Electronic Supplementary Material (ESI) for Analytical Methods.

Supplementary Data

Second-order fluorimetric approach based on a boron dipyrromethene (BODIPY) tetraamide derivative for Hg(II) chemosensing in water and fish samples

Valeria A. Lozano,[a] Arsenio Muñoz de la Peña,[b,*] Graciela M. Escandar[a,*]

[a] *Instituto de Química Rosario (CONICET–UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531 (2000) Rosario, Argentina.*

[b] *Department of Analytical Chemistry, University of Extremadura, 06006, Badajoz, España.*

Theory

The PARAFAC model

In the PARAFAC model, the second–order data for the I_{cal} training matrices, each of them as a $J\times K$ matrix $\mathbf{X}_{i,cal}$ (J and K are the number of data points in each mode), are joined with the unknown sample matrix $\mathbf{X}_u$ into a three-way data array $\boldsymbol{X}$, whose dimensions are $[(I_{cal}+1)\times J\times K]$. This three-way data array has an internally mathematical structure called trilinear and the fitting of a trilinear three-way array to the PARAFAC model provides unique solutions. Uniqueness implies that the estimated PARAFAC model cannot be rotated without a loss of fit as opposed to two-way analysis where one may rotate scores and loadings without changing the fit of the model.[1] If the array $\boldsymbol{X}$ is trilinear, each responsive component can be explained in terms of three vectors $\mathbf{a}_n$, $\mathbf{b}_n$ and $\mathbf{c}_n$, which collect the relative concentrations $[(I_{cal}+1)\times 1]$ for component n, and the profiles in both modes ($J\times 1$) and ($K\times 1$), respectively. The PARAFAC model[2] can be defined as:

$$\boldsymbol{X}_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} + \boldsymbol{E}_{ijk} \qquad (1)$$

in which N is the total number of responsive components, a_{in} is the relative concentration of component n in the ith sample, and b_{jn} and c_{kn} are the intensities at the j and k variables, respectively. The values of $\boldsymbol{E}_{ijk}$ are the elements of the matrix array $\boldsymbol{E}$, which contains the variation not captured by the model. The column vectors $\mathbf{a}_n$, $\mathbf{b}_n$ and $\mathbf{c}_n$ are collected into the corresponding score matrix $\mathbf{A}$ and loading matrices $\mathbf{B}$ and $\mathbf{C}$.

The decomposition of $\boldsymbol{X}$ by Eq. (1), usually accomplished through an alternating least–squares minimization scheme,[1,3] retrieves the profiles in both data modes ($\mathbf{B}$ and $\mathbf{C}$) and relative concentrations ($\mathbf{A}$) of individual components in the $(I_{cal}+1)$ mixtures,

whether they are chemically known or not, constituting the basis of the second–order advantage.

Some relevant issues concerning the application of PARAFAC model to the calibration of three-way data have to be considered:

Initialization of the algorithm. Different strategies to manage this step include the use of vectors given by GRAM[4] (generalized rank annihilation method), known spectral profiles of pure components, or loadings giving the best fit after a small number of PARAFAC runs with a few iterations. These alternatives can be found in Bro's PARAFAC package[5] and are conveniently implemented in the MVC2 graphical interface.[6]

Determination of the number of responsive components. Several methods can be applied to estimate the number of responsive components (*N*). CORCONDIA, a useful diagnostic tool which considers the PARAFAC internal parameter known as core consistency,[7] involves the study of the structural model based on the data and the estimated parameters of gradually augmented models. If the addition of more components does not considerably improve the fit, the model could be considered as suitable, and the core consistency parameter significantly drops from a value of ca. 50. The evaluation of the PARAFAC residual error, i.e. the standard deviation of the elements of the array $\boldsymbol{E}$ in Eq. (1), which decreases with increasing *N* until it stabilizes at a value compatible with the instrumental noise, can be considered as another useful technique. The value of *N* can be established as the smallest number of components for which the residual error is not statistically different than the instrumental noise.[8]

Restrictions during the least–squares fit. With the aim of obtaining physically interpretable profiles, the alternating least–squares PARAFAC fitting can be constrained by several available restrictions. For instance, non–negativity restrictions in all three modes allow the fit to converge to the minimum with physical meaning from the several minima which may exist in certain cases.

Identification of specific components. The estimated profiles retrieved by the model have to be compared with those for standard solutions of the analytes of interest in order to identify the chemical components under investigation, since the order in which they are sorted can be different between samples, i.e., it depends on their contribution to the overall spectral variance.

Calibration of the model to obtain absolute concentrations in unknown samples. Due to the fact that the three–way array decomposition provides relative values (**A**), absolute analyte concentrations can only be obtained after calibration. Calibration is carried out by regression of the set of standards with known analyte concentrations (contained in an $I_{\mathrm{cal}}\times 1$ vector **y**), and regression of the first I_{cal} elements of column $\mathbf{a}_n$ against **y** (provided they correspond to the I_{cal} samples):

$$k = \mathbf{y}^{+} \times [a_{1,n} \mid \ldots \mid a_{I\mathrm{cal},n}] \qquad (2)$$

in which '+' implies taking the pseudo–inverse. Then, for each test sample, the unknown relative concentration of n has to be converted to absolute by division of the last element of column $\mathbf{a}_n$ $[a_{(I\mathrm{cal}+1)n}]$ by the slope of the calibration graph k:

$$y_{\mathrm{u}} = a_{(I\mathrm{cal}+1)n} / k \qquad (3)$$

References

1 R. Bro, *Chemom. Intell. Lab. Syst.*, 1997, **38**, 149–171.

2 S. Leurgans and R. T. Ross, *Stat. Sci.*, 1992, **7**, 289–319.

3 P. Paatero, Chemom. Intell. Lab. Syst., 1997, **38**, 223–242.

4 E. Sánchez and B.R. Kowalski, Anal. Chem., 1986, **58**, 496–499.

5 http://www.models.life.ku.dk/algorithms.

6 A. C. Olivieri, H. L. Wu and R. Q. Yu, *Chemom. Intell. Lab. Syst.*, 2009, **96**, 246–251.

7 R. Bro and H. A. L. Kiers, *J. Chemom.*, 2003, **17**, 274–286.

8 A. Muñoz de la Peña, A. Espinosa-Mansilla, D. González-Gómez, A. C. Olivieri and H. C. Goicoechea, *Anal. Chem.*, 2003, **75**, 2640–2646.