CHEMOMETRIC DISCRIMINATION BETWEEN SMOKED AND NON-SMOKED PAPRIKA SAMPLES. QUANTIFICATION OF PAHs IN SMOKED PAPRIKA BY FLUORESCENCE-U-PLS/RBL

Olga Monago-Maraña, Teresa Galeano-Díaz* and Arsenio Muñoz de la Peña.

Department of Analytical Chemistry and Research Institute on Water, Climate Change and Sustainability (IACYS), University of Extremadura, Badajoz 06006, Spain.

* Corresponding author. E-mail address: tgaleano@unex.es, phone: +34 924289779

Abstract

This study presents a strategy for differentiating paprika obtained by means of different drying systems. The differentiation is performed using spectroscopic fluorescence in combination with multivariate analysis. The two groups of samples (smoked or non smoked paprika) are classified according to the content of some of their fluorescent compounds presented in each group, among which several polycyclic aromatic hydrocarbons (PAHs) are included. These compounds are characteristic in smoked food. The full information of excitation – emission matrices (EEMs) is processed with the aid of unsupervised parallel factor analysis (PARAFAC), PARAFAC supervised by linear discriminant analysis (LDA), and discriminant unfolded partial least-squares (DU-PLS). The last algorithm allows an adequate classification of unknown paprika samples. Besides, the quantification of several PAHs in paprika was performed by means of unfolded partial least-squares with residual bilinearization (U-PLS/RBL). On this way, three (fluorene, phenantrene and anthracene) out of the five (fluorene, phenantrene, anthracene, pyrene and chrysene) selected analytes were quantified.

Keywords: paprika, classification, (PARAFAC)-LDA, DU-PLS, U-PLS/RBL, fluorescence

1. Introduction

Food smoking is an old and traditional technological process widely applied to many foodstuffs such as meat, fish and cheese, not only for the special organoleptic profiles that it confers, but also due to the inactivating effect of smoke and heat on enzymes and microorganisms (Ledesma et al. 2015). Today, smoking technology mainly uses the special effects of various sensory active components (phenol derivatives, carbonyls, organic acids and their esters, lactones, pyrazines, pyrols and furan derivatives), contained in smoke, for aromatization of meat products, to make food with a specific organoleptic profile, widely demanded on the market (Simko 2002).

Paprika is a product obtained from dehydrated and milled fruits of certain varieties of red peppers (*Capsicum annum L.*). There are different drying systems to obtain this product. Thus, for example, in Spain, there are two main areas where this product is obtained, in La Vera (Extremadura) and Murcia. In the first one, peppers are smoked-dried (oak or holm wood fire), while in Murcia, among other places, peppers are sundried (Bartolomé et al. 2011).

Smoking process provided to paprika samples a characteristic flavour and smell. However, this kind of treatment may produce the presence of unwanted compounds in food, such as polycyclic aromatic hydrocarbons (PAHs), which present carcinogenic, mutagenic and bioaccumulative capacities (Purcaro et al. 2013).

Although there are several kinds of pattern recognition methods to be applied in food science, they essentially differ in the way they achieve the classification. Two main types of methods are commonly distinguished: those focused on discrimination among classes, for example, linear discriminant analysis (LDA) or discriminant unfolded partial least-squares (DU-PLS); and those oriented towards modelling classes, such as soft independent modelling of class analogy (SIMCA), among others. Discriminating techniques are used to build models based on all the categories concerned in the discrimination, whereas disjoint class-modelling methods create a separate model for each category. One of the drawbacks of discriminating methods is that samples are always classified into one of the given categories, even if they do not belong to any of them.

Class-modelling methods consider those objects that fit the model for a category as part of the model, and classify as non-members those that do not (Berrueta et al. 2007).

These techniques has been amply employed in the classification of food samples according to their physical and chemical properties, their production processes, their spectroscopic properties and so on. In this sense, fluorescence coupled with these multivariate analysis techniques have been commonly used in the last years in the food classification (Berrueta et al. 2007; Sádecká and Tóthová 2007; Sikorska et al. 2008; Azcarate et al. 2015; Borrás et al. 2015; Da Silva et al. 2015; Ledesma et al. 2015; Lenhardt et al. 2015; Sahar et al. 2016). Specifically, chemometric techniques have been employed in the authentication and determination of contaminants in condiments, where paprika is included. However, no studies are found about classification according to the drying system of paprika (Di Anibal et al. 2015; Reinholds et al. 2015). Hitherto, fluorescence coupled to PARAFAC-LDA and DU-PLS for food sample classification have been used in very few studies (Azcarate et al. 2015).

On the other hand, if we focus on the use of spectroscopic techniques in combination with chemometric algorithms to quantify PAHs, we found several recent examples of quantification of PAHs in food and drinks. In the last years, Bortolato et al. 2008 (Bortolato et al. 2008) have quantified benzo(a)pyrene and dibenzo[a,h]anthracene in waters, by means of excitation – emission fluorescence spectroscopy assisted by chemometrics; Ferreto et al. 2014 (Ferretto et al. 2014) have also quantified five PAHs in marine water using excitation – emission matrices (EEMs) and parallel factor analysis (PARAFAC), and Alarcón et al., 2013 (Alarcón et al. 2013) have determined PAHs, by means of EEMs, unfolded partial least-squares/residual bilinearization (U-PLS/RBL), and PARAFAC, in edible oils. However, in the case of paprika samples, no studies have been found with these techniques.

With this background, the aims of this study were investigating the usefulness of chemometrics in order to differentiate paprika samples according to their drying system and, taking into account the presence of PHAs in smoked paprika, quantifying them in this kind of samples, by means of EEMs, in combination with multivariate chemometric tools.

2. Materials and methods

2.1. Chemical reagents and samples

Stocks of PAHs (Fluorene (Flu), Phenantrene (Phe), Anthracene (Ant), Pyrene (Pyr) and Chrysene (Chr)) were obtained from Sigma (Sigma-Aldrich Química, S.A., Madrid). Each individual standard solution was prepared in acetonitrile (ACN) and stored at 4 ºC until use.

LC-grade acetonitrile solvent was purchased from Sigma (Sigma-Aldrich Química, S.A., Madrid). LC-grade iso-hexane and diethyl ether were acquired from Panreac (Panreac Química, S.A.U., Barcelona). High-purity water was obtained from a Milli-Q water system (Millipore S.A.S., Molsheim, France). Sep-Pak Plus Silica cartridges of 690 mg were obtained from Waters (Waters Corp., Milford, MA, USA).

Samples of smoked paprika sample are part of the Spanish Protected Designation of Origin (PDO) *"Pimentón de La Vera"* and they were obtained from Regulatory Council of the Denomination of Origin *"Pimentón de La Vera"* and the non-smoked paprika samples were obtained from local markets. The origin of the non-smoked paprika samples was not available although in the label reports packaging in Spain.

2.2. Instrumentation and software

In order to obtain the fluorescence excitation-emission matrices, a Cary Eclipse VARIAN spectrofluorimeter equipped with two Czerny-Turner monochromators, a xenon light source and a photomultiplier tube, as detector, was employed. A 1.0 cm quartz cell was used. Data acquisition was performed with the Cary Eclipse software.

The software package The Unscrambler® v6. 11 (CAMO A/S Olav Tryggvasonsgt, N-7011, Trondheim, Norway) was used for the experimental design.

Second order analysis of data was done using MatLab R2008a (MATLAB Version 7.6, The Marhworks, Natick, Massachusetss, 2010) and the MVC2 routines developed by Oliveri, Wu and Yu (Olivieri et al. 2009). An in house MatLab routine was used for LDA calculations (Kemsley 1998).

2.3. Fluorescence excitation-emission matrices

To obtain fluorescence excitation-emission matrices (EEMs), excitation wavelengths were increased from 230 to 350 at 5 nm steps; for each excitation wavelength, the emission spectrum was obtained in the range 270-500 nm at 1 nm steps. The instrumental parameters used were as follow: photomultiplier voltage of 550 V and slit widths of 5 nm.

2.4. Calibration and test sets for U-PLS/RBL analysis

To assess the ability of the U-PLS/RBL model in the determination of a mixture of PAHs in paprika, a 18-standards set was built for Flu calibration, and a 22-standards set was built for Phe, Ant, Pyr and Chr calibration. The analyte concentrations were corresponded with a Fractional Factorial Design and they were between 0 – 40 µg L^{-1} for Flu, 0 – 150 µg L^{-1} for Phe, between 0 – 40 µg L^{-1} for Ant, between 0 – 40 µg L^{-1} for Pyr and between 0 – 15 µg L^{-1} for Cry. Samples were prepared in acetonitrile taking the corresponding volume of the stock solutions.

Moreover, a set of 15 samples were prepared for validation of the method, with concentrations different from those employed for calibration, but within their corresponding calibration ranges. EEMs were measured as it is indicated in the section 2.3.

2.5. Pretreatment of sample

In order to extract the analytes from paprika samples, 0.2 g precisely weighed aliquot of this product was extracted with 10 mL of diethyl ether for 10 min in an ultrasonic bath. The extract solution was centrifuged for 10 min and evaporated to dryness. The residue was suspended in 5 mL of iso-hexane and loaded on a silica cartridge. Then the PAHs were eluted from the cartridge with 7 mL of iso-hexane. This extract together with the 5 mL fraction initially percolated were combined, evaporated to dryness and reconstituted in 5 mL of ACN. In the case of smoked paprika a dilution was employed before registering EEMs, however, the non-smoked samples were registered without dilution.

2.6. Chemometric algorithms

2.6.1. PARAFAC

PARAFAC is one of several decomposition methods for multi-way data, which decompose the array into sets of scores and loadings that hopefully describes the data in a more condensed form than the original data array (Bro 1997). Because of the multi-way nature of the data, and the particular constraints of the PARAFAC model, the solution is unique. What this means in a practical application is that, ideally, the loading of each factor in each mode represents a pure component contribution to the fluorescence of the mixture (the fluorescent components recovered by PARAFAC may actually represent discrete species, covarying species, interacting pairs or sets of species, or instrumental artefacts). The number of components found are, therefore, only approximately equal to the actual number of fluorescent chemical species (Hall and Kenny 2007). A PARAFAC model of a three-way array is given by three loading matrices, A, B and C with elements a_{in}, b_{jn}, c_{kn}, respectively, where n indicate the component number (Bro 1997). The trilinear model is found to minimize the sum of squares of the residuals, $e_{ijk,}$ in the model

$$x_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} + e_{ijk} \qquad (1)$$

where x_{ijk} is the fluorescence intensity for sample i at the emission wavelength j and excitation wavelength k and e_{ijk} indicates an element of the array E, which collects the variability not accounted by the model. For a given component n, the elements a_{in}, b_{jn} and c_{kn} are arranged in the score vector a_n (whose elements are directly proportional to its concentration in each sample) and the loading vectors b_n and $c_{n,}$ which estimate its emission and excitation profiles. The array of EEMs data is fitted to eq. 1 by least-squares.

2.6.2. LDA

LDA is probably the most frequently supervised pattern recognition method used. It is based on the determination of linear discriminant functions, which maximize the ratio of between-class variance and minimize the ratio of within-class variance using linear combinations of the original variables to achieve class discrimination (Berrueta et al. 2007; Borrás et al. 2015; Muñoz de la Peña et al. 2016).

In LDA, categories are supposed to follow a multivariate normal distribution and be linearly separated. LDA can be considered, as PCA, as a feature reduction method in the sense that both, LDA and PCA, determine a smaller dimension hyperplane on which the points will be projected from the higher dimension. However, whereas PCA selects a direction that retains maximal structure among the data in a lower dimension, LDA selects a direction that achieves maximum separation among the given classes. The latent variable obtained in LDA is a linear combination of the original variables. This function is called canonical variate (CV), ant its values are the roots. Being k classes, k-1 canonical variates can be determined if the number of variables is larger than k (Berrueta et al. 2007).

With the A score matrix of PARAFAC and the I x g dummy matrix Y of binary digits representing the group assignments (g is the number of categories), the best representation is obtained if the ratio of the between-class variance Bc matrix and the within-class variance Wc matrix is maximized. Suitable expressions for the matrices Bc and Wc are given by the following expressions (Arruda et al. 2003):

$$B_C = (g-1)^{-1} A^T Y (Y^T Y)^{-1} Y^T A \qquad (2)$$

$$W_c = (I-g)^{-1} [A^T A - (g-1) B_c] \qquad (3)$$

The canonical variate (CV) scores contain the successively maximized ratios between-groups variance/within-groups variance. They are obtained by PCA of the matrix (Wc^{-1} Bc) and projection of the data matrix A onto the first loadings. The samples are then plotted on a two- or three-dimensional space defined by the first CV scores of each sample.

2.6.3. DU-PLS

U-PLS was originally developed for multivariate calibration purposes (Indahl 2014; Azcarate et al. 2015), however, it has been also employed for the classification of samples. The main difference between U-PLS and discriminant U-PLS (DU-PLS) consists in the building of the dependent variable y. For model calibration purposes, the variable y contains concentration values. For discriminant analysis purposes, y contains a coding integer representing the class label of the samples. PLS regression is conducted between the instrumental response in X block (built with the unfolded original second-order matrix data) and the class label in y block using training samples, and the optimal number of latent variables is chosen based on the error range by cross-validation. The final model for A latent variables is used to predict the class label in the test set according to the following:

$$y_{test} = {t_{test}}^{T} v \qquad (4)$$

where y_{test} is the label class predicted, t_{test}^{T} are the scores of test samples obtained by projection of x_{test} onto the training loadings, and v is the vector of the regression coefficients. In the ideal case scenario, the calculated values of y_{test}, for two classes of samples, are 1 or 2; in practice, y_{test} values are often close to 1 or 2. Therefore, in order to assign a test sample to a given class, it is necessary to establish thresholds for the y_{test} predicted values. The threshold is defined as the value that minimizes the number of false positives and false negatives.

3. Results and discussion

3.1. Preliminary considerations

Taking into account a previous study performed (data send to publish), with the sample treatment described in the section 2.4., it can be secured that PAHs are present in smoked paprika extracts. For this reason, the target analytes in this study were the majority PAHs present in paprika samples: Fluorene, Phenantrene, Anthracene, Crysene and Pyrene. EEMs of each PAHs were registered with the selected conditions indicated in the section 2.3. and they are shown in the Figure 1. Besides, in this figure, EEMs of a smoked and a non-smoked paprika samples are shown. It can be observed that smoked paprika presents fluorescence intensity in the same zone than PAHs. The PARAFAC, PARAFAC-LDA and DU-PLS analysis which are shown in the after sections follow the same strategies than Muñoz de la Peña et al. (Muñoz de la Peña et al. 2016).

3.2. PARAFAC analysis

Twelve EEMs of each group of paprika samples studied were registered in the selected conditions, as it is indicated in the section 2.3. Spectral decomposition of EEMs was performed via PARAFAC with all matrices registered. PARAFAC was first applied without supervision. Non-negativity constraints were applied on all three modes for the estimation of the model.

The number of principal components was estimated according to the core consistency diagnostic (CORCORDIA) (Bro and Kiers 2003) and the analysis of residuals (Bro 1997). Thus, the number of optimum components was four. Figure 2 shows the excitation – emission loadings corresponding to the different components found. According to the shape of the different loadings, only the first one could be related with a combination of the different PAHs, which exhibit fluorescence intensity in this zone. The fourth loading presents fluorescence intensity in the same zone that Fluorene, but the shape of the EEM does not correspond with Fluorene EEM.

Taking into account that four components were the optimum, scores of one of these four components was removed to make the corresponding plots. The removal order was: firstly, the scores corresponding to the fourth component, secondly, the scores corresponding to the third component, thirdly, the scores corresponding to the second component and, finally, the scores corresponding to the first component. In all cases, the samples were clustering in two groups. Figure 3 shows the tridimensional plots of PARAFAC scores of 1, 2 and 3 components, such as an example of the classification, for each group of samples investigated. Besides, the projections of the 95% ellipses over the different planes defined by the corresponding axes to offer a better visualization of the formed groups. The prediction interval for the multivariate normal distribution yielded an ellipse consisting of x vectors satisfying the following equation:

$$(x - \mu)^T \Sigma^{-1}(x - \mu) \leq \chi_k^2(p) \qquad (5)$$

where μ is the mean, Σ is the covariance matrix and $\chi^2{}_k(p)$ is the quantile function for probability p of the χ^2 distribution with k degrees of freedom, where k is the dimension of the data. The axes are defined by the eigenvectors of the variance matrix and the radius of each axis is equal to 2.796 times the square root of the corresponding eigenvalue. The value 2.796 is obtained from the square root of the χ^2 distribution with three degrees of freedom and 95 % confidence interval (Slotani 1964).

In a previous study, one differentiation was performed due to the fluorescence signal of paprika sample without treatment (Monago Maraña et al. 2016). However, the differentiation could not be attributed to the same components because the sample treatment was different and the loading shape was also different. In this case, it is known that some of components present in this extract are PAHs, furthermore, these compounds exhibit fluorescence in the working excitation – emission wavelengths.

3.3. PARAFAC-LDA

Usually, applying a supervised technique, as LDA is, improves the screening capabilities (Muñoz de la Peña et al. 2016). In this case, the results obtained for the discrimination between smoked and non-smoked paprika were similar to the previous case (PARAFAC). In the Figure 4, it is shown these results obtained, with the same procedure that in the previous case, removing the scores corresponding to one of four each time. Two clearly defined clusters appears in both regions, one corresponding to the smoked paprika and other one corresponding to the non-smoked paprika samples.

No significant differences are found respect the PARAFAC analysis. Also, it can be said that there is a clear difference between both groups according to the first component, which was previously related to the presence of PHAs. Thus, it is a fact that both groups can be differentiated by the presence of PAHs in the case of smoked paprika because of these compounds are formed in the smoked drying system.

3.4. DU-PLS

In the case of DU-PLS, the regions employed were the same that the previous cases. The number of optimum latent variables (h) was estimated via the leave-one-sample-out cross-validation approach (Haaland and Thomas 1988) using a 24-samples set (12 of each group of paprika samples studied). The optimum number of latent variables were those corresponding to the model given a PRESS value (PRESS value is defined as PRESS = Σ ($c_{i,act}$ − $c_{i,pred}$)^2) statistically no different to the minimum PRESS value (F-ratio probability falling below 0.75). Hence, one factor was found. This fact could mean that samples are differentiated according to one of the components present in them. For the discriminant analysis, the variable y of the model contains a coding integer representing the class label of the sample. In this case, the labels were 1 or 2. However, when unknown samples are predicted, they are classified as 1 or 0. It can be explained due to the fact that only one component was found as optimum, so the model predicts the samples as the presence or not of this component. A good prediction of the unknown samples was found, as can be observed in the Figure 5. Hence, this strategy can be useful to predict if some samples have been smoked dried or not. The confidence interval for each category was estimated as the product of the calculated standard deviations of the results for the training samples and the Student t-value with n-1 degrees of freedom for each category. These confidence intervals were 1.09 ± 0.33 and 0.06 ±0.15 for smoked and non-smoked categories, respectively. In the case of training samples, 100 % of smoked paprika samples and 92 % (11 out of 12) of non-smoked paprika samples were well classified. For unknown samples, 88 % of smoked paprika samples and 100 % of non – smoked paprika samples were correctly classified.

3.5. U-PLS/RBL analysis

Because the presence of PAHs in smoked paprika samples has been demonstrated, the quantification of these analytes (Flu, Phe, Ant, Pyr and Chr) using multiway chemometrics was intended. Thus, U-PLS/RBL algorithm was employed to achieve this aim.

Taking into account the region of fluorescence of each compound (Figure 1), two initial regions were stablished. One corresponding to the analysis of Flu, and another one corresponding to the rest of analytes.

Thus, two calibration sets were constructed. In the case of Flu, a set of 18 calibration samples were employed and, in the case of Phe, Ant, Pyr and Chr, a set of 22 calibration samples was employed, as it is described in the section 2.4. The range of each calibration curve was chosen according to the real concentration determined in the samples by means of a LC-FLD method previously developed (data send to publish).

Firstly, the cross-validation and the Haaland and Thomas criterion (Haaland and Thomas 1988) was used to choose the optimum number of factors as it was said before, in the previous section.

With the aim of validating the proposed method, a set of tests samples containing a mixture of Phe, Ant, Pyr and Chr, in the same range of concentrations that the calibration samples, were analysed. In the case of Flu, it was not necessary to build a validation set because of it was the only analyte present in its range of calibration. In the case of Phenantrene, the range of wavelengths to quantify it (λ_{exc} = 320 – 340 nm, λ_{em} = 350 – 400 nm) was chosen according to the selectivity of this range, with the aim to avoid the presence of matrix interferences in real samples. Table 1 shows the optimum number of factors for each analyte, in their range of wavelengths. Also, in the Table 2, figures of merit of this methodology are shown (Olivieri and Escandar 2000).

In order to get further insight into the accuracy and precision of the algorithm analyzed, nominal versus found concentration values of the test samples were compared by application of the EJCR (Elliptical Joint Confidence Region) test (Riu and Rius 1997; Del Rio et al. 2001). The corresponding plots are shown in the Figure 6. The prediction values for all analytes are in good agreement with the nominal values. Besides, all confidence regions contain the ideal point of unit slope and zero intercept (indicating accuracy). These results are confirmed with the statistical results, with very satisfactory values for the root mean square error of prediction (RMSEP) and relative error of prediction (REP) for the four analytes taking into account other similar studies (Bortolato et al. 2008; Alarcón et al. 2013). These results were 2.9 (Phe), 1.1 (Ant), 1.1 (Cry) µg mL^{-1} and 0.70 (Pyr) for RMEP and 4 (Phe), 5 (Ant), 5 (Cry), 7 (Pyr) % for REP. Taking into account these good results, this methodology was employed for the quantification of these analytes in real paprika samples.

In this case, it was necessary to assess a number of unexpected components to be employed in the RBL procedure (Olivieri and Escandar 2000), taking into account the presence of matrix interferences, as it can be appreciated in the Figure 1. This number of unexpected components was different according to the analyte. The new factors are shown in Table 1.

In the case of Flu, Phe and Ant, good results were found and their concentrations were well-correlated with those found by a LC-FLD method, previously developed. However, in the case of Pyr and Chr, only 6 or 7 samples were well-correlated. This fact could be due to the low concentration of these analytes and the presence of the interferences. Table 3 shows the correlation between results obtained by both methods. These results corresponding to the smoked samples.

In order to stablish the LOD and LOQ for real samples, a non-smoked sample, whit a low concentration of PAHs was extracted according to the described procedure. The procedure was applied five times with the same sample, and the concentration of each analyte was predicted with these algorithms. The limit of detection (LOD) and quantification (LOQ) were calculated as three and ten times the standard deviation of the different extractions, respectively. With this, the LOD of this method and samples, for the different analytes, were 2 μg L^{-1} (Flu), 18 μg L^{-1} (Phe), 4 μg L^{-1} (Ant), 18 μg L^{-1} (Pyr) and 12 μg L^{-1} (Cry) and the LOQ were 8 μg L^{-1} (Flu), 60 μg L^{-1} (Phe), 13 μg L^{-1} (Ant), 60 μg L^{-1} (Pyr) and 40 μg L^{-1} (Cry). These samples were register without a previous dilution due to their low concentration. Taking into account these results, only the smoked samples were quantified, because the non-smoked samples presented PAHs concentrations lower than LOQ of the method.

4. Conclusions

EEMs in combination with different chemometric tools have been employed to demonstrate the successful discrimination between paprika samples obtained by different drying systems. On the one hand, PARAFAC (unsupervised technique) has allowed discriminating and classifying paprika samples. Also, on the other hand, good results have been obtained with PARAFAC-LDA (supervised technique). In the case of DU-PLS, its ability to distinguish smoked or non-smoked paprika was assayed and unknown samples were well-classified.

Finally, a method based on EEMs coupled to U-PLS/RBL has been employed to quantify Fluorene, Phenantrene and Anthracene in smoked paprika samples. Results obtained showed good correlations with a previous developed LC-method.

Acknowledgments

Financial support was provided by the Ministerio de Economía y Competitividad of Spain (Project CTQ2014-52309-P) and the Junta de Extremadura (GR15090-Research Group FQM003), both co-financed by the European FEDER funds. Olga Monago Maraña is grateful to the Ministerio de Educación, Cultura y Deporte of Spain for a FPU grant (Resolución de 18 de noviembre de 2013, de la Secretaría de Estado de Educación, Formación Profesional y Universidades, BOE n◦ 279, de 21/11/13, reference number FPU13/02249).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alarcón F, Ɓaez ME, Bravo M, et al (2013) Feasibility of the determination of polycyclic aromatic hydrocarbons in edible oils via unfolded partial least-squares/residual bilinearization and parallel factor analysis of fluorescence excitation emission matrices. Talanta 103:361–370. doi: 10.1016/j.talanta.2012.10.080

Arruda A, Goicoechea HC, Santos M, et al (2003) Solid-liquid extraction room temperature phosphorimetry and pattern recognition for screening polycyclic aromatic hydrocarbons and polychorinated biphenyls in water samples. Environ Sci Technol 37:1385 – 1391.

Azcarate SM, De Araújo Gomes A, Alcaraz MR, et al (2015) Modeling excitation-emission fluorescence matrices with pattern recognition algorithms for classification of Argentine white wines according grape variety. Food Chem 184:214–219. doi: 10.1016/j.foodchem.2015.03.081

Bartolomé T, Coleto JM, Velázquez R (2011) Pimentón de la Vera: un caso paradigmático de la Denominación de Origen Protegida. In: Lucas MR, Saraiva M, Rosa A (eds) A qualidade. Numa perspectiva multi e interdisciplinary. Lisboa, Portugal: Ediçoes Sílabo, Lda., pp 117 – 125

Berrueta LA, Alonso-Salces RM, Héberger K (2007) Supervised pattern recognition in food analysis. J Chromatogr A 1158:196–214. doi: 10.1016/j.chroma.2007.05.024

Borrás E, Ferré J, Boqué R, et al (2015) Data fusion methodologies for food and beverage authentication and quality assessment - A review. Anal Chim Acta 891:1–14. doi: 10.1016/j.aca.2015.04.042

Bortolato SA, Arancibia JA, Escandar GM (2008) Chemometrics-assisted excitation-emission fluorescence spectroscopy on nylon membranes. Simultaneous determination of benzo[a]pyrene and dibenz[a,h]anthracene at parts-per-trillion levels in the presence of the remaining EPA PAH priority pollutants as int. Anal Chem 80:8276–8286. doi: 10.1021/ac801458a

Bro R (1997) PARAFAC. Tutorial and applications. Chemom Intell Lab Syst 38:149–171. doi: 10.1016/S0169-7439(97)00032-4

Bro R, Kiers HAL (2003) A new efficient method for determining the number of components in PARAFAC models. J Chemom 17:274–286. doi: 10.1002/cem.801

Da Silva CET, Filardi VL, Pepe IM, et al (2015) Classification of food vegetable oils by fluorimetry and artificial neural networks. Food Control 47:86–91. doi: 10.1016/j.foodcont.2014.06.030

Del Rio FJ, Riu J, Rius FX (2001) Graphical criterion for the detection of outliers in linear regression taking into account errors in both axes. Anal Chim Acta 446:49–58.

Di Anibal C V., Rodríguez MS, Albertengo L (2015) Synchronous fluorescence and multivariate classification analysis as a screening tool for determining Sudan I dye in culinary spices. Food Control 56:18–23. doi: 10.1016/j.foodcont.2015.03.010

Ferretto N, Tedetti M, Guigue C, et al (2014) Identification and quantification of known polycyclic aromatic hydrocarbons and pesticides in complex mixtures using fluorescence excitation-emission matrices and parallel factor analysis. Chemosphere 107:344–353. doi: 10.1016/j.chemosphere.2013.12.087

Haaland DM, Thomas E V. (1988) Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. Anal Chem 60:1193–1202. doi: 10.1021/ac00162a020

Hall GJ, Kenny JE (2007) Estuarine water classification using EEM spectroscopy and PARAFAC-SIMCA. Anal Chim Acta 581:118–124. doi: 10.1016/j.aca.2006.08.034

Indahl UG (2014) The geometry of PLS1 explained properly: 10 key notes on mathematical properties of and some alternative algorithmic approaches to PLS1 modelling. J Chemom 28:168–180. doi: 10.1002/cem.2589

Kemsley EK (1998) A genetic algorithm (GA) approach to the calculation of canonical variates (CVs). TrAC - Trends Anal Chem 17:24–34. doi: 10.1016/S0165-9936(97)00085-X

Ledesma E, Rendueles M, Díaz M (2015) Spanish smoked meat products: Benzo(a)pyrene (BaP) contamination and moisture. J Food Compos Anal 37:87–94. doi: 10.1016/j.jfca.2014.09.004

Lenhardt L, Bro R, Zekovic I, et al (2015) Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey. Food Chem 175:284–291. doi: 10.1016/j.foodchem.2014.11.162

Monago Maraña O, Bartolomé García T de J, Galeano Díaz T (2016) Characterization of Spanish Paprika by Multivariate Analysis of Absorption and Fluorescence Spectra. 49 (8):1184 – 1197. doi: 10.1080/00032719.2015.1089257

Muñoz de la Peña A, Mujumdar N, Heider EC, et al (2016) Nondestructive Total Excitation–Emission Fluorescence Microscopy Combined with Multi-Way Chemometric Analysis for Visually Indistinguishable Single Fiber Discrimination. Anal Chem 88:2967–2975. doi: 10.1021/acs.analchem.6b00264

Olivieri AC, Escandar GM (2000) Practical Three-Way Calibration.

Olivieri AC, Wu HL, Yu RQ (2009) MVC2: A MATLAB graphical interface toolbox for second-order multivariate calibration. Chemom Intell Lab Syst 96:246–251. doi: 10.1016/j.chemolab.2009.02.005

Purcaro G, Moret S, Conte LS (2013) Overview on polycyclic aromatic hydrocarbons: Occurrence, legislation and innovative determination in foods. Talanta 105:292–305. doi: 10.1016/j.talanta.2012.10.041

Reinholds I, Bartkevics V, Silvis ICJ, et al (2015) Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review. J Food Compos Anal 44:56–72. doi: 10.1016/j.jfca.2015.05.004

Riu J, Rius FX (1997) Method comparison using regression with uncertainties in both axes. TrAC - Trends Anal Chem 16:211–216. doi: 10.1016/S0165-9936(97)00014-9

Sádecká J, Tóthová J (2007) Fluorescence Spectroscopy and Chemometrics in the Food Classification - a Review. Czech J Food Sci Vol 25:159–173.

Sahar A, Rahman U ur, Kondjoyan A, et al (2016) Monitoring of thermal changes in meat by synchronous fluorescence spectroscopy. J Food Eng 168:160–165. doi: 10.1016/j.jfoodeng.2015.07.038

Sikorska E, Khmelinskii I, Sikorski M (2008) Fluorescence methods for analysis of beer. Elsevier Inc.

Simko P (2002) Determination of polycyclic aromatic hydrocarbons in smoked meat products and smoke flavouring food additives. J Chromatogr B Anal Technol Biomed Life Sci 770:3–18. doi: 10.1016/S0378-4347(01)00438-8

Slotani M (1964) Tolerance regions for a multivariate normal population. Ann Inst Stat Math 16:135–153. doi: 10.1007/BF02868568

Figure captions

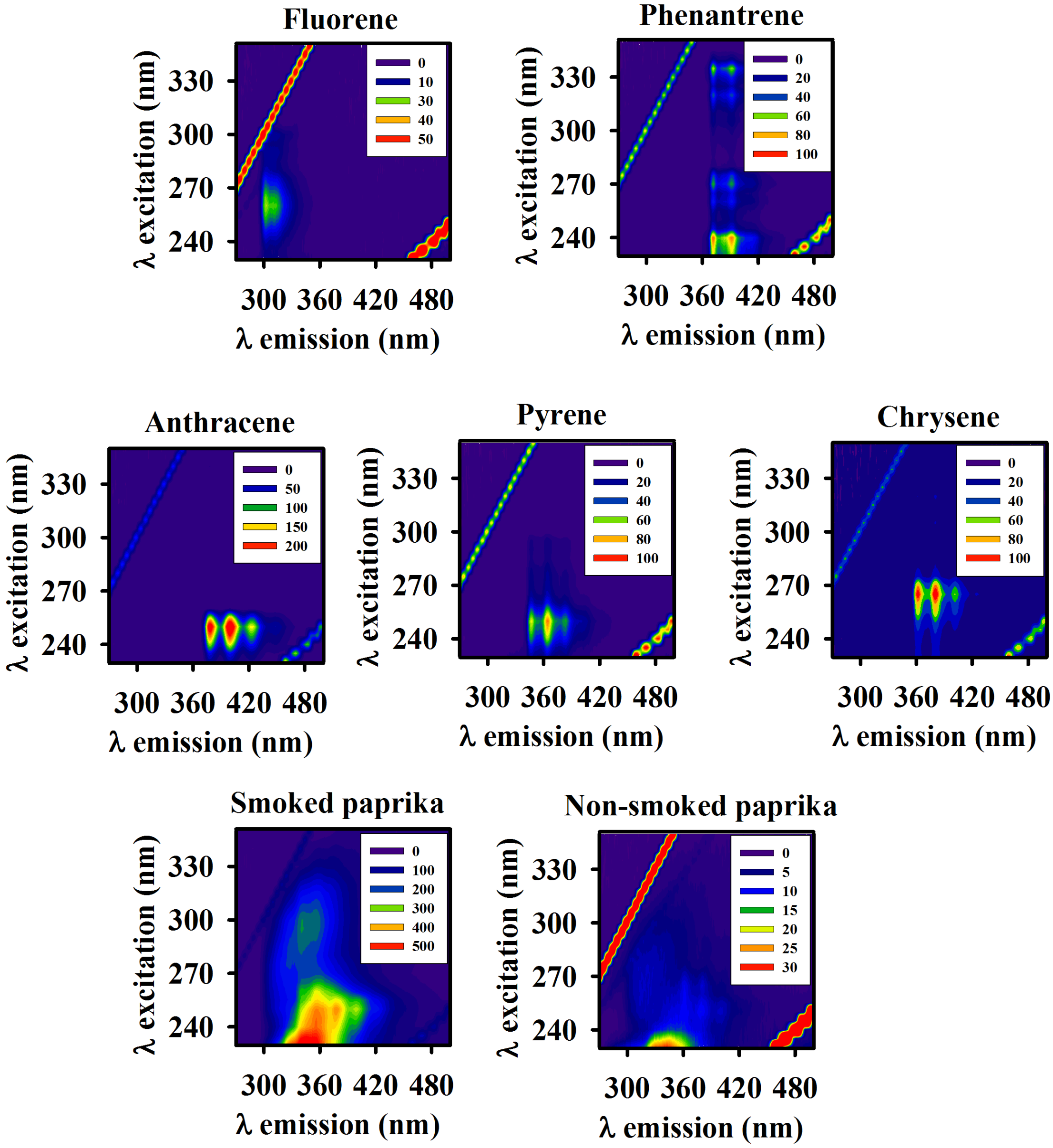

Figure 1. Excitation – emission matrices corresponding to Fluorene (10 µg mL^{-1}), Phenantrene (100 µg mL^{-1}), Anthracene (20 µg mL^{-1}), Pyrene (100 µg mL^{-1}) and Chrysene (100 µg mL^{-1}), a smoked paprika sample, non-smoked paprika sample (sample registered without previous dilution).

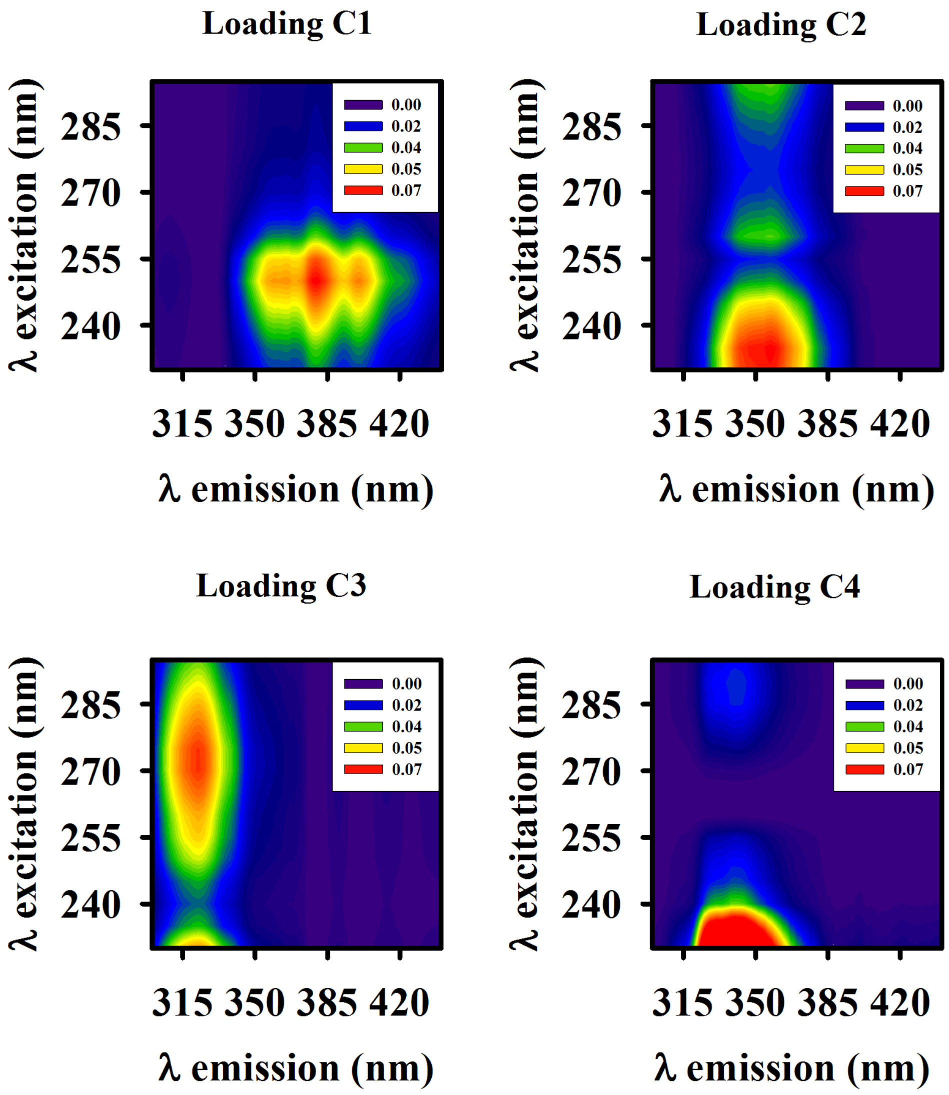

Figure 2. Structures of the four PARAFAC components (loadings corresponding to different components) obtained by multiplying the corresponding vectors.

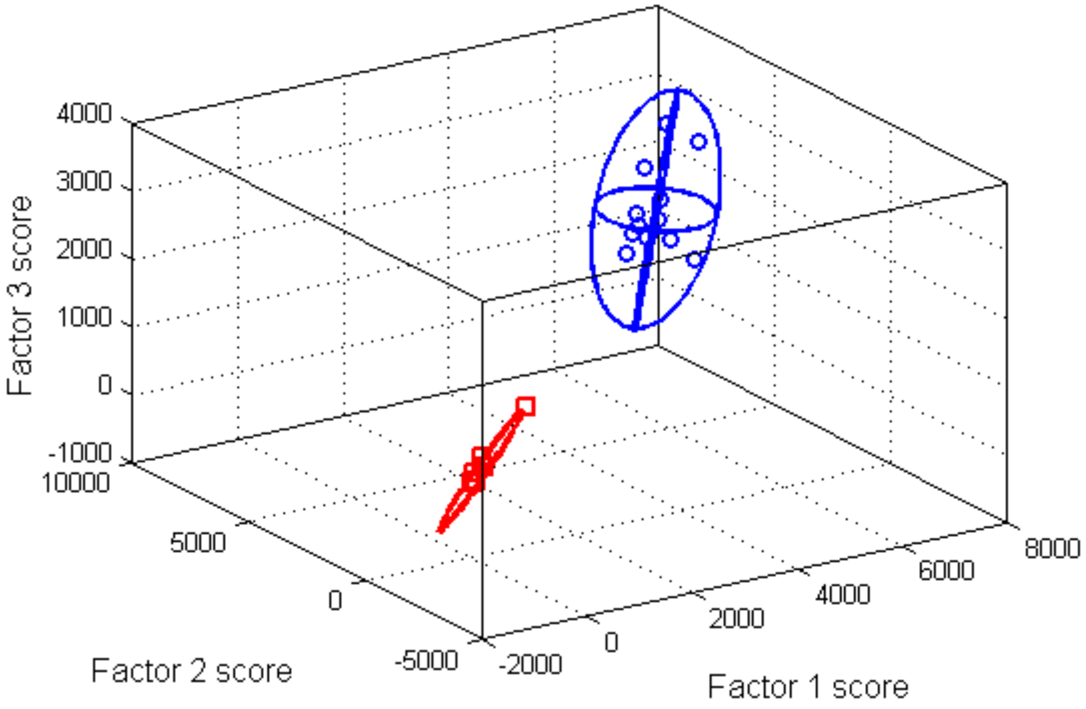

Figure 3. PARAFAC scores (3 first model's components) for 24 samples (12 corresponding to smoked paprika and 12 corresponding to non-smoked paprika). The three-dimensional projection of the 95% confidence ellipse of the data collected from each type of paprika is included to facilitate visualization of the obtained results.

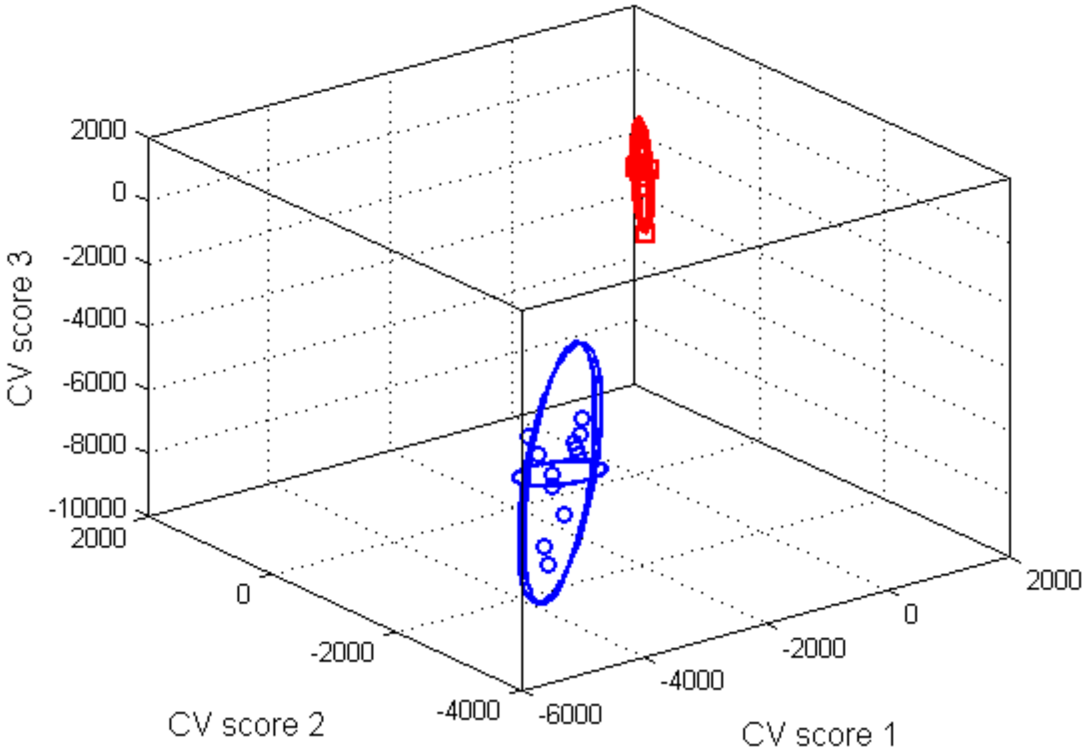

Figure 4. LDA CV scores (3 first model's components) for 24 samples (12 corresponding to smoked paprika and 12 corresponding to non-smoked paprika). The three-dimensional projection of the 95% confidence ellipse of the data collected from each type of paprika is included to facilitate visualization of the obtained results.

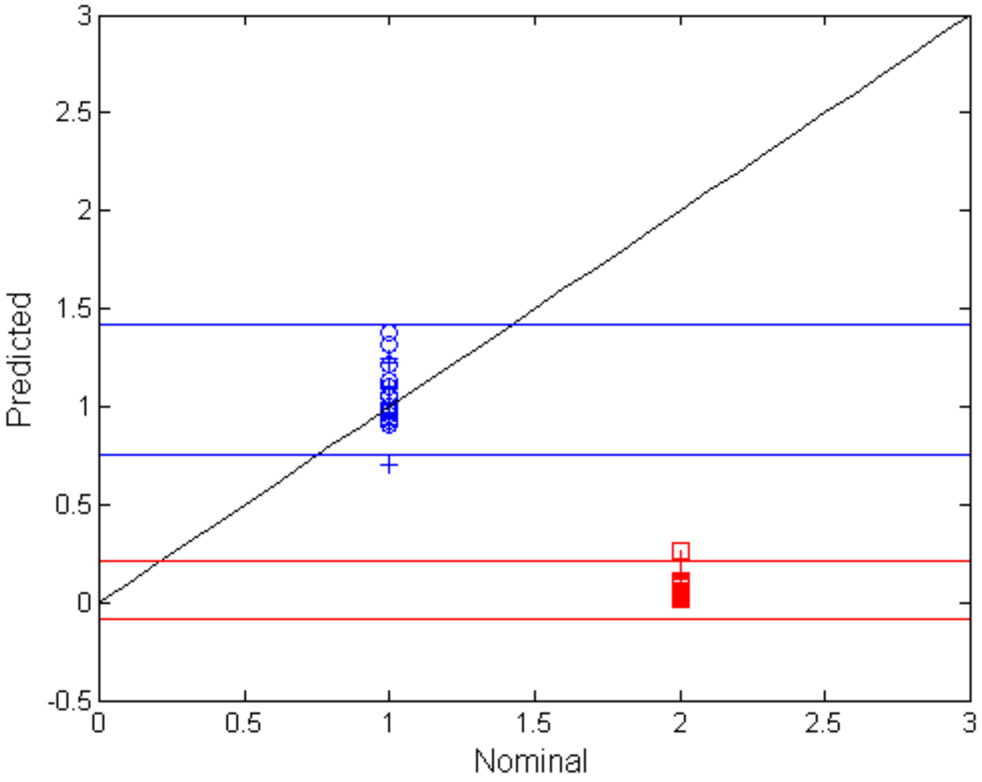

Figure 5. Plot of the DU-PLS (1 component model) predicted vs nominal coded values for 21 smoked paprika samples (12 calibration samples = blue circles; 9 validation samples = blue crosses) and 21 non-smoked paprika samples (12 calibration samples = red squares; 9 validation samples = red crosses).

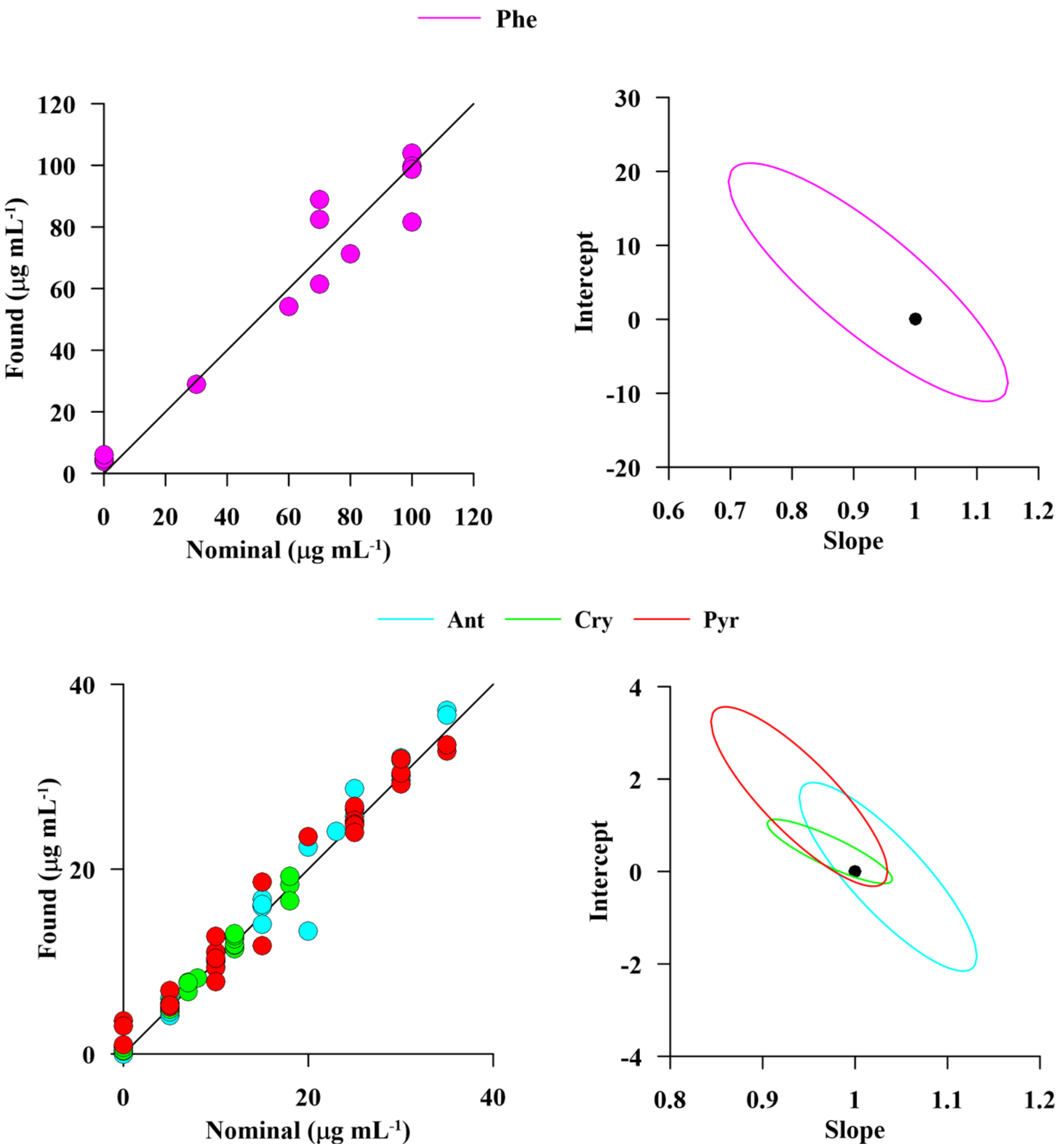

Figure 6. Plots of Phe (pink), Ant (blue), Cry (green) and Pyr (red) predicted concentrations as a function of the nominal values (left) and the corresponding elliptical joint regions (at 95% confidence level) for the slopes and intercepts of the regressions (right). Theoretical point (intercept = 0, slope = 1) is marked in the figure by the black point.

Table 1. Optimum number of factors for each analyte in their range of wavelengths in the U-PLS/RBL analysis and number of unexpected components found for each analyte in real samples.

Analyte	λ_{exc} (nm)	λ_{emis} (nm)	Components	RBL
Flu	250 - 275	300 - 350	2	2
Phe	320 - 340	350 - 400	1	2
Ant	240 - 260	395 - 410	5	1
Chr	250 - 275	355 - 410	5	2
Pyr	235 - 255	345 - 380	3	2

Table 2. Figures of merit for the different analytes using U-PLS/RBL (Olivieri and Escandar 2000).

	Flu	Phe	Ant	Cry	Pyr
SEN	9.1	1.3	4.7	6.7	1.9
γ	12	1.6	2.8	5.7	0.93
LOD	0.27	2.1	1.2	0.58	3.6
LOQ	0.80	6.2	3.5	1.7	11

SEN: Sensitivity (AU mL ng^{-1}); γ: Analytical sensitivity (mL ng^{-1}); LOD: limit of detection (ng mL^{-1}); LOQ: limit of quantification (ng mL^{-1}).

Table 3. Concentrations (mg kg^{-1}) obtained for each analyte by both methods and the error percentages between both methods.

Fluorene			Phenantrene			Anthracene		
HPLC-FLD-MCR-ALS	U-PLS/RBL	% E	HPLC-FLD-MCR-ALS	U-PLS/RBL	% E	HPLC-FLD-MCR-ALS	U-PLS/RBL	% E
1.91	1.98	3.6	11.00	12.24	11.3	2.47	2.59	4.9
2.01	2.19	8.9	11.81	11.67	1.2	2.64	2.76	4.5
2.95	3.38	13.4	16.69	12.51	25.0	4.14	3.87	6.5
3.48	2.23	35.9	13.04	13.89	6.5	2.95	2.81	4.7
2.09	2.00	4.5	10.41	11.82	13.5	2.37	2.43	2.5
1.83	2.00	9.3	11.27	9.92	11.9	2.54	2.97	16.9
2.70	2.45	9.3	16.50	9.13	44.7	4.23	3.15	25.6
2.51	2.88	14.7	16.63	11.92	28.3	4.29	4.11	4.2
2.52	2.30	8.7	14.97	13.34	10.9	3.13	2.96	5.4
2.17	1.93	11.1	12.16	12.16	0	2.83	2.34	17.3
1.77	1.75	2.0	9.80	11.46	16.9	2.30	2.19	4.8
2.29	1.88	17.9	18.89	19.19	1.6	4.33	3.01	30.5
1.57	1.23	21.6	11.48	10.17	11.4	2.44	1.97	19.3
1.78	2.43	36.5	12.10	14.16	17	2.74	2.87	4.7
1.98	2.24	13.1	12.50	12.39	0.88	2.79	2.93	5.0
1.86	1.74	6.5	10.92	10.63	2.7	2.37	1.93	18.5
2.63	2.67	1.5	10.00	9.00	9.0	2.06	1.97	4.4
2.26	3.07	35.8	18.56	17.25	7.1	4.36	3.09	29.1
2.30	3.13	36	17.27	17.53	1.5	4.00	3.66	8.5
1.43	2.03	41.9	13.53	13.10	3.2	3.14	2.66	15.3
2.22	3.10	39.6	14.76	15.78	6.9	3.32	2.82	15.1