Article

Green Chemistry Method Based on PARAFAC EEM Data Modeling for Benzo[a]pyrene Quantitation in Distilled Spirit

Amanda C. Silva,[a] Licarion Pinto,[a] Adriano A. Gomes[b] and Mario C. U. Araujo [ORCID] *,[a]*

*[a]Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA),
Departamento de Química, Centro de Ciências Exatas e da Natureza,
Universidade Federal da Paraíba, CP 5093, 58051-970 João Pessoa-PB, Brazil*

*[b]Laboratório de Quimiometria e Instrumentação Analítica (LQIA),
Instituto de Química, Universidade Federal do Rio Grande do Sul,
Avenida Bento Gonçalves, 9500, 91509-900 Porto Alegre-RS, Brazil*

Benzo[a]pyrene (BaP) is often used as a marker of polycyclic aromatic hydrocarbons (PAHs) in beverages. This marker is often quantified by chromatographic methods, which require sample preparations involving the use of reagents, solvents, extraction, pre-concentration, and/or cleanup steps. In this study, a new method for quantification of BaP in cachaças (liquors) that does not use any sample preparation was developed. Interferents in cachaças were overcome using excitation-emission matrices data modeling with parallel factor analysis (PARAFAC). The recoveries ranged from 93.20 to 101.13%, and the relative error of prediction and limit of detection were, respectively, estimated at 2.66% and 2.88 ng mL^{-1}. The proposed method is inexpensive and less time consuming than other approaches described in the literature, uses no reagents, solvents or extraction, has no pre-concentration or cleanup steps, contributing to green analytical chemistry.

Keywords: benzo[a]pyrene, cachaça, excitation-emission matrices, parallel factor analysis, polycyclic aromatic hydrocarbons

Introduction

Polycyclic aromatic hydrocarbons (PAHs) are a class of chemical compounds that have at least two combined aromatic rings and are generally formed by incomplete combustion of organic matter.[1,2] Due to their carcinogenic potential, and increasing exposure, there is growing global concern. PAHs are highly lipophilic because of their aromatic polynuclear structure; they can be quickly absorbed by inhalation, ingestion, or dermal contact.[3,4] PAHs also react with deoxyribonucleic acid (DNA) causing mutations.[5] Human contamination occurs through ingestion of food and beverages containing PAHs. Brazilian sugarcane spirit or cachaça, a typical distilled spirit widely consumed in Brazil and exported to many countries may contain considerable amounts of PAHs due to its production and aging processes, which involve burning sugar cane and storage in wooden barrels.[6]

Although in several foods the limits of PAHs are established by international committees and organizations,

no recommendation is made with respect to maximum levels in distilled spirits.[7] The same is true for benzo[a]pyrene (BaP), which is considered a marker for both the occurrence, and carcinogenic effects of PAHs in beverages and foods.[7-9] The European Commission[10] recommends 10 ng g^{-1} as the maximum allowable value for BaP in foodstuffs. However, the organization does not mention a maximum value for distilled spirits. Therefore, in this work it will be used the limit for foodstuffs as a hit limit.

Reports can be found in the literature concerning PAH quantitation in distilled beverages, the majority of them use gas or liquid chromatography.[11-17] Because of its rigid molecular structure, BaP displays fluorescent properties, thus as an alternative to chromatographic techniques, fluorescent spectroscopy can be used to develop sensitive analytical methods for BaP quantitation.[18] In addition, total fluorescence spectroscopy based on multiple emission spectra at several excitation wavelengths (second order data) allows developing multi-way calibration models, where each sample generates an excitation-emission matrix (EEM).[18,19] This type of calibration model is used for reliable predictions in the presence of un-calibrated constituents; this property

*e-mail: mariougulino@gmail.com

is known as the second order advantage; it can be used to develop methods with little or no sample preparation, no reagents and no consumption of solvents.[20]

In the literature, several works report the use of second order data combined with multiway calibration methods, such that the second order advantage is successful in analysis of complex samples.[21-23] In the particular case of EEM data, where trilinearity properties tend to be fulfilled, parallel factor analysis (PARAFAC) is an appropriate chemometric multiway tool to achieve the second order advantage. PARAFAC helps to obtain quantitative and qualitative information about an analyte as well as the un-calibrated constituents that affect measurement selectivity, i.e., both analyte concentration and the pure spectral profile in both excitation and emission modes for the analyte and other fluorophores present in the sample.[24,25]

The PARAFAC model assumes that a three-way array ($\underline{\mathbf{X}}_{I \times J \times K}$) formed by $\mathbf{I}$ EEM matrix sized $\mathbf{J} \times \mathbf{K}$ can be decomposed as in equation 1:

$$\underline{\mathbf{X}} = \sum_{n=1}^{N} \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c} + \underline{\mathbf{E}} \qquad (1)$$

where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are loading vectors of the PARAFAC model obtained by means of alternating least squares (ALS), $\underline{\mathbf{E}}$ is a three-way array of the unmodeled information, $\otimes$ is the Kronecker product, and N is the number of factors in the PARAFAC model. The information stored in the $\mathbf{a}$ vector is directly related to concentration, and can be used for analyte prediction by means of a pseudo-univariate calibration approach. When compared to other multiway methods for trilinear data modeling, PARAFAC has the attractive advantage of uniqueness, i.e., it does not have rotational freedom.[26]

In this study, a new method for quantification of BaP in cachaças is proposed as alternative to chromatographic methods, which require sample preparations involving the use of reagents, solvents, extraction, pre-concentration and/or cleanup steps. To overcome interferents in cachaças, an excitation-emission matrices data modeling with parallel factor analysis was employed. Since no sample preparation is performed, the proposed method does not use reagents or solvents, is faster, less expensive, less susceptible to errors and leads to better recoveries than other approaches described in the literature.[21,22]

Experimental

Chemicals and stock solutions

All chemicals used in this work were of analytical grade. All PAHs were purchased from Sigma-Aldrich. Acetonitrile and ethanol were acquired from TEDIA. Milli-Q water (Millipore) was used in all experiments. The stock solutions 1200 mg L^{-1} of BaP, 3000 mg L^{-1} of acenaphthene (ACL) and of phenanthrene (PHN); 2000 mg L^{-1} of pyrene (PYR) and of fluoranthene (FLT) and 800 mg L^{-1} of anthracene (ANT) were prepared in acetonitrile and stored in amber bottles at –20 °C.

Calibration standard solutions and validation set samples

Ten calibration standards ranging from 10 to 100 ng mL^{-1} of BaP in 45% ethanol solution were prepared by appropriated diluting of stock solutions and used to construct PARAFAC model.

To evaluate the PARAFAC model, a validation set was prepared based on a Taguchi design, where each sample is a mixture of six target PAHs, BaP (the analyte), and five interferents (ACL, PYR, ANT, PHN and FLT). Concentration levels of the analyte and interferents in Taguchi design (Table 1) were chosen according to the limits established by European Commission.[10] The validation set was prepared to mimic possible interference scenarios for quantitation of BaP in Brazilian cachaças.

Recovery study in cachaça

Six cachaças from different manufacturers were purchased in the local market. To remove possible solid particles, the samples were filtered using a cellulose acetate filter of 0.22 μm before analysis. To assess the accuracy of the proposed method, a recovery study was carried out with spiked cachaça in three concentration levels of BaP 25, 55 and 85 ng mL^{-1}.

EEMs acquisition

Spectrofluorimetric measurements were performed using a Fluorolog-3 (Horiba Jobin Yvon Inc.), equipped with two single-grade monochromators, one for excitation, another for emission; a 450 W Xenon discharge lamp and a photomultiplier. A Hellma quartz cuvette with dimensions of $1.0 \times 0.2 \times 3.5$ cm was used for all measurements, and the excitation path was 1.0 cm. The slit band widths for the excitation and emission monochromators were respectively fixed at 3 and 2 nm, and the integration time was 0.05 s. EEMs were recorded from 275 to 375 nm (excitation) and from 330 to 530 nm (emission), both with resolutions of 2 nm, generating a 51×101 matrix (excitation × emission variables). For all cases, the EEMs were recorded in triplicate.

Table 1. Concentration of PAHs for each synthetic sample

Sample/PAH	BaP / (ng mL^{-1})	PYR / (ng mL^{-1})	ACL / (ng mL^{-1})	ANT / (ng mL^{-1})	FLT / (ng mL^{-1})	PHN / (ng mL^{-1})
1	25	25	240	75	200	200
2	40	25	390	120	350	400
3	55	25	540	165	500	600
4	70	25	690	210	650	800
5	85	25	840	255	800	1000
6	40	40	540	75	650	1000
7	55	40	690	120	800	200
8	70	40	840	165	200	400
9	85	40	240	210	350	600
10	25	40	390	255	500	800
11	55	55	840	75	350	800
12	70	55	240	120	500	1000
13	85	55	390	165	650	200
14	25	55	540	210	800	400
15	40	55	690	255	200	600
16	70	70	390	75	800	600
17	85	70	540	120	200	800
18	25	70	690	165	350	1000
19	40	70	840	210	500	200
20	55	70	240	255	650	400
21	85	85	690	75	500	400
22	25	85	840	120	650	600
23	40	85	240	165	800	800
24	55	85	390	210	200	1000
25	70	85	540	255	350	200

BaP: benzo[*a*]pyrene; PYR: pyrene; ACL: acenaphthene; ANT: anthracene; FLT: fluoranthene; PHN: phenanthrene.

Softwares

The PARAFAC model was carried out in the MatLab® environment using the N-way toolbox coupled with the MVC2 graphical user interface.[27-29]

To remove the Rayleigh and Raman scatterings, all EEMs were previously treated using a MatLab routine, which is based on interpolation approach.[30,31]

Results and Discussion

Data set

The EEM contour surface displays in Figure 1a the Rayleigh and Raman scattering, which strongly affects the PARAFAC model's fit. Such phenomenon was removed using a MatLab routine, which is based on an interpolation approach proposed by Bahram *et al.*[31] As can be seen

in Figure 1b, after this pretreatment, no deformation of BaP's characteristic signal was observed, a key point for maintenance of data trilinearity. Thus, before any PARAFAC calculations, all EEM data were previously treated using the MatLab routine.

Validation set samples and PARAFAC treatment

The BaP fluorescent signal was strongly overlapped by ACL, PYR, ANT, PHN and FLT, as can be seen in Figure 2. In addition, when many fluorophores are present in samples, quenching and/or inner filter effect may occur, which becomes a challenge from an analytical point of view. Since other PAHs can be found in cachaça, PARAFAC models were built and evaluated to study these problems employing a validation set.

As can be seen in Figure 3a, the explained variance increases significantly to the sixth factor (indicated by
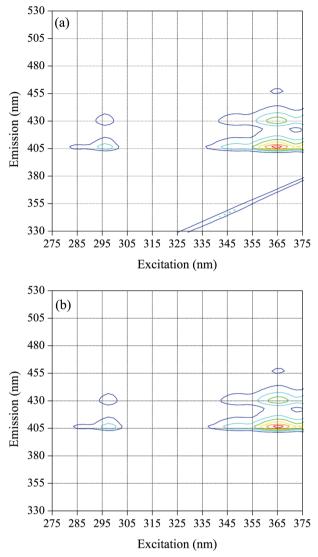
Figure 1. Contour plots of 100 ng mL⁻¹ of BaP standard solution (a) before and (b) after scattering correction.
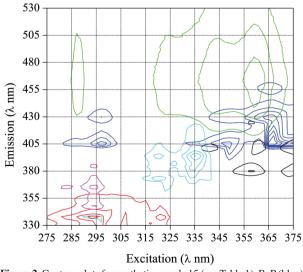

Figure 2. Contour plots for synthetic sample 15 (see Table 1). BaP (blue); ACL (red); PYR (cyan); ANT (black); PHN (magenta) and FLT (green).

an arrow), which is related to the six target PAHs. Thus, the three-way EEM array was decomposed by PARAFAC, under non-negativity constraint, with six factors.

The retrieved spectral profile of BaP in PARAFAC modeling presents great similarity to the recorded spectrum (Figure 3b). This suggests that the BaP profile was correctly retrieved and can be quantified in the presence of the ACL, PYR, ANT, PHN and FLT interferents; quenching and/or inner filter effects were not detected.

Beyond simple visual comparison, it is also possible to estimate the similarity between the recorded (s_1) and retrieved (s_2) spectra by means of the degree of overlap (S_{12}), defined as equation 2:[32]

$$S_{12} = \frac{\|s_1^T s_2\|}{\|s_1\| \|s_2\|} \tag{2}$$

$\|.\|$ denotes the Euclidian norm, and the T superscript indicates a transpose of vector s_1. It is important to remember that S_{12} is the cosine of the angle between vectors s_1 (recorded spectrum) and s_2 (retrieved spectrum). If the angle between s_1 and s_2 is zero, the cosine is 1, i.e., the recovered spectrum is exactly the same as recorded. However, if S_{12} is zero, this suggests that the vectors s_1 and s_2 are orthogonal, indicating that the recorded and retrieved spectra are completely different. In this case, the obtained values for S_{12} were 0.9917 and 0.9969 for excitation and emission spectra, respectively, indicating a great similarity between recorded and PARAFAC model retrieved profile, suggesting that the correct analyte profile was retrieved.

Afterwards, BaP quantitation was carried out by least squares (LS) fit between the nominal concentration in standards and the PARAFAC scores. This model is generally referred to as a pseudo-univariate calibration.[20] Using this approach with validation set, a root mean square error (RMSE) of 1.84 ng mL⁻¹ was found, which corresponds to a relative prediction error (REP) of 3.34%. This suggests that the proposed methodology achieved good accuracy in the presence of five target interferents.

In Figure 3c is shown the predicted *versus* nominal concentration plot, where the points are close to the ideal line (bisectrix) in agreement with the low RMSE and REP obtained values.

In Figure 3d is displayed the elliptical joint confidence region (EJCR) that corresponds to the joint confidence interval for the slope and intercept of the linear fit between the nominal and predicted concentrations (see Figure 3c). As can be seen, the EJCR contains the ideal point (1 and 0), this means that the PARAFAC model has no bias at 95% confidence level. In addition, a model with good fit allows the development of methodologies with favorable figures of merit. The observed sensitivity (SEN)
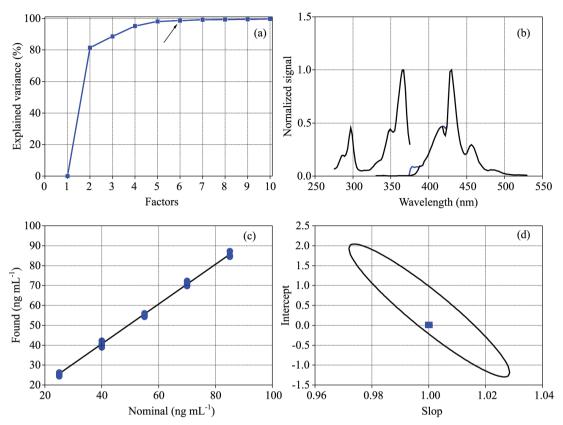
Figure 3. (a) Plots of explained variance *versus* number of factors; (b) recorded (blue line) and retrieved (black line); (c) excitation-emission spectra (both normalized) of BaP; predicted *versus* nominal concentration; (d) elliptical joint confidence regions (EJCR).

and analytical sensitivity (γ^{-1}) were, respectively, 2.02×10^4 and 0.20 ng mL^{-1}, allowing an estimate for a limit of detection (LOD) of 3.85 ng mL^{-1}, which is smaller than the recommended maximum value of BaP given by the European Commission for foodstuffs (10 ng g^{-1}),[10] which was used due to the lack of established limit for beverages.

Cachaça analysis

In this work, the analyzed cachaças were all produced in copper alembic stills, with alcoholic grade ranging between 38 and 42%. Sample C1 is commercially sold in aluminum cans and the other cachaças in glass bottles. In Figure 4, the contour plot for the six analyzed cachaças is displayed.

As can be seen in Figure 4, cachaça presents fluorescence that interferes with the BaP signal (see Figure 1b). This suggests that in order to use a univariate calibration approach, it would be necessary to perform an extraction or other cleanup procedure to circumvent the presence of these interferents. Such sample treatments may lead to analyte losses that result in worse recoveries. Moreover, when using such procedures and different samples are analyzed, unexpected constituents may appear and interfere in BaP quantitation. Therefore, to overcome these drawbacks, the EEM three-way array of cachaças were decomposed by PARAFAC under the same

conditions as described for the validation set. Yet, due to the complexity of the matrix, the number of factors was re-evaluated by general inspection of the explained variance and CORE consistency (CORCONDIA) diagnostic.[33] This result is shown in Figure 5a.

It is observed in Figure 5a that after the third factor (indicated by an arrow), explained variance does not increase and the CORE consistency value remains very close to 100%. From the fourth factor, the CORE consistency value is below 50%, indicating low trilinearity for the data. Therefore, three factors were chosen for the PARAFAC cachaça models.

The excitation and emission profiles retrieved by PARAFAC are, respectively, displayed in Figures 5b and 5c. The maximum fluorescent intensities were observed at 287, 331 and 353 nm in excitation, and 330, 372, and 432 nm in emission modes. In Figure 5d, the scores for the PARAFAC models are shown, which are directly related to concentration. It was possible to identify these three constituents (fluorophores), found in the analyzed cachaça in different amounts. This identification does not impair BAP quantification thanks to the second-order advantage of the multiway calibration that allows the quantification of the analyte even in the presence of uncalibrated constituents as demonstrated by the recovery study.
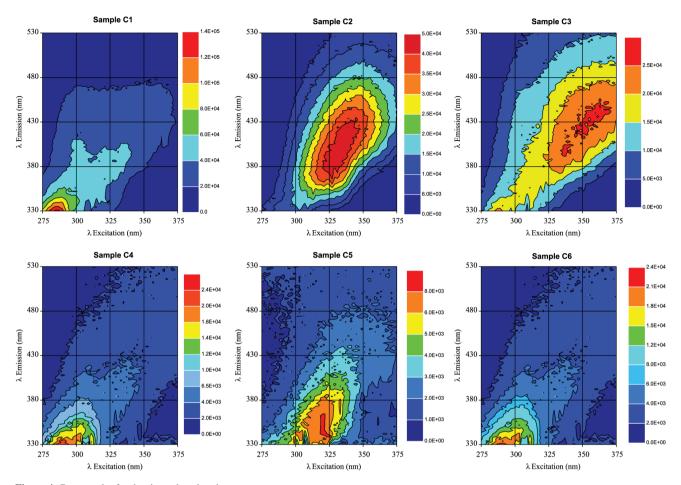
Figure 4. Contour plot for the six analyzed cachaças.

To assess the accuracy of the proposed method, a recovery study was carried out. For this purpose, the cachaças were spiked at three concentration levels, as described in the "Recovery study in Cachaça" sub-section and results obtained are summarized in Table 2. As seen in Table 2, the predicted concentrations are in accordance with the BaP spikes, and the recovery ranged from 93.20 to 101.13%. The RMSE and REP for BaP quantitation in the cachaça were, respectively, 1.46 ng L^{-1} and 2.66%. It is important to mention that the relative standard deviation (RSD) value for each spike level *per* sample suggests a methodology with good reproducibility and without matrix effect.

Conclusions

A new method with no sample handling was proposed for quantification of BaP in cachaças. Interferents were overcome by PARAFAC EEM data modeling, and the second order advantage was successfully explored. This new approach for cachaça quality control does not suffer from inner filter or quenching effects for PAHs at the studied concentration levels, and no matrix effect

was identified in the cachaças analyzed. In addition, no reagents, solvents, extraction, pre-concentration or cleanup steps were used, which made the procedure less time consuming, and cheaper than other approaches described in the literature, it also avoided analyte losses, and thus led to better recoveries. The proposed method showed good accuracy, reproducibility and a smaller LOD than the maximum value recommended by an international regulatory agency. Therefore, the proposed methodology has potential to be extended for analyses of other distilled spirits such as vodka, rum and whiskey, etc., due to the second order advantage of the EEM data modeling by PARAFAC.
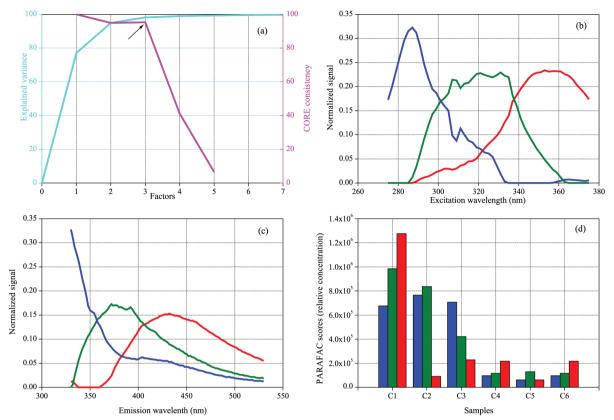
Acknowledgments

Figure 5. Results for cachaça with PARAFAC decomposition. (a) CORCONDIA (■) and the explained variance (■) *versus* number of factor plots; (b) excitation; (c) emission; (d) profiles retrieved by PARAFAC and relative concentration for factor 1 (■), factor 2 (■) and factor 3 (■).

Table 2. BaP prediction in spiked cachaças

Sample	Taken / (ng mL⁻¹)	Found / (ng mL⁻¹)	RSD / %	REC / %
C1	–	> LOD	–	–
	25.00	23.96	2.20	95.84
	55.00	52.00	3.12	94.54
	85.00	83.80	1.64	98.59
C2	–	> LOD	–	–
	25.00	24.53	4.72	98.12
	55.00	53.13	0.20	96.6
	85.00	84.93	2.15	99.92
C3	–	> LOD	–	–
	25.00	23.67	3.37	94.68
	55.00	53.41	1.40	97.11
	85.00	85.96	2.12	101.13
C4	–	> LOD	–	–
	25.00	23.3	0.24	93.20
	55.00	52.66	1.74	95.74
	85.00	81.81	0.69	96.25
C5	–	> LOD	–	–
	25.00	24.41	3.77	97.64
	55.00	53.22	1.10	96.76
	85.00	84.10	0.01	98.94
C6	–	> LOD	–	–
	25.00	24.34	1.93	97.36
	55.00	54.00	2.98	98.18
	85.00	82.05	0.65	96.53

RSD: relative standard deviation; REC: recovery; LOD: limit of detection.

References

1. German Federal Environment Agency (UBA); *Polycyclic Aromatic Hydrocarbons: Harmful to the Environment! Toxic! Inevitable*? Available at http://www.bsnc.nl/wp-content/uploads/2015/10/Polycyclic-Aromatic-Hydrocarbons-why-the-ban.pdf, accessed on July 15, 2018.

2. World Health Organization (WHO); *Air Quality Guidelines*, 2ⁿᵈ ed.; WHO Regional Office for Europe: Copenhagen, 2000.

3. Abdel-Shafy, H. I.; Mansour, M. S. M.; *Egypt. J. Pet.* **2016**, *25*, 107.

4. Houessou, J. K.; Benac, C.; Delteil, C.; Camel, V.; *J. Agric. Food Chem.* **2005**, *53*, 871.

5. Rengarajan, T.; Rajendran, P.; Nandakumar, N.; Lokeshkumar, B.; Rajendran, P.; Nishigaki, I.; *Asian Pac. J. Trop. Biomed.* **2015**, *3*, 182.

6. Galinaro, C. A.; Cardoso, D. R.; Franco, D. W.; *J. Agric. Food Chem.* **2007**, *55*, 3141.

7. Riachi, L. G.; Santos, Â.; Moreira, R. F. A.; Maria, C. A. B.; *Food Chem.* **2014**, *149*, 159.

8. Kumar, K.; Mishra, A. K.; *Talanta* **2013**, *117*, 209.

9. Anastasio, A.; Mercogliano, R.; Vollano, L.; Pepe, T.; Cortesi, M. L.; *J. Agric. Food Chem.* **2004**, *52*, 4452.

10. European Commission (EC); *Commission Regulation (EC) No. 1881/2006 of 19 December 2006 Setting Maximum Levels*

for Certain Contaminants in Foodstuffs, Official Journal of the European Union, Brussels, 2006. Available at http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:364:00 05:0024:EN:PDF, accessed on July 15, 2018.

11. Menezes, H. C.; Paulo, B. P.; Paiva, M. J. N.; Barcelos, S. M. R.; Macedo, D. F. D.; Cardeal, Z. L.; *Microchem. J.* **2015**, *118*, 272.

12. Tfouni, S. A. V.; Machado, R. M. D.; Camargo, M. C. R.; Silvia, H. P.; Vitorino, S. H. P.; Vicente, E.; Toledo, M. C. F.; *Food Chem.* **2007**, *101*, 334.

13. Cacho, J. I.; Campillo, N.; Viñas, P.; Hernández-Córdoba, M.; *Food Chem.* **2016**, *190*, 324.

14. Porto, C.; Moret, S.; *Food Chem. Toxicol.* **2007**, *45*, 2069.

15. Ding, Y. S.; Ashley, D. L.; Watson, C. H.; *J. Agric. Food Chem.* **2007**, *55*, 5966.

16. Rodriguez-Acuña, R.; Perez-Camino, M. C.; Cert, A.; Moreda, W.; *J. Agric. Food Chem.* **2008**, *56*, 10428.

17. Raters, M.; Matissek, R.; *J. Agric. Food Chem.* **2014**, *62*, 10666.

18. Elcoroaristizabal, S.; Juan, A.; García, J. A.; Durana, N.; Alonso, L.; *Chemom. Intell. Lab. Syst.* **2014**, *132*, 63.

19. Airado-Rodriguez, D.; Galeano-Diaz, T.; Duran-Meras, I.; Wold, J. P.; *J. Agric. Food Chem.* **2009**, *57*, 1711.

20. Olivieri, A. C.; *Anal. Methods* **2012**, *4*, 1876.

21. Cañas, A.; Richter, P.; Escandar, G. M.; *Anal. Chim. Acta* **2014**, *852*, 105.

22. Goicoechea, H. C.; Calimag-Williams, K.; Campiglia, A. D.; *Anal. Chim. Acta* **2012**, *717*, 100.

23. Elcoroaristizabal, S.; Juan, A.; García, J. A.; Elorduy, I.; Durana, N.; Alonso, L.; *J. Chemom.* **2014**, *28*, 260.

24. Bro, R.; *Chemom. Intell. Lab. Syst.* **1997**, *38*, 149.

25. Baunsgaard, D.; Nørgaard, L.; Godshall, M.; *J. Agric. Food Chem.* **2000**, *48*, 4955.

26. Andersen, C. M.; Bro, R.; *J. Chemom.* **2003**, *17*, 200.

27. Andersson, C. A.; Bro, R.; *Chemom. Intell. Lab. Syst.* **2000**, *52*, 1.

28. Olivieri, A. C.; Wu, H.; Yu, R.; *Chemom. Intell. Lab. Syst.* **2009**, *96*, 246.

29. http://www.iquir-conicet.gov.ar/esp/div5.php?area=12, accessed on July 15, 2018.

30. http://www.models.life.ku.dk/algorithms, accessed on July 15, 2018.

31. Bahram, M.; Bro, R.; Stedmon, C.; Afkhami, A.; *J. Chemom.* **2006**, *20*, 99.

32. Moreira, E. D. T.; Pinto, L.; Gomes, A. A.; Goicoechea, H. C.; Araújo, M. C. U.; *J. Braz. Chem. Soc.* **2015**, *26*, 1573.

33. Bro, R.; Kiers, H. A. L.; *J. Chemom.* **2003**, *17*, 274.