanel, S. R.; Dubey, B. *Sci. Total Environ*., **2017**, *574*, pp 1379–1388 (https://doi.org/10.1016/j.scitotenv.2016.08.055).
10. Oularbi, L.; Turmine, M.; Salih, F. E.; El Rhazi, M. *J. Environ. Chem. Eng*., **2020**, 103774 (https://doi.org/10.1016/j.jece.2020.103774).
11. Zaidi, S. A. *Int. J. Electrochem. Sci*., **2013**, *8*, pp 11337–11355.
12. Kalambate, P. K.; Rawool, C. R.; Srivastava, A. K. *New J. Chem.,* **2017***, 41* (15), pp 7061–7072 (https://doi.org/10.1039/c7nj00101k).
13. Yang, Z.; Tian, J.; Yin, Z.; Cui, C.; Qian, W.; Wei, F. *Carbon*, **2019**, *141*, pp 467-480 (https://doi.org/10.1016/j.carbon.2018.10.010).
14. Seman, R. N. A. R.; Azam, M. A.; Mohamad, A. A. *Renewable Sustainable Energy Rev.*, **2017**, *75*, pp 644-659 (https://doi.org/10.1016/j.rser.2016.10.078).
15. Wildgoose, G. G.; Banks, C. E.; Leventis, H. C.; Compton, R. G. *Microchim. Acta*, **2006**, *152*, pp 187–214 (https://doi.org/10.1007/s00604-005-0449-x).
16. Li, C.; Thostenson, E. T.; Chou, T. W. *Compos. Sci. Technol.*, **2008**, *68* (6), pp 1227-1249 (https://doi.org/10.1016/j.compscitech.2008.01.006).
17. Oliveira, T. M. B. F.; Morais, M. *Appl. Sci*., **2018**, *8*, 1925 (https://doi.org/10.3390/app8101925)
18. Kumar, V.; Lee, G.; Monika; Choi, J.; Lee, D. J. *Polymer*, **2020**, *190* (2), 122221 (https://doi.org/10.1016/j.polymer.2020.122221).
19. Pumera, M. *Chem. Eur. J*., **2009**, *15*, pp 4970–4978 (https://doi.org/10.1002/chem.200900421).
20. Mazloum-Ardakani, M.; Beitollahi, H.; Ganipour, B.; Naeimi, H.; Nejati, M. *Bioelectrochemistry,* **2009**, *75* (1), pp 1-8 (https://doi.org/10.1016/j.bioelechem.2008.11.006).
21. Afshar, S.; Zamani, H. A.; Maleh, H. K. *J. Pharm. Biomed. Anal.*, **2020**, *188*, 113393 (https://doi.org/10.1016/j.jpba.2020.113393).
22. Bijad, M.; Maleh, H. K.; Farsi, M.; Shahidi, S. A. *Food Anal. Meth.*, **2017**, *10*, pp 3773-3780 (https://doi.org/10.1007/s12161-017-0933-z).
23. Valentini, F.; Amine, A.; Orlanducci, S.; Terranova, M. L.; Palleschi, G. *Anal. Chem*., **2003**, *75* (20), pp 5413-5421 (https://doi.org/10.1021/ac0300237).
24. Khodabakhshi, S.; Fulvio, P. F.; Andreoli, E. *Carbon*, **2020**, *162*, pp 604-649 (https://doi.org/10.1016/j.carbon.2020.02.058).
25. Arduini, F.; Cinti, S.; Mazzaracchio, V.; Scognamiglio, V.; Amine, A.; Moscone, D. *Biosens. Bioelectron*, **2020**, *156*, 112033 (https://doi.org/10.1016/j.bios.2020.112033).
26. Vicentini, F. C.; Ravanini, A. E.; Figueiredo-Filho, L. C.; Iniesta, J.; Banks, C. E.; Fatibello-Filho, O. *Electrochim. Acta*, **2015**, *157*, pp 125–133.
27. Cabral, M. F.; Pedrosa, V. A.; Suffredini, H. B.; Moreno, R. B.; Mattoso, L. H.; Gonçalves, P. S.; Machado, S. A. S. *Zaštita Materijala*, **2006**, *47* (4), pp 41-45.
28. Liu, X.; Liu, J. *View*, **2021**, *2* (1), 20200102 (https://doi.org/10.1002/VIW.20200102).
29. Siddeeg, S. M. *Int. J. Electrochem. Sci*, **2020**, 15, pp 599-612 (https://doi.org/10.20964/2020.01.61).
30. Beitollahi, H.; Safaei, M.; Tajik, S. *Anal. Bioanal. Chem. Res*., **2019**, *6* (1), pp 81-96 (https://doi.org/10.22036/ABCR.2018.142219.1229).
31. Matysik, F. M. *Anal. Bioanal. Chem*., **2003**, *375*, pp 33–35 (https://doi.org/10.1007/s00216-002-1635-x).
32. Barragan, J. T. C.; Kubota, L. T. *Electrochim. Acta*, **2020**, *341*, 136048 (https://doi.org/10.1016/j.electacta.2020.136048).
33. Deroco, P. B.; Fatibello Filho, O.; Arduini, F.; Moscone, D. *Electroanalysis*, **2019**, *31* (11), pp 2145–2154 (https://doi.org/10.1002/elan.201900042).
34. Li, Y. Y.; Kang, P.; Wang, S. Q.; Liu, Z. G.; Li, Y. X.; Guo, Z. *Sensors and Actuators*, **2021**, *327*, 128878 (https://doi.org/10.1016/j.snb.2020.128878).
35. Sajid, M.; Nazal, M. K.; Mansha, M.; Alsharaa, A.; Jillani, S. M. S.; Basheer, C. *TrAC, Trends Anal. Chem*., **2016**, *76*, pp 15-29 (https://doi.org/10.1016/j.trac.2015.09.006).
36. Mosleh, M.; Ghoreishi, S. M.; Masoum, S.; Khoobi, A. *Sens. Actuators, B*, **2018**, *272*, pp 605-611 (https://doi.org/10.1016/j.snb.2018.05.172).

37. Cuéllar, M.; Pfaffen, V.; Ortiz, P. I. *J. Electroanal. Chem.*, **2016**, *765*, pp 37-44 (https://doi.org/10.1016/j.jelechem.2015.07.050).
38. Varol, T. O.; Perk, B.; Avci, O.; Hakli, O.; Xue, C.; Li, Q.; Anik, U. *Measurement*, **2019**, *147*, 106867 (https://doi.org/10.1016/j.measurement.2019.106867).
39. Zhao, G.; Wang, H.; Liu, G.; Wang, Z. *Sens. Actuators, B*, **2016**, *235*, pp 67-73 (https://doi.org/10.1016/j.snb.2016.05.051).
40. Avlyanov, J. K. *Synth. Met.*, **1999**, *102* (1–3), pp 1272-1273 (https://doi.org/10.1016/S0379-6779(98)01468-4).
41. Domenech, S. C.; Bendo, L.; Mattos, D. J. S.; Borges Jr., N. G.; Zucolotto, V.; Mattoso, L. H. C.; Soldi, V. *Polym. Compos.*, **2009**, *30* (7), pp 897-906 (https://doi.org/10.1002/pc.20630).
42. Zucolotto, V.; Avlyanov, J.; Gregório Jr., R.; Mattoso, L. H. C. *J. Appl. Polym. Sci.*, **2004**, *94* (2), pp 553–557 (https://doi.org/10.1002/app.20952).
43. Zucolotto, V.; Avlyanov, J.; Mattoso, L. H. C. *Polym. Compos.*, **2004**, *25* (6), pp 617-621 (https://doi.org/10.1002/pc.20056).
44. Malha, S. I. R.; Mandli, J.; Ourari, A.; Amine, A. *Electroanal*., **2013**, *25*, pp 2289–2297 (https://doi.org/10.1002/elan.201300257).
45. Zhou, H.; Chen, H.; Luo, S.; Lu, G.; Wei, W.; Kuang, Y. *J. Solid. State Electrochem*., **2005**, *9*, pp 574–580 (https://doi.org/10.1007/s10008-004-0594-x).
46. Blinova, N. V.; Stejskal, J.; Trchová, M.; Prokeš̆, J.; Omastová, M. *Eur. Polym. J*., **2007**, *43*, pp 2331–2341 (https://doi.org/10.1016/j.eurpolymj.2007.03.045).
47. Hawley, M. D.; Tatawawa, S. V.; Piekarsk, S.; Adams, R. N. *J. Anal. Chem. Soc*, **1967**, *89* (2), pp 447-450.
48. Bacil, R. P.; Chen, L.; Serrano, S. H. P.; Compton, R. G. *Phys. Chem. Chem. Phys.*, **2020**, *22*, pp 607-614 (https://doi.org/10.1039/C9CP05527D).
49. Schindler, S.; Bechtold, T. *J. Electroanal. Chem.*, **2019**, *836*, pp 94-101 (https://doi.org/10.1016/j.jelechem.2019.01.069).
50. Siqueira Jr, J. R.; Gasparotto, L. H. S.; Crespilho, F. N.; Carvalho, A. J. F.; Zucolotto, V.; Oliveira Jr, O. N. *J. Physical. Chem. B*, **2006**, *110* (45), pp 22690-22694 (https://doi.org/10.1021/jp0649089).
51. Brett, A. M. O.; Brett, M. A. *Eletroquímica: princípios, métodos e aplicações*. Livraria Almedina, Coimbra, **1996.**
52. Santos, P. M.; Sandrino, B.; Moreira, T. F.; Wohnrath, K.; Nagata, N.; Pessoa, C. A. *J. Braz. Chem. Soc.*, **2007**, *18* (1), pp 93-99 (https://dx.doi.org/10.1590/S0103-50532007000100010).
53. Kim, D. S.; Kang, E. S.; Baek, S.; Choo, S. S.; Chung, Y. H.; Lee, D.; Min, J.; Kim, T. H. *Sci. Rep.*, **2018**, *8*, 14049 (https://doi.org/10.1038/s41598-018-32477-0).
54. Cheng, J.; Wang, X.; Nie, T.; Yin, L.; Wang, S.; Zhao, Y.; Wu, H.; Mei, H. *Anal. Bioanal. Chem.*, **2020**, *412*, pp 2433–2441 (https://doi.org/10.1007/s00216-020-02455-5).

SUPPORTING INFORMATION

S1 – Scanning Electron Microscopy (SEM) Morphological analysis of pristine Eeonomers® by SEM were performed on a LEICA / Stereoscan - 440 equipment and also on Zeiss equipment - DSM 960. All composites were cryogenically fractured and analyzed in the fracture region. The studies on the morphology of composites aimed to dispersion analysis of Eeonomers® particles in the elastomeric matrix (SEBS). SEM analyzes were performed for pure Eeonomers® (in powder) and in films of cryogenically fractured composites.

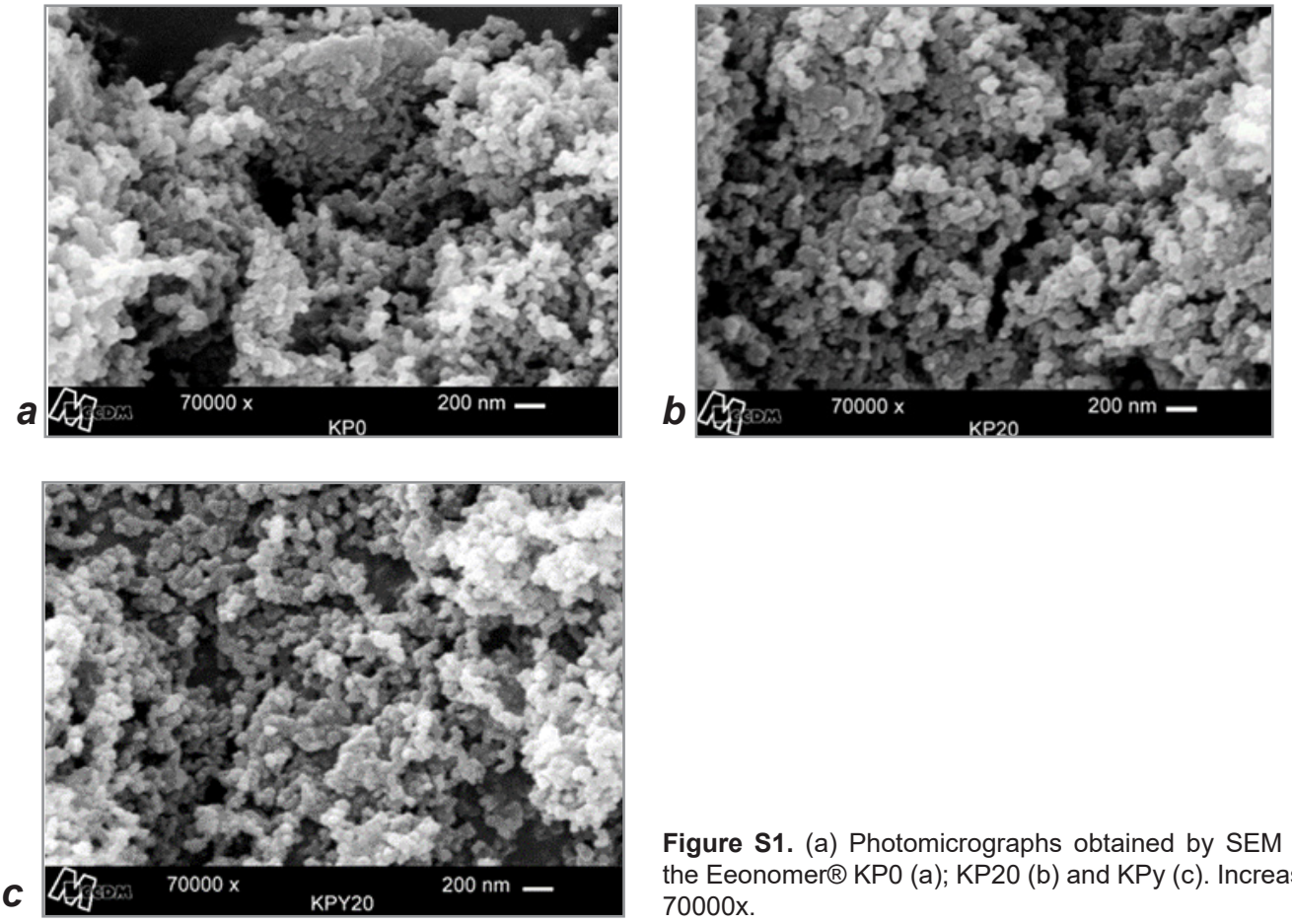


Figure S1. (a) Photomicrographs obtained by SEM from the Eeonomer® KP0 (a); KP20 (b) and KPy (c). Increase at 70000x.

S2 – Eonix Co provides Eeonomers® compounds in a fix proportion of conducting polymers, such as 3,5, 20, 40, 60, 200 %. Consequently, in this work the effects of the amount of conductive polymer on the electrochemical signals of dopamine at the carbon black electrode in the presence and absence of ascorbic acid were examined. With a low proportion of Pani or PPy, the carbon black powder is free to conduct an electric current. Since when the material contains a high proportion of polymer in its composition, the carbon black is prevented from conducting an electric current by thick layers of polymer, the which are hardly doped. Therefore, in the cyclic voltammograms performed at pH = 1.0 of KP3.5 (3.5% PANI), faradaic currents related to the redox processes of chemical doping of polyaniline in the emerald state can be observed, promoting an increase in conductivity. At KP20 (20% PANI) only an increase in capacitive current is observed and at KP40 (40% PANI), a resistive profile can be observed in addition to the capacitive current. Due to these electrochemical behaviors, KP20 was chosen to be investigated in EPC.

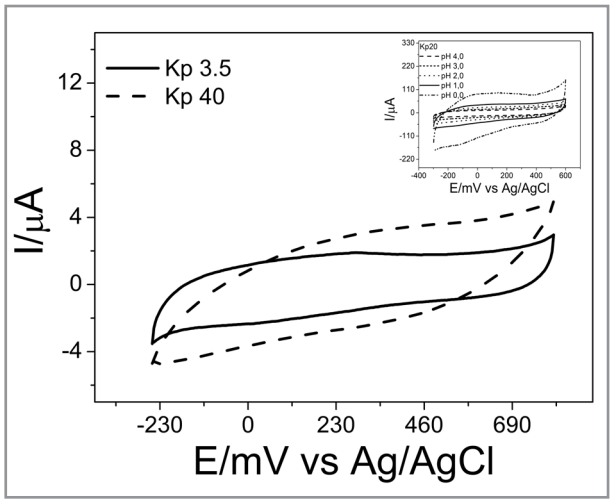
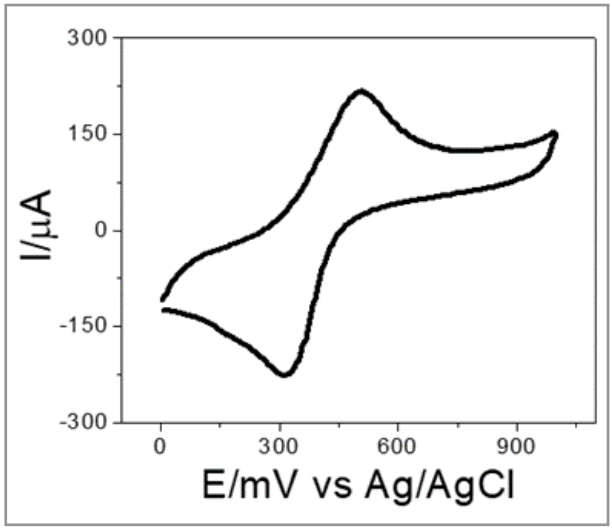


Figure S2. Cyclic voltammograms for CPE/KP3.5 and CPE/KP40 at 30 mV s^{-1} scan rate. Supporting electrolyte: NaCl 0.5 mol L^{-1} at pH 1.0. Inset: cyclic voltammograms of CPE/KP20 in different pH (0 -4.0).



S3 –The oxidation of DA on a Pt electrode occurred at higher potentials (Epa = 504 mV) in relation to electrodes modified with polymers (Kp20 and Kpy20). Another disadvantage on the use of Pt electrode is that DA adsorbs on the surface of this metal, which may lead to a fouling effect.

Figure S3. Cyclic voltammogram for the Pt electrode in 6,0 mmol.L^{-1} of DA. Supporting electrolyte: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 30 mV s^{-1}.

S4 – We observed that oxidation of DA on a Pt electrode occurred at higher potentials (Epa = 504 mV). Another disadvantage on the use of Pt electrode is that DA adsorbs on the surface of this metal, which may lead to a fouling effect.

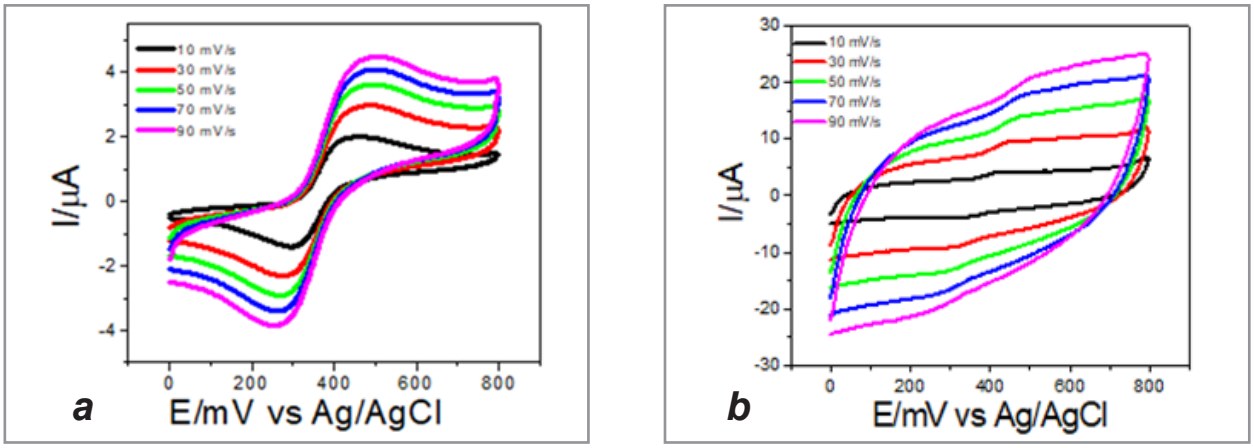


Figure S4. Cyclic voltammograms for the KP0PE (***a***), KPPE (***b***), KPyPE (***c***) in 50 μmol L^{-1} of DA. Supporting electrolyte: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 10, 30, 50. 70 and 90 mV s^{-1}.

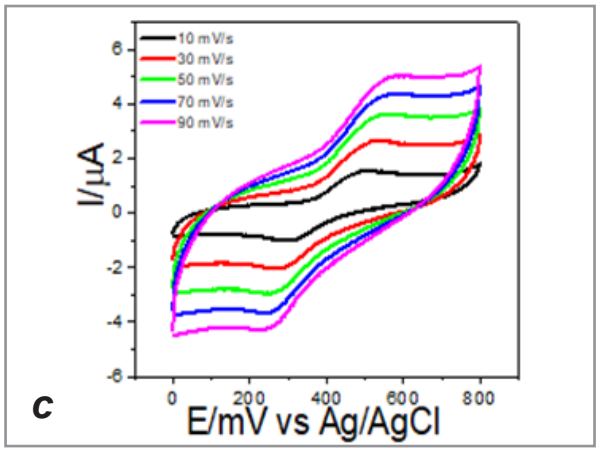


Figure S4 continuation. Cyclic voltammograms for the KP0PE (***a***), KPPE (***b***), KPyPE (***c***) in 50 μmol L^{-1} of DA. Supporting electrolyte: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 10, 30, 50, 70 and 90 mV s^{-1}.

S5 – The study of the stability of the electrodes modified with polymers (Kp20 and Kpy20) was carried out in the same way as the electrodes of GR, KP0 and NTCPM. Cyclic voltammograms and specific graphs of the correlation between Ipa and Epc *vs* the number of cycles, are illustrated.

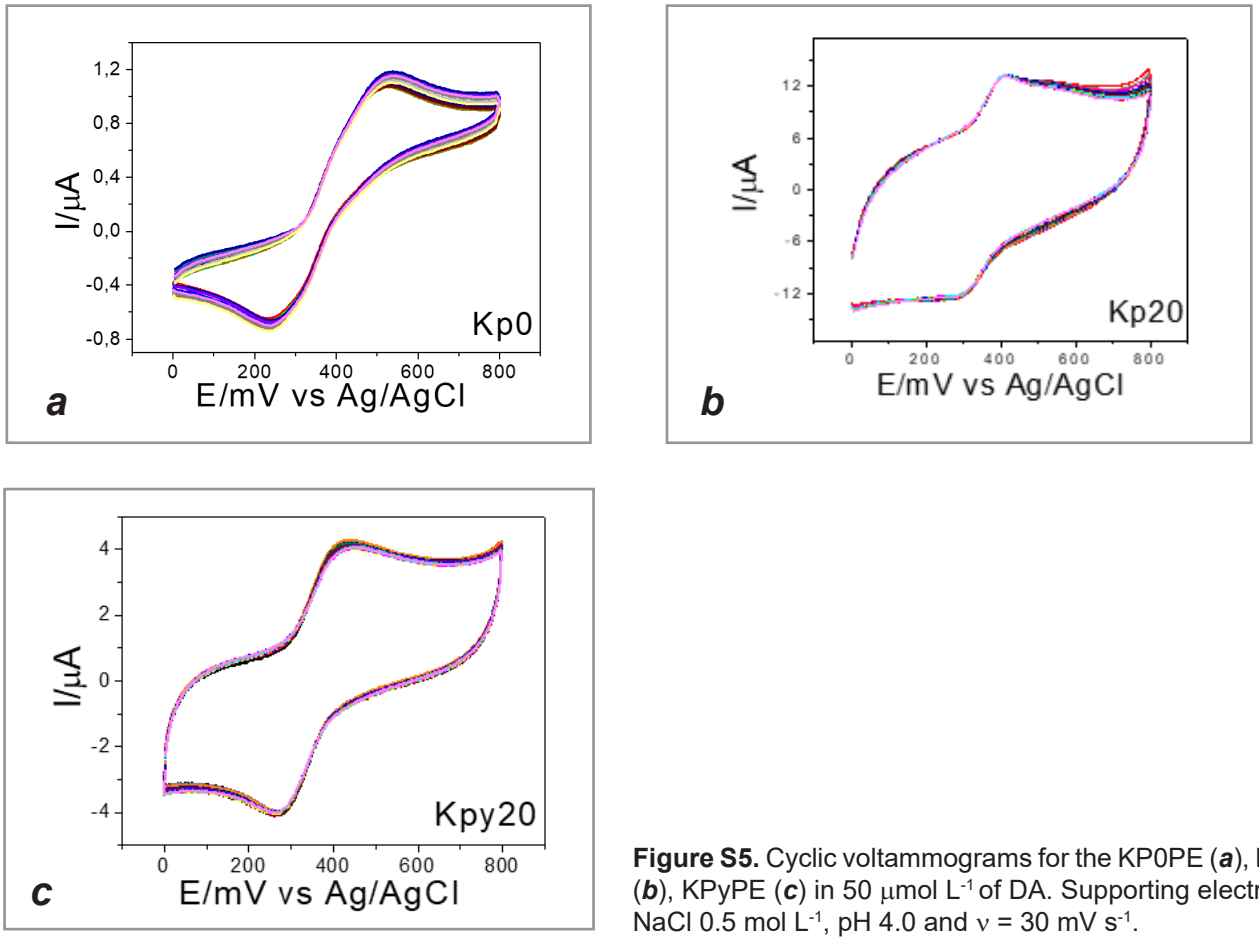


Figure S5. Cyclic voltammograms for the KP0PE (***a***), KPPE (***b***), KPyPE (***c***) in 50 μmol L^{-1} of DA. Supporting electrolyte: NaCl 0.5 mol L^{-1}, pH 4.0 and ν = 30 mV s^{-1}.

S6 – All electrodes (GR, Kp0, NTCPM, Kp20 and Kpy20) were studied individually to analyze the electrochemical behavior of each one before the simultaneous addition of dopamine and ascorbic acid. From the concentration of 1.0 mmol L^{-1}, the KP20 electrode began to detect ascorbic acid, but at this concentration, the voltammogram still showed a small faradaic current from dopamine. The other electrodes from the concentration of 0.1 mmol L^{-1} have already detected ascorbic acid.

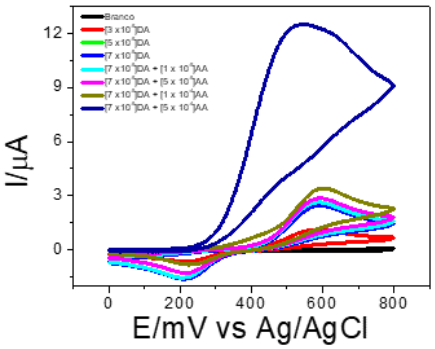
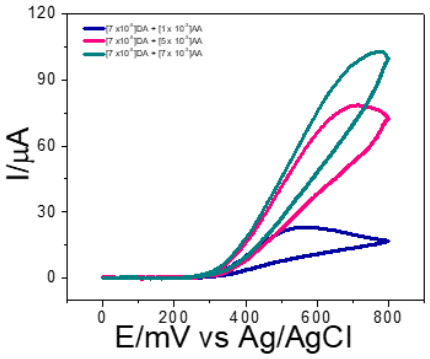
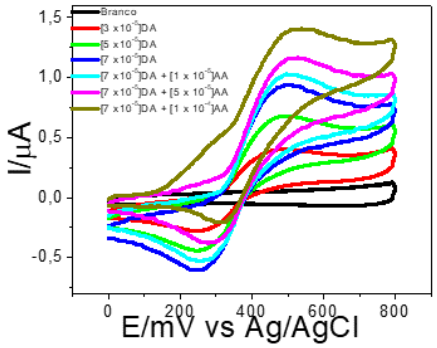
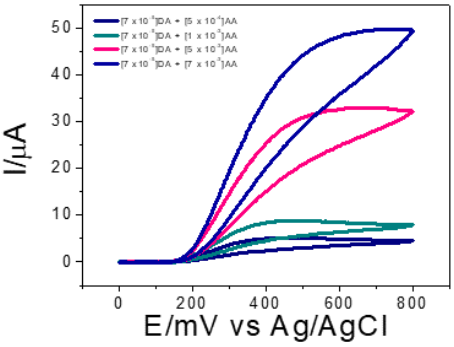
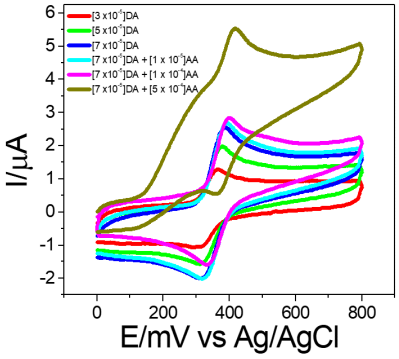
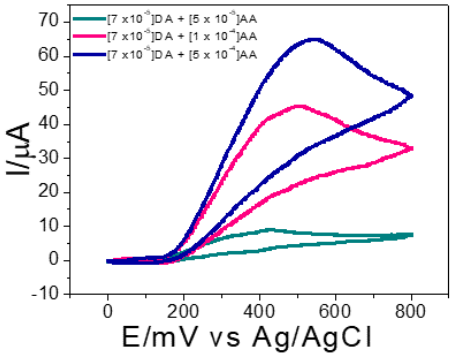
CPE	*Ipa vs E*	*Ipa vs E*
GRPE		
KP0PE		
CNTPE		

Figure S6. Cyclic voltammograms performed with KPPE, KPyPE, KP0PE, GRPE and CNTPE in DA (3.0 x 10^{-5} to 7.0 x 10^{-5} mol L^{-1}). Once the dopamine concentration was fixed at 7.0 x 10^{-5} mol L^{-1}, the concentrations of ascorbic acid varied from 1.0 x 10^{-5} mol L^{-1} to 7.0 x 10^{3} mol L^{-1}.

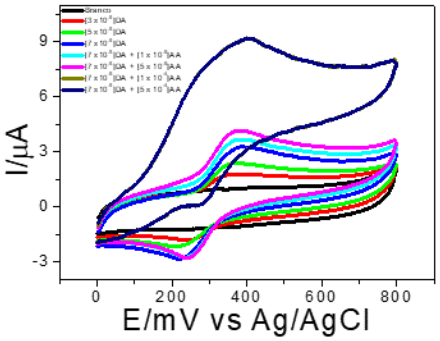
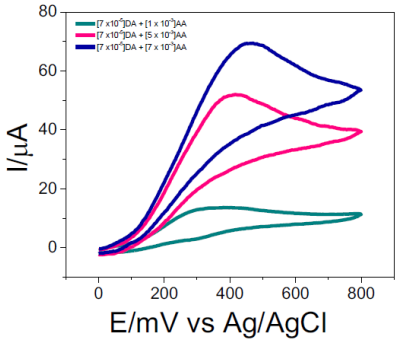
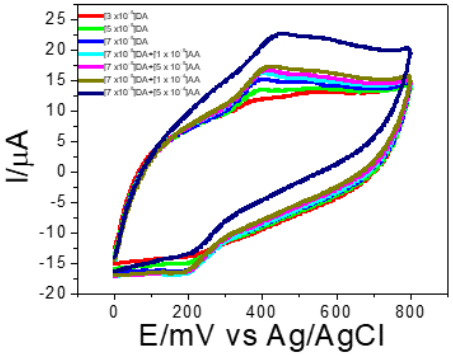
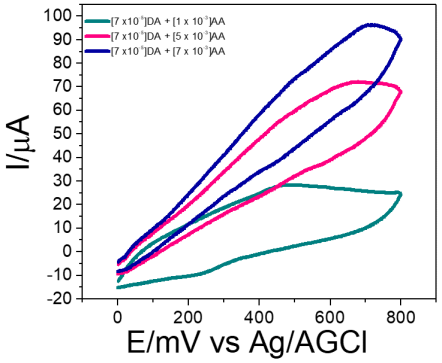
CPE	*Ipa vs E*	*Ipa vs E*
KPy20		
KP20		

Figure S6 continuation. Cyclic voltammograms performed with KPPE, KPyPE, KP0PE, GRPE and CNTPE in DA (3.0 x 10^{-5} to 7.0 x 10^{-5} mol L^{-1}). Once the dopamine concentration was fixed at 7.0 x 10^{-5} mol L^{-1}, the concentrations of ascorbic acid varied from 1.0 x 10^{-5} mol L^{-1} to 7.0 x 10^{3} mol L^{-1}.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 198-206
doi: 10.30744/brjac.2179-3425.feature-chemometrics

BrJAC

FEATURE

Chemometrics & Fast Analytics: A New Scenario in Business Intelligence

Guilherme Post Sabin[1,2*], Leandro Wang Hantao[2], Jane Kelly Finzi[3], Luiz Felipe Merenholz de Aquino[4], Nadir Hermes[5] and Cleber Oliveira Soares[6]

[1]OpenScience; [2]University of Campinas; [3]Waters Corporation; [4]Astro34; [5]British American Tobacco and [6]Brazilian Ministry of Agriculture, Livestock and Food Supply

A new analytical scenario

We are in the world of information. Data intelligence is the great asset of our time and it will be no different in our area of expertise. Analytical chemistry, driven by chemometrics, will follow the same path and provide unprecedented and essential information for the sustainability of companies.

In recent years, trends in analytical chemistry have emerged with important advances in speed, cost, intelligence and simplicity. In fact, it is a great "virtuous circle", where fast analysis generates increased analytical capacity and lower operating costs, large amounts and a high quality of data produce extraordinary databases, chemical data science opens up possibilities for augmented intelligence, and finally, in the real world, if the solution is not simple and robust, it will probably not go any further.

OpenScience works in partnership with the main players in the market. The idea is to bring the successful experience of applied research in tobacco business to leverage open innovation programs in many segments of industry, food, agribusiness and other bioeconomic issues. It is important to say that these initiatives have the support of an exceptional group of professionals who contribute to this reality, with many national and international references in analytical chemistry, chemometrics and businesses. They are researchers, professors, senior managers, R&D directors, entrepreneurs, innovation managers, government leaders, and technology providers, as well as chemical data scientists.

The next leap in chemical intelligence

The best way to move on to the new scenario is to take a step back and review some paradigms in applied analytical chemistry. Historically, we have learned to think in a univariate way: "one thing at a time". In business, it is no different; complex problems are often divided and addressed as their parts. Based on this concept, the questions demand traditional analytical methods which are selective, slow and laborious. For the same reason, they are disconnected from their original complexity, and often underestimated, after major simplifications. Consequently, what we have achieved is only a poor approximation of reality.

On the other hand, through a more holistic assessment, we realize that food, beverages and ingredients have complex chemical signatures that can provide information, with a cause/effect, of each stage of the business. The chemical composition related to food can reveal its sensory attributes, nutritional value, and special products, among others. Therefore, a deeper analysis of molecular markers can guide plant breeding, management practices and the influence of environmental factors. In addition, chemical, physical and microbiological processes leave "chemical marks" that can be measured using advanced analytical techniques and chemical intelligence.

In this way, the new approaches provide comprehensive and bidirectional analysis (i.e., top-down and bottom-up), eliminate slow and costly sample preparation steps, and can provide a large amount of chemical data. Therefore, we can transform chemical data into business information through chemical data science. Figure 1 presents the summarized scheme of the analytical approaches to business intelligence.

Cite: Sabin, G. P.; Hantao, L. W.; Finzi, J. K.; de Aquino, L. F. M.; Hermes, N.; Soares, C. O. Chemometrics & Fast Analytics: A New Scenario in Business Intelligence. *Braz. J. Anal. Chem.,* 2021, *8* (32), pp 198–206. doi: http://dx.doi.org/10.30744/brjac.2179-3425.feature-chemometrics

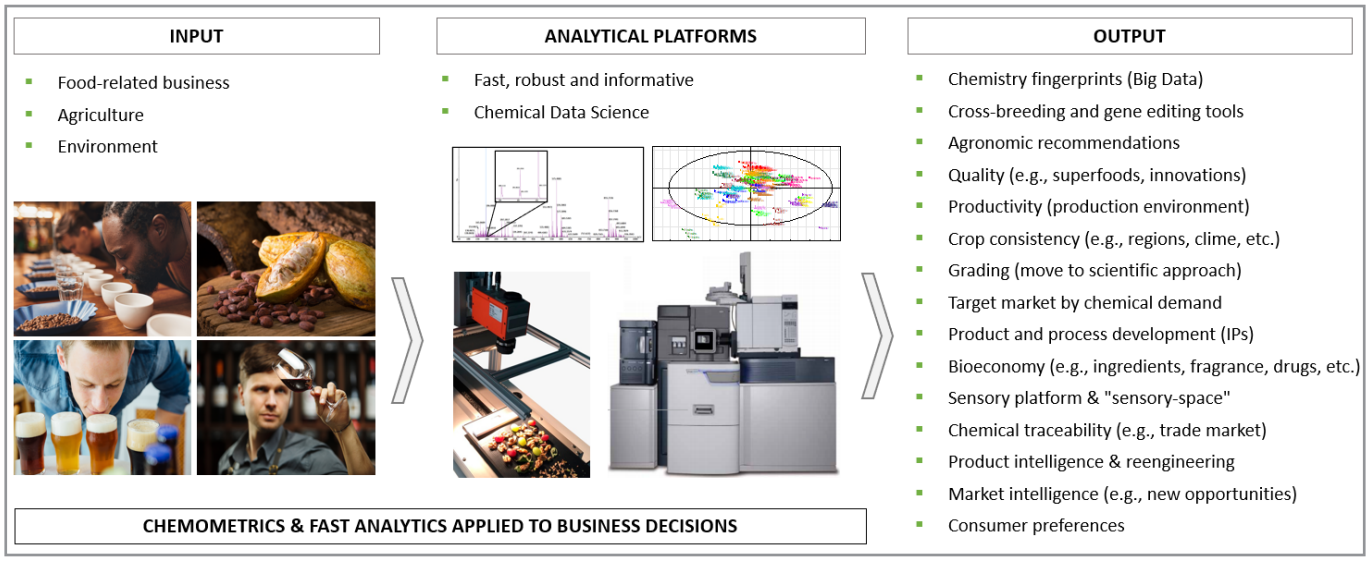


Figure 1. Analytical approaches to business intelligence.

Affording this scenario, it is important to pay attention to the level of complexity involved in choosing the technique and extracting the chemical information associated with the business objectives. It is not a simple data science work. Many current attempts are simply frustrated by the lack of technical knowledge and, mainly, by the absence of a culture of innovation and market, which makes it possible to understand the problems and their solutions in a realistic and scientifically robust way.

Open innovation in analytical chemistry

In this new scenario, our goal is to build a modern analytical chemistry, to leverage the agribusiness bioeconomy and create new action standards for food-related business. For this, we believe that analytical chemists will be the new protagonists in business intelligence. Therefore, we must be prepared to influence our business environments, through articulation between important stakeholders, connecting scientific knowledge to social demands, in a transparent and efficient manner.

In open science and innovation, it is important to be guided in an aggregating environment between the science and technology institutions, startups and big companies that seek products and services in an assertive and sustainable way. However, the success of these initiatives depends on correct managerial decisions and both public and private policies of investing in science, technology and innovation.

In the next sections, we will present some areas of opportunity that range from farming to the consumer. Furthermore, it is not just abstract, but issues that we have been privileged to innovate in the past two decades in real business scenarios.

New analytical tools in agriculture and food

The genetic potential of a crop depends on a complexity of factors which define the characteristics of the production: the phenotype. In pre-breeding work, analytical tools play an important role in the discovery of new marker compounds and in rapid screening for genetic crossbreeding. The time window for crossing is generally short and marker profiles are often complex. Thus, high-throughput tools are usually necessary. In this sense, new approaches to high resolution mass spectrometry in real-time have been emerging, which guarantee high analytical capacity with unprecedented information. An example where fast screening tools were used to patent new hybrid tobacco plants with innovative taste profile has been previously reported [1].

The concept can be used for various applications where it is necessary to find compositions or properties relevant to quality, productivity or resistance of food and ingredients to pests and diseases, and lots of

applications. These approaches are also important for monitoring new gene editing programs. In both cases, the idea is to recognize the latent variables which pinpoint the root cause for genetic improvement and, at the same time, avoid undesirable new compounds via the expression of new metabolic pathways, ensuring quality and environmental safety necessary for the expected objective.

From a scientific point of view, we can define agricultural strategies and practices based on experimental design and cause/effect relationships in management. Environmental factors, soil fertility, and other characteristics can define new production standards, treatments and processes used in the food generation. This scenario aims to use new-to-the-world chemical simulation tools based on target production, carbon sustainability and return on agricultural investments.

After defining the target metabolic profiles for production, we can apply the concept to the crop quality index. The objective is to quantify the quality of production for each farmer or group of producers. This can be used to verify the farmer's assertiveness to apply the technological package, or assess regional production trends. In the same way, it is possible to categorize raw materials and even determine the value of the product quality using objective methods of analysis. An example is the modernization of product classification guidelines, often based on sensory and subjective characteristics. In addition, we can use multivariate management tools and assess the factors that truly impact the value of commodities and return on investment for rural producers, strengthening national production. These previous concepts are some examples of the tobacco industry and patented analytical platforms [2].

Product and process intelligence

After business and scientific definitions, we can implement new tools for product inspection, traceability and certifications in real-time or at high throughput. In this case, the development of an analytical approach to estimate a wide range of quality parameters can increase the capacity of material inspection to much higher levels than currently reported. For example, the present spectroscopic imaging techniques can simultaneously monitor dozens of parameters and provide an accurate dashboard of products and processes performance. The new approaches open up opportunities for extensive sample monitoring due to low analysis costs and high analytical capacity. This is a clear opportunity for many segments. In this way, the new analytical tools are important to leverage the digital transformation and innovation in both agriculture and industry environments. The following publications from British American Tobacco describe online systems using hyperspectral imaging [3,4]. Let's see the point of view of real-time analytical technologies in the highlighted box.

Real-time analytical technologies

Hyperspectral imaging technologies are critical to accelerate business decisions on quality control and automation for Industry 4.0. According to Luiz Felipe Merenholz de Aquino, who provides this type of technology in Brazil: "chemical imaging is a mandatory tech for leading companies. In the Agro sector, we help producers of feed, food, fruits to check quality parameters in real time such as sugars, fat, moisture and many others, allowing automated machines to select and separate products according to their quality. In the pharmaceutical industry, the efficiency of the mixture of active ingredients is checked in real-time, ensuring products to meet QC standards."

However, spectroscopy may not always be satisfactory. In industry, an extremely important area is related to product and competitor intelligence. The use of advanced instrumental techniques of chemical fingerprints, combined with predictive models and databases, can determine the composition of complex formulations. In other words, the composition of products, quality of ingredients, or even the processes used can be evaluated through the chemical composition of the products. In this case, the most promising technologies (High Resolution Mass Spectrometry – HRMS) are being improved to provide incredibly fast and powerful molecular data. On the other hand, some of these approaches are described in the industry for crop quality and the monitoring of toxic compounds [5,6].

In addition to product quality or formulations, we can create objective sensory assessment tools. The subjective sensory panels, through specialists, are generally expensive, have low analytical capacity and can often expose the worker to unnecessary work risks. In some cases, they can be biased or have high subjectivity and imprecision. Thus, employing chemical fingerprint methods can dramatically increase the ability to evaluate products on a routine basis, allowing the appropriate use of specialized human resources to practice more qualified activities, for example, products with sensory innovation [7]. In addition, a deep dive in molecular knowledge was used to understand the chemical contributions and sensory characterizations. In the next paper, you can see an example of this platform [8].

Once fast and objective methods have been implemented, accurate sensory information can be used to determine the "sensory space" of a product: sensory 4.0. Through chemical signatures, products and processes can be objectively understood and protected based on chemical characteristics that define the novelty of the process or product in intellectual property [9].

Thus, it is possible to project quality through product categories, differences in origin and competitive positioning. A new area of opportunity purposes to the possibility of using the databases generated by the above technologies, to design (*in silico*) new products with desired characteristics. These approaches can drastically reduce the cycle of developing new products and bringing competitive advantages to companies. The same concept can be used to recommend best practices, processes and product formulations. An augmented intelligence point of view is shown in the highlighted box.

Augmented intelligence point of view

According to Guilherme Post Sabin, founder of OpenScience: "Artificial intelligence has been gaining ground in all areas of activities. However, our proposal goes further. We want to value human experience, empirical knowledge, the learning accumulated over the years, in the art of producing food. The proposal is to bring augmented intelligence to business, that is, to enhance the human capacity to make plans, analysis and make decisions, driven by artificial intelligence of chemical data. As protagonists on this path, we believe that our approaches based on chemical intelligence can interact to people and transform the way we produce food and do business!"

Finally, the new analytical platforms can be used for product intelligence, market mapping and consumer preferences. Once consumer preferences and market opportunities are defined, it is possible to describe the action standards in product development or reengineering. In addition, raw material management, harvest volumes and service to markets can be more accurately predicted when simulated by a chemical database suitable for the definition of commodity logistics, national brands, premium products, and trade market relationship, among others.

Analytical trends and main motivations of stakeholders

What is the future of applied analytical chemistry? This is an issue that we will address shortly. For now, let's think about the different angles from which we can look at analytical chemistry, from the point of view of the professional environment in which we are inserted in the different areas; this is the first clue to try to answer the question.

Therefore, if you are an academic researcher, a reference in the field of analytical chemistry and chemometrics, you will probably be concerned with showing the best performance in terms of speed, precision and analytical cost to obtain information with less effort and cost than is currently possible (i.e., analytical democratization). It is likely that you are focused on the technique and method used to obtain the best information, and on the analytical advantages until the moment at which you provide the result. Let's see the academic point of view in the highlighted box.

Academic point of view

According to Professor Leandro Wang Hantao, University of Campinas, recent advances in separation science have imposed new requirements for method development, which has ultimately altered the essence of contemporary analytical chemistry. Advances in column technology and multidimensional instrumentation, including high-resolution mass spectrometers, have led to the achievement of unprecedented peak capacities and fast analyses. Today, fast analyses can be routinely performed with the development of modern and powerful analytical instruments. This reality challenges the fundamental pillars of method development. In the near future, if not now, sampling and sample preparation and data processing will be the limiting steps in modern analytical methods. Today, these two steps are already considered the most time-consuming stages and responsible for most of the errors in an analytical measurement.

Please let us consider the following scenario. Considering that an analyst can obtain a chemical fingerprint in under 10 seconds, it is still feasible to consider traditional liquid-liquid and solid-liquid extractions. Researchers will need to revise the requirements for sampling and sample preparation to improve sample throughput, while preserving the accuracy and precision of the method. Today, sampling and sample preparation comprise an amazing scope of techniques and methods used for the isolation of analytes in complex matrices. Numerous formats and configurations are available, ranging from immobilized sorbent-extractions, 3D-printed sampling devices, acoustic-assisted sampling, laser assisted desorption/ionization, just to name a few. It is imperative that sample preparation bypasses the use of solvents and other consumables to support a sustainable work environment. Also, automation is key to improve productivity and feature attractive operational expenditure.

Analytical chemistry has evolved too much to be restrained for QA/QC purposes only. This field of research is the embodiment of applied chemistry and it will gain much attention in improving business intelligence in smart-industries. How are researchers going to tackle the hundreds of files generated weekly, if not daily, in their laboratory? How can we extract the most meaningful information from chemical measurements and establish statistically valid models and interpretations of such “big data”? We are incapable of quickly determining multivariate behaviors in data using univariate methods. Multivariate data analysis using chemometrics will be unavoidable in modern analytical chemistry. There are endless opportunities for chemometric processing to improve the analytical workflow in laboratories, even in the seemingly distant applications of water research, like effect-directed analysis.

Similarly, a technology company will want to produce equipment or software that is simple and robust to use and will also believe that a higher throughput and lower operating cost approach will increase the analytical capacity to measure more and better. This certainly benefits the customer. But what kind of business intelligence does the customer need? Will this company need to update its technology base? These are non-trivial questions and are worth thinking about! So, let’s see the cutting-edge technological point of view in the highlighted box.

Cutting-edge technologies point of view

According to Jane Finzi, Market Development Lead at Waters Corporation, to unlock the potential of science today, more than an instrument of excellent technical precision is necessary; we need to drive greater collaboration to help to uncover the needs of the next generation of scientific advances. With the collaboration of customers, the development of complete solutions according to the real needs of the market is faster and more efficient. Creation of new products, development of prototypes and continuous improvement of existing solutions are built with the help of customers who are in different institutions around the globe, such as Universities, Research Centers, Hospitals and Industry, with the goal of delivering benefit to the customer and society. In addition to the most diverse collaborations, Waters created Immerse, a new innovation and research laboratory, which includes the Waters Innovation Response team, whose team works closely with several researchers to provide the technology and experience necessary to make new ones advances in science a reality. In this way, new-to-the-world analytical platforms based on high resolution mass spectrometry are giving us unprecedented multidimensional information, as well as real-time possibilities for business intelligence.

However, if you are a consumer goods company or government institution linked to a country's bioeconomy, you will think about Market Intelligence. This is very clear for the food, energy and agribusiness industry. However, intelligence will also be part of the analytical service labs and in the pharmaceutical industry through PAT and QbD. Finally, this subject has already been discussed by the major laboratories of clinical analyses and personalized medicine. Let's hear the views of a senior scientist who works in a Fast Moving Consumer Goods (FMCG) industry.

FMCG's senior scientist point of view

Nadir Hermes, Senior Scientist at British American Tobacco, describes this new scenario: "The century-old tobacco industry has been transformed in the last years. Consumers increasingly have a choice between a single agricultural product (cigarette) and new categories of tobacco and nicotine products. A trigger of this transformation was the access to evolving new technologies coupled with the continuing historical commitment of tobacco product manufacturers to provide adult consumers with a wide range of potentially reduced risk products as compared to continued cigarette smoking. The new product portfolio includes vapor, tobacco free nicotine pouches and tobacco heating products that have shown a reduction in emissions of 90-95%* of certain harmful compounds found in cigarette smoke [10,11], and which have demanded, from an analytical perspective, more sensitive and robust methods, and shorter turnarounds. A new set of methods for the new categories had to be implemented, at the speed required by the pace of the innovation and scheduled product launches. The dynamic of the new market, fast technology changes and speed of change in response to consumer demand required an Analytics Strategy/Program able to identify sound innovative capabilities to quickly support all stages of product development, manufacturing and quality control. The advances in the traditional Analytical Chemistry approaches can cope with the given challenges, but not necessarily at the right pace to win in this new market. All advantages of chemometrics and fast analytics approaches to support business intelligence is clearly the way to follow."

Note:

*Comparison of smoke from a scientific standard reference cigarette (approximately 9 mg tar) and emissions from glo™ and Vype ePen™ products, in terms of the average of the 9 harmful components the World Health Organisation recommends to reduce in cigarette smoke.

Now, it is worth reflecting on the following: what kind of information can chemical intelligence bring us? For this, we need to understand the chemistry and type of information that each technique can provide to assist in this new scenario. From a technological point of view, we must increasingly migrate to top-down analytical platforms. This will have an impact on today's laboratories, which are usually bottom-up.

Furthermore, to complete the professional scenario, we cannot forget the current model of open innovation where startups, small companies and analytical intelligence consultants sit, who can operate with agility and low cost. Also, it is true that both scientists and impresarios want to act in this segment, but there is a big gap of understanding between them. It is necessary to convince investors that current investments are insufficient, and that chemical intelligence is the most scalable model that we can produce in this area. Any serious scientist who knows business will say: science takes time and is worth investing in!

Moreover, we must remember that success depends on the established metrics. Interestingly, accelerators of open innovation are increasingly parameterized "inside the box". Therefore, we need to understand the problem and propose the goals and metrics that will lead us to the solutions we need in the near future; otherwise, we will lose the race! Let's see analytical intelligence according Ministry of Agriculture, Livestock and Food Supply in the highlighted box.

Analytical intelligence - new paradigm for value chains

According to Cleber Oliveira Soares, Director of Innovation at the Ministry of Agriculture, Livestock and Supply: "For this decade, drivers powered by innovation will be levers to raise the bar for food security and humanity development on the globe. This new logic will expand the prominent position of analytical science in relation to technological, economic, social and environmental aspects. Countless value chains will have to transform.

In agriculture, food and the bioeconomy the drivers associated with sustainability supported by the analysis of the product's life cycle – from field to table – with efficient use of inputs will make a difference in the production equation. The bioeconomy and the use of advanced biology technologies will improve genetic, zootechnical and health aspects. For the same purpose, the use of biological engineering tools using RNA, DNA, proteins and subunits of molecules for vaccines and other bio-applications. Bio-based inputs will make up this contemporary arsenal enhanced by the use of probiotics, biofilms, sanitizers, enzymes, proteins to prevent pathogens and parasites, and may even be used to mitigate greenhouse gases.

Transformations that do not require digital. Increasingly, digital and analytics will make all the difference in the contemporary economic sectors. The increase in global demand for food brings another innovative perspective - advances and opportunities in food tech. Innovations in processes such as conservation, maturation, traceability, intelligent packaging and other applications will also be common in production chains and their products. A new paradigm is born supported by analytical intelligence!"

In conclusion, let's return to the initial question: what is the future of applied analytical chemistry? In this direction, there are no surprises, although someone may disagree. As in all other areas of our society: "follow the funding!" The industry will be more successful and sustainable when it knows how to use chemical intelligence! Technology companies which see this opportunity will focus on world-class comprehensive platforms. Universities will continue to discover the foundations for the greater effectiveness of new scientific and technological solutions. Finally, the business style startups will continue to evolve towards self-knowledge and maturity, discovering the confusing paths between opportunities and purpose!

Acknowledgement

Guilherme Post Sabin thanks Waters Technologies do Brasil Ltda. for the strong support, Prof. Leandro Wang Hantao for the academic partnership and Prof. Ronei Jesus Poppi (*in memoriam*), for background in chemometrics and his reference as a human being.

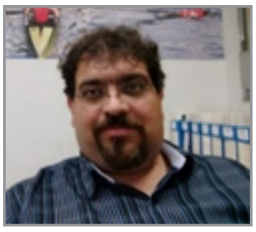
REFERENCES

1. Pontes, O. S.; Pulcinelli, C. E. US 2020/0229370 A1, **2020**.
2. Pontes, O. S.; Sabin, G. P.; da Silva, J. R. P.; Dias, J. C.; Kaiser, S. WO2018007789A1, **2018**.
3. Marcelo, M. C. A.; Soares, F. L. F.; Ardila, J. A.; Dias, J. C.; Pedó, R.; Kaiser, S.; Pontes, O. F. S.; Pulcinelli, C. E.; Sabin, G. P. *Anal. Methods*, **2019**,*11*, pp 1966-1975 (https://doi.org/10.1039/C9AY00413K).
4. Soares, F. L. F.; Marcelo, M. C. A.; Porte, L. M. F.; Pontes, O. F. S.; Kaiser, S. *Microchem. J.*, **2019**, *151*, 104225 (https://doi.org/10.1016/j.microc.2019.104225).
5. Kaiser, S.; Dias, J. C.; Ardila, J. A.; Soares, F. L. F.; Marcelo, M. C. A.; Porte, L. M. F.; Gonçalves, C.; Canova, L. S.; Pontes, O. F. S.; Sabin, G. P. *Talanta*, **2018**, *190*, pp 363–374 (https://doi.org/10.1016/j.talanta.2018.08.007).
6. Kaiser, S.; Soares, F. L. F.; Ardila, J. A.; Marcelo, M. C. A.; Dias, J. C.; Porte, L. M. F.; Gonçalves, C.; Pontes, O. F. S.; Sabin, G. P. *Chem. Res. Toxicol.*, **2018**, *31*, pp 964–973 (https://doi.org/10.1021/acs.chemrestox.8b00154).
7. Soares, F. L. F.; Marcelo, M. C. A.; Dias, J. C.; Juliano, L. C.; Porte, L. M. F.; Canova, L. S.; Ardila, J. A.; Pontes, O. F. S.; Sabin, G. P.; Kaiser, S. *J. Chemom.*, **2020**, *34*, e3297 (https://doi.org/10.1002/cem.3297).
8. Schwanz, T. G.; Bokowski, L. V. V.; Marcelo, M. C. A.; Jandrey, A. C.; Dias, J. C.; Maximiano, D. H.; Canova, L. S.; Pontes, O. F. S.; Sabin, G. P.; Kaiser, S. *Talanta*, **2019**, *202*, pp 74–89 (https://doi.org/10.1016/j.talanta.2019.04.060).

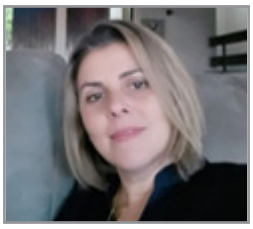
9. Benjak, D.; Field, P.; Glesse, A.; Link, M. US 2016/0249674 A1. **2016**.
10. Forster, M.; Fiebelkorn, S.; Yurteri, C.; Mariner, D.; Liu, C.; Wright, C.; McAdam, K.; Murphy, J.; Proctor, C. *Regul. Toxicol. Pharmacol*., **2017**, *93*, pp 14-33 (https://doi.org/10.1016/j.yrtph.2017.10.006).
11. Margham, J.; McAdam, K.; Forster, M.; Liu, C.; Wright, C.; Mariner, D.; Proctor, C. *Chem. Res. Toxicol*., **2016**, *29*, pp 1662-1678 (https://doi.org/10.1021/acs.chemrestox.6b00188).

Key Players and Stakeholders

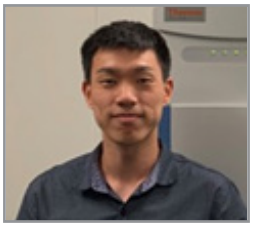
Cleber Oliveira Soares, PhD
Executive with 20 years of experience in Science, Technology & Innovation, Strategic Management, Agriculture & Business, developed in Brazilian agribusiness, with significant experience at Embrapa and IICA - Inter-American Institute for Agricultural Cooperation, currently at the Brazilian Ministry of Agriculture, Livestock and Food Supply.



Guilherme Post Sabin, PhD
Founder of OpenScience and senior scientist with 19 years' experience in analytical science for business and academic settings. In 2017, he won the "Sir Charles Ellis Award", in recognition of the best scientist in global R&D at the British American Tobacco. In 2018, he decided to work widely through chemical data science for big companies. Also, he is visiting professor at Unicamp and co-advisor for chemometrics and fast analytics applied to business intelligence.



Jane Kelly Finzi, MSc
Market Development lead at Waters, she is responsible for business development in Brazil. Demonstrated history of working in the research industry. Strong sales professional skilled in Good Laboratory Practice (GLP), Liquid Chromatography-Mass Spectrometry (LC-MS), Protein and small molecule Chemistry, Validation, and GC-MS. LC-MS Hardware knowledge and field service experience.



Leandro Wang Hantao, PhD
Professor of Chemistry at the Institute of Chemistry, State University of Campinas (IQ-UNICAMP). His studies aim to sample preparation for analysis of organic compounds, chromatographic techniques, mass spectrometry and data processing. Among the awards, the 2018 Power List "TOP 40 UNDER 40" (Analytical Scientist) and 2019 The "John Phillips Award" stand out.

Luiz Felipe Merenholz de Aquino
Entrepreneur who dedicated his life to advanced analytical technologies since his youth and has always loved to participate in high end projects and research. He has founded ASTRO34 in 2003 with the purpose of selling scientific instruments like hyperspectral cameras. Today he is dedicating his time to develop solutions to serve Agro, Pharma, and Food among others.

Nadir Hermes, PhD

Senior Scientist – PRRP* Analytical at British American Tobacco (UK). Nadir is about to complete 30 years of experience in Analytical Chemistry. He started his career working for the University of Santa Cruz do Sul (RS, Brazil), where he was technical responsible for the Analytical Centre and Professor of Analytical Chemistry in the last years. After 17 years, he joined Souza Cruz Company (BAT Brazil). In 13 years at BAT, one of the remarkable achievements was the transformation of the labs he managed or have technically supported. The views expressed in this article do not necessary represent the views of BAT.

*PRRP – Potentially Reduced Risk Product.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 207–230

BrJAC

SPONSOR REPORT

This section is dedicated for sponsor responsibility articles.

Consolidated Analysis of Soil Contaminants
Four-fold increase in the sample throughput with GC-Orbitrap

Aaron Lamb, Dominic Roberts, and Cristian Cojocariu

Thermo Fisher Scientific, Runcorn, UK

This report was extracted from the Thermo Scientific Application Note 10731

The purpose of this study was to assess the quantitative performance and advantages of PAHs and PCBs using the Thermo Scientific™ Orbitrap Exploris™ GC in addition to screening of unknown soil contaminants.

Keywords: Analytical environmental testing, polychlorinated biphenyls (PCBs), polyaromatic hydrocarbons (PAHs), persistent organic pollutants (POPs), QuEChERS, targeted quantification, screening, unknowns, gas chromatography, high resolution mass spectrometry, full scan (FS), sensitivity, Orbitrap Exploris GC, electron ionization (EI), chemical ionization (CI), Chromeleon, Compound Discoverer.

INTRODUCTION

Polychlorinated biphenyls (PCBs) and polyaromatic hydrocarbons (PAHs) are toxic organic compounds that can contaminate soils, air, sediments, and water as a result of natural and anthropogenic processes. PCBs and PAHs are resistant to environmental degradation and can be transported over long distances. Moreover, due to their lipophilicity these chemicals can undergo biomagnification and accumulation in the food chain and can pose significant health risks to humans. Their toxicity even at very low concentrations means that their presence in the environment needs to be monitored so that the risk of uptake of these compounds into to the food chain and subsequently into human populations is minimized.

More recently it has become apparent that oxidized and substituted derivatives of PAHs (such as oxy and methyl PAHs) have similar or increased toxicities compared to non-substituted versions; therefore, governments have already began monitoring them in soil and particulate matter [1,2]. Nitrogen, sulfur, and oxygen-containing polyaromatic heterocycles (NSO-PAHs) are another class of compounds that have gained interest due to their ubiquitous presence in the environment and lack of data on their toxicities [2,3].

PCBs and PAHs (and derivatives) are typically analyzed by gas chromatography (GC) coupled to mass spectrometry (MS). The challenges for the analysis of PAHs and PCBs are the requirement for complicated and costly sample preparation such as Sohxlet extraction. Often long chromatographic separations (>40 min per sample) are required, which overall will result in low sample throughput and high cost of analysis.

To comprehensively characterize an environmental sample, multiple methods are employed for both the sample preparation and GC-MS analysis of these compounds. Having multiple chromatographic methods for the same sample increases the requirement for both labor and instrumentation. Multiple methods and chemists to review the process and report the data add to the time and cost of analysis.

In this report, a consolidated approach for the rapid and cost-effective analysis of sixteen EPA PAHs, seven marker PCBs, three oxyPAHs, ten methylPAHs, and nine NSO-PAHs in soil samples using a sensitive HRMS instrument was employed. For this, a modified QuEChERS sample extraction and clean up was investigated. Chromatographic separation of target compounds was optimized for a <20 min/sample method and detection was achieved using the Orbitrap Exploris GC system.

The evaluation of system robustness and method suitability for PAH and PCB GC-MS analysis was outside of the scope of this application but is discussed in a supporting technical note (TN10728).

EXPERIMENTAL

Sample preparation

Calibration standards containing 45 native PCB, PAHs, methyl PAHs, oxyPAHs, PANHs, PASHs, and PAOHs at twelve concentration levels (Appendix 1 – Table 1), and 14 (^{13}C-labeled) internal standards (Appendix 2 – Table 2), were acquired from Fisher Scientific, AccuStandards, and Wellington Laboratories Inc. (Ontario, Canada).

For the calculation of MDLs and LOQs QuEChERS soil extract was spiked at 0.5, 1.0, 1.5, 2.5, and 5.0 pg/µL. Soil was freeze dried, homogenized, and sieved prior to a modified QuEChERS extraction and clean up procedure. A summary of the QuEChERS methodology can be seen in a recent application note (AN10720).

GC-MS analysis

An Orbitrap Exploris GC instrument equipped with the ExtractaBrite™ electron ionization source was used for this analysis. This configuration allows vent-free column changes and ionization source maintenance in under 2 minutes representing a 98% time saving versus traditional venting approaches, which take up to 4 hours. This is achieved using state of the art NeverVent technology, which increases laboratory productivity through the minimization of instrument downtime.

Liquid injections of the sample extracts were performed using a Thermo Scientific™ TriPlus™ RSH series autosampler and chromatographic separation was achieved by a Thermo Scientific™ TraceGOLD™ TG-5 SilMS 30 m × 0.25 mm i.d. × 0.25 µm film (P/N 26096-1420) capillary column. Additional details of instrument parameters are displayed in Tables 1 and 2. Full details of all consumables used can be found in the Thermo Scientific™ AppsLab™ library.

Table 1. GC conditions *Full list of consumables and instrument can be found in the AppsLab library*

TRACE 1310 GC parameters	
Injection volume (µL)	1.0
Liner	Single gooseneck with glass wool LinerGOLD™ (P/N 453A1925-UI)
Inlet (ºC)	300
Inlet module and mode	SSL, Splitless
Splitless time (min)	1.0
Split flow (mL/min)	50.0
Septum purge flow (mL/min)	5.0
Carrier gas, flow rate (mL/min)	He, 1.2
Oven temperature program	
Temperature 1 (ºC)	40
Hold time (min)	1.0
Temperature 2 (ºC)	285
Rate (ºC/min)	28
Hold time (min)	0
Temperature 3 (ºC)	305
Rate (ºC/min)	3
Hold time (min)	0
Temperature 4 (ºC)	350
Rate (ºC/min)	30
Hold time (min)	5
Total GC run time (min)	20

Table 2. Mass spectrometer conditions

Orbitrap Exploris GC EI GC-MS parameters	
Transfer line (ºC)	320
Ion source (ionization type)	ExtractaBrite (EI)
Ion source (ºC)	350
Electron energy (eV)	70
Emission current (µA)	50
Acquisition mode	Full scan (FS)
Mass range (*m/z*)	50–550
Mass resolution	60,000 (FWHM @ *m/z* 200, scan speed 7.4 Hz)
Lock mass (*m/z*)	207.03235

Orbitrap Exploris GC CI GC-MS parameters	
Transfer line (ºC)	320
Ion source (ionization type)	ExtractaBrite (PCI)
Reagent gas type	10% ammonia in methane
Flow rate (mL/min)	0.6
Ion source (ºC)	190
Electron energy (eV)	70
Emission current (µA)	100
Acquisition mode	Full scan (FS)
Mass range (*m/z*)	65–690
Mass resolution (FWHM @ *m/z* 200)	60,000 (scan speed 7.4 Hz)
Lock mass	None

Data processing

Data were acquired using full scan (FS) mode, processed, and reported using Thermo Scientific™ Chromeleon™ 7.3 chromatography data system (CDS). Additional screening of unknowns was performed using Compound Discover software. Thermo Scientific™ Compound Discoverer™ software, version 3.2, was also used for spectral deconvolution, NIST library searching, and compound identification using the EI and CI nodes.

RESULTS AND DISCUSSION

Chromatography, selectivity, and linearity were evaluated using solvent based standards. Assessment of sensitivity (as matrix detection limits and limits of quantitation), recovery, and selectivity were performed in soil using a modified QuEChERS extraction method, which is described in the experimental section.

Chromatography

All compounds were analyzed in <20 min and excellent separation of the critical pairs was obtained for the 16 EPA PAH standard (i) phenanthrene/anthracene, (ii) benzo(a)anthracene/chrysene, (iii) benzo(b) fluoranthene/benzo(k)fluoranthene (Figure 1, A-D). As expected, with fast multiresidue methods of this nature, some coelution did occur in which case the data was reported as a sum of the combined area (ex: included (i) 1-ethylnapthanalene/2-ethylnapthalene, (ii) 1,3-dimethylnapthalene/1,6-dimethylnapthalene). Due to the superior inertness of the TraceGOLD silphenylene GC columns, excellent peak shape was observed for all compounds including the strongly basic compound quinoline which had a European Pharmacopeia (EP) asymmetry value of 1.0 [3].

Due to the diversity of sample matrices with various degrees of complexity, selectivity can be challenging in GC-MS analysis of soils. An example of sample complexity is shown in Figure 1, E-F as an overlay of the TIC EI full scan of a sonicated unspiked QuEChERS soil extract (top chromatogram) and of a FS XIC (bottom chromatogram) showing the incurred residues.

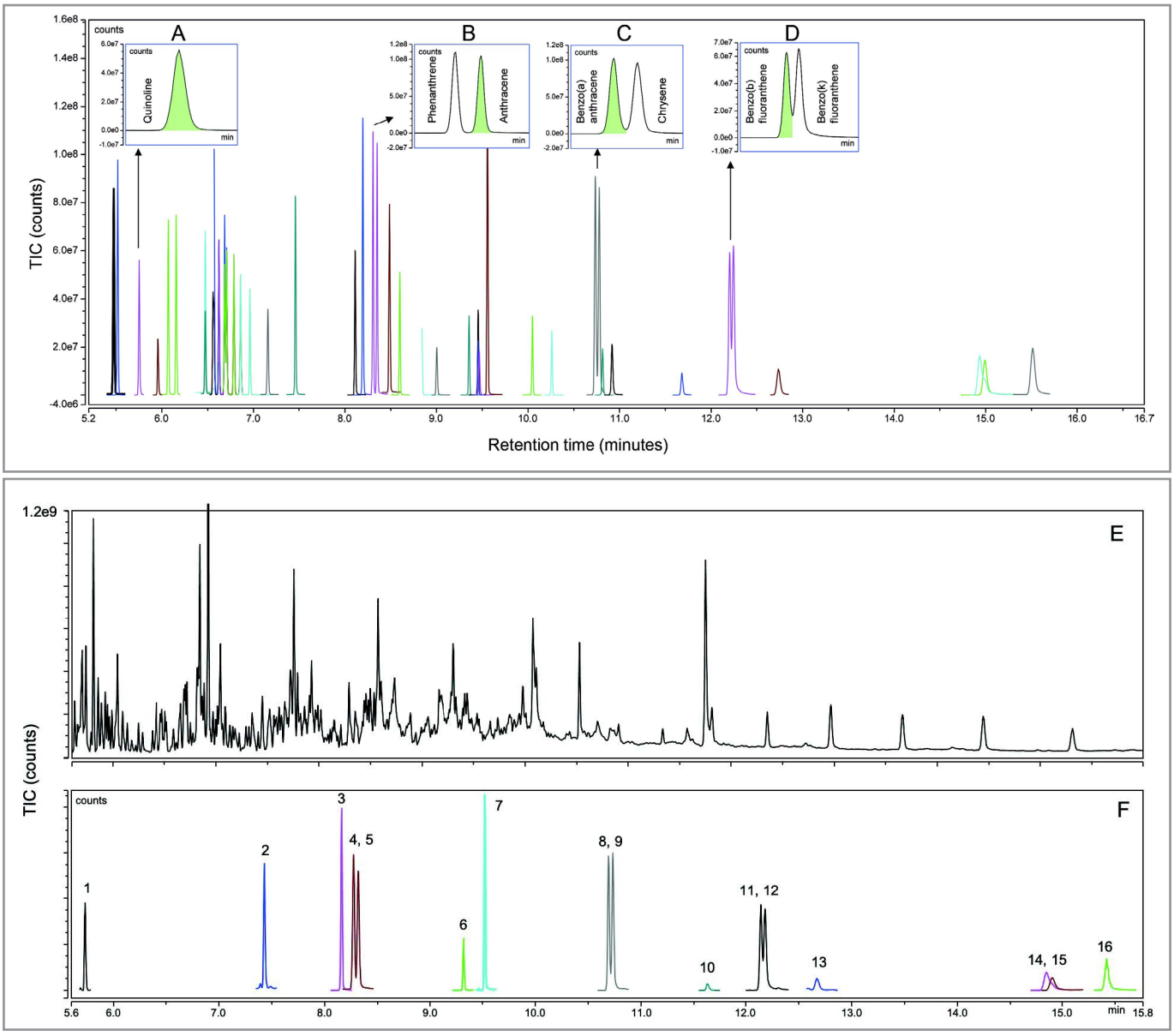


Figure 1. Example chromatograms showing overlaid native PAHs and PCBs FS XICs for a 50 pg/µL (50 pg on column (OC)) solvent standard in *n*-hexane with excellent chromatographic peak shapes for all compounds in <20 min. A) Peak shape for nitrogen containing polyaromatic heterocycle quinoline with peak asymmetry of 1.0; (B) Resolution of critical components phenanthrene and anthracene with EP resolution of 1.5; (C) Resolution of critical components benzo(a)anthracene and chrysene with chromatographic resolution of 1.3; (D) Resolution of critical components benzo(b)fluoranthene and benzo(k)fluoranthene with EP resolution of 1.0. (E) QuEChERS soil extract unspiked, FS, *m/z*=50–550; (F) QuEChERS soil extract unspiked, native incurred residue XICs; Compounds: 1= Quinoline, 2=Fluorene, 3=Dibenzothiophene, 4, 5=Phenanthrene/Anthracene 6=Fluoranthene, 7=Pyrene, 8, 9=Benzo[a]anthracene,Chrysene, 10=5,12-Napthacenequinone, 11, 12=Benzo[b/k]fluoranthene, 13=Benzo[a] pyrene, 14=Indeno[1,2,3-cd]pyrene, 15=Dibenzo[a,h]anthracene, 16=Benzo[ghi]perylene. C^{13}-labeled internal standards were not displayed to show native peak shapes clearly.

Sensitivity: determination of method detection limits (MDLs)

To practically assess the MDLs, n=18 replicate injections of the lowest serially diluted matrix-matched standard (0.5, 1.0, 2.5 pg/uL) with a peak area % RSD of <15% were used. The MDL was then calculated by considering the injected amount, peak area % RSD, and t-score of 2.567, corresponding to 17 (n-1) degrees of freedom at the 99% confidence interval (Figure 2). The MDL values calculated ranged from 118 to 475 fg on column (corresponding to 0.1–0.5 µg/kg in sample).

Sensitivity: determination of limit of quantitation (LOQ)

Method LOQs were calculated using serially diluted matrix-matched standards at 0.5, 1.0, 2.5, and 5.0 pg/µL. Eighteen (n=18) replicate injections of each of the diluted standards ranging between 0.5 pg/µL and 5.0 pg/µL were performed (0.5–5.0 µg/kg in sample) (Appendix 3 – Table 3).

The criteria used to assess individual LOQs were:

- Ion ratios within ±30% of the expected values calculated as an average across a calibration curve ranging from 0.1 to 500 pg/µL (corresponding to 0.1–500 µg/kg in sample, Figure 3)
- Peak area repeatability of <15% RSD

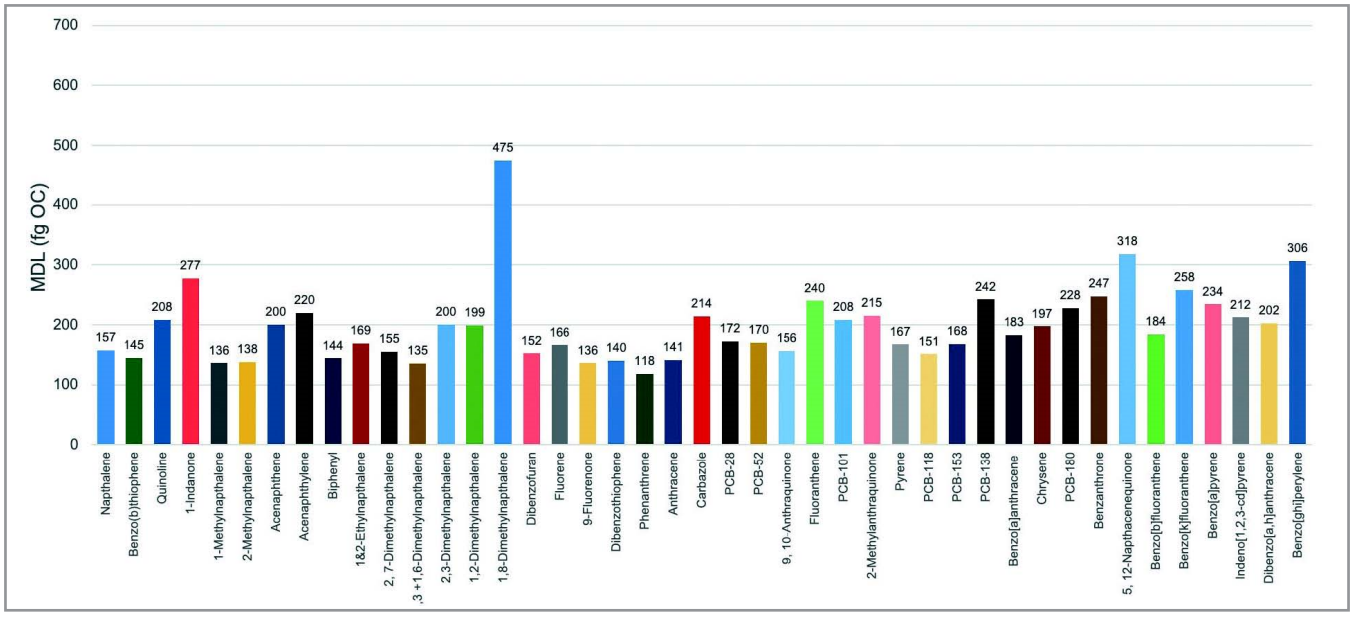


Figure 2. Graph showing individual MDLs (as detectable fg on column) for 45 native PCB, PAH, methyl PAH, oxyPAH, and NSO-PAHs calculated from n=18 replicate injections of the lowest serially diluted matrix-matched standards. *1,8-Dimethyl naphthalene 1.0 pg OC had a peak area % RSD >15% so the nearest standard 2.5 pg OC was used giving a higher MDL; however, by using a lower amount OC ~1.5 pg the true MDL value would be expected to be lower.

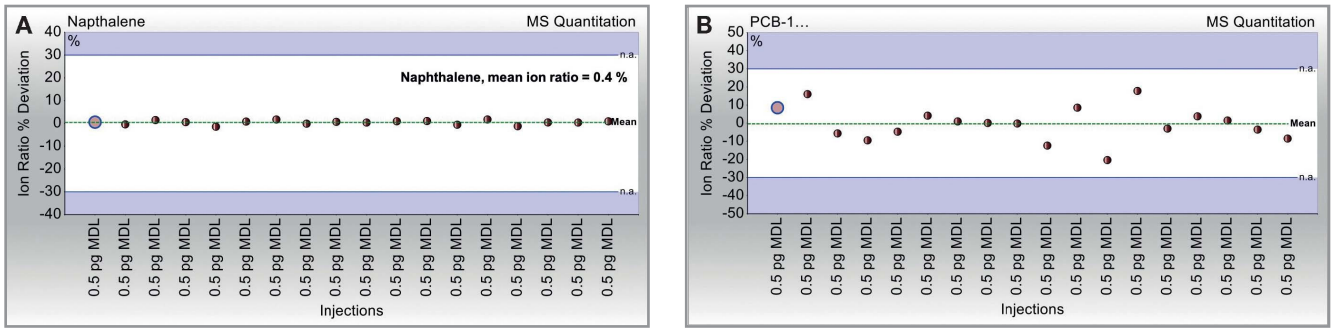


Figure 3. Graphs showing ion ratio consistency for selected PAHs and PCBs. (A) Naphthalene; (B) PCB 118, over n=18 replicate injections at the LOQ level. The average ion ratio % deviation calculated from the calibration range is displayed as a green dotted line in the center. The ±30% upper and lower ion ratio tolerance windows are also defined, and for all PAHs and PCBs the ion ratio % deviation for injections were within specification. This also illustrates how using Chromeleon CDS interactive charts allows the user to easily handle and interpret MS data.

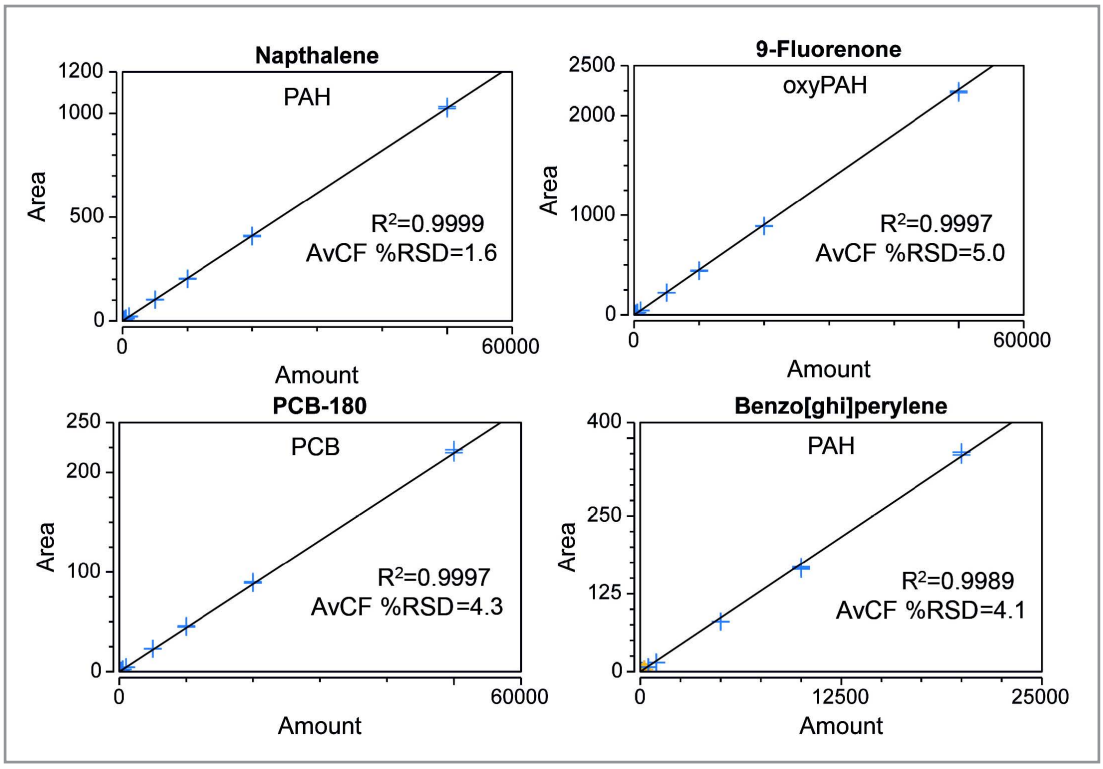
Recoveries

Seven replicate QuEChERS extractions, performed on soil spiked with deuterated internal standards at 50 ng/g added prior to extraction, were used to assess the compound recovery (details of sample preparation given in a recent application note (AN10720). Triphenyl phosphate at 100 ng/g was added post extraction and used as internal standard to adjust for potential injection variability (Appendix 5 – Table 5). All compounds show good recoveries with the average values of 79% (Appendix 5 – Table 5). Lower boiling point compounds, such as naphthalene-d8, had lower recoveries that could be explained by losses during the solvent evaporation phase. Although the recovery of such compounds is low, precision of measurement over n=7 replicate extractions was <15% RSD for all compounds and the majority being <5%. This clearly demonstrates that the QuEChERS extraction and dSPE procedure method is highly reproducible and therefore suitable for analytical testing laboratories. The total sample preparation time was <2 hours, which compared to typical Soxhlet extractions of 24–48 hours, and is a significant time (and cost) savings of 10–20×.

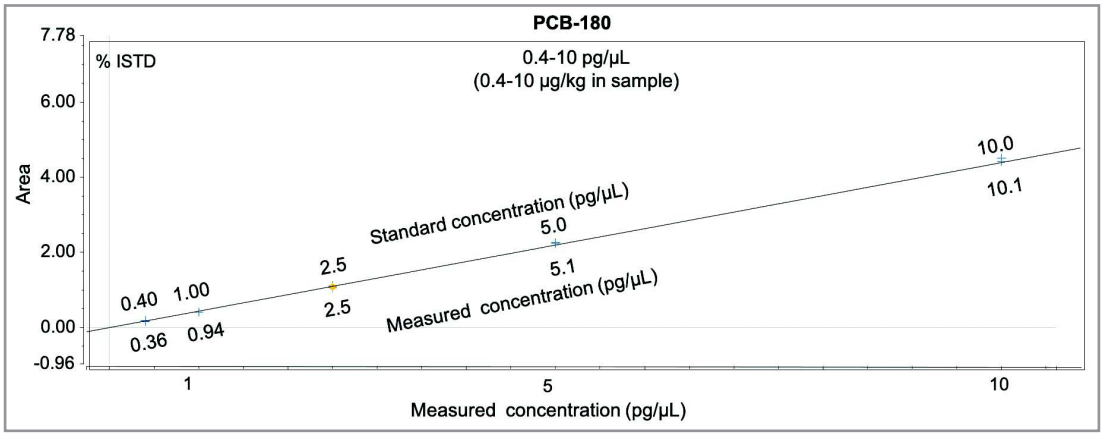
Linearity

Linearity was determined using solvent standards at concentrations 0.1–500 pg/µL. The calibration of each compound was performed using the linear/average calibration factor function in Chromeleon CDS (AvCF) over three injections at each concentration level (Figure 4).

All compounds show excellent linear responses with coefficients of determination R^2 ≥0.995, and average calibration factors %RSD across the calibration range being <13%. The R^2 values ranged from 0.9951 to 1.0000 with an average value of 0.999. (Appendix 4 – Table 4).



A

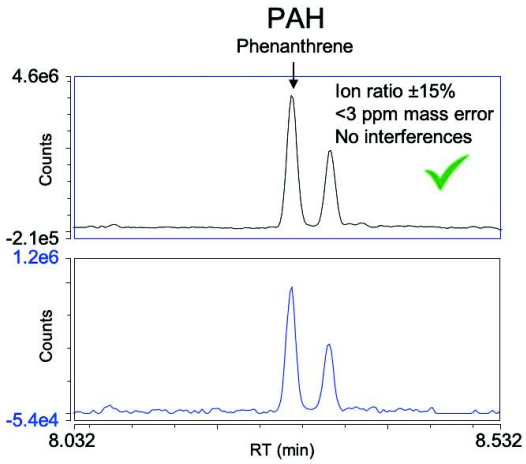


B

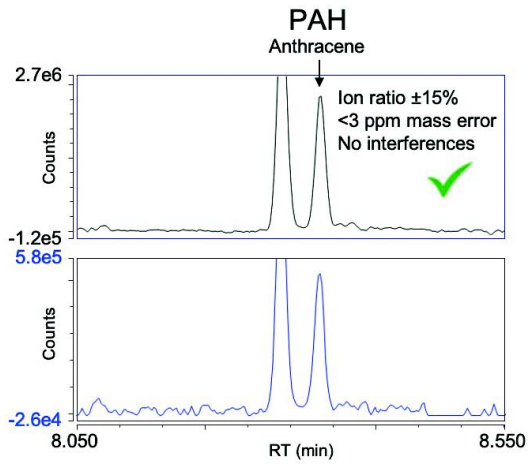
Figure 4. (A) Linearity of example PAHs and PCBs as demonstrated using solvent-based calibration curves ranging from 0.1 to 500 pg/µL (corresponding to 0.1–500 µg/kg in sample). Average calibration factor function (AvCF) was used in Chromeleon CDS and three replicate injections at each concentration with internal standard adjustment were performed. Coefficient of determination (R^2) and average calibration factor values (AvCF %RSD) are displayed. (B) A magnified region of the calibration for PCB 180 ranging from 0.4 to 10 pg/µL is shown (corresponding to 0.4–500 µg/kg in sample) showing excellent precision and accuracy for triplicate injections per point.

Quantification of PAHs and PCBs in QuEChERS soil extracts

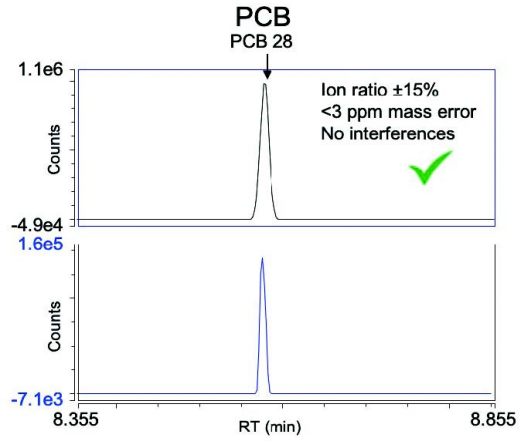
Soil samples, extracted as described in AN10720, were analyzed for their native incurred residues. The quantitative performance of the method in terms of sensitivity and selectivity is highlighted below with examples of low level native incurred residues (Figure 5).



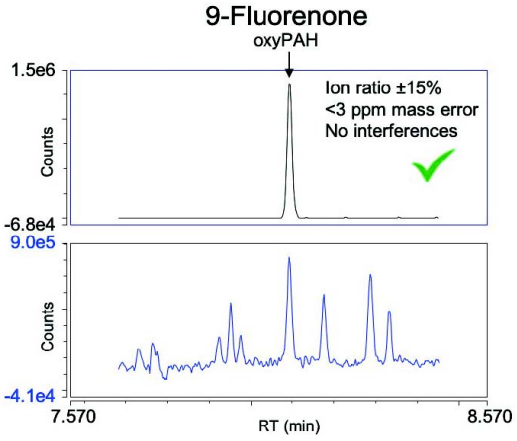
Amount found in sample	1.3 µg/kg
Ion ratio % deviation	0.1%
Measured mass (m/z)	178.07771
Theoretical mass (m/z)	178.07770
Chemical formula	$C_{14}H_{10}$
Mass error (ppm)	0.1



Amount found in sample	0.7 µg/kg
Ion ratio % deviation	5.6%
Measured mass (m/z)	178.07779
Theoretical mass (m/z)	178.07770
Chemical formula	$C_{14}H_{10}$
Mass error (ppm)	0.5



Amount found in sample	0.6 µg/kg
Ion ratio % deviation	0.7%
Measured mass (m/z)	255.96074
Theoretical mass (m/z)	255.96078
Chemical formula	$C_{12}H_7Cl_3$
Mass error (ppm)	0.2



Amount found in sample	0.8 µg/kg
Ion ratio % deviation	4.8%
Measured mass (m/z)	180.05698
Theoretical mass (m/z)	180.05697
Chemical formula	$C_{13}H_8O$
Mass error (ppm)	0.1

Figure 5. Examples of FS XIC chromatograms (quantification in black, and confirmation ions in blue) for phenanthrene in soil (top left), anthracene in soil (top right), PCB-28 in soil (bottom left), and 9-fluroenone in soil (bottom right). Below each of the FS XIC chromatograms the following is annotated: (i) amount found in sample as µg/kg, (ii) ion ratio deviation from the calibration average, (iii) measured mass (*m/z*), (iv) theoretical mass (*m/z*), (v) chemical formula, and (vi) mass error (ppm).

In summary, the results obtained in these experiments demonstrate that a consolidated compound class method using a modified QuEChERS sample preparation can be used to quantify PAHs and PCBs in soils. In the case of PCB-28, low levels of incurred residues of 0.6 µg/kg were detected and quantified within an ion ratio deviation from the calibration of only 0.7% and a mass error of the theoretical exact mass of 0.2 ppm with minimal matrix interferences all while in FS.

Screening for additional soil contaminants

The advantages of acquiring data in FS with high resolution and accurate mass were leveraged through retrospective analysis of samples and additional screening of unknown contaminants with confirmation by chemical ionization (CI). The Compound Discoverer platform includes a streamlined workflow for GC EI data allowing for extraction, deconvolution, and putative identification of the unknowns based on mass spectral library matching (NIST 2017). The software first performed spectral deconvolution above a customizable signal to noise (S/N) followed by compounds detection and grouping to consider compounds that elute at the same retention time (within ±6 s window). The deconvoluted spectra were then searched against mass spectral libraries (such as NIST), and the hits were scored based on the total score derived from a combination of library search index (SI) score and presence/absence of the molecular ions as well as percentage of fragment ions that can be explained from the NIST elemental composition. The use of a retention index acquired under the same conditions used for sample analysis helped to increase the confidence in compound identification. Compounds detected with NIST SI scores >750 can be seen in (Figure 6A). With the Compound Discoverer browser an overlaid XIC of the peak eluting at 10.95 min (*m/z* 136.07579) was identified as the top hit versus NIST library (Figure 6B). The peak was putatively identified as pyriproxyfen with a SI score of 953; however, the molecular ion of *m/z* 321.135945 was not observed, which demonstrates the requirement for additional chemical ionization and mass accuracy confirmation of molecular ions. Full results of the EI NIST matches for the deconvoluted data can be found in the (Appendix 6 – Table 6).

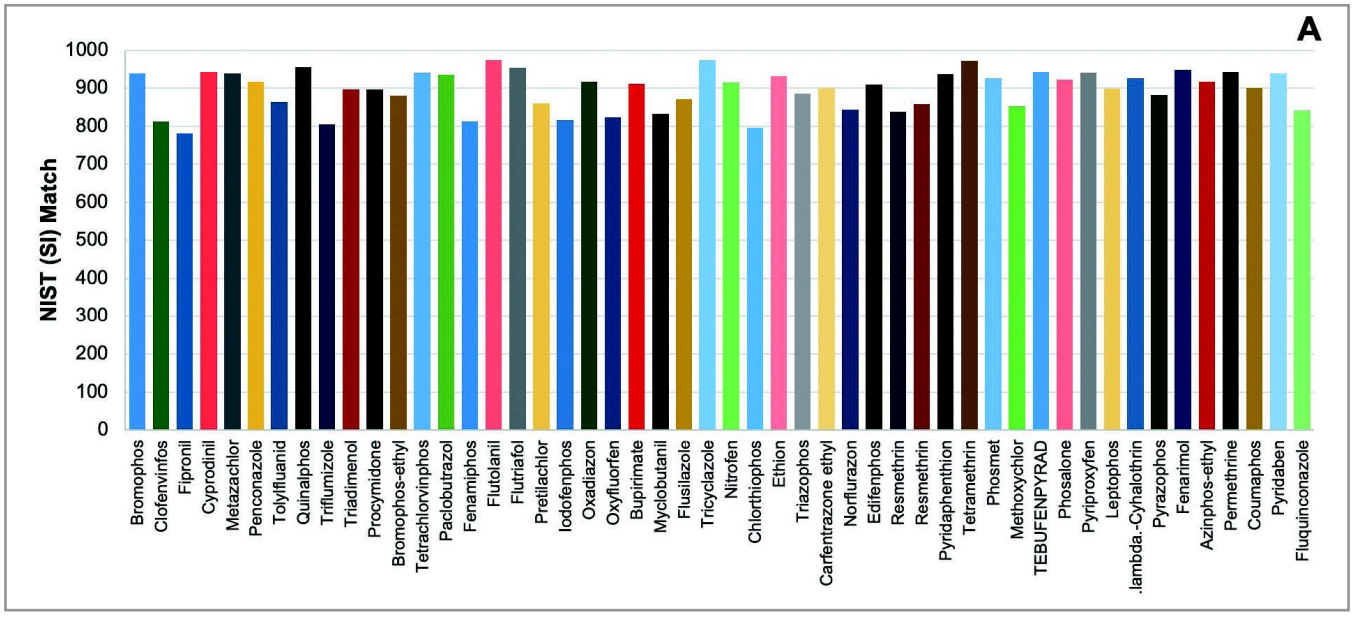


Figure 6. (A) Example NIST SI match scores for compounds detected in the deconvoluted EI spectra QuEChERS soil extract spiked at 100 pg/µL. (B) Compound Discoverer software EI spectrum of a spiked QuEChERS soil extract – deconvoluted versus NIST library of the peak eluting at 10.95 min (*m/z* 136.07579), with the structure of the top SI match pyriproxyfen from the result table.

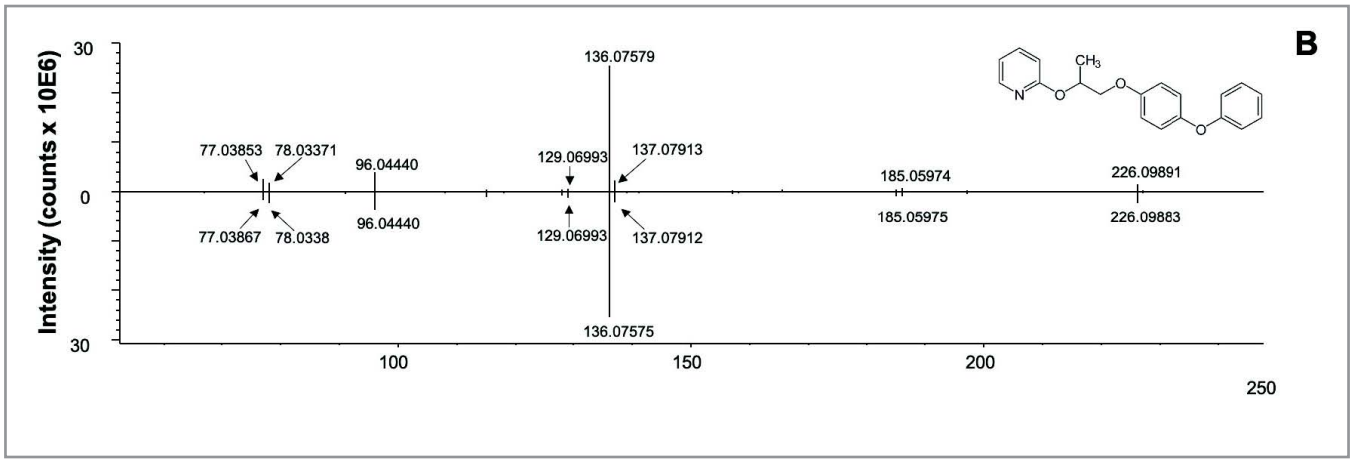


Figure 6 continued. (A) Example NIST SI match scores for compounds detected in the deconvoluted EI spectra QuEChERS soil extract spiked at 100 pg/µL. (B) Compound Discoverer software EI spectrum of a spiked QuEChERS soil extract – deconvoluted versus NIST library of the peak eluting at 10.95 min (*m/z* 136.07579), with the structure of the top SI match pyriproxyfen from the result table.

Full-scan data for blank and spiked QuEChERS soil extracts (100 pg/µL) were analyzed with Compound Discoverer 3.2 software for putative identification of peaks. A complete workflow was used to identify compounds with a high degree of confidence using deconvoluted EI spectra based on search index scores (SI) and confirmation of the corresponding molecular ion and or adducts using positive chemical ionization. FS data was acquired using Chromeleon 7.3 in EI and PCI modes at 60,000 FWHM resolution and then imported in Compound Discoverer 3.2 software. The software was used to deconvolute, align, and filter the peaks to putatively identify the compounds using mass spectral library match (NIST 17). The power of the deconvolution algorithms become clear when overlaying both the FS TIC and deconvoluted spectra for analytes eluting in a crowded area of the chromatogram (Figure 7).

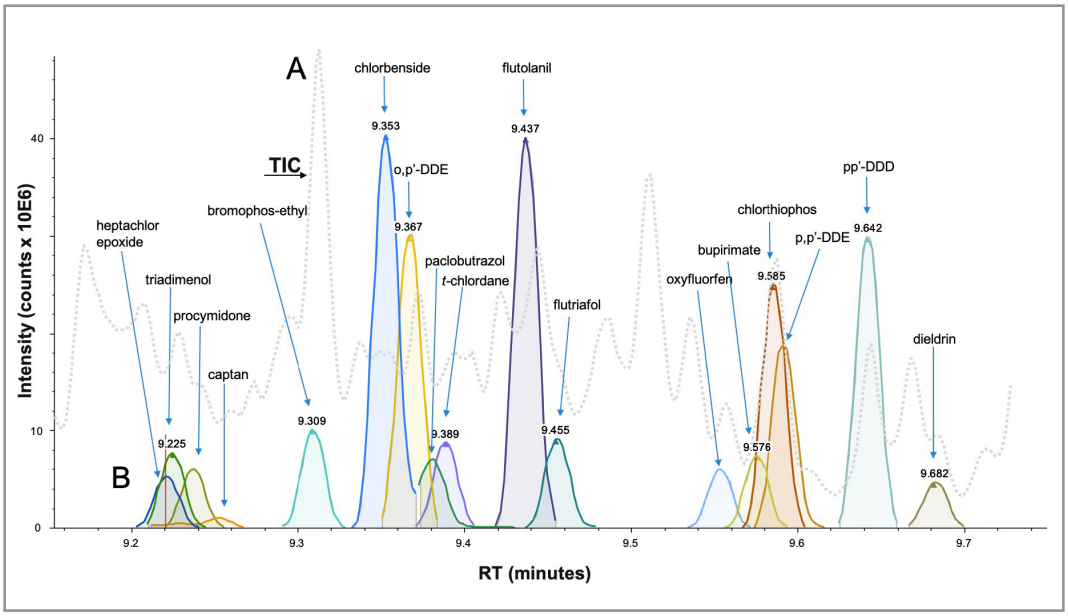


Figure 7. (A) Overlaid FS (*m/z* = 50–550) TIC for a soil QuEChERS extract spiked with pesticides at 100 pg/µL. (B) Compound Discoverer 3.2 software deconvoluted EI spectrum showing closely eluting compounds extracted from the complex TIC FS data.

Confirmation of suspect contaminants using positive chemical ionization

Further confirmation used in the identification of compounds was achieved by assessing the PCI spectra to identify the elemental composition of the parent ion by looking at common adducts. In PCI experiments using methane as the reagent gas, three adducts are typically observed: [M+H]+, [M+C2H5]+, and [M+C3H5]+. An example was shown for a peak at 9.44 min, which was identified as flutolanil versus the NIST library; however, the molecular ion *m/z* 323.11276 was not giving a significant response (Figure 8 A). When looking at the PCI data for this compound a significant boost in the molecular ion was observed with minimal mass error of 0.09 ppm (Figure 13B). Two additional adducts [M+H]+ and [M+C2H5]+ were also observed with ppm mass errors of -0.1 and -0.03 ppm respectively. Full results of the PCI confirmation can be found in Appendix 7 – Table 7.

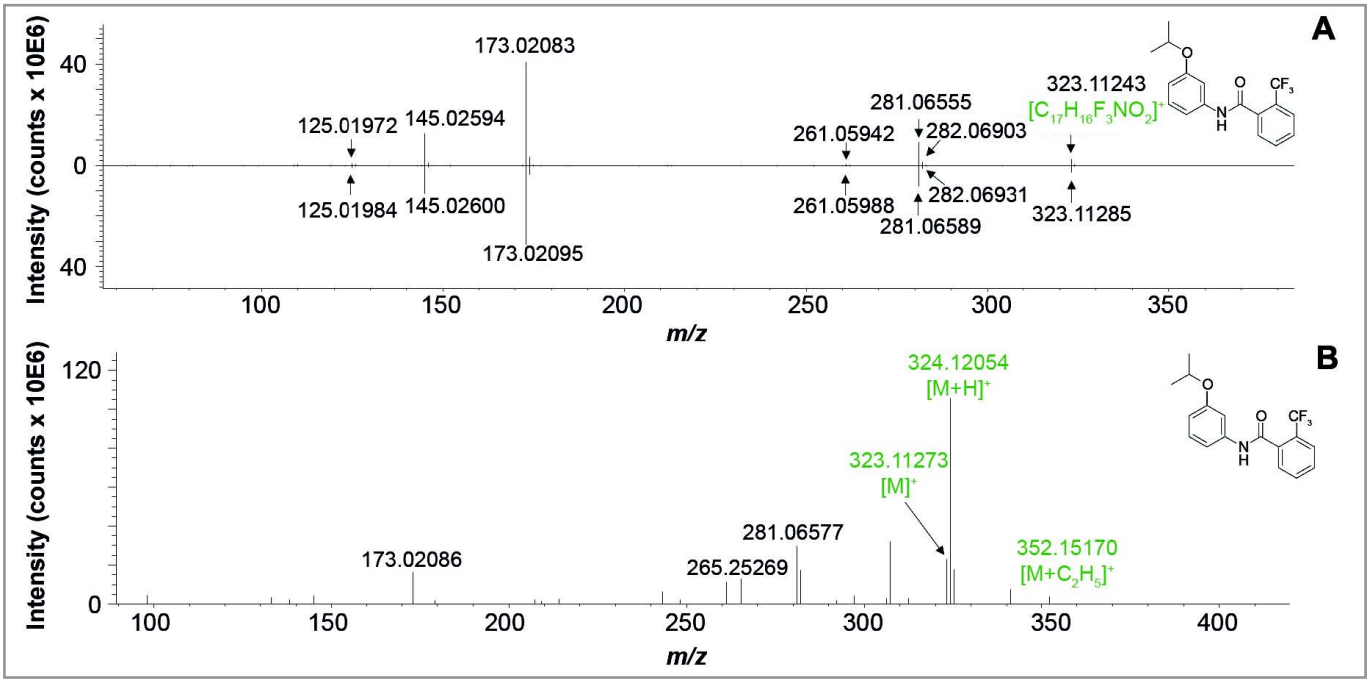


Figure 8. (A) Compound Discoverer software EI spectrum of a spiked QuEChERS soil extract – deconvoluted versus NIST library of the peak eluting at 9.437 min (*m/z* 323.11243), with the structure from the top SI match flutolanil from the result table. (B) PCI mass spectrum for flutolanil displaying adducts $[M+H]^+$ and $[M+C_2H_5]^+$ used for confirmation of this compound in conjunction with the EI data.

CONCLUSIONS

The results of the experiments presented demonstrate that modified QuEChERS methods and the TriPlus RSH autosampler in combination with the Orbitrap Exploris GC provides an ideal solution for analytical testing laboratories looking to improve productivity and deliver confident results.

- Comprehensive method consolidation with chromatographic separation and overall analytical performance was achieved for the analysis of PAHs and PCBs in soil in <20 min.
- Increased throughput of up to 20× can be realized by using a modified QuEChERS method compared to tradition Sohxlet extraction methods, saving cost and time.
- Femtogram level sensitivity was achieved using the Orbitrap Exploris GC, with the MDLs values calculated for 45 native compounds ranging from 115 to 475 fg OC (corresponding to 0.1–0.5 µg/kg in sample).

- LOQs ranged from 0.5 to 5.0 µg/kg in soil as determined from n=18 repeat injections of the lowest serially diluted matrix-matched standard that satisfied the acceptance criteria defined below:
  - Ion ratios within ±30% of the expected values calculated as an average across a calibration curve ranging from 0.1 to 500 pg/µL (equivalent to 0.4–5.0 µg/kg in sample)
  - Peak area repeatability of <15% RSD
- Linearity was achieved across a calibration range of 0.1–500 pg/µL (corresponding to 0.1–500 µg/kg in soil) showed coefficient of determination values of R^2 ≥0.995, and residuals <13%.
- All compounds show good recovery overall with the average internal standard recovery being 79%, and precision of the seven technical replicate extractions <5% RSD.
- Quantitative performance with soil samples was excellent as demonstrated by the closeness of the ion ratios and mass error compared to expected values when used for confirmation of low-level incurred residues in soil such as PAHs, PCBs, and oxyPAHs.
- Rapid change-over from EI (for spectral library search) to softer ionization such as PCI (for molecular ion confirmation using adduct information) is possible.
- The streamlined GC-EI data processing workflow with Compound Discoverer software allows for quick extraction, deconvolution, and identification of unknown compounds.

REFERENCES

1. European Committee for Standardization. CEN/TR 16998. Ambient air – *Report on nitro- and oxy-PAHs - Origin, toxicity, concentrations and measurement methods*, **2016**.
2. Sun, Z.; Zhu, Y.; Zhuo, S.; Liu, W. Z.; Zheng, E. Y.; Wang, X.; Xing, B.; Tao, S. Occurrence of nitro- and oxy-PAHs in agricultural soils in eastern China and excess lifetime cancer risks from human exposure through soil ingestion. *Environ Int*., **2017**, *108*, pp 261-270.
3. Anderson, J. T.; Achten, C. Time to say Goodbye to the 16 EPA PAHs? Toward an up-to-date use of PACs for environmental purposes. *Polycyclic Aromatic Compounds*, **2015**, *35*, pp 330–354.
4. Thermo Fisher Scientific. Application note 21736, **2017**: *Ultra-inert low bleed GC columns with advanced silphenylene polymer technology*. Available at: https://assets.thermofisher.com/TFS-Assets/CMD/Application-Notes/an-21735-gc-ms-ultra-inert-columns-an21735-en.pdf
5. Hites, R. A. Polybrominated diphenyl ethers in the environment and in people: a meta-analysis of concentrations. *Environ. Sci. Technology*, **2004**, *38*, p 945.
6. UN Environment Programme. Stockholm Convention. *Guidance for the inventory of polybrominated diphenyl ethers (PBDEs) listed under the Stockholm Convention on POPs*. Available at: http://chm.pops.int/Implementation/NationalImplementationPlans/Guidance/GuidancefortheinventoryofPBDEs/tabid/3171/Default.aspx [Accessed May 8, 2018].
7. Fernandes, A.; White, S.; D'Silva, K.; Rose, M. Simultaneous determination of PCDDs, PCDFs, PCBs and PBDEs in food. *Talanta*, **2004**, *63*, pp 1147–1155.
8. Thermo Fisher Scientific. Technical note 51797, **2009**: *Timed SRM: improved capabilities for multi-target compound analysis*. Available at: https://tools.thermofisher.com/content/sfs/brochures/AN51797-TimedSRM-Multitarget-Compound-Analysis.pdf

Appendix 1 – Table 1. Details of 45 native compounds analyzed, including compound type, CAS number, and calibration range

Native standard	Compound type	CAS Number	Calibration range (ng/mL)
Napthalene	PAH	91-20-3	
Benzo(b)thiophene	PASH	95-15-8	
Quinoline	PANH	91-22-5	
1-Indanone	PAOH	83-33-0	
2-Methylnapthalene	methylPAH	91-57-6	
1-Methylnapthalene	methylPAH	90-12-0	
Biphenyl	aromatic	92-52-4	
Acenaphthylene	PAH	208-96-8	
1-Ethylnapthalene	methylPAH	1127-76-0	
2-Ethylnapthalene	methylPAH	939-27-5	
Acenaphthene	PAH	83-32-9	
2,7-Dimethylnapthalene	methylPAH	582-16-1	
1,3-Dimethylnapthalene	methylPAH	575-41-7	
1,6-Dimethylnapthalene	methylPAH	575-43-9	
2,3-Dimethylnapthalene	methylPAH	581-40-8	
1,2-Dimethylnapthalene	methylPAH	573-98-8	
1,8-Dimethylnapthalene	methylPAH	569-41-5	
Dibenzofuran	PAOH	132-64-9	
Fluorene	PAH	86-73-7	
9-Fluorenone	oxyPAH	486-25-9	
Dibenzothiophene	PASH	132-65-0	
Phenanthrene	PAH	85-01-8	
Anthracene	PAH	120-12-7	0.1–500
Carbazole	PAOH	86-74-8	
PCB-28	PCB	7012-37-5	
PCB-52	PCB	35693-99-3	
9,10-Anthraquinone	PAOH	84-65-1	
Fluoranthene	PAH	206-44-0	
PCB-101	PCB	37680-73-2	
2-Methylanthraquinone	PAOH	84-54-8	
Pyrene	PAH	129-00-0	
PCB-118	PCB	31508-00-6	
PCB-153	PCB	35065-27-1	
PCB-138	PCB	35065-28-2	
Benzo[a]anthracene	PAH	56-55-3	
Chrysene	PAH	218-01-9	
PCB-180	PCB	35065-29-3	
Benzanthrone	oxyPAH	82-05-3	
5,12-Napthacenequinone	oxyPAH	1090-13-7	
Benzo[b]fluoranthene	PAH	205-99-2	

Table 1 continued. Details of 45 native compounds analyzed, including compound type, CAS number, and calibration range

Native standard	Compound type	CAS Number	Calibration range (ng/mL)
Benzo[k]fluoranthene	PAH	207-08-9	
Benzo[a]pyrene	PAH	50-32-8	
Indeno[1,2,3-cd]pyrene	PAH	193-39-5	0.1–500
Dibenzo[a,h]anthracene	PAH	53-70-3	
Benzo[ghi]perylene	PAH	191-24-2	

Appendix 2 – Table 2. Details of the 14 internal standards, including compound type, CAS number, and concentration (suffix "L" indicates mass-labeled)

Internal standard	Compound type	CAS Number	Concentration (ng/mL)
Napthalene-d-8	PAH	1146-65-2	
Dibenzofuran-d8	PAOH	93952-04-6	
9-Fluorenone-d8	oxyPAH	137219-34-2	
Pyrene-d-10	PAH	1718-52-1	
PCB-28L	PCB	7012-37-5	
PCB-52L	PCB	35693-99-3	
PCB-101L	PCB	37680-73-2	100
PCB-118L	PCB	31508-00-6	
PCB-153L	PCB	35065-27-1	
PCB-138L	PCB	35065-28-2	
PCB-180L	PCB	35065-29-3	
Quinoline-d7	PANH	34071-94-8	
o-Terphenyl	aromatic	84-15-1	
Perylene-d-12	PAH	1520-96-3	

Appendix 3 – Table 3. Method LOQs were determined from the lowest serially diluted spiked QuEChERS extract solution prepared as detailed in the experimental section, which pass the criteria. Eighteen replicate injections of each of the diluted standards ranging between 0.5 pg/µL and 5.0 pg/µL were performed. The criteria used to assess individual LOQs were (i) measured ion ratio (IR) ±30% compared to the target ion ratio calculated from the average ion ratio across the calibration range and (ii) peak area <15 % RSD.

Compound	Injected amount (pg OC)	Min IR % dev	Max IR % dev	Mean IR % dev	Peak area % RSD	LOQ (pg OC)	LOQ (µg/kg)
Napthalene	0.5	-1.4	1.7	0.4	4.5%	0.5	0.5
Benzo(b)thiophene	1.0	-0.8	-13.3	1.0	5.7%	1.0	1.0
Quinoline	1.0	10.9	-0.7	1.0	8.1%	1.0	1.0
1-Indanone	2.5	-10.9	13.0	2.6	6.2%	2.5	2.5
1-Methylnapthalene	0.5	6.9	10.3	8.8	2.2%	0.5	0.5
2-Methylnapthalene	0.5	4.6	7.3	6.0	2.2%	0.5	0.5
Acenaphthene	0.5	-5.8	13.5	4.4	5.6%	0.5	0.5
Acenaphthylene	0.5	-10.2	14.7	2.8	6.6%	0.5	0.5

Table 3 continued. Method LOQs were determined from the lowest serially diluted spiked QuEChERS extract solution prepared as detailed in the experimental section, which pass the criteria. Eighteen replicate injections of each of the diluted standards ranging between 0.5 pg/µL and 5.0 pg/µL were performed. The criteria used to assess individual LOQs were (i) measured ion ratio (IR) ±30% compared to the target ion ratio calculated from the average ion ratio across the calibration range and (ii) peak area <15 % RSD.

Compound	Injected amount (pg OC)	Min IR % dev	Max IR % dev	Mean IR % dev	Peak area % RSD	LOQ (pg OC)	LOQ (µg/kg)
Biphenyl	0.5	-12.0	1.6	-5.2	3.6%	0.5	0.5
1 & 2-Ethylnapthalene	0.5	-9.1	9.9	3.0	2.9%	0.5	0.5
2, 7-Dimethylnapthalene	0.5	5.6	14.7	10.0	2.2%	0.5	0.5
1,3 & 1,6-Dimethylnapthalene	0.5	10.4	17.5	13.7	3.3%	0.5	0.5
2,3-Dimethylnapthalene	0.5	-13.3	13.2	0.3	10.2%	0.5	0.5
1,2-Dimethylnapthalene	0.5	-24.2	16.1	-12.1	5.8%	0.5	0.5
1,8-Dimethylnapthalene	2.5	-12.5	11.3	-3.1	3.2%	2.5	2.5
Dibenzofuran	0.5	-13.9	2.1	-5.4	2.0%	0.5	0.5
Fluorene	1.0	11.1	5.4	1.0	6.5%	1.0	1.0
9-Fluorenone	1.0	9.1	-2.3	1.0	5.3%	1.0	1.0
Dibenzothiophene	1.0	9.7	-4.4	1.0	5.4%	1.0	1.0
Phenanthrene	0.5	-13.3	2.9	-5.6	2.6%	0.5	0.5
Anthracene	1.0	11.5	0.8	1.0	5.5%	1.0	1.0
Carbazole	1.0	11.3	-1.0	1.0	8.3%	1.0	1.0
PCB-28	2.5	-11.7	9.5	-3.6	3.7%	2.5	2.5
PCB-52	1.0	12.1	0.4	1.0	6.6%	1.0	1.0
9,10-Anthraquinone	1.0	14.6	8.8	1.0	6.1%	1.0	1.0
Fluoranthene	0.5	-11.3	13.7	4.5	6.7%	0.5	0.5
PCB-101	0.5	-27.9	15.3	-12.9	8.3%	0.5	0.5
2-Methylanthraquinone	1.0	13.5	-1.4	1.0	8.4%	1.0	1.0
Pyrene	0.5	-12.2	2.6	-4.1	3.6%	0.5	0.5
PCB-118	0.5	-20.4	17.8	-0.3	7.0%	0.5	0.5
PCB-153	1.0	19.5	1.5	1.0	6.6%	1.0	1.0
PCB-138	1.0	13.2	1.2	1.0	9.4%	1.0	1.0
Benzo[a]anthracene	1.0	6.8	-0.6	1.0	7.1%	1.0	1.0
Chrysene	1.0	10.7	0.0	1.0	7.7%	1.0	1.0
PCB-180	0.5	-24.2	24.7	-4.5	9.5%	0.5	0.5
Benzanthrone	2.5	-12.3	12.0	1.3	6.3%	2.5	2.5
5,12-Napthacenequinone	2.5	-12.2	12.8	0.4	8.1%	2.5	2.5
Benzo[b]fluoranthene	1.0	6.3	-2.7	1.0	7.2%	1.0	1.0
Benzo[k]fluoranthene	1.0	7.6	-3.9	1.0	10.1%	1.0	1.0
Benzo[a]pyrene	1.0	8.3	-3.1	1.0	9.1%	1.0	1.0
Indeno[1,2,3-cd]pyrene	1.0	-7.8	-14.6	1.0	8.3%	1.0	1.0
Dibenzo[a,h]anthracene	2.5	-6.4	9.6	1.4	4.6%	2.5	2.5
Benzo[ghi]perylene	2.5	-9.9	11.8	-0.2	3.0%	2.5	2.5

Appendix 4 – Table 4. Coefficient of determination (R2) and residual average response factor (% RSD)

Compound	Compound type	R^2	AVCF % RSD
Naphthalene	PAH	0.9999	1.6
Acenaphthylene	PAH	0.9987	5.4
Acenaphthene	PAH	0.9995	4.0
Biphenyl	PAH	0.9998	2.6
Fluorene	PAH	0.9981	9.0
Phenanthrene	PAH	0.9995	3.8
Anthracene	PAH	0.9981	4.3
Fluoranthene	PAH	0.9998	3.0
Pyrene	PAH	0.9997	3.2
Benzo[a]anthracene	PAH	0.9999	1.7
Chrysene	PAH	0.9997	3.1
Benzo[b]fluoranthene	PAH	0.9998	2.6
Benzo[k]fluoranthene	PAH	0.9994	4.5
Benzo[a]pyrene	PAH	0.9987	5.4
Indeno[1,2,3-cd]pyrene	PAH	0.9964	9.3
Dibenzo[a,h]anthracene	PAH	0.9978	7.3
Benzo[ghi]perylene	PAH	0.9989	5.1
1-Methylnapthalene	methylPAH	1.0000	1.1
2-Methylnapthalene	methylPAH	0.9999	1.8
2,7-Dimethylnapthalene	methylPAH	0.9999	1.5
1,3 & 1,6-Dimethylnapthalene	methylPAH	0.9999	2.0
2,3-Dimethylnapthalene	methylPAH	0.9999	1.8
1,2-Dimethylnapthalene	methylPAH	0.9993	4.5
1,8-Dimethylnapthalene	methylPAH	0.9998	2.6
PCB-28	PCB	0.9997	2.5
PCB-52	PCB	0.9991	2.8
PCB-101	PCB	0.9998	3.2
PCB-118	PCB	0.9998	3.7
PCB-153	PCB	0.9998	1.6
PCB-138	PCB	0.9991	2.8
PCB-180	PCB	0.9997	4.3
Benzo(b)thiophene	PASH	0.9998	3.2
Dibenzothiophene	PASH	0.9988	3.7
1 & 2-Ethylnapthalene	ethylPAH	0.9996	3.7
Quinoline	PANH	0.9988	4.0
1-Indanone	PAOH	0.9993	4.7
Dibenzofuran	PAOH	0.9993	5.3
Carbazole	PAOH	0.9980	4.7
9,10-Anthraquinone	PAOH	0.9951	12.9
2-Methylanthraquinone	PAOH	0.9981	6.5

Table 4 continued. Coefficient of determination (R2) and residual average response factor (% RSD)

Compound	Compound type	R^2	AVCF % RSD
9-Fluorenone	oxyPAH	0.9997	5.0
Benzanthrone	oxyPAH	0.9985	6.0
5,12-Napthacenequinone	oxyPAH	0.9963	9.6
	Min	0.9951	1.1
	Max	1.0000	12.9
	Mean	0.9991	4.2

Appendix 5 – Table 5. QuEChERS soil extraction IS % recovery data

Compound	Extraction internal standard spiked recovery %							Mean	STDEV	% RSD
	1	2	3	4	5	6	7			
Napthalene d8	70.6	69.3	68.4	69.1	69.1	71.0	64.8	69	2.015	2.9%
Quinoline d7	72.0	71.7	69.9	74.8	72.4	71.9	69.3	72	1.792	2.5%
Dibenzofuran d8	82.0	82.8	80.4	80.7	79.8	82.4	77.4	81	1.847	2.3%
9-Fluorenone d8	81.8	81.4	78.2	82.5	79.1	79.9	78.7	80	1.676	2.1%
PCB 28L	89.2	89.5	90.9	93.7	86.0	92.7	91.2	91	2.546	2.8%
PCB 52L	89.7	88.5	87.0	85.2	85.1	86.5	83.8	87	2.063	2.4%
PCB 101L	85.2	82.6	82.7	78.0	81.6	79.5	80.3	81	2.375	2.9%
Pyrene d10	92.9	91.0	90.0	86.7	88.7	86.0	85.7	89	2.734	3.1%
PCB 118L	83.9	82.1	80.0	78.8	80.0	79.3	77.8	80	2.067	2.6%
PCB 153L	83.0	82.1	78.2	76.9	79.0	77.0	76.1	79	2.685	3.4%
PCB 138L	84.9	83.5	82.6	78.1	82.2	81.9	80.5	82	2.168	2.6%
PCB 180L	75.2	73.8	71.3	68.8	71.9	72.3	70.8	72	2.0735	2.9%
Perylene d12	63.6	63.6	62.3	64.4	66.7	70.4	68.1	66	2.9116	4.4%

Appendix 6 – Table 6. Compound Discoverer 3.2 software QuEChERS soil extract deconvoluted EI data NIST search index

Compound name	Reference RT [min]	Measured (m/z)	NIST formula	NIST theoretical (*m/z*)	Mass error (ppm)	Area	Calculated RI	RI Delta	NIST formula	Score	SI	RSI
Benzyl alcohol	4.466	79.05409	C6H7	79.05423	1.78	16069986	1039	5	C7H8O	95.7	937	987
Mevinphos	6.543	127.01549	C2H8O4P	127.01547	0.13	17076195	1429	16	C7H13O6P	99	960	962
Pebulate	6.729	128.10695	C7H14NO	128.10699	0.32	8595860	1469	0	C10H21NOS	95.2	895	936
Phthalimide	6.788	147.03149	C8H5NO2	147.03148	0.10	1242332	1482	0	C8H5NO2	99.4	968	968
Methacrifos	6.874	180.00058	C5H9O3PS	180.00045	0.71	6569371	1501	6	C7H13O5PS	98.5	931	932
Chloroneb	6.945	190.96625	C7H5Cl2O2	190.96611	0.72	17004439	1517	0	C8H8Cl2O2	98.2	960	981
Benzene, pentachloro-	7.084	249.84847	C6HCl5	249.84859	0.50	24762934	1549	0	C6HCl5	94.2	943	973
Tecnazene	7.384	202.87970	C5HCl4	202.87974	0.18	4337238	1609	0	C6HCl4NO2	94.3	975	988
Propachlor	7.411	120.08082	C8H10N	120.08078	0.36	9254651	1612	0	C11H14ClNO	97.2	938	964
Diphenylamine	7.501	169.08841	C12H11N	169.0886	1.13	24878507	1624	2	C12H11N	95.6	954	983
Cycloate	7.537	83.08540	C6H11	83.08553	1.48	12902977	1628	8	C11H21NOS	96.4	859	874
Chlorpropham	7.587	127.01830	C6H6ClN	127.01833	0.20	9390559	1634	0	C10H12ClNO2	97.9	953	976
Trifluralin	7.590	264.02240	C8H5F3N3O4	264.02267	1.01	10518127	1635	0	C13H16F3N3O4	96.9	855	860
Benfluralin	7.612	292.05356	C10H9F3N3O4	292.05397	1.40	11572988	1637	0	C13H16F3N3O4	98.3	915	924
Sulfotep	7.640	293.99060	C6H16O5P2S2	293.99089	0.98	5705056	1641	0	C8H20O5P2S2	98.9	945	948
Phorate	7.764	75.02623	C3H7S	75.0263	0.89	9945770	1656	0	C7H17O2PS3	93.4	858	914
Pentachloroanisole	7.934	264.83575	C6Cl5O	264.83568	0.28	10759890	1677	0	C7H3Cl5O	98.6	950	954
Botran	7.936	123.99490	C6H3ClN	123.99485	0.35	2725458	1677	0	C6H4Cl2N2O2	96	886	907
Atrazine	7.963	200.06975	C7H11ClN5	200.06975	0.00	8074011	1681	67	C8H14ClN5	97.5	947	948
Clomazone	8.022	125.01531	C7H6Cl	125.01525	0.41	24300346	1688	0	C12H14ClNO2	95.9	884	907
Terbuthylazine	8.073	214.08533	C8H13ClN5	214.0854	0.34	9737849	1694	82	C9H16ClN5	98.5	969	969
Diazinone	8.102	137.07097	C7H9N2O	137.07094	0.22	13998989	1698	90	C12H21N2O3PS	98.5	924	926
Propyzamide	8.113	172.95569	C7H3Cl2O	172.95555	0.82	17602860	1699	85	C12H11Cl2NO	98.2	942	944
Fonofos	8.148	108.98717	C2H6OPS	108.98715	0.17	23490821	1708	73	C10H15OPS2	98.3	950	956
Pyrimethanil	8.177	198.10248	C12H12N3	198.10257	0.48	45520948	1715	0	C12H13N3	94.9	918	978
Isazophos	8.203	118.98820	C2H2ClN3O	118.98809	0.90	9010911	1722	0	C9H17ClN3O3PS	98.1	905	915
Disulfoton	8.221	88.03407	C4H8S	88.03412	0.56	8860789	1727	0	C8H19O2PS3	97.2	921	951
Chlorothalonil	8.239	265.87787	C8Cl4N2	265.87806	0.72	19236284	1732	0	C8Cl4N2	97.1	966	971
Anthracene	8.252	178.07787	C14H10	178.0777	0.93	764555	1736	70	C14H10	97.8	909	914
Triallate	8.287	268.03238	C10H16Cl2NOS	268.03242	0.14	6379150	1745	80	C10H16Cl3NOS	97.4	896	897
Dibutyl phthalate	8.386	149.02332	C8H5O3	149.02332	0.03	5881147	1772	181	C16H22O4	98.1	915	916

Table 6 continued. Compound Discoverer 3.2 software QuEChERS soil extract deconvoluted EI data NIST search index

Compound name	Reference RT [min]	Measured (m/z)	NIST formula	NIST theoretical (*m/z*)	Mass error (ppm)	Area	Calculated RI	RI Delta	NIST formula	Score	SI	RSI
Propanil	8.464	160.97940	C6H5Cl2N	160.97936	0.28	19471408	1793	0	C9H9Cl2NO	98.7	955	960
Chloropyriphos-methyl	8.514	285.92539	C7H7Cl2NO3PS	285.92558	0.69	20128971	1807	72	C7H7Cl3NO3PS	97.9	940	941
Transfluthrin	8.523	163.01637	C7H3F4	163.01654	1.02	11807916	1809	0	C15H12Cl2F4O2	97.3	897	927
Vinclozoline	8.526	212.00269	C10H8Cl2N	212.00283	0.68	3926786	1810	0	C12H9Cl2NO3	97.6	921	942
Alachlor	8.561	160.11224	C11H14N	160.11208	1.05	8630775	1820	74	C14H20ClNO2	92.8	849	907
Tolclofos-methyl	8.570	264.98505	C9H11ClO3PS	264.98496	0.34	30985887	1822	74	C9H11Cl2O3PS	97	848	848
Fenchlorphos	8.643	284.93015	C8H8Cl2O3PS	284.93033	0.66	22859648	1843	0	C8H8Cl3O3PS	97.4	887	973
Pirimiphos methyl	8.682	290.07211	C10H17N3O3PS	290.07228	0.56	12041239	1853	79	C11H20N3O3PS	98.1	913	914
Fenitrothion	8.726	260.01404	C9H11NO4PS	260.01409	0.20	6776685	1866	0	C9H12NO5PS	97.7	911	918
Malathion	8.755	124.98213	C2H6O2PS	124.98206	0.55	9227531	1874	0	C10H19O6PS2	99	953	954
Linuron	8.786	61.05217	C2H7NO	61.05222	0.67	996941	1882	0	C9H10Cl2N2O2	93.7	754	767
Dichlofluanid	8.801	123.01375	C6H5NS	123.01372	0.22	10765859	1886	71	C9H11Cl2FN2O2S2	98.3	923	924
Pentachlorothioanisole	8.839	295.83633	C7H3Cl5S	295.83631	0.08	10450278	1897	58	C7H3Cl5S	98.1	937	954
Parathion	8.894	96.95074	H2O2PS	96.95076	0.27	4484103	1913	75	C10H14NO5PS	97.9	907	912
DCPA	8.901	300.87985	C9H3Cl4O3	300.88013	0.93	23819724	1915	72	C10H6Cl4O4	98.5	944	947
Triadimefon	8.918	208.02710	C9H7ClN3O	208.02722	0.56	4235155	1919	77	C14H16ClN3O2	97.7	896	898
9,10-Anthracenedione	8.949	208.05202	C14H8O2	208.05188	0.65	8399424	1928	55	C14H8O2	97.5	903	919
Pirimiphos ethyl	8.956	168.05891	C7H10N3S	168.05899	0.48	9261908	1930	0	C13H24N3O3PS	98.4	928	945
Isopropalin	9.011	238.08212	C10H12N3O4	238.08223	0.46	11131988	1946	0	C15H23N3O4	97.8	892	902
Bromophos	9.027	330.87711	C8H8BrClO3PS	330.87687	0.72	18301347	1951	68	C8H8BrCl2O3PS	97.9	938	949
Clofenvinfos	9.060	266.93747	C8H6Cl2O4P	266.93753	0.22	423034	1960	0	C12H14Cl3O4P	94.6	813	831
Fipronil	9.080	366.94272	C11H4Cl2F3N4OS	366.94295	0.62	1478777	1966	0	C12H4Cl2F6N4OS	95.6	781	781
Cyprodinil	9.093	224.11832	C14H14N3	224.11822	0.41	40372063	1969	0	C14H15N3	97.3	943	971
Metazachlor	9.117	132.08084	C9H10N	132.08078	0.50	11264282	1976	68	C14H16ClN3O	98.6	939	941
Penconazole	9.137	158.97646	C7H5Cl2	158.97628	1.09	18713951	1982	74	C13H15Cl2N3	98.3	917	918
Tolylfluanid	9.161	137.02948	C3H12Cl2F	137.02946	0.15	13660572	1989	74	C10H13Cl2FN2O2S2	96.6	864	894
Quinalphos	9.201	146.04755	C8H6N2O	146.04746	0.56	10014184	2000	77	C12H15N2O3PS	98.4	956	959
Triflumizole	9.214	205.99829	C8H4ClF3N	205.99789	1.96	3132804	2004	83	C15H15ClF3N3O	96	805	806
Triadimenol	9.225	112.05049	C4H6N3O	112.05054	0.42	8170623	2008	78	C14H18ClN3O2	97.9	897	897
Procymidone	9.238	96.05702	C6H8O	96.05697	0.58	6299842	2012	0	C13H11Cl2NO2	96.9	897	916
Bromophos-ethyl	9.309	302.84604	C7H4BrCl2O2S	302.84574	0.97	10486942	2034	0	C10H12BrCl2O3PS	96.8	880	888

Table 6 continued. Compound Discoverer 3.2 software QuEChERS soil extract deconvoluted EI data NIST search index

Compound name	Reference RT [min]	Measured (m/z)	NIST formula	NIST theoretical (*m/z*)	Mass error (ppm)	Area	Calculated RI	RI Delta	NIST formula	Score	SI	RSI
Tetrachlorvinphos	9.342	328.92981	C10H9Cl3O4P	328.92985	0.13	8515425	2044	0	C10H9Cl4O4P	97.8	940	952
Paclobutrazol	9.382	125.01531	C7H6Cl	125.01525	0.41	7793995	2057	0	C15H20ClN3O	98.5	935	979
Fenamiphos	9.417	55.05417	C4H7	55.05423	1.11	8959158	2068	0	C13H22NO3PS	93.4	813	896
Flutolanil	9.437	173.02083	C8H4F3O	173.02088	0.27	42560841	2074	0	C17H16F3NO2	99.5	974	976
Flutriafol	9.455	123.02410	C7H4FO	123.02407	0.26	9903083	2080	50	C16H13F2N3O	98.5	953	960
Pretilachlor	9.508	162.12763	C11H16N	162.12773	0.62	10905234	2096	0	C17H26ClNO2	95.3	860	902
Iodofenphos	9.510	376.86569	C8H8ClIO3PS	376.86595	0.68	21511075	2097	61	C8H8Cl2IO3PS	93.1	816	857
Oxadiazon	9.530	174.95873	C6H3Cl2NO	174.95862	0.60	11688803	2103	0	C15H18Cl2N2O3	98.3	916	918
Oxyfluorfen	9.554	252.03917	C13H7F3O2	252.03927	0.38	6287263	2111	85	C15H11ClF3NO4	95.9	824	826
Bupirimate	9.576	208.14444	C11H18N3O	208.14444	0.01	7679148	2118	0	C13H24N4O3S	98.2	911	915
Myclobutanil	9.583	179.02463	C8H6ClN3	179.02448	0.85	7298452	2120	0	C15H17ClN4	96.1	832	891
Flusilazole	9.592	233.05919	C13H11F2Si	233.05926	0.30	20153005	2123	73	C16H15F2N3Si	90.3	871	891
Tricyclazole	9.614	189.03554	C9H7N3S	189.03552	0.10	3484770	2129	43	C9H7N3S	97.1	973	981
Nitrofen	9.759	282.97965	C12H7Cl2NO3	282.97975	0.37	4113285	2175	60	C12H7Cl2NO3	97	915	922
Chlorthiophos	9.782	268.92563	C7H7ClO3PS2	268.92573	0.36	4905455	2182	0	C11H15Cl2O3PS2	95.5	796	815
Ethion	9.837	230.97318	C5H12O2PS3	230.97315	0.09	17039535	2199	73	C9H22O4P2S4	98.1	932	970
Triazophos	9.934	162.06630	C8H8N3O	162.06619	0.68	4817402	2229	75	C12H16N3O3PS	97.6	886	887
Carfentrazone ethyl	9.974	312.05896	C13H9F3N3O3	312.05905	0.29	7221687	2242	0	C15H14Cl2F3N3O3	97.9	900	922
Norflurazon	10.055	303.03787	C12H9ClF3N3O	303.03808	0.67	5679397	2267	0	C12H9ClF3N3O	96.9	843	845
Carbophenothion	10.062	156.98746	C6H6OPS	156.98715	1.97	11686197	2269	62	C11H16ClO2PS3	98.4	941	946
Edifenphos	10.102	109.01067	C6H5S	109.01065	0.17	11653922	2282	0	C14H15O2PS2	94.3	910	920
Resmethrin	10.214	128.06212	C10H8	128.06205	0.53	3084053	2315	0	C22H26O3	96.4	838	852
Pyridaphenthion	10.422	199.08656	C12H11N2O	199.08659	0.13	4151296	2375	0	C14H17N2O4PS	98.8	937	942
Tetramethrin	10.481	164.07062	C9H10NO2	164.07061	0.08	16344579	2392	0	C19H25NO4	99.4	972	976
Phosmet	10.539	160.03935	C9H6NO2	160.0393	0.30	28822343	2408	0	C11H12NO4PS2	98.4	926	986
Methoxychlor	10.574	227.10657	C15H15O2	227.10666	0.39	20536768	2417	69	C16H15Cl3O2	94.2	852	895
Tebufenpyrad	10.609	171.03200	C6H8N2O2P	171.03179	1.22	13061973	2426	76	C18H24ClN3O	96.7	942	956
Phosalone	10.852	182.00035	C8H5ClNO2	182.00033	0.10	11687589	2490	0	C12H15ClNO4PS2	97	922	938
Pyriproxyfen	10.883	136.07579	C8H10NO	136.07569	0.73	30668503	2498	71	C20H19NO3	97.9	940	953
Leptophos	10.885	171.00285	C7H8OPS	171.0028	0.32	13220533	2499	51	C13H10BrCl2O2PS	95.6	899	913
.lambda.-Cyhalothrin	10.921	181.06485	C13H9O	181.06479	0.33	12337809	2507	85	C23H19ClF3NO3	98.2	925	928
Pyrazophos	11.064	221.07954	C10H11N3O3	221.07949	0.24	6802604	2540	0	C14H20N3O5PS	97.6	881	923
Fenarimol	11.177	138.99460	C7H4ClO	138.99452	0.57	7620002	2567	52	C17H12Cl2N2O	97.9	947	950

Table 6 continued. Compound Discoverer 3.2 software QuEChERS soil extract deconvoluted EI data NIST search index

Compound name	Reference RT [min]	Measured (m/z)	NIST formula	NIST theoretical (*m/z*)	Mass error (ppm)	Area	Calculated RI	RI Delta	NIST formula	Score	SI	RSI
Azinphos-ethyl	11.214	132.04440	C8H6NO	132.04439	0.10	6897902	2575	0	C12H16N3O3PS2	97.7	916	960
Permethrine	11.503	183.08044	C13H11O	183.08044	0.02	14135701	2639	0	C21H20Cl2O3	98.8	942	952
Pyridaben	11.578	147.11690	C11H15	147.11683	0.48	25077978	2655	0	C19H25ClN2OS	98.5	938	944
Fluquinconazole	11.583	340.03946	C16H8ClFN5O	340.03959	0.39	19132298	2656	62	C16H8Cl2FN5O	87.6	842	9

Appendix 7 – Table 7. Compound Discoverer 3.2 software QuEChERS soil extract PCI confirmation data for [M+], [M+H], [M+C_2H_5], and [C_3H_5] with associated ppm mass error (where detected)

Name	Reference RT [min]	NIST Formula	$[M]^+$	Mass error (ppm)	$[M+H]^+$	Mass error (ppm)	$[M+C_2H_5]^+$	Mass error (ppm)	$[M+C_3H_5]^+$	Mass error (ppm)
Benzyl alcohol	4.466	C7H8O	108.05697	0.83	109.06534	0.88	137.09.664	0.19	149.09664	0.04
Mevinphos	6.543	C7H13O6P	224.04443		225.05280	-0.16	253.08410	-1.36	265.08410	
Pebulate	6.729	C10H21NOS	203.13384		204.14221	0.68	232.17351	0.33	24.417.351	
Phthalimide	6.788	C8H5NO2	147.03148		148.03986	-0.22	176.07116		188.07116	
Methacrifos	6.874	C7H13O5PS	240.02158	0.06	241.02996	0.15	269.06126	-0.25	281.06126	-0.2
Chloroneb	6.945	C8H8Cl2O2	205.98959	0.5	206.99796	1.42	235.02926	-0.26	247.02926	
Benzene, pentachloro-	7.084	C6HCl5	247.85154	0.21	248.85992	2.63	276.89122		288.89122	
Tecnazene	7.384	C6HCl4NO2	258.87559	0.53	259.88397	0.74	287.91527		299.91527	
Propachlor	7.411	C11H14ClNO	211.07584	0.39	212.08422	0.24	240.11552	0.50	252.11552	-1.15
Diphenylamine	7.501	C12H11N	169.08860	0.65	170.09698	0.19	198.12828	0.50	210.12828	0.57
Cycloate	7.537	C11H21NOS	215.13384		216.14221	0.45	244.17351	0.13	256.17351	
Chlorpropham	7.587	C10H12ClNO2	213.05511	0.21	214.06348	1.39	242.09478		254.09478	
Trifluralin	7.590	C13H16F3N3O4	335.10874	-0.34	336.11712	-0.14	364.14842		376.14842	
Benfluralin	7.612	C13H16F3N3O4	335.10874	-0.34	364.14842	-0.14	336.11712		37.614.842	
Sulfotep	7.640	C8H20O5P2S2	322.02219	-0.02	323.03057	-0.46	351.06187	-0.25	363.06187	-0.04
Phorate	7.764	C7H17O2PS3	260.01228	-0.24	261.02066	0.10	289.05196		301.05196	
Pentachloroanisole	7.934	C7H3Cl5O	277.86210	-0.02	278.87048	2.27	306.90178		318.90178	
Botran	7.936	C6H4Cl2N2O2	205.96443	0.34	206.97281	0.94	235.00411		24.700.411	
Atrazine	7.963	C8H14ClN5	215.09322	0.46	216.10160	0.44	244.13290	-0.01	256.13290	0.45
Clomazone	8.022	C12H14ClNO2	239.07076		240.07913	0.11	268.11043	0.43	280.11043	-1.17

Table 7 continued. Compound Discoverer 3.2 software QuEChERS soil extract PCI confirmation data for [M+], [M+H], [M+C_2H_5], and [C_3H_5] with associated ppm mass error (where detected)

Name	Reference RT [min]	NIST Formula	$[M]^+$	Mass error (ppm)	$[M+H]^+$	Mass error (ppm)	$[M+C_2H_5]^+$	Mass error (ppm)	$[M+C_3H_5]^+$	Mass error (ppm)
Terbuthylazine	8.073	C9H16ClN5	229.10887	0.34	230.11725	0.10	258.14855	-0.37	270.14855	-0.27
Diazinone	8.102	C12H21N2O3PS	304.10050	-0.09	305.10888	-0.84	333.14018	-0.37	345.14018	-0.62
Propyzamide	8.113	C12H11Cl2NO	255.02122	0.43	256.02960	-0.02	284.06090	0.84	296.06090	-1.29
Fonofos	8.148	C10H15OPS2	246.02964	-0.11	247.03802	-0.26	275.06932	-0.40	287.06932	-0.41
Pyrimethanil	8.177	C12H13N3	199.11040	0.99	200.11878	0.17	228.15008	-0.11	240.15008	-0.28
Isazophos	8.203	C9H17ClN3O3PS	313.04113	-0.09	314.04950	-0.37	342.08080	-0.17	354.08080	-0.17
Disulfoton	8.221	C8H19O2PS3	274.02793	-0.74	275.03631	-0.28	303.06761		315.06761	
Chlorothalonil	8.239	C8Cl4N2	263.88101	-0.22	264.88939	1.09	292.92069	-0.58	304.92069	
Anthracene	8.252	C14H10	178.07770	0.03	179.08608	-0.18	207.11738		219.11738	
Triallate	8.287	C10H16Cl3NOS	303.00127	304.00965	-0.33	332.04095	-0.07		344.04095	
Dibutyl phthalate	8.386	C16H22O4	278.15126	279.15964	-0.06	307.19094	319.19094			
Propanil	8.464	C9H9Cl2NO	217.00557	0.06	218.01395	0.21	246.04525	-0.82	258.04525	
Chloropyriphos-methyl	8.514	C7H7Cl3NO3PS	320.89443	0.46	321.90281	-0.29	349.93411	-0.20	361.93411	0.03
Transfluthrin	8.523	C15H12Cl2F4O2	370.01450	-0.04	371.02288	-0.40	399.05418	0.74	411.05418	
Vinclozoline	8.526	C12H9Cl2NO3	284.99540	-0.55	286.00378	1.29	314.03508	-0.34	326.03508	
Alachlor	8.561	C14H20ClNO2	269.11771	0.16	270.12608	0.20	298.15738	0.21	310.15738	
Tolclofos-methyl	8.570	C9H11Cl2O3PS	299.95381	300.96218	-0.21	328.99349	-0.28		340.99349	-0.27
Fenchlorphos	8.643	C8H8Cl3O3PS	319.89919	320.90756	0.10	348.93886	-0.27		360.93886	0.08
Pirimiphos methyl	8.682	C11H20N3O3PS	305.09575	0.41	306.10413	-0.16	334.13543	0.22	346.13543	0.32
Fenitrothion	8.726	C9H12NO5PS	277.01683	0.42	278.02521	0.25	306.05651	0.30	318.05651	
Malathion	8.755	C10H19O6PS2	330.03552	331.04389	0.02	359.07519	-0.50		371.07519	
Linuron	8.786	C9H10Cl2N2O2	248.01138	0.08	249.01976	0.48	277.05106		289.05106	
Dichlofluanid	8.801	C9H11Cl2FN2O2S2	331.96175	-0.08	332.97013	0.39	361.00143		373.00143	
Pentachlorothioanisole	8.839	C7H3Cl5S	293.83926	0.23	294.84764	1.63	322.87894	0.30	334.87894	-0.4
Parathion	8.894	C10H14NO5PS	291.03248	0.27	292.04086	0.35	320.07216		332.07216	
DCPA	8.901	C10H6Cl4O4	329.90147	0.33	330.90985	0.37	358.94115	-0.89	370.94115	
Triadimefon	8.918	C14H16ClN3O2	293.09256	294.10093	-0.06	322.13223	2.74		334.13223	
9,10-Anthracenedione	8.949	C14H8O2	208.05188	0.71	209.06026	0.86	237.09156	0.33	249.09156	

Table 7 continued. Compound Discoverer 3.2 software QuEChERS soil extract PCI confirmation data for [M+], [M+H], [M+C_2H_5], and [C_3H_5] with associated ppm mass error (where detected)

Name	Reference RT [min]	NIST Formula	$[M]^+$	Mass error (ppm)	$[M+H]^+$	Mass error (ppm)	$[M+C_2H_5]^+$	Mass error (ppm)	$[M+C_3H_5]^+$	Mass error (ppm)
Pirimiphos ethyl	8.956	C13H24N3O3PS	333.12705	0.24	334.13543	-0.34	362.16673	0.27	374.16673	-0.1
Isopropalin	9.011	C15H23N3O4	309.16831	310.17668	0.45	338.20798	350.20798			
Bromophos	9.027	C8H8BrCl2O3PS	363.84867	364.85705	0.54	392.88835	0.33		404.88835	0.92
Clofenvinfos	9.060	C12H14Cl3O4P	357.96898	358.97736	-0.11	387.00866	399.00866			
Fipronil	9.080	C12H4Cl2F6N4OS	435.93816	436.94653	-0.10	464.97783	476.97783			
Cyprodinil	9.093	C14H15N3	225.12605	1.74	226.13443	0.40	254.16573	-0.07	266.16573	0.2
Metazachlor	9.117	C14H16ClN3O	277.09764	0.86	278.10602	-0.23	306.13732		318.13732	
Penconazole	9.137	C13H15Cl2N3	283.06375	284.07213	-0.10	312.10343	324.10343			
Tolylfluanid	9.161	C10H13Cl2FN2O2S2	345.97740	-1.96	346.98578	-0.07	375.01708		387.01708	
Quinalphos	9.201	C12H15N2O3PS	298.05355	-0.05	299.06193	-0.10	327.09323	-0.16	339.09323	-0.31
Triflumizole	9.214	C15H15ClF3N3O	345.08503	346.09340	-0.31	374.12470	386.12470			
Triadimenol	9.225	C14H18ClN3O2	295.10821	296.11658	-0.77	324.14788	336.14788			
Procymidone	9.238	C13H11Cl2NO2	283.01614	-0.08	284.02451	-0.17	312.05581	-0.39	324.05581	-0.16
Bromophos-ethyl	9.309	C10H12BrCl2O3PS	391.87997	392.88835	0.04	420.91965	0.36		432.91965	1.75
Tetrachlorvinphos	9.342	C10H9Cl4O4P	363.89871	364.90708	-0.21	392.93838	404.93838			
Paclobutrazol	9.382	C15H20ClN3O	293.12894	294.13732	-0.13	322.16862	334.16862			
Fenamiphos	9.417	C13H22NO3PS	303.10525	-0.3	304.11363	1.41	332.14493	1.84	344.14493	
Flutolanil	9.437	C17H16F3NO2	323.11276	0.09	324.12114	-0.09	352.15244	0.03	364.15244	
Flutriafol	9.455	C16H13F2N3O	301.10212	302.11050	-0.37	330.14180	342.14180			
Pretilachlor	9.508	C17H26ClNO2	311.16466	-0.66	312.17303	-0.06	340.20434	0.50	352.20434	0.62
Iodofenphos	9.510	C8H8Cl2IO3PS	411.83480		412.84318	-0.04	440.87448	-0.26	452.87448	-0.01
Oxadiazon	9.530	C15H18Cl2N2O3	344.06890	0.02	345.07728	-0.12	373.10858	1.81	385.10858	
Oxyfluorfen	9.554	C15H11ClF3NO4	361.03232	-1.34	362.04070	-0.02	390.07200		402.07200	
Bupirimate	9.576	C13H24N4O3S	316.15636	0.13	317.16474	-0.03	345.19604	-0.51	357.19604	-0.47
Myclobutanil	9.583	C15H17ClN4	288.11363		289.12200	-0.26	317.15330		329.15330	
Flusilazole	9.592	C16H15F2N3Si	315.09978	0.35	316.10816	0.20	344.13946	1.59	356.13946	
Tricyclazole	9.614	C9H7N3S	189.03552	0.96	190.04390	0.70	218.07520		230.07520	
Nitrofen	9.759	C12H7Cl2NO3	282.97975	-0.57	283.98813	-1.32	312.01943		324.01943	

Table 7 continued. Compound Discoverer 3.2 software QuEChERS soil extract PCI confirmation data for [M+], [M+H], [M+C_2H_5], and [C_3H_5] with associated ppm mass error (where detected)

Name	Reference RT [min]	NIST Formula	$[M]^+$	Mass error (ppm)	$[M+H]^+$	Mass error (ppm)	$[M+C_2H_5]^+$	Mass error (ppm)	$[M+C_3H_5]^+$	Mass error (ppm)
Chlorthiophos	9.782	C11H15Cl2O3PS2	359.95718	0.17	360.96556	0.28	388.99686	0.22	400.99686	-0.33
Ethion	9.837	C9H22O4P2S4	383.98707	0.19	384.99544	0.29	413.02674		425.02674	
Triazophos	9.934	C12H16N3O3PS	313.06445	-1.35	314.07283	0.18	342.10413	0.02	354.10413	
Carfentrazone ethyl	9.974	C15H14Cl2F3N3O3	411.03588	0.16	412.04426	0.15	440.07556	2.15	452.07556	
Norflurazon	10.055	C12H9ClF3N3O	303.03808	0.52	304.04645	0.41	332.07775	-0.37	344.07775	
Edifenphos	10.102	C14H15O2PS2	310.02456	-0.21	311.03294	0.33	339.06424		351.06424	
Resmethrin	10.214	C22H26O3	338.18765	0.5	339.19602	0.40	367.22732		379.22732	
Resmethrin	10.258	C22H26O3	338.18765	0.5	339.19602	0.40	367.22732		379.22732	
Pyridaphenthion	10.422	C14H17N2O4PS	340.06412	-0.63	341.07249	-0.09	369.10379	-0.99	381.10379	
Tetramethrin	10.481	C19H25NO4	331.17781		332.18619	-0.34	360.21749		372.21749	
Phosmet	10.539	C11H12NO4PS2	316.99399		318.00236	-0.14	346.03366		358.03366	
Methoxychlor	10.574	C16H15Cl3O2	344.01321		345.02159	0.14	373.05289	0.86	385.05289	0.23
Tebufenpyrad	10.609	C18H24ClN3O	333.16024	0.15	334.16862	0.22	362.19992	-0.37	374.19992	
Phosalone	10.852	C12H15ClNO4PS2	366.98632	-0.29	367.99469	0.32	396.02599		408.02599	
Pyriproxyfen	10.883	C20H19NO3	321.13595		322.14432	-0.09	350.17562		362.17562	
Leptophos	10.885	C13H10BrCl2O2PS	409.86941		410.87778	0.22	438.90908	0.15	450.90908	
.lambda.-Cyhalothrin	10.921	C23H19ClF3NO3	449.10001		450.10838	-0.05	478.13968		490.13968	
Pyrazophos	11.064	C14H20N3O5PS	373.08558	-1.03	374.09396	0.10	402.12526	0.09	414.12526	
Fenarimol	11.177	C17H12Cl2N2O	330.03212	-1.22	331.04050	-0.52	359.07180		371.07180	
Azinphos-ethyl	11.214	C12H16N3O3PS2	345.03652		346.04490	-0.31	374.07620		386.07620	
Permethrine	11.503	C21H20Cl2O3	390.07840		391.08678	0.07	419.11808		431.11808	
Coumaphos	11.563	C14H16ClO5PS	362.01391	0.55	363.02229	0.56	391.05359	-0.88	403.05359	
Pyridaben	11.578	C19H25ClN2OS	364.13706	-1.4	365.14544	0.55	393.17674		405.17674	
Fluquinconazole	11.583	C16H8Cl2FN5O	375.00845		376.01682	0.86	404.04812	-0.67	416.04812	

This Sponsor Report is the responsibility of Thermo Fisher Scientific.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 231–236

BrJAC

SPONSOR REPORT

PDF

This section is dedicated for sponsor responsibility articles.

Overcome unexpected interferences and accelerate environmental analysis using triple quadrupole ICP-MS

Bhagyesh Surekar and Daniel Kutscher
Thermo Fisher Scientific, Bremen, Germany

This report was extracted from the Thermo Scientific Application Note 44484

The goal of this study was to highlight the benefits of triple quadrupole ICP-MS for the analysis of elemental contaminants in waters, waste, soils, and sediments.

Keywords: Environmental analysis, ICP-MS, triple quadrupole ICP-MS, high-throughput analysis, interference removal

INTRODUCTION

Toxic elements, such as arsenic, cadmium, mercury, and lead, need to be monitored in the environment to protect the surrounding nature from contamination and keep water resources safe. Therefore, the analysis of drinking waters, surface waters and waste waters, as well as monitoring the level of potential contaminants in solid waste or sewage sludges, is an important task. Regulated methods with applicable sample preparation procedures [1] and stringent protocols to assure system suitability and quality control have been established globally. ICP-MS is the preferred analytical solution for trace element determination at the lowest concentration levels. Modern instruments offer comprehensive interference removal and innovative solutions to increase sample throughput. Moreover, although the handling of samples containing more than 0.3% (m/v) of total dissolved solids has been historically a challenge for ICP-MS, today´s instruments can be easily configured for the direct analysis of solutions containing 3% TDS (for example undiluted sea water) [2] or more. Special sampling accessories, like automatic dilution systems, allow for prescriptive dilution of each sample without any user interaction and therefore open new possibilities to analyze a wide range of sample types under optimized conditions.

Single quadrupole ICP-MS systems are commonly used in many laboratories, combining consistent interference removal with short measurement times, and hence enabling robust, high-throughput analysis. Interference removal is accomplished by using a collision/reaction cell (CRC) filled with helium in order to remove the most common polyatomic interferences. Reactive gases, such as hydrogen, oxygen, or ammonia are occasionally used to eliminate certain interferences (e.g., H_2 for reduction of $^{40}Ar_2^+$ on ^{80}Se), but often, the formation of new interferences in the collision cell is possible, as other ions present in the ion beam can lead to uncontrolled side reactions.

One way to overcome these interference challenges is to use triple quadrupole ICP-MS (TQ-ICP-MS) instrumentation [3], such as the Thermo Scientific™ iCAP™ TQe ICP-MS. TQ-ICP-MS instruments eliminate interferences using reactive gases in the CRC (often referred to as quadrupole 2 (Q2), independent of the number of rods). Interference removal is accomplished as analyte and interferent react differently with the reactive gas and can be separated from each other in the analyzing quadrupole (Q3). Suppression of side reactions is accomplished through a mass filtration step in the first quadrupole (Q1), added axially in front of the CRC (Q2).

EXPERIMENTAL

An iCAP TQe ICP-MS system was used for all analysis. On each day of measurement, performance of the instrument was verified using an automatic performance report wizard and tuned using the autotune routines in case performance criteria were not met. The optimized parameters for the analysis are summarized in Table 1.

Table 1. Instrument configuration and operating parameters

Parameter	Value
Nebulizer	Borosilicate glass MicroMist (Glass Expansion, Australia), 400 µL min^{-1}, pumped at 40 rpm
Spray chamber	Quartz cyclonic spray chamber cooled at 3 °C
Injector	2.5 mm quartz
Interface	Ni sampler and Ni skimmer with 3.5 mm insert
Forward power	1,550 W
Nebulizer gas	1.00 mL min^{-1}
TQ-O_2	O_2 at 0.34 mL min^{-1} (Low gas flow) or 0.8 mL min^{-1} (High gas flow)
SQ-KED	He at 4.3 mL min^{-1} with 3 V KED

The selection of the optimum measurement mode for each analyte can be accomplished easily using the Reaction Finder Method Development Assistant included in the Thermo Scientific™ Qtegra™ Intelligent Scientific Data Solution (ISDS) Software. For single mode analysis, all settings made by Reaction Finder were modified manually.

RESULTS AND DISCUSSION

False positive results caused by the presence of interfering elements can lead to a significant disruption to the daily workplan in a busy laboratory. During the required troubleshooting, the instrument affected will have to stop productive analysis, and affected samples will need to be repeated. Kinetic energy discrimination (KED) allows the full mass range (typically lithium to uranium) to be analyzed, with consistent removal of polyatomic interferences in the most affected area of the mass range (*m/z* 40–100).

To illustrate the process of interference removal on a triple quadrupole ICP-MS, Figure 1 contains an overview on the analysis of ^{75}As and ^{202}Hg, demonstrating the two potential pathways of measurement: Either in a mass shift reaction (where the analyte is reactive to a specific gas), or an on-mass reaction (where the interference reacts while the analyte is detected on its original *m/z* ratio).

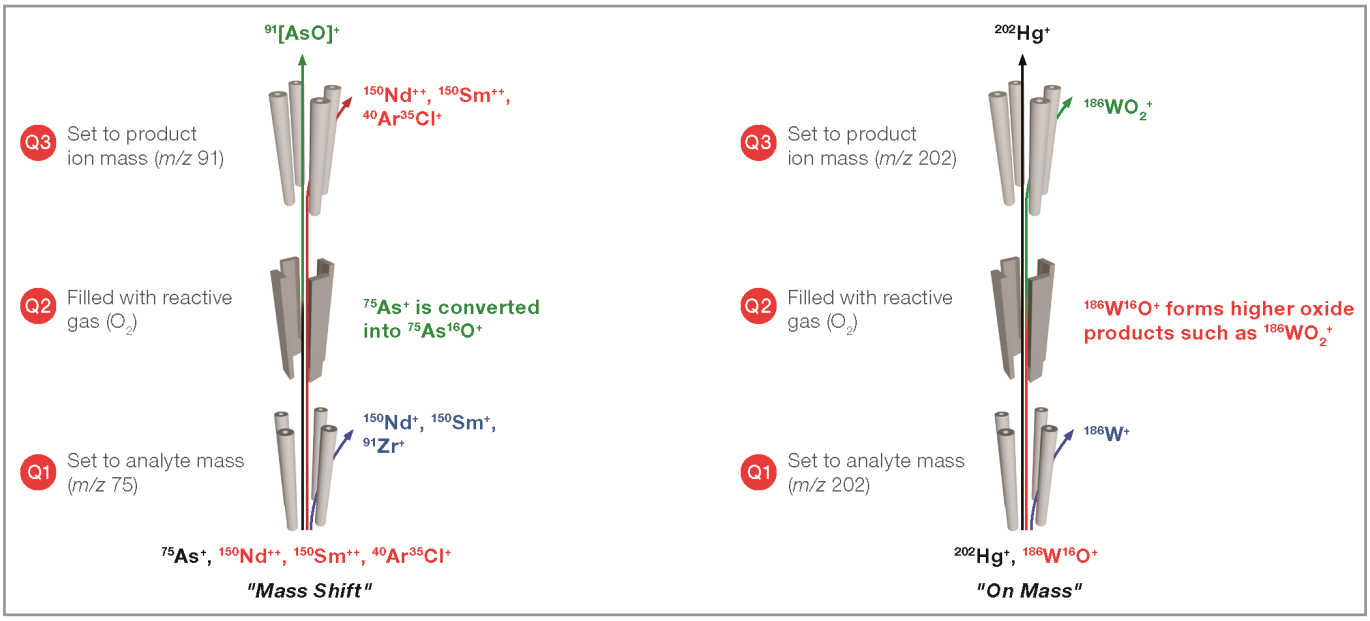


Figure 1. Interference removal using a mass shift reaction (e.g., ^{75}As) or an on-mass reaction (e.g., ^{202}Hg) on a triple quadrupole ICP-MS.

For some analytes, predominantly measured in on-mass mode, the use of a higher flow rate of the reaction gas provides lower backgrounds and hence superior interference removal. One example for this is cadmium, where the potential interferent molybdenum (leading to formation of $^{95}Mo^{16}O+$ on $^{111}Cd^+$) can be present in elevated concentrations, potentially requiring a higher flow of the reactive gas O_2 for interference removal.

Arsenic is a critical contaminant and under regulation in a wide variety of sample types. Whereas the most predominant interference on arsenic is caused by chlorine ($^{40}Ar^{35}Cl^+$ and $^{40}Ca^{35}Cl^+$ interfering on the only available isotope ^{75}As), bias can also arise from the potential presence of rare earth elements, specifically neodymium and samarium, through the formation of doubly charged ions ($^{150}Nd^{++}$ and $^{150}Sm^{++}$). The latter cannot be removed by KED and are even elevated in this measurement mode. The resulting concentrations for arsenic (measured in different modes) in a series of samples is shown in Table 2.

Table 2. Results obtained for the analysis of arsenic in a variety of sample matrices. No arsenic was added to any of the solutions.

Sample	SQ-KED [µg L^{-1}]	TQ-O_2 [µg L^{-1}]	SQ-O_2 [µg L^{-1}]	Comment
100 mg L^{-1} Ca, 5% HCl	0.06	0.00	0.2	KED is generally suitable, but the combination of calcium and chlorine leads to a small bias. SQ-O_2 shows a significant bias due to the formation of $^{43}Ca^{16}O_2^+$
10 µg L^{-1} REE Mix	1.05	0.00	0.02	KED shows (as expected) false positive result.
100 µg L^{-1} REE Mix	11.97	0.00	0.04	SQ-O_2 is less effective compared to TQ-O_2 to remove these interferences.
0.5 mg L^{-1} Zr	0.00	0.00	148.72	No mass filtration in Q1 leads to a completely wrong result in SQ-O_2 mode. A false positive is observed due to unreacted $^{91}Zr^+$.

As can be seen from the data, chlorine-based polyatomic interferences are generally well controlled using KED and will not cause major bias in most samples. However, the potential presence of rare earth elements can create a significant false positive result already at low levels. The use of oxygen and a subsequent oxidation of $^{75}As^{+}$ to $^{75}As^{16}O^{+}$ fully removes these interferences. The control of other ions entering the cell is essential, as can be seen from the results observed in SQ-O_2 mode (with no mass filtration before the CRC). Here, new interferences are formed, leading to increased backgrounds, as observed in the presence of calcium. The analysis of a solution containing 0.5 mg L^{-1} of zirconium shows another critical reason for having a mass filtration stage before the CRC. Although zirconium effectively reacts with oxygen to form ZrO (and higher oxides), the reaction does not completely remove all zirconium from (in this case) *m/z* 91, the resulting product ion mass of $^{75}As^{16}O^{+}$.

Mercury is another example of a critical element frequently required for analysis. The high mass range is often considered interference free in ICP-MS. However, this statement is only valid relative to the low and mid mass range, where abundant polyatomic interferences based on argon species cause significant backgrounds. As the difference in size between a polyatomic interference (consisting of at least two atoms) and an analyte ion is significantly lower for higher masses, KED is less effective, and hence, even small contributions can lead to a noticeable signal increase. For mercury, the presence of tungsten can lead to a potential false positive result. This is highlighted in Table 3, summarizing the results of the analysis of a series of solutions containing increasing amounts of tungsten. Even for a concentration of tungsten as low as 5 µg L^{-1}, a false positive bias for mercury is obtained that could potentially exceed the applicable reporting limit in a regulated method (0.1 µg L^{-1} is set in a variety of regulations). Note that all major isotopes of mercury (e.g., ^{199}Hg or ^{200}Hg) would be similarly affected by comparable interferences from other isotopes of tungsten. Other isotopes, such as ^{196}Hg, ^{198}Hg, or ^{204}Hg are not useful due to potential overlaps with gold (^{197}Au), which is commonly used as a stabilizer for mercury, and an isobaric overlap with ^{204}Pb. At the same time, analysis of mercury using TQ-O_2 mode fully removes the potential interferences with no bias observed. It is important to note here that the use of O_2 as a reactive gas on a single quadrupole ICP-MS instrument would lead to an immediate reaction of W^{+} ions (and hence formation of WO^{+}). Only after completely eliminating W+ from the ion beam can the selective oxidation of WO^{+} into higher oxides to remove the interference on mercury be achieved.

Table 3. Measured concentrations for different isotopes of mercury when analyzing solutions containing increasing concentrations of tungsten. Isotopic abundance of mercury isotopes is noted in parentheses. n.d. - not detected

	^{199}Hg (16.84%)		**^{200}Hg (23.13%)**		**^{201}Hg (13.22%)**		**^{202}Hg (29.60%)**	
Concentration of W in solution [µg L^{-1}]	Result SQ-KED [µg L^{-1}]	Result TQ-O_2 [µg L^{-1}]	Result SQ-KED [µg L^{-1}]	Result TQ-O_2 [µg L^{-1}]	Result SQ-KED [µg L^{-1}]	Result TQ-O_2 [µg L^{-1}]	Result SQ-KED [µg L^{-1}]	Result TQ-O_2 [µg L^{-1}]
5	0.154	n.d.	0.186	n.d.	0.070	n.d.	0.120	n.d.
10	0.360	n.d.	0.436	n.d.	0.170	n.d.	0.287	n.d.
20	0.891	n.d.	1.088	n.d.	0.425	n.d.	0.720	n.d.
50	1.549	n.d.	1.882	n.d.	0.734	n.d.	1.241	n.d.
Interference caused by	$^{183}W^{16}O^{+}$		$^{184}W^{16}O^{+}$		$^{184}W^{17}O^{+}$, $^{184}W^{16}O^{1}H^{+}$		$^{186}W^{16}O^{+}$	

Impact on sample throughput

A common challenge is the need to analyze a high number of samples to keep up with the backlog and achieve return on investment. TQ-ICP-MS systems are therefore less commonly found in analytical testing

laboratories, as each specialized set of measurement parameters (e.g., for multiple reactive gases) adds to the total method runtime. In contrast, SQ-ICP-MS allow for the analysis of samples using one or two acquisition modes only.

To achieve high throughput with full interference removal, the use of a single mode method using triple quadrupole technology and oxygen as a reactive cell gas is a great alternative. In this case, the iCAP TQe ICP-MS is operated in a single mode for all elements. This allows for effective removal of all interferences across the entire mass range.

Table 4 compares the resulting data acquisition times for a method analyzing a total of 29 individual masses (simulating a common set of analytes and typical internal standards), with different methods typically used on single and triple quadrupole ICP-MS. In all cases, an identical sample delivery system was used (a standard autosampler) with identical uptake and rinse times (30 seconds each). From the turnover time per sample, the total runtime of a batch containing 100 samples was calculated and compared between different methods.

Table 4. Comparison of acquisition time per sample, total runtime for 100 samples, and interference removal capability for different ICP-MS methods

	STD or KED Single Mode*	KED and H_2*	TQ-O_2 Single Mode	TQ-ICP-MS using Reaction Finder defaults
Acquisition time per sample not including uptake/wash [s]	91	139	121	125
Total runtime [min] and time saving for 100 samples	251 + 0 minutes	331 + 80 minutes	301 + 50 minutes	308 minutes + 57 minutes
Polyatomic interferences removed	No/Yes	Yes	Yes	Yes
Doubly charged interferences removed (e.g., REE)	No	No	Yes	Yes
Confidence all interferences are under control	No	No	Yes	Yes

*Data acquired on a Thermo Scientific™ iCAP™ RQ ICP-MS

As can be seen, a method based on single mode analysis has a clear potential for the realization of time savings, especially compared to a method combining the use of KED with H_2. Here, the sample turnover time is not only extended because of the need to evacuate and flush the CRC with two different gases (He, for KED operation, and H_2), but additionally through the longer duration of this process when adding a gas with a high flow rate (often in the range of 7 to 10 mL min^{-1} for H_2). Importantly, the single mode analysis using the TQ-O_2 mode includes an increase of the gas flow rate for some analytes, which requires a short stabilization time, but significantly shorter as compared to a full gas exchange in the CRC. With no change of the gas flow rate, identical turnover times could be achieved. Even though the measurement time is increased over a single mode approach using KED, the use of a triple quadrupole ICP-MS allows the full removal of all types of interferences (including those previously highlighted) and increases the confidence in the data generated.

CONCLUSION

As demonstrated in this note, the use of TQ ICP-MS with a combination of helium and oxygen as collision/reaction gases is a powerful approach to address common and potential challenges found in environmental analysis. With this technology, existing and established methods can be transferred and extended for improved interference removal on key analytes and new options become available. There are many benefits, not only in research driven laboratories, but also for laboratories tasked with screening high numbers of samples in a high-throughput environment.

A reduction of the data acquisition time of more than 10% can be achieved compared to methods based on single quadrupole ICP-MS using different collision/reaction gases.

This can be achieved with no or negligible impact on the detection limits. The iCAP TQe ICP-MS enables complete removal of many commonly encountered interferences, eliminating the need to re-analyze the samples and ultimately increasing confidence in data quality.

With no additional investment in laboratory infrastructure on top of what is required for a single quadrupole instrument (only the common gases helium and oxygen are required for interference removal), unnecessary interruptions to the daily laboratory routine caused by unexpected false positives can be avoided.

REFERENCES

1. Thermo Scientific Technical Note 44483: *Sample preparation techniques for AAS, ICP-OES and ICP-MS for regulated testing laboratories*. https://assets.thermofisher.com/TFS-Assets/CMD/Technical-Notes/tn-44483-aas-icp-oes-icp-ms-sample-prep-regulated-test-tn44483-en.pdf
2. Thermo Scientific Application Note 44417: *Direct analysis of trace elements in estuarine waters using triple quadrupole ICP-MS*. https://assets.thermofisher.com/TFS-Assets/CMD/Application-Notes/an-44417-icp-ms-trace-elements-esturine-waters-an44417-en.pdf
3. Thermo Scientific Smart Note 44452: *iCAP TQ ICP-MS*. https://assets.thermofisher.com/TFS-Assets/CMD/brochures/sn-44452-icp-ms-triple-quad-sn44452-en.pdf

Find out more at thermofisher.com/TQ-ICP-MS

This Sponsor Report is the responsibility of Thermo Fisher Scientific.

Brazilian Journal of Analytical Chemistry
2021, Volume 8, Issue 32, pp 237– 240

BrJAC

SPONSOR REPORT

PDF

This section is dedicated for sponsor responsibility articles.

Utilizing Direct Mercury Analysis for Mercury Detection in Botanical Extracts, Vitamins, Minerals and Dietary Supplements

Mercury in Nutraceutical Samples Utilizing Direct Mercury Analysis for Mercury Detection in Botanical Extracts, Vitamins, Minerals and Dietary Supplements

This report was extracted from the Milestone Industry Report DMA–80 *evo* / Nutraceutical

With the expansion of the global nutraceutical market, the spotlight on the analysis of its raw materials is ever increasing. Testing of nutraceutical products for heavy metals like lead, arsenic, cadmium and mercury has gained tremendous attention, and the low level limits in these products makes the analysis particularly challenging. Analytical chemists have to rely on techniques like CVAA and ICP-MS, which involve a time consuming and a labor-intensive sample preparation step. Direct mercury analysis on the other hand, as described by EPA 7473, is an alternative method to traditional techniques that requires no sample prep and delivers results in as little as 6 min per sample. This makes it significantly faster with comparable or better recoveries than CVAA and ICP-MS.

INTRODUCTION

The nutraceutical industry covers a broad spectrum of products including botanical extracts, vitamins, minerals and dietary supplements. A rise in the level of health consciousness among consumers has led to an exponential growth of the nutraceutical market. However, the industry has come under scrutiny to ensure control over toxic heavy metals such as mercury. Given the volatile nature of mercury and the industry's need to test it regularly at low concentrations, sample preparation can often become complicated and confusing. Traditional techniques used to analyze mercury in nutraceuticals involve a sample digestion step followed by CVAA or ICP-MS. Although effective, sample preparation requires manpower, equipment, handling and disposing large amounts of acid resulting in hours for completion.

Alternatively, the U.S. EPA developed a method 7473 for rapid determination of mercury in solids and aqueous samples without sample preparation. This method, known as direct mercury analysis, uses an integrated sequence of thermal decomposition followed by catalytic conversion, amalgamation and atomic absorption spectrophotometry.

The main benefits of direct mercury analysis include:

- Reduced Sample Turnaround (6 Minutes)
- No Sample Preparation
- Reduced Hazardous Waste Generation
- Reduction of Analytical Errors
- General Cost Savings (70% versus CVAA)

EXPERIMENTAL

Instrument

The DMA-80evoDirect Mercury Analyzer from Milestone, as referenced in EPA Method 7473, was used in this study (Figure 1).

Figure 1. Milestone's DMA-80 evo

The DMA-80 *evo* features a circular, stainless steel, interchangeable 40 position autosampler for virtually limitless throughput and can accommodate both nickel (500 mg) and quartz boats (1500 uL) depending on the requirements of the application. It operates from a single phase 110/220V, 50/60 Hz power supply and requires regular grade oxygen as a carrier gas.

As the process does not require the conversion of mercury to mercuric ions, both solid and liquid matrices can be analysed without the need for acid digestion or other sample preparation. The fact that zero sample preparation is required also eliminates all hazardous waste generation.

All results, instrument parameters including furnace temperatures, are controlled and saved with easy export capabilities to Excel or LIMS.

Principles of operation

Direct mercury analysis incorporates the following sequence: Thermal Decomposition, Catalytic Conversion, Amalgamation, and Atomic Absorption Spectrophotometry. Controlled heating stages are implemented to first dry and then thermally decompose a sample introduced into a quartz tube. A continuous flow of oxygen carries the decomposition products through a hot catalyst bed where halogens, nitrogen, and sulphur oxides are trapped.

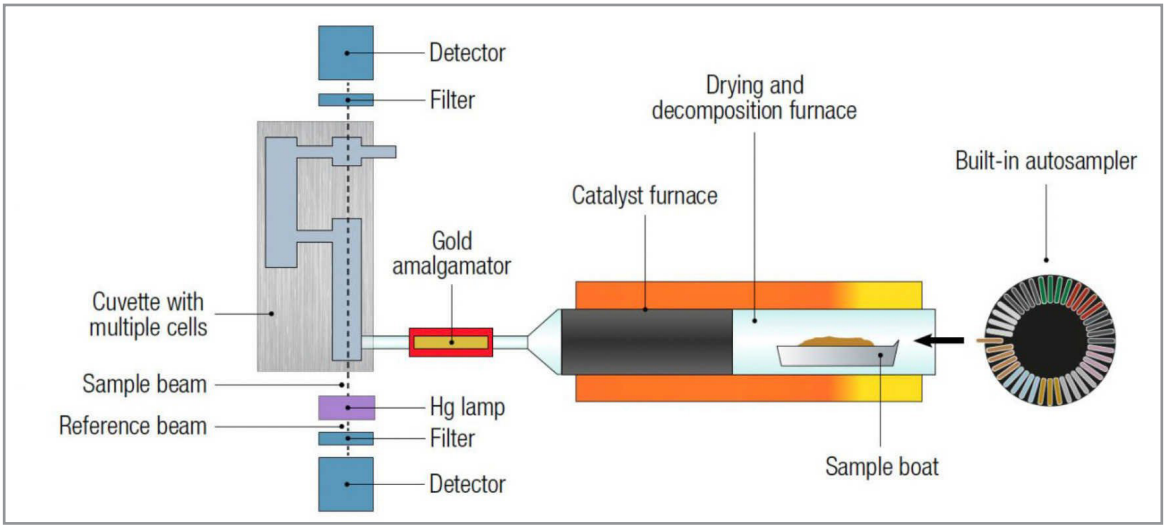


Figure 2. An Internal Schematic of Milestone's DMA-80 evo.

All mercury species are reduced to Hg(0) and are then carried along with reaction gases to a gold amalgamator where the mercury is selectively trapped. All non-mercury vapors and decomposition products are flushed from the system by the continuous flow of gas. The amalgamator is subsequently heated and releases all trapped mercury to the single beam, fixed wavelength atomic absorption spectrophotometer. Absorbance is measured at 253.7 nm as a function of mercury content.

EXPERIMENTAL DISCUSSION

To test the efficiency of the DMA-80 *evo*, three commonly available nutraceutical samples –Valerian Root, Ginkgo Biloba and Glucosamine Chondroitin were spiked by a solution having a mean mercury concentration of ~10.2ppband were run in the instrument to test for spike recoveries. Also, Gingko SRM 3248 was analysed to test if its mercury concentration falls in the NIST certified range.

Calibration

The DMA-80 *evo* can be calibrated using aqueous standards or Standard Reference Materials (SRM). The DMA-80 *evo* used for this experiment had a tricell spectrophotometer and covered adynamic range of 0.0015-1200 ng Hg. Each cell was calibrated using different volumes of 1 ppm and 0.1 ppm stock solutions, prepared from an NIST traceable 1000 ppm stock solution (VHG Labs).

Operating conditions

The DMA-80's operating conditions for all analyses are shown in Table 1.

Table 1. Analysis Operating Parameters

Parameter	Setting
Drying Temp/Time	30 seconds to 200 °C
Decomposition Ramp	90seconds to 660 °C
Decomposition Hold	90 seconds at 660 °C
Catalyst Temp	565 °C
Purge Time	60 seconds
Amalgamation Time	12 seconds at 900 °C
Recording Time	30 seconds
Oxygen Flow	120 mL/min

RESULTS

The concentrations mentioned in Table 2 are mean values obtained after running duplicates for each sample. The 2 concentrations obtained for the Ginkgo SRM –3248 were 0.2544 ppb and 0.2673 ppb respectively. These concentrations were not only in the certify ed range of mercury concentration, but also had an RSD of 0.05%, representing the accuracy and reproducibility of the DMA-80 *evo* at low mercury concentrations. The recovery data mentioned in the Table 2 suggests efficient spike recoveries.

Table 2. Results

Sample	Concentration (ppb)	Expected Conc. (ppb)	Recovery (%)
Valerian Root	4,7447	-	-
Gingko Biloba	2,0537	-	-
Glucosamine Chondroitin	1,5817	-	-
Spike	10,2750	10	102,75
Valerian Root (spiked)	15,5589	15,0197	103,59
Gingjko Biloba (spiked)	12,8810	12,3287	104,48
Glucosamine Chondroitin (spiked)	12,4578	11,8567	105,07
Gingko SRM 3248	0,2609	0,271 +/-0,034	96,27

CONCLUSION

A nutraceutical testing laboratory is required to analysedifferent matrices accurately and quickly while keeping operating costs under control. The DMA-80evois an excellent tool as it yields results in ~6 min/ sample and proves to be proficient, matrix-independent and cost-effective while completely eliminating the challenges of sample preparation posed by conventional mercury analysis techniques.

Further reading

Please visit our Hg info center for complete access to application notes, technical papers, as well as links to valuable resources for mercury testing.

Go to www.milestonesrl.com/dma-80

To learn more about mercury and other related topics, feel free to visit these websites:

- EPA Method 7473: http://www.epa.gov/waste/hazard/testmethods/sw846/pdfs/7473.pdf
- ASTM Method D6722-01: http://www.astm.org/Standards/D6722.htm
- EPA Mercury: http://www.epa.gov/mercury/
- Methyl Mercury: http://en.wikipedia.org/wiki/Methylmercury
- Mercury in Fish: http://www.epa.gov/waterscience/fish/advice/mercupd.pdf
- Mercury in Coal: http://energy.er.usgs.gov/health_environment/mercury/

This Sponsor Report is the responsibility of Milestone.

ABOUT MILESTONE

At Milestone we help chemists by providing the most innovative technology for metals analysis, direct mercury analysis and the application of microwave technology to extraction, ashing and synthesis. Since 1988 Milestone has helped chemists in their work to enhance food, pharmaceutical and consumer product safety, and to improve our world by controlling pollutants in the environment.

20º ENQA | 8º CIAQA

Encontro Nacional de Química Analítica
20th Brazilian Meeting on Analytical Chemistry

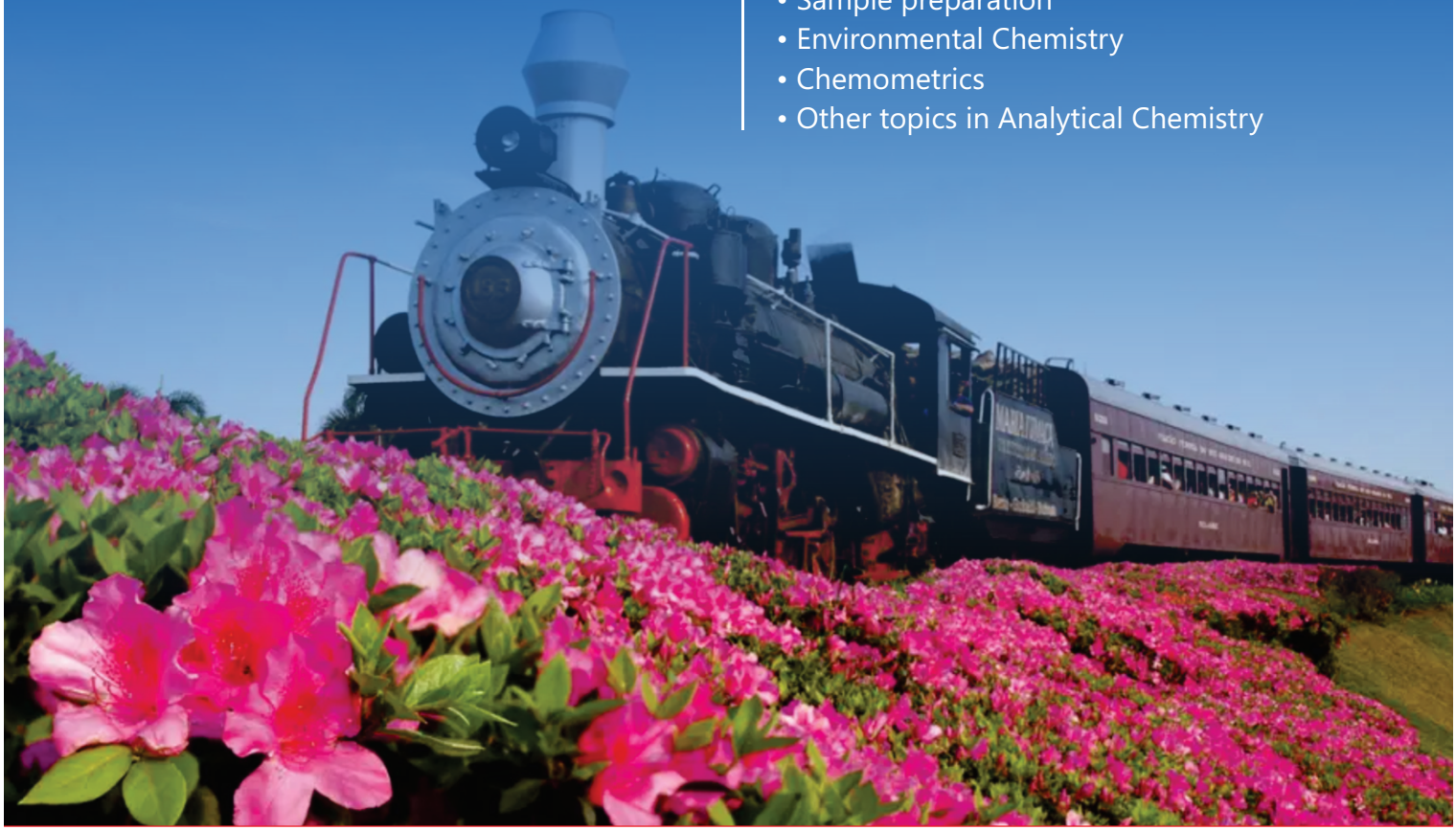
Congresso Ibero-Americano de Química Analítica

September • 25 to 28, 2022 • Dall'Onder Grande Hotel • Bento Gonçalves • RS • Brazil

These events are organized with short courses, conferences, oral and poster sessions, workshops, awards and commercial exhibitions.

Event topics

- Electrochemistry and Electroanalysis
- Atomic Spectrometry
- Mass Spectrometry
- Analytical Instrumentation
- Separation methods
- Sample preparation
- Environmental Chemistry
- Chemometrics
- Other topics in Analytical Chemistry



For more information: www.enqa.com.br

Held by: Sociedade Brasileira de Química, UFSM

Sponsorship – Diamond: Anton Paar, ThermoFisher Scientific

Sponsorship - Gold: Agilent, Astro 34, Metrohm, analitica

Sponsorship – Silver: HORIBA, Waters

Sponsorship - Bronze: ALLCROM, SHIMADZU

Organized by aconteceeventos científicos

RELEASE

Orbitrap Exploris GC Mass Spectrometer

Expanding analytical power with simplicity

The Orbitrap Exploris GC mass spectrometer provides high-confidence detection and precise quantitation in complex matrices, with minimal method development for maximum productivity. As one of the next-generation Thermo Scientific™ mass spectrometers, the Orbitrap Exploris GC mass spectrometer brings together exceptional levels of mass resolution, sensitivity, speed, and linear dynamic range without detector saturation to deliver accurate results even for your most challenging samples.

Orbitrap Exploris GC-MS key benefits

Evolved to exceed the demands of everyday testing, the Orbitrap Exploris GC mass spectrometer simplifies operations and delivers consistently accurate results. This powerful GC-MS system can help you**:**

- ***Power laboratory productivity***

  Simplify analytical workflows with the compact Orbitrap Exploris GC mass spectrometer, which brings the versatility of full-scan, high-resolution accurate-mass data to screening and quantitation. With proven robustness and reliability across all applications, the system delivers accurate results sample after sample.

- ***Ensure certainty in results***

  Reduce time spent evaluating data and increase confidence with exceptional levels of selectivity, sensitivity, and linear dynamic range that come together to deliver accurate results, in all sample matrices. Full-scan, high-resolution MS data allows multipoint compound identification with spectral matching, isotope patterns, retention indices, and elemental compositions to reduce time to results.

- ***Maximize uptime***

  Deliver results on time and with ease. With intuitive instrument control and method templates, the system is fully accessible to all members of your analytical team and provides complete confidence that the system is always operating at maximum performance. Integrated informatics solutions with Thermo Scientific Chromeleon Chromatography Data System (CDS) ensure a fast and efficient route to reporting.

- ***Explore new opportunities***

  Expand analytical capability and open up new business opportunities such as sample profiling, increased analytical scope, and unknown compound identification services. Full scan high-resolution accurate-mass allows multiple questions to be asked of a sample to give a deeper understanding of every sample you inject.

thermo scientific

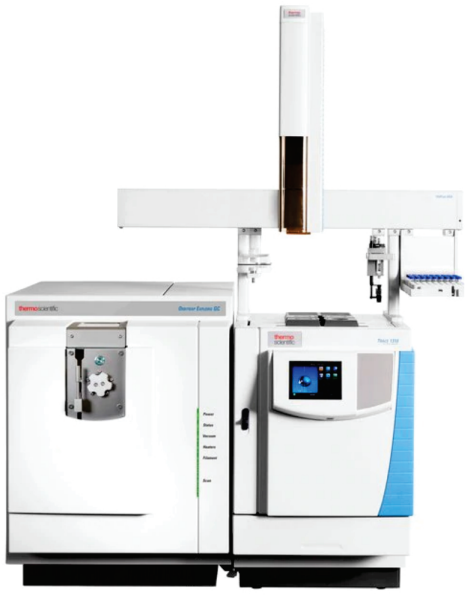
Efficiency transformed

Expanding analytical power with simplicity

Orbitrap Exploris GC mass spectrometer

New opportunities for Analytical testing

Simplify operations and deliver consistently accurate results in analytical testing using the Thermo Scientific™ Orbitrap™ Exploris™ GC Mass Spectrometer. This GC-MS system offers a unique depth of analysis and unparalleled accuracy and precision to improve everyday quantitation, screening and sample profiling. It provides the flexibility to keep pace with ever-changing regulations and explore new analytical opportunities through increased scope and method consolidation. Simple instrument set up and intelligent informatics enable users of all abilities to easily access data-rich information.



Orbitrap™ Exploris™ GC Mass Spectrometer

VIDEO WEBSITE

ThermoFisher SCIENTIFIC

RELEASE

Thermo Scientific iCAP RQ ICP-MS

Simplicity, productivity and robustness for routine labs in agribusiness

This innovative single quadrupole (SQ) ICP-MS is the ideal trace elemental analyzer for a wide range of several agricultural samples.

The Thermo Scientific™ iCAP™ RQ ICP-MS is particularly suited to the analysis of food, simultaneously determining trace level contaminants and macro level nutrients. This system provides a simple, robust, multielemental analysis method for major and minor concentrations of elements in food.

The simultaneously determination of all elements of interest in a wide range of food samples can be done efficiently and quickly by the Thermo Scientific™ iCAP™ RQ ICP-MS.

Simplicity guarantees user-friendly operation

- Quick connect sample introduction system
- Easy access cones via bench-height drop-down door
- Open geometry architecture for easy peripheral connection
- Intuitive platform software for seamless workflows

Productivity delivers more analysis in less time

- Less training for software and hardware
- Compatibility with automation and sample handling systems
- Comprehensive interference removal in single measurement mode
- Reduced drift and operator intervention

Robustness ensures high up-time and low maintenance

- Improved matrix tolerance interface
- All new state-of-the-art electronics
- New sturdy design RF generator
- Reliable hot and cold plasma operation

Comprehensive interference removal assures data accuracy, while our innovative helium Kinetic Energy Discrimination (He KED) technology enables measurement of all analytes in a single mode.

Our highly effective QCell collision/reaction cell, combined with unique flatapole design reduces BECs even further than He KED alone, through the clever, dynamic application of low mass cut off (LMCO).

Intuitive Thermo Scientific™ Qtegra™ Intelligent Scientific Data Solution™ (ISDS) software delivers all the support features essential to any lab, while containing all the flexibility needed to achieve the most challenging applications.

ThermoFisher SCIENTIFIC

Complete trace elemental analysis

Innovation inspired by real laboratory needs lies at the core of our trace elemental analysis instruments. The iCAP RQ ICP-MS part of the new Thermo Scientific™ iCAP™ Qnova™ Series ICP-MS, delivers the accuracy, robustness and ease of use essential to every high-throughput lab. Driven by our user-friendly Qtegra™ Intelligent Scientific Data Solution™ (ISDS) software, this powerful new system is ideal for labs up-grading to ICP-MS for the first time. The new Element Finder™ plug-in for Qtegra ISDS software, automates wavelength selection, simplifying method development for interference free analysis with the iCAP 7000 Plus Series ICP-OES.

maximum productivity

VIDEO **WEBSITE**

• powerful solutions at thermofisher.com/TEA

Thermo Scientific™ iCE™ 3000 Series AAS

Thermo Scientific™ iCAP™ 7000 Plus Series ICP-OES

Thermo Scientific™ iCAP™ RQ ICP-MS

Thermo Scientific™ ELEMENT 2™ & XR™ HR-ICP-MS
ELEMENT GD™ PLUS GD-MS

RELEASE

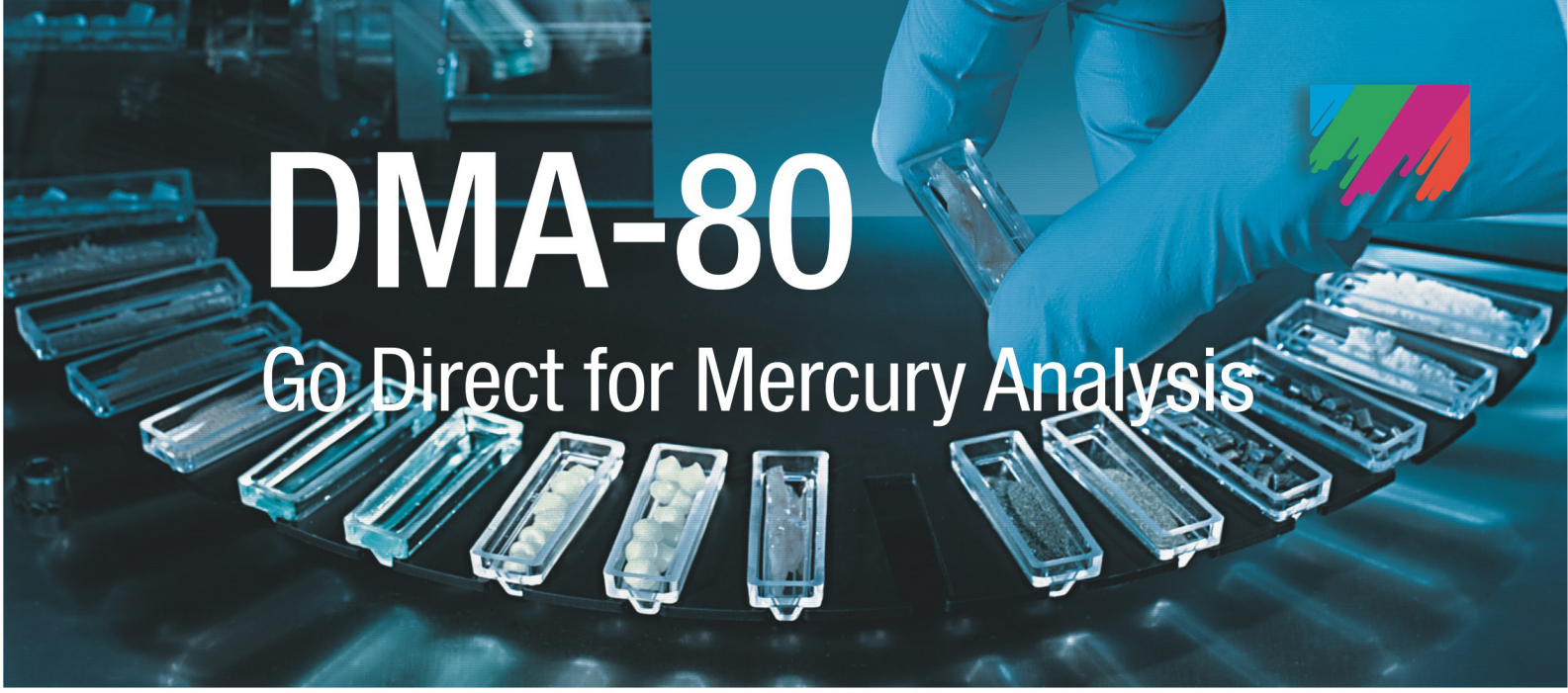
Save Time — Go Direct for Mercury Analysis

The Milestone DMA-80 has been at the forefront of direct total mercury analysis for almost two decades

The challenges in mercury determination are well known to analysts who often face a number of issues connected with tedious sample preparation process or the mercury analysis steps causing memory effect even after long cleaning cycles performed using latest generation of ICP-MS or cold vapor systems.

The DMA-80 mercury determination system can analyze any matrix (solid, liquid or gas) without any pre-treatment or chemical additions in as few as 6 minutes in full compliance with EPA method 7473. With thousands of units installed, Milestone defines the benchmark for direct mercury determination in a wide range of industries: environmental, food energy, cement, cosmetics, agriculture, mining and petrochemical.

- **Improved design** from the outside out for superior performance event at the ppt level.
- **Easy-to-use technology** with a single long-lasting calibration for all matrices.
- **No memory effect** thanks to the revolutionary autoblank feature.
- **Milestone Connect** for 30-years of experience at your fingertips
- **Fully compliant** with US EPA Method 7473, ASTM method D-6722-01 and D-7623-10.
- **Maximize your ROI** with an easy-to-maintain system and reliable components.



DMA-80

Go Direct for Mercury Analysis

SIX MINUTE, PREP-FREE MERCURY ANALYSIS

If youre looking for unsurpassed processing speed, efficiency and precision for mercury samples, its time for the Milestone DMA-80offering more than double the productivity of traditional cold vapor techniques at a fraction of the cost.

NO SAMPLE PREP REQUIRED

Eliminating the need for any wet chemistry sample treatment prior to analysis.

WIDEST DYNAMIC RANGE

Capable of achieving a detection limit of 0.001 to 30,000 ng of mercury

RESULTS IN JUST 6 MINUTES

Dramatically improves your laboratorys turnaround time and productivity.

UP TO 70% COST REDUCTION

Compared to traditional mercury analysis techniques such as CV-AAS.

Watch Videos, Download Brochures & Application Notes at:

VIDEO

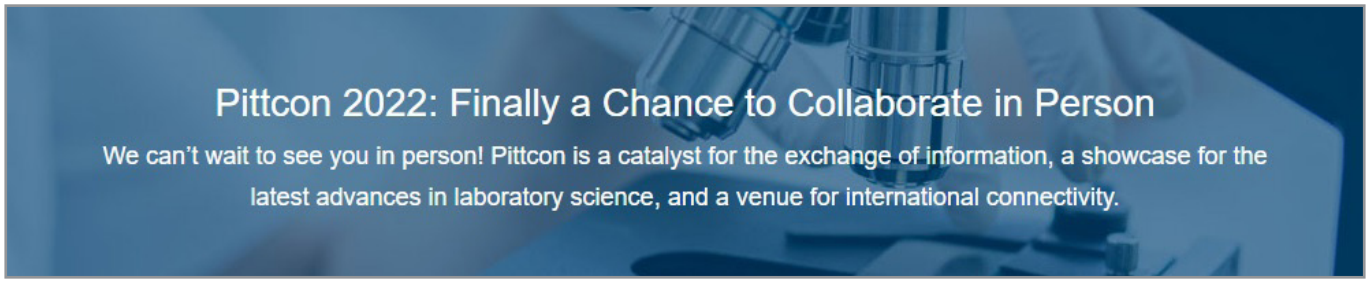
WEBSITE

DMA-80

DMA-1

MILESTONE

RELEASE



Reconnect with colleagues, feel the instruments, and experience the sights. Celebrate bringing together a world of knowledge to impact, enrich, and inspire the future of science.

Pittcon is a friendly, welcoming environment where analytical chemists at all professional levels meet. It is a platform for sharing ideas and cooperating to form new ones. Here, you will find that spark that drives your research, your career, and above all, your scientific perspective forward.

What to Expect
Our goal is to advance scientific endeavor through collaboration, bringing together a world of knowledge to impact, enrich, and inspire the future of science. Pittcon is a catalyst for the exchange of information, a showcase of the latest advances in laboratory science, and a venue for international connectivity.

Technical Program
One of the foremost analytical chemistry conferences, Pittcon's technical program provides participants with direct access to the latest research and developments from top scientists and innovators found throughout the world.

Short Courses
With over 60 from which to choose, these courses cover significant topics in bioanalytics, cannabis, energy, enviroscience, food science, forensics, industry, life sciences, nanotechnology, materials science, pharmaceuticals, and many more.

Exposition
Pittcon's transnational exposition gives you the opportunity to see the latest laboratory instrumentation, participate in demonstrations and product seminars, talk directly with technical experts, and find solutions to all your laboratory challenges.

WHEN
Technical Program Dates: Sunday, March 6 – Wednesday, March 9, 2022
Short Courses Dates: Saturday, March 5 – Wednesday, March 9, 2022
Exposition Dates: Monday, March 7 – Wednesday, March 9, 2022

VENUE: Georgia World Congress Center, Atlanta, GA, USA

For more information visit https://pittcon.org/conferees/

PITTCON® CONFERENCE & EXPO

March 5-9, 2022
Georgia World Congress Center
Atlanta, GA USA

Pittcon: Catalyzing a World of Collaborative Science

Pittcon is a dynamic, transnational exposition and comprehensive technical conference, a venue for presenting the latest advances in research and scientific instrumentation, and a platform for continuing education and career-enhancing opportunity

Exposition

Technical Program

Short Courses

Networking Sessions

Employment Bureau

Learn More

RELEASE

SelectScience® Pioneers online Communication and Promotes Scientific Success since 1998

SelectScience® promotes scientists and their work, accelerating the communication of successful science. SelectScience® informs scientists about the best products and applications through online peer-to-peer information and product reviews. Scientists can make better decisions using independent, expert information and gain easy access to manufacturers.

SelectScience® informs the global community through Editorial, Q&A and Application Articles, Featured Topics, Event Coverage, Video and Webinar programs.

Some recent contributions from SelecScience® to the scientific community

Application eBook: ***Best practices for protein characterization excellence***

Recent advancements in the field of biologics, whether in terms of research, development or production, have been greatly attributed to real-time label-free (RT-LF) technologies. This biosensor-based technology is especially crucial for the characterization of molecular interactions across various stages in biopharmaceutical discovery, development and production. Download this eBook **here**

Interview: ***Food safety: Tackling the global PFAS contamination crisis with mass spec***

In this exclusive interview, SelectScience® speaks with Marco Zhong, Director of the National Reference Laboratory for Food Contact Material, Guangdong, to find out more about the perfluoroalkyl and polyfluoroalkyl substances (PFAS) contamination crisis and how these commonly used synthetic chemicals, found everywhere from cookware to food packaging, have been linked to cancer, liver damage, obesity and other health risks. Read this Interview **here**

Webinars: ***Advance your research skills and knowledge with these 9 upcoming webinars***

SelectScience® hosts dozens of informative and insightful free webinars for scientists every month, featuring world-class speakers at the very forefront of their respective fields. In this regular feature, SelectScience® highlights the events you won't want to miss over the next week or so, as well as some of our top on-demand webinars. Access **here**

Video: ***Optimized fat oxidation stability testing***

Fat and oil oxidation stability is a vital parameter for determining shelf-life. In this video, uncover the benefits of the OXITEST from Velp Scientifica, and hear from researchers at the University of Basilicata as they share their experiences and results. Watch this video **here**

SelectScience® is the leading independent online publisher connecting scientists to the best laboratory products and applications.

Access 2 Million+ Decision Makers

- Working with Scientists to Make the Future Healthier.
- Informing scientists about the best products and applications.
- Connecting manufacturers with their customers to develop, promote and sell technologies.

SelectScience®
The Fastest Way to Expert Opinion

RELEASE

CHROMacademy is the leading provider of eLearning for analytical science

CHROMacademy *helps scientific organizations acquire and maintain excellence in their laboratories*.

For over 10 years, CHROMacademy has increased knowledge, efficiency and productivity across all applications of chromatography. With a comprehensive library of learning resources, members can improve their skills and knowledge at a pace that suits them.

CHROMacademy covers all chromatographic applications – HPLC, GC, mass spec, sample preparation, basic lab skills, and bio chromatography. Each paradigm contains dozens of modules across theory, application, method development, troubleshooting, and more. Invest in analytical eLearning and supercharge your lab.

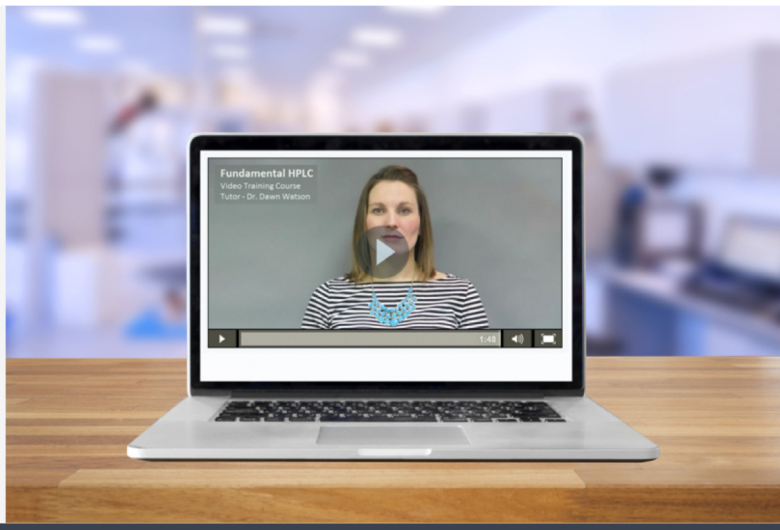
For more information, please visit www.chromacademy.com/

LC GC's CHROMacademy
powered by crawford scientific

CHROMacademy Lite members have access to less than 5% of our content. Premier members get so much more !

Video Training courses

Fundamental HPLC
Fundamental GC
Fundamental LCMS
Fundamental GCMS
HPLC Method Development
GC Method Development



Ask the Expert

We are always on hand to help fix your instrument and chromatographic problems, offer advice on method development, help select a column for your application and more.

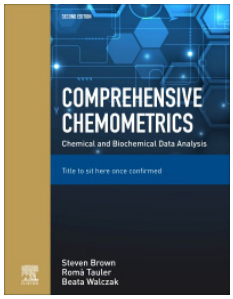
To find out more about Premier Membership contact:

Glen Murry: +1 732.346.3056 | Glen.Murry@ubm.com
Peter Romillo: +1 732.346.3074 | Peter.Romillo@ubm.com

www.chromacademy.com

The worlds largest e-Learning website for analytical scientists

NOTICES OF BOOKS

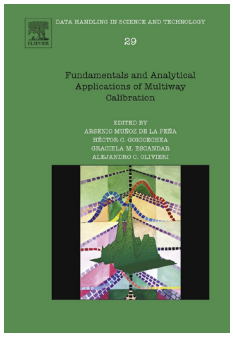


Comprehensive Chemometrics / Chemical and Biochemical Data Analysis, 2nd Ed.

Steven Brown, Roma Tauler, and Beata Walczak, Editors

May 2020. Publisher: Elsevier

This book features expanded and updated coverage, along with new content that covers advances in the field. Subject of note include updates in the fields of multidimensional and megavariate data analysis, omics data analysis, big chemical and biochemical data analysis, data fusion and sparse methods. Many chapters from the previous edition are updated, but there are also many new chapters on the latest developments. Read more

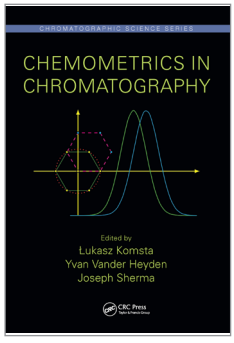


Fundamentals and Analytical Applications of Multiway Calibration, Volume 29, 1st Edition

Alejandro C. Olivieri, Graciela Escandar, Héctor C. Goicoechea, Arsenio Muñoz de la Peña, Series Volume Editors

August 2015. Publisher: Elsevier

This book presents researchers with a set of effective tools they can use to obtain the maximum information from instrumental data. It includes the most advanced techniques, methods, and algorithms related to multi-way calibration and the ways they can be applied to solve actual analytical problems. Read more

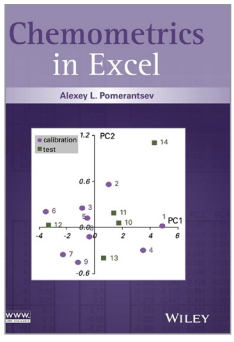


Chemometrics in Chromatography

Łukasz Komsta, Yvan Vander Heyden, Joseph Sherma, Editor

2018. Publisher: CRC Press, Taylor & Francis Group

Chemometrics significantly extends the possibilities of chromatography and with the technological advances of the personal computer and continuous development of open-source software, many laboratories are interested in incorporating chemometrics into their chromatographic methods. This book is an up-to-date reference that presents the most important information about each area of chemometrics used in chromatography, demonstrating its effective use when applied to a chromatographic separation. Read more



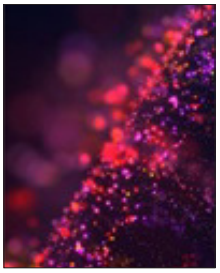
Chemometrics in Excel

Alexey L. Pomerantsev, Author

May 2014. Publisher: John Wiley & Sons, Inc.

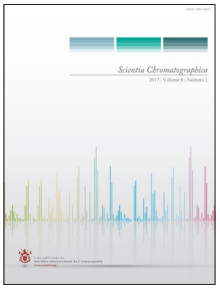
This booh provides an easy explanation of the fundamentals, methods, and applications of chemometrics. Acts as a practical guide to multivariate data analysis techniques. Explains the methods used in Chemometrics and teaches the reader to perform all relevant calculations. Presents the basic chemometric methods as worksheet functions in Excel. Includes Chemometrics Add In for download which uses Microsoft Excel® for chemometrics training. Online downloads include workbooks with examples. Read more

PERIODICALS & WEBSITES



American Laboratory
American Laboratory® is a platform that addresses basic research, clinical diagnostics, drug discovery, environmental, food and beverage, forensics and other markets, and combines in-depth articles, news, and video to deliver the latest advances in their fields. **Featured Article:** ***LABTips: Particle Characterization and Testing*** Posted: May 14, 2021. Regardless of application, knowing the size and shape of particles allows manufacturers to control product effectiveness and maintain customer satisfaction. Read more

LCGC
Chromatographyonline delivers practical, nuts-and-bolts information to help scientists and lab managers become more proficient in the use of chromatographic techniques and instrumentation, thereby making laboratories more productive and businesses around the world more successful.
Article: ***The Importance of Complete Overlapping of Analyte and Internal Standard Peaks in Eliminating Matrix Effects with LC–MS*** Authors: Hewavitharana, A.K.; Shaw, P.N.; Smyth, H.D.C.; Samaranayake, L.P.; Bandara, H.M.H.N. (Peer-Reviewed Article). Read more



Scientia Chromatographica
Scientia Chromatographica is the first and to date the only Latin American scientific journal dedicated exclusively to Chromatographic and Related Techniques (Mass Spectrometry, Sample Preparation, Electrophoresis, etc.). With a highly qualified and internationally recognized Editorial Board, it covers all chromatography topics (HPLC, GC, SFC) in all their formats, in addition to discussing related topics such as “The Pillars of Chromatography”, Quality Management, Troubleshooting, Hyphenation (GC-MS, LC-MS, SPE-LC-MS/MS) and others. It also provides columns containing general information, such as: calendar, meeting report, bookstore, etc. Read more

Select Science
SelectScience® promotes scientists and their work, accelerating the communication of successful science. SelectScience® informs scientists about the best products and applications through online peer-to-peer information and product reviews. Scientists can make better decisions using independent, expert information and gain easy access to manufacturers. SelectScience® informs the global community through Editorial, Features, Video and Webinar programs. Read more

Spectroscopy
With the Spectroscopy journal, scientists, technicians, and lab managers gain proficiency through unbiased, peer-reviewed technical articles, trusted troubleshooting advice, and best-practice application solutions. **Article:** ***An X-ray Fluorescence (XRF) Analysis of a Molecular Layer Deposition (MLD) Method Used in Producing Cement from Phosphogypsum*** Authors: Dandan Gao, Yubing Liu, Ping Dai, Juanjuan Lu, Chuannan Luo (Peer-Reviewed Research). Read more

EVENTS 2021 – It is suggested to consult the event's official website for updates.

July 20 – 21, 2021
Road to Chemometrics in Analytical Chemistry (CAC) 2022 // ZOOM Conference
http://cac2020.sciencesconf.org

August 30 – September 3, 2021
International Conference on Electronic Materials (2021 IUMRS-ICEM) and the XIX Brazilian Materials Research Society Meeting (XIX B-MRS) – VIRTUAL
https://www.sbpmat.org.br/19encontro/

September 13 – 24, 2021
44th Annual Meeting of the Brazilian Chemical Society (RASBQ) – VIRTUAL
http://www.sbq.org.br/44ra/

October 4 – 8, 2021
20th IUPAB Congress, 45th Annual SBBf Meeting, and 49th Annual SBBq Meeting – VIRTUAL
http://iupab2020.sbbq.org.br/interna-278/home

October 11 – 15, 2021
34th Latin American Congress of Chemistry – CLAQ 2020; 18th Latin American Congress of Chromatography – COLACRO; 10th Colombian Congress of Chromatography – COCOCRO; 4th Colombian Congress of Biochemistry and Molecular Biology - C2B2
Convention Center, Cartagena de Indias, Colombia
https://claq2020.com/en/bienvenida/

October 18 – 21, 2021
Metrology 2021 – Online
https://metrologia2021.org.br/

November 16 – 19, 2021
60th Brazilian Chemistry Congress – Virtual
http://www.abq.org.br/cbq/

November 24 – 27, 2021
Analitica Congress – Virtual
https://www.analiticanet.com.br/

November 30 – December 2, 2021
FCE Pharma
São Paulo, SP, Brazil
https://www.fcepharma.com.br/o-evento

November 30 – December 2, 2021
Analitica Latin America Expo
São Paulo, SP, Brazil
https://www.analiticanet.com.br/

December 12 – 16, 2021
XXIII International Mass Spectrometry Conference (IMSC 2021)
Sheraton Grand Rio Hotel & Resort, Rio de Janeiro, RJ, Brazil
https://www.imsc2020.com/

EVENTS 2022 – It is suggested to consult the event's official website for updates.

May 25 – 28, 2022
XXII Brazilian Congress of Toxicology (CBTox 2022)
Balneário Camboriú, SC, Brazil
https://www.cbtox2021.com.br/

June 4 – 8, 2022
18th International Conference on Electroanalysis (ESEAC 2022)
Vilnius, Lithuania
http://www.eseac2020.com/

September 25 – 28, 2022
20th National Meeting on Analytical Chemistry (20th ENQA) & 8th Ibero-American Congress of Analytical Chemistry (8th CIAQA)
Bento Gonçalves, RS, Brazil
https://enqa.com.br/

GUIDELINES FOR AUTHORS

Scope

The *Brazilian Journal of Analytical Chemistry* (BrJAC) is dedicated to the diffusion of significant and original knowledge in all branches of Analytical Chemistry and Bioanalytics. The BrJAC is addressed to professionals involved in science, technology and innovation projects at universities, research centers and in industry.

Professional Ethics

Manuscripts submitted for publication in BrJAC cannot have been previously published or be currently submitted for publication in another journal.

The BrJAC does not consider submissions of manuscripts that have been posted to preprint servers prior to submission, and will withdraw from consideration any papers posted to those servers prior to publication.

The submitted manuscripts are the full responsibility of the authors. Manipulation/invention/omission of data, duplication of publications, the publication of papers under contract and confidentiality agreements, company data, material obtained from non-ethical experiments, publications without consent, the omission of authors, plagiarism, the publication of confidential data and undeclared conflicts of interests are considered serious ethical faults.

The BrJAC discourages and restricts the practice of excessive self-citation by the authors.

The BrJAC does not practice coercive citation, that is, it does not require authors to include references from BrJAC as a condition for achieving acceptance, purely to increase the number of citations to articles from BrJAC without any scientific justification.

Misconduct will be treated according to the COPE's recommendations (https://publicationethics.org/) and the Council of Science Editors White Paper on Promoting Integrity in Scientific Journal Publications (https://www.councilscienceeditors.org/).

For more detailed information on the BrJAC's ethics and integrity policy, please see the "About us" menu at www.brjac.com.br

Manuscripts that can be submitted to BrJAC

- **Articles:** Full descriptions of an original research finding in Analytical Chemistry. Articles undergo double-blind full peer review.
- **Reviews:** Articles on well-established subjects, including critical analyses of the bibliographic references and conclusions. Manuscripts submitted for publication as Reviews must be original and unpublished. Reviews undergo double-blind full peer review.
- **Technical Notes:** Concise descriptions of a development in analytical methods, new techniques, procedures or equipment falling within the scope of the BrJAC. Technical notes also undergo double-blind full peer review.
- **Letters:** Discussions, comments, suggestions on issues related to Analytical Chemistry, and consultations to authors. Letters are welcome and will be published at the discretion of the BrJAC editor-in-chief.

Documents Preparation

It is highly recommended that authors download and use the templates to create their four mandatory documents to avoid the suspension of a submission that does not meet the BrJAC guidelines.

Download templates here

Cover Letter

In addition to the usual content of a Cover Letter to the Editor-in-Chief of the journal, the submitting author must declare any conflicts of interest on behalf of all co-authors and their agreement with the copyright policy of the BrJAC. Please read about conflicts of interest and copyright in the About us menu at www.brjac.com.br

It is the duty of the submitting author to inform their collaborators about the manuscript content and obtain their permission for submission and eventual publication.

The Cover Letter must be signed by the submitting author.

Title Page

The Title Page must contain information for each author: full name, affiliation, full postal address, e-mail, phone number, ORCID and information on the contribution of each author to the work. Acknowledgments must be entered on the Title Page. The submitting author must sign the Title Page.

Novelty Statement

The Novelty Statement must contain clear and succinct information about what is new and innovative in the study in relation to previously related works, including the works of the authors themselves.

Manuscript

It is highly recommended that authors download the Manuscript template and create their manuscript in this template, keeping the layout of this file.

- **Language**: **English** is the language adopted by BrJAC. The correct use of English is of utmost importance. In case the Editors and Reviewers consider the manuscript to require an English revision, the authors will be required to send an English proofreading certificate before the final approval of the manuscript by BrJAC.
- **Required items**: the manuscript must include a title, abstract, keywords, and the following sections: Introduction, Materials and Methods, Results and Discussion, Conclusion, and References.
- **Identification of authors**: as the BrJAC adopts a double-blind review, the manuscript file must **NOT** contain the authors' names, affiliations nor acknowledgments. Full details of the authors and their acknowledgements should be on the Title Page.
- **Layout**: the lines in the manuscript must be numbered consecutively and double-spaced.
- **Graphics and Tables**: must appear close to the discussion about them in the manuscript. For **figures** use **Arabic** numbers, and for **tables** use **Roman** numbers.
- **Permission to use content already published**: to use figures, graphs, diagrams, tables, etc. identical to others previously published in the literature, even if these materials have been published by the same submitting authors, a publication permission from the publisher or scientific society holding the copyrights must be requested by the submitting authors and included among the documents uploaded in the manuscript management system at the time of manuscript submission.
- **Chemical nomenclature, units and symbols**: should conform to the rules of the International Union of Pure and Applied Chemistry (IUPAC) and Chemical Abstracts Service. It is recommended that, whenever possible, the authors follow the International System of Units, the International Vocabulary of Metrology (VIM) and the NIST General Table of Units of Measurement. Abbreviations are not recommended except for those recognized by the International Bureau of Weights and Measures or those recorded and established in scientific publications. Use L for liters. Always use superscripts rather than /. For instance: use mg mL^{-1} and NOT mg/mL. Leave a space between numeric values and their units.
- **References throughout the manuscript:** the references must be cited as **numbers in square brackets [ ]**. It is recommended that references older than 5 (five) years be avoided, except in relevant cases. Include references that are accessible to readers.
- **References item:** include a DOI (as a hyperlink) whenever it is available. **References should be thoroughly checked for errors by the authors before submission.**

See how to format the References item below:

Journals

1. Orlando, R. M.; Nascentes, C. C.; Botelho, B. G.; Moreira, J. S.; Costa, K. A.; Boratto, V. H. M. *Anal. Chem*., **2019**, *91* (10), pp 6471-6478 (https://doi.org/10.1021/acs.analchem.8b04943).
   - *Publications with more than 10 authors, list the first 10 authors followed by a semicolon and et al.*
   - *Titles of journals must be abbreviated as defined by the Chemical Abstracts Service Source Index (http://cassi.cas.org/search.jsp).*

Books

2. Burgot, J.-L. *Ionic Equilibria in Analytical Chemistry*. Springer Science & Business Media, New York, **2012**, Chapter 11, p 181.
3. Griffiths, W. J.; Ogundare, M.; Meljon, A.; Wang, Y. Mass Spectrometry for Steroid Analysis. In: Mike, S. L. (Ed.). *Mass Spectrometry Handbook*, v. 7 of Wiley Series on Pharmaceutical Science and Biotechnology: Practices, Applications and Methods. John Wiley & Sons, Hoboken, N.J., **2012**, pp 297-338.

Standard methods

4. International Organization for Standardization. ISO 26603. Plastics — *Aromatic isocyanates for use in the production of polyurethanes — Determination of total chlorine*. Geneva, CH: ISO, **2017**.

Master's and doctoral theses or other academic literature

5. Dantas, W. F. C. *Application of multivariate curve resolution methods and optical spectroscopy in forensic and photochemical analysis*. Doctoral thesis, **2019**, Institute of Chemistry, University of Campinas, Campinas, SP, Brazil.

Patents

6. Trygve, R.; Perelman, G. US 9053915 B2, June 9, **2015**, Agilent Technologies Inc., Santa Clara, CA, US.

Web pages

7. http://www.chromedia.org/chromedia [Accessed 10 January 2019].

Unpublished source

8. Viner, R.; Horn, D. M.; Damoc, E.; Konijnenberg, A. *Integrative Structural Proteomics Analysis of the 20S Proteasome Complex* (WP-25). Poster presented at the XXII International Mass Spectrometry Conference (IMSC 2018) / August 26-31, **2018**, Florence, IT.
9. Author, A. A. *J. Braz. Chem. Soc*., in press.
10. Author, B. B., **2019**, submitted for publication.
11. Author, C. C., **2019**, unpublished manuscript.

Note: *Unpublished results may be mentioned only with the express authorization of the author(s). Personal communications can be accepted in exceptional circumstances.*

Manuscript Submission

The BrJAC uses an online manuscript manager system for the submission of manuscripts. This system guides authors stepwise through the entire submission process.

Submit manuscripts at www.brjac.com.br

The submitting author must add all co-authors to the Authors section of the manuscript manager system. After submission, all co-authors will receive an alert and will have the opportunity to confirm whether or not they are co-authors.

Four documents are mandatorily uploaded by the submitting author: Cover letter, Title Page, Novelty Statement and the Manuscript. Templates for these documents are available at www.brjac.com.br

The four documents mentioned above must be uploaded into the manuscript manager system as Word files. The manuscript Word file will be converted by the system to a PDF file which will be used in the double-blind peer review process.

All correspondence, including notification of the Editor's decision and requests for revision, is sent by e-mail to the submitting author through the manuscript manager system.

Review process

Manuscripts submitted to the BrJAC undergo an initial check for compliance with all of the journal's guidelines. Submissions that do not meet the journal's guidelines will be suspended and an alert sent to the corresponding author. The authors will be able to resend the submission within 30 days. If the submission according to the journal's guidelines is not made within 30 days, the submission will be withdrawn on the first subsequent day and an alert will be sent to the corresponding author.

Manuscripts that are in accordance with the journal's guidelines are submitted for the analysis of similarities by the iThenticate software.

The manuscript is then forwarded to the Editor-in-Chief who will check whether the manuscript is in accordance with the journal's scope and will analyze the similarity report issued by iThenticate.

If the manuscript passes the screening described above, it will be forwarded to an Associate Editor who will also analyze the iThenticate similarity report and invite reviewers.

Manuscripts are reviewed in double-blind mode by at least 2 reviewers. A larger number of reviewers may be used at the discretion of the Editor. As evaluation criteria, the reviewers employ originality, scientific quality, contribution to knowledge in the field of Analytical Chemistry, the theoretical foundation and bibliography, the presentation of relevant and consistent results, compliance with the BrJAC's guidelines, clarity of writing and presentation, and the use of grammatically correct English.

Note: In case the Editors and Reviewers consider the manuscript to require an English revision, the authors will be required to send an English proofreading certificate, by the ProofReading Service or equivalent service, before the final approval of the manuscript by the BrJAC.

The 1st-round review process usually takes around 5-6 weeks. If the manuscript is not rejected but requires corrections, the authors will have one month to submit a corrected version of the manuscript. In another 3-4 weeks, a new decision on the manuscript may be presented to the corresponding author.

The manuscripts accepted for publication are forwarded to the article layout department. Minor changes to the manuscripts may be made, when necessary, to adapt them to BrJAC guidelines or to make them clearer in style, respecting the original content. The articles are sent to the authors for approval before publication. Once published online, a DOI number is assigned to the article.

Copyright

When submitting their manuscript for publication, the authors agree that the copyright will become the property of the Brazilian Journal of Analytical Chemistry, if and when accepted for publication. The submitting author declares in the Cover Letter, on behalf of all other co-authors, to consent to this transfer.

The copyright comprises exclusive rights of reproduction and distribution of the articles, including reprints, photographic reproductions, microfilms or any other reproductions similar in nature, including translations.

Final Considerations

Whatever the nature of the submitted manuscript, it must be original in terms of methodology, information, interpretation or criticism.

With regard to the contents of published articles and advertisements, the sole responsibility belongs to the respective authors and advertisers; the BrJAC, its editors, editorial board, editorial office and collaborators are fully exempt from any responsibility for the data, opinions or unfounded statements.