SUPPLEMENTARY MATERIAL

Theory

MCR-ALS

Multivariate curve-resolution (MCR) coupled to alternating least-squares (ALS) is capable of handling data matrices with varying component profiles in one of the data dimensions. It can be thought that this makes it especially suitable for the convenient processing of kinetic-spectrophotometric matrix data; tipically many examples of successful application of MCR-ALS for resolving this kind of matrix data have been published in scientific literature .[1-3]

In MCR-ALS, an augmented data matrix is created from a group of data matrices for several samples. We consider matrices of size *JxK*, where *J* is the number of data points in the spectral dimension it means the number of wavelengths and *K* is the number of data points in the kinetic or temporal dimension it means the number of reaction times. This mathematical resource allows this algorithm to resolve linear dependence data. The augmented matrix could be constructed assembling data matrices following columns direction, rows direction or both directions simultaneously. In general, augmentation can be performed in either direction, depending on the type of experiment being analyzed.

However, usually the matrices are augmented in the mode suspected to lose the linearity. In the present case, the mode of augmentation should be the spectral one, since identical data profiles were obtained in spectral dimension correspond to unreacted HCF. An augmented matrix **D** of size J(I+1) x K was developed by joining calibration matrices $\mathbf{X_{c,i}}$ and the unknown matrix $\mathbf{X_u}$.[4, 5]

Therefore, the bilinear decomposition of the augmented matrix **D** was performed according to the expression:

$$\mathbf{D} = \mathbf{S_{aug}}\,\mathbf{G^T} + \mathbf{E} \qquad (1)$$

where the rows of **D** contain the spectra measured for different samples at several values of the temporal dimension, $\mathbf{S}_{aug}$ contains spectra for intervenient species, **G** contains the temporal or kinetic profiles and **E** is a matrix of residual not fitted by the model. The property dimensions for **D, S, G** and **E** are (I + 1)JxK, (I+1)JxN, KxN, and (I + 1)JxK respectively (being I the number of calibration matrices , J the number of wavelengths , K number of temporal data and N is the number of responsive components). As can be seen, **D** contains data for the I different samples.

The iterative ALS procedure aims at minimizing the Frobenius norm of || **E** ||, and was initialized using an initial estimation of the kinetic profiles of pure components used for calculating $\hat{S}$ ($\hat{S}$ means the estimation of **S**) as :

$$\hat{S} = \mathbf{D}\,\mathbf{G}^{T+} \qquad (2)$$

where '+' means the pseudoinverse of the matrix $\mathbf{G}^T = [\mathbf{G}(\mathbf{G}_T\mathbf{G})^{-1}]$.

Using the matrix $\hat{S}$ (equation 2) and the original matriz **D**, the matriz **G** was re-estimated by least squares as :

$$\hat{\mathbf{G}} = (\hat{S}^{+}\,\mathbf{D})^{T} \qquad (3)$$

The generalized inverse of **G** can be obtained only if the kinetic profiles of the sample components are different, as in the present case. The generalized inverse of S could be obtained because the matrix augmentation broke the linear dependence of the individual spectral profiles.

Finally, **E** was calculated applying equation (1) using matrix **D** and the estimation of **G** and $\hat{\mathbf{G}}$**.** These steps could be implemented in an alternating least squares cycles , so that in each iteration new S and **G** matrices were obtained. During the iterative recalculation of **S** and **G** a series of constraints were applied to improve these solutions, to give them a physical meaning and to limit their possible number for the same data fitting, such as 1) non-negativity for spectral and kinetic profiles, 2) stoichiometric relations among different chemical species in equilibrium or in kinetics. Iterations

continued until an optimal solution was obtained that fulfills the postulated constraints and the established convergence criteria.

After MCR-ALS decomposition of matrix **D**, concentration information contained in **S** can be used for quantitative predictions, by first defining the analyte concentration score as the area under the profile for the ith. sample.

Calibration samples are always within those employed to build the augmented matrix **D**. Their associated scores can be used to build a pseudo-univariate calibration graph against the nominal analyte concentrations. Prediction of analyte concentration in unknowns then proceeds by interpolation of the corresponding analyte scores in the calibration graph .[4]

U-PLS

Unfolded partial least squares (U-PLS) operates in a similar way to partial least squares-1 (PLS-1), except that second-order data are first vectorized or unfolded along one of the data dimensions, and then a conventional partial least-squares (PLS) model is built using these unfolded data and the nominal analyte concentrations. [6,7]

The **I** calibration data matrices are first vectorized into JK×1 vectors, and then a usual PLS model is built using these data together with the vector of calibration concentrations y (size I×1). This provides a set of loadings **P** and weight loadings **W** (both of size JK×A, where A is the number of latent factors), as well as regression coefficients **v** (size A×1)

The parameter A can be selected by techniques such as leave-one-out cross-validation .[8]

Notice that PLS is a latent variable method, and hence no prior information as to the spectral or time evolution of the analyte are in principle required for its successful operation.

If no unexpected components are present in the test sample, **v** could be used to estimate the analyte concentration according to

$$y_u = \mathbf{t_u}^T \mathbf{v} \quad (4)$$

where $\mathbf{t_u}$ is the test sample score, obtained by projecting the vectorized data for the test sample vec(**Xu**) onto the space of the A latent factors:

$$\mathbf{t}_u = (\mathbf{W}^T \mathbf{P})^{-1} \mathbf{W}^T \text{vec}(\mathbf{X}_u) \quad (5)$$

where vec(·) implies the vectorization operator.

N-PLS

Multiway regression methods such as N-PLS extend the tradicional PLS algorithm to higher orders, using the multidimensional structure of the data for model building and prediction .[6]

The tridimensional matrix **X** (*I* x *J* x *K*) is decomposed in a series of triads. In the case of three-way data, the model is given by the following equation:

$$x_{ijk} = \sum_{f=1}^{N} t_i w_j^J w_k^K + e_{ijk} \quad (6)$$

where x_{ijk} is the variation of absorbance intensity for sample *i* at wavelength wavelength j and time k, N is the number of components, t is an element of the score matrix **T** and, two w are elements of the two loading matrices **W**, one in spectral dimension $\mathbf{w^J}$ (*J* x 1) and the other for temporal dimension $\mathbf{w^K}$ (*K* x 1), and e_{ijk} is a residue not fitted by the model. The model finds the scores yielding maximum covariance with analyte concentrations as the dependent variable. The advantage of using N-PLS over bidimensional regression is a stabilization of the decomposition involved in Eq. (6), which potentially gives increased interpretability and better predictions

The algorithm used the data matrix of the I calibration samples within the concentration vector Y (Ix1) for obtaining loading and the regression coefficient v (size Ax1). As U-PLS, if no unexpected

components are present in the test sample, **v** could be used to estimate the analyte concentration according to

$$y_u = \mathbf{t_u}^T \mathbf{v} \quad (7)$$

Cross-validation can also be employed to estimate the number of calibration latent variables. [8]

N-PLS/RBL and U-PLS/RBL

If unexpected constituents occur in a test sample, neither the U-PLS nor N-PLS scores for the latter sample can be used for analyte prediction using the trained model. In this case, it is necessary to resort to a technique which is able to: (1) detect the new sample as an outlier, indicating that further actions are necessary before prediction, and (2) isolate the contribution of the unexpected component from that of the calibrated analytes, in order to recalculate appropriate scores for the test sample. U-PLS and N-PLS will consider a sample as an outlier if the residuals of the test data reconstruction (s_p) are abnormally large in comparison with the typical instrumental noise.

$$s_p = \| \mathbf{e}_p \| / (JK-A)^{1/2} = \| \mathrm{vec}(\mathbf{X}_u) - \mathbf{P} (\mathbf{W}^T \mathbf{P})^{-1} \mathbf{W}^T \mathrm{vec}(\mathbf{X}_u) \| / (JK-A)^{1/2} =$$

$$= \| \mathrm{vec}(\mathbf{X}_u) - \mathbf{P}\, \mathbf{t}_u \| / (JK-A)^{1/2} \quad (8)$$

$\| \mathbf{.} \|$ indicates the Euclidean norm

In such a case, residual bilinearization can be employed to model the presence of unexpected sample components using principal component analysis (PCA) or singular value decomposition (SVD), which allows one to estimate profiles for the unexpected components in the three data dimensions.[7]

For a single unexpected component the expression is :

$$\mathbf{vec(X_u) = P\, t_u + vec[}g_{\mathbf{unx}}\, \mathbf{b_{unx}\, (c_{unx})^T]} + \mathbf{e_u} \quad (9)$$

Where $\mathbf{b}_{unx}$ and $\mathbf{c}_{unx}$ are the left and right eigenvectors of **E**p and g_{unx} is a scaling factor appropriate for SVD analysis:

$$(\boldsymbol{g}\mathbf{unx}, \mathbf{bunx}, \mathbf{cunx}) = \mathbf{SVD}_1(\mathbf{E}_p) \quad (10)$$

where **E**p is the J x K matrix obtained after reshaping the JK x1 $\mathbf{e}_p$ vector of eq (8) and SVD_1 indicates taking the first principal component.

The RBL procedure consists in keeping constant the matrix of calibration loadings (P), and varying the test sample scores ($\mathbf{t_u}$) until the norm of the residual vector $\| \mathbf{e}_u \|$ is minimized in eq (9) using a Gauss-Newton procedure, so that a final value of t_u vector is obtained and applied for calculating the analyte concentration using eq.(8). So, the number of unexpected components (N_{unx}) can be determined by comparing the final residuals $\boldsymbol{s}\mathbf{u}$ with the instrumental noise level, with $\boldsymbol{s}\mathbf{u}$ given by:

$$S_u = \frac{\|e_u\|}{\left[JK - \left(A + N_{uns}\right)\right]^{1/2}} \quad (11)$$

where **e**u is calculated from Eq. (9). Typically, a plot of s_u computed for trial values of N_{unx} will show decreasing values, starting at s_p when $N_{unx} = 0$, until it stabilizes at a value compatible with the experimental noise, allowing to locate the correct number of unexpected components. [7, 9] However, some reports that have been recently published in scientific literature suggest that the number of unexpected components could be determined by compare value of the property of interest (sugar concentration in the present case) with those obtained applying a referente method .[10]

Once the RBL step is finished, and the correct test sample scores have been found, they are employed to provide the analyte concentration as is regularly done in all PLS models.

PLS/RBL-LD

However, this classical RBL procedure is not appropriate when the unexpected components have profiles that are identical to the analyte profile in one of the data dimensions.[11,12] In these cases a new RBL procedure for linear dependency (RBL-LD) is proposed . The underlying idea is similar to that of the classical RBL method: to minimize the norm of the residual vector $\mathbf{e}_u$, computed by fitting the test data to the sum of the relevant contributions (i.e., the part that is modeled by the current calibration and the contribution from the interfering agents) but taking into account the identical profiles in one of the data dimensions. This can be done either by modeling the residuals with MCR-ALS or with PARAFAC with linear dependency (PARALIND) instead of using SVD or PCA. [12]

Software

All routines employed to carry out the calculations described in this paper were written in MATLAB 7.0.[14] A Second-order multivariate calibration toolbox January 2013 For assistance read the document 'mvc2_gui_manual.pdf' and Chemom. Intell. Lab. Syst. 96 (2009) 246 -251[15] as well as an MCR-ALS Multivariate curve resolution- alternating least-squares written by Alejandro Olivieri Department of Analytical Chemistry University of Rosario Argentina were used in the persent work Moreover, the N-PLS code is available on the internet at http://www.models.life.ku.dk/source/ PLS/RBL is available from the authors on request, including a useful graphical user interface for data input and parameter setting, of the type already described for first-order multivariate calibration [22, 29, 56], and which also works for PARAFAC.

U-PLS/RBL is available at www.chemometry.com , including a graphical user interface. [6.7, 15] MCR-ALS was implemented using the graphical interface provided by Prof..Romá Tauler in his web page http://www.ub.edu/mcr/welcome.html. [16]

3.4 Figures of merit

Figures of merit such as sensitivity (SEN), analytical sensitivity (γn) and limit of detection (LODn) are regularly employed for method comparison.

MCR-ALS: Sensitivity can be calculated as:

$$SEN = S_n \left\{ \partial_n^T \left[Z_{Cal}^T \left(\mathbf{I} - Z_{unx} Z_{unx}^+ \right) Z_{cal} \right]^{-1} \partial_n \right\}^{-\frac{1}{2}}$$

Being: S_n the slope of MCR-ALS pseudo-univariate plot divided by the number of data point in each individual data matrix in the augmented mode, δn column vector (size N_{cal}x1) where N_{cal} is the number of MCR-ALS components present in the calibartion set, Z_{cal} profiles in the nonaugmented data mode for the components present in the calibration set, Z_{unk} profiles in the nonaugmented data mode for the unexpected sample components present in the tests samples.[7] Limit of detection and quantitation can be computed using the traditional univariate approach.[17]

U-PLS and N-PLS

Sensitivity can be calculated in cases when the second order adavantage operates as follows:

$$\mathrm{SEN} = \left\{ \left[\left(\mathbf{B}_{sus}{}^{\mathrm{T}} \mathbf{P}_{b,uns} \mathbf{B}_{sus} \right) * \left(\mathbf{C}_{sus}{}^{\mathrm{T}} \mathbf{P}_{c,uns} \mathbf{C}_{sus} \right) \right]^{-1} \right\}_{nn}^{-1/2}$$

where SEN is the sensitivity for component n, s_n is the integrated total signal for component n at unit concentration, B_{sus} and C_{sus} are the matrices containing the profiles for all suspected components (i.e., those present in the training set of samples) in each data dimension, 'nn' implies selecting the

(n,n) element corresponding to the n th. analyte of interest, '*' implies the Hadamard matrix product, and the projection matrices $P_{b,uns}$ and $P_{c,uns}$ are given by:

$$\mathbf{P}_{b,uns} = \mathbf{I} - \mathbf{B}_{uns}\mathbf{B}_{u+ns}$$

$$\mathbf{P}_{c,uns} = \mathbf{I} - \mathbf{C}_{uns}\mathbf{C}_{u+ns}$$

where B_{uns} and C_{uns} contain the profiles for the unsuspected components as columns.

Notice that when the second-order advantage is employed, this equation implies that SEN for component n is sample-specific and cannot be defined for the multivariate method as a whole. In such cases an average value for a set of samples can be estimated and reported .[18]

In the case of PLS/RBL, the appropriate expresión for the estimation of sensitivity is : [7, 18]

$$SEN_n = 1 / \| (\mathbf{P}_{eff}^{+})^T \mathbf{v} \|$$

where **v** is the (A×1) latent vector of regression coefficients for the PLS model, and **P**eff is a matrix given by:

$$\mathbf{P}_{eff} = (\mathbf{P}_{c,uns} \otimes \mathbf{P}_{b,uns})^T \mathbf{P}$$

where P is the (JK×A) loading matrix provided by the PLS model, $P_{c,uns}$ and $P_{b,uns}$ have the same meaning as above, and ⊗ implies the Kronecker product.

More useful than SEN seems to be the analytical sensitivity γ_n, defined, in analogy with univariate calibration, as the quotient between SEN and the instrumental noise level. Its inverse establishes the minimum difference of concentration which can be appreciated across the lineal range, and is independent on instrument or scale.[7] So that, the analytical sensitivity is suitable for comparing analytical methods based on different response nature.

Moreover, the limit of detection (LOD) can be calculated as a interval obtaining the lower and an upper limits of it , propopsed by Olivieri et al . in a recently publication [19]

$$LOD_{min} = 3.3\ [SEN^{-2}\ var(x) + h_{0min}\ SEN^{-2}\ var(x) + h_{0min}\ var(y_{cal})]^{1/2}$$

$$LOD_{max} = 3.3\ [SEN^{-2}\ var(x) + h_{0max}\ SEN^{-2}\ var(x) + h_{0max}\ var(y_{cal})]^{1/2}$$

Being, $h_{0\min} = \dfrac{\bar{y}_{cal}^2}{\sum_{i=1}^{I} y_i^2}$ where y_i is the centered concentration for the ith calibration sample ,

meanwhile the upper limit can be estimated as $h_{0max} = \max(h_{0cal})$, in which

$h_{0cal} = h_{cal} + h_{0min}\left[1 - \left(\dfrac{y_{cal}}{\bar{y}_{cal}}\right)^2\right]$ where h_{cal} and y_{cal} are the leverage and (centered) analyte concentration

of a generic calibration sample. LOD_{min} and LOD_{max} depend on the leverage, which is a function of the calibration score matrix **T** so that the limits of the LOD will depend on the calibration design and the number of calibration latent variables.[19]

209

210 **References**

211 **[1]** A.Espinosa-Mansilla , A.Muñoz de la Peña , H.C. Goicoechea , A.C. Olivieri A. C., "Two
212 multivariate strategies applied to three-way kinetic spectrophotometric data for the determination of
213 mixtures of the pesticides carbaryl and chlorpyrifos". *Appl. Spectrosc.,* 58 (2004) 83-90.

214 **[2]** M. De Luca , G. Ragno, G. Ioele, R. Tauler "Multivariate curve resolution of incomplete fused
215 multiset data from chromatographic and spectrophotometric analyses for drug photostability studies".
216 *Anal. Chim. Acta***,** 837 (2014) 31–37.

217 **[3]** P. Henrique Março, R. J. Poppi, , I. S. Scarminio, R. Tauler **"**Investigation of the pH effect and UV
218 radiation on kinetic degradation of anthocyanin mixtures extracted from *Hibiscus acetosella" Food*
219 *Chem.***,** 125 (2011) 1020–1027

220 **[4]** R.Tauler, "Multivariate curve resolution applied to second order data". *Chemom. Intell. Lab. Syst*.,
221 30 (1995) 133–146.

222 **[5]** R.Tauler "Comments on a recently published paper 'some surprising properties of multivariate
223 curve resolution-alternating least squares (MCR-ALS) algorithms"*J. Chemom*., 24 (2010) 87-90**.**

224 **[6]** G.M.Escandar, H.C. Goicoechea, A.Muñoz De La Peña, A.C. Olivieri, "Second- and higher-order
225 data generation and calibration: A tutorial". *Anal. Chim. Acta*., 806 (2014) 8 – 26.

226 **[7]** A.Olivieri, G.Escandar, "Practical Three way calibration". Ed. Elsevier Inc, UK (2014)

227 **[8]** D.M. Haaland, E.V. Thomas, "Partial least-squares methods for spectral analyses. 1. Relation to
228 other quantitative calibration methods and the extraction of qualitative information *Anal. Chem*., 60
229 (1988) 1193–1202

230 **[9]** M.D.Gil García, M.J.Culzoni, M.M. De Zan, R.Santiago Valverde., M.Martínez Galera.,
231 H.C.Goicoechea, "Solving matrix effects exploiting the second-order advantage in the resolution and
232 determination of eight tetracycline antibiotics in effluent wastewater by modelling liquid

chromatography data with multivariate curve resolution-alternating least squares and unfolded-partial least squares followed by residual bilinearization algorithms: II. Prediction and figures of merit" *J Chromatogr A.,* 1179 (2008) 106-124.

[10] J.W.Batista Braga , R.L. Carneiro , R.J Poppi . "Evaluation of the number of factors needed for residual bilinearization in BLLS and UPLS models to achieve the second-order advantage". *Chemomr. Intell. Lab. Syst*., 100 (2010) 99-109.

[11] M.J.Culzoni, H.C.Goicoechea, G.A. Ibañez, V.A. Lozano, N.R. Marsili, A.C. Olivieri, A.P. Pagani, "Second-order advantage from kinetic-spectroscopic data matrices in the presence of extreme spectral overlapping". *Anal. Chim. Acta,* 614 (2008) 46–57.

[12] V.A. Lozano, G.A. Ibañez, A.C.Olivieri, " Second-Order Analyte Quantitation under Identical Profiles in One Data Dimension. A Dependency-Adapted Partial Least-Squares/Residual Bilinearization Method". *Anal. Chem.,* 82 (2010) 4510-4519.

[13] R. Bro, R. Harshman, N. D. Sidiropoulos. "Modeling multi-way data with linearly dependent loadings". *J.Chemom.,* 23 (2009) 324-340

14] MATLAB 7.0, The Mathworks, Natick, Massachussets, USA, 2007**]**

[15] A. C. Olivieri , H-L Wu, R-Q. Yu, "MVC2: A MATLAB graphical interface toolbox for second-order multivariate calibration". *Chemom. Intell. Lab.Syst.,* 96 (2009) 246-251.

[16] J. Jaumot, R. Gargallo, A. de Juan, R. Tauler, "An user friendly interface for MCR-ALS : a new tool for Multivariate Curve Resolution in MATLAB". *Chemom. Intell. Lab. Syst*., 76 (2005) 101-110

[17] J.Saurina, C.Leal, R.Compañó, M.Granados, M.Dolors Prat, R.Tauler, "Estimation of figures of merit using univariate statisticsfor quantitative second order multivariate curve resolution". *Anal.Chim.Acta*, 432 (2001) 241-251

[18] A.C.Olivieri, "Analytical figures of merit from univariate to multivariate calibration ". *Chem.Rev.,* 114(2014)5358-5378

[19] F.Allegrini , A.C. Olivieri , "IUPAC-Consistent Approach to the Limit of Detection in Partial Least- Squares Calibration", *Anal. Chem*. 86 (2014) 7858−7866.