JOURNAL OF CHEMOMETRICS
J. Chemometrics 2005; **19**: 583–592
Published online 27 June 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/cem.964

A closed-form expression for computing the sensitivity in second-order bilinear calibration

Alejandro C. Olivieri[1]* and Nicolaas (Klaas) M. Faber[2]†

[1]Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario S2002LRK, Argentina
[2]Chemometry Consultancy, Rubensstraat 7, 6717 VD Ede, The Netherlands

Received 22 December 2005; Revised 23 March 2006; Accepted 27 March 2006

A general expression is derived for estimating the sensitivity of second-order bilinear calibration models, particularly parallel factor analysis (PARAFAC) and bilinear least-squares (BLLS), whether the second-order advantage is required or not. In the latter case, the sensitivity is correctly estimated either if the advantage is achieved by processing the unknown sample together with the calibration set (PARAFAC), or by post-calibration residual bilinearization (BLLS). The expression includes, as special cases, the sensitivity expressions already discussed by Messick, Kalivas and Lang (MKL) and by Ho, Christian and Davidson (HCD). The former one is the maximum achievable sensitivity in a given calibration situation, where all components are present in the calibration set of samples. The latter approach gives the lowest possible sensitivity, corresponding to only calibrating the analyte of interest, leaving the remaining components as uncalibrated constituents of the unknown sample. In intermediate situations, that is more than one calibrated analyte and presence of unexpected components in the unknown sample, only the present approach is able to provide a satisfactory sensitivity parameter, in close agreement with previously described Monte Carlo numerical simulations. Copyright © 2006 John Wiley & Sons, Ltd.

KEYWORDS: second-order bilinear calibration; sensitivity; net analyte signal; parallel factor analysis; bilinear least-squares

1. INTRODUCTION

Figures of merit such as the sensitivity are important in developing, comparing and assessing the reliability of analytical methodologies. Relevant references dealing with figures of merit in first- and higher-order multivariate calibration can be found in the specific literature [1–22].

The estimation of sensitivity and other figures of merit for second-order multivariate calibration models has become an active area of chemometric research. Within these models, the so-called bilinear calibration models are of great interest, because there are special relationships among the data that lead to the second-order advantage [23]. The latter property permits the determination of a calibrated component in the presence of unexpected sample components, and is of paramount importance in the field of complex sample analysis. Relevant second-order methodologies are based on: (1) the use of latent variables, such as unfolded partial least-squares (PLS) [24], where 'unfolded' refers to working with previously vectorized data matrices [25], multi-way PLS (nPLS) [26] and PLS combined with residual bilinearization (RBL) [27,28], (2) alternating least-squares (ALS), such as parallel factor analysis (PARAFAC) [29], self-weighted alternating trilinear decomposition (SWATLD) [30], and multivariate curve resolution (MCR-ALS) [31], (3) direct least-squares, such as bilinear least-squares (BLLS) [32–34] in its several variants and (4) eigenvector-eigenvalue techniques, such as the generalized rank annihilation method (GRAM) [35]. Among the above cited methods, those which exploit the second-order advantage are PARAFAC, SWATLD, GRAM, MCR-ALS and the combinations BLLS/RBL and PLS/RBL. It is important to note that the RBL procedure only works when data for the unexpected components are bilinear, because a relevant step in the former technique involves the modelling of their signals by singular value decomposition [27]. On a different note, all of the above mentioned methods are able to handle multiple calibration samples, except GRAM, which works with a

*Correspondence to: A. C. Olivieri, Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario (S2002LRK), Argentina.
E-mail: aolivier@fbioyf.unr.edu.ar
†Chemometry consultancy.
Contract/grant sponsors: Consejo Nacional de Investigaciones Científicas y Técnicas; Agencia Nacional de Promoción Científica y Tecnológica; Universidad Nacional de Rosario.

single calibration sample (TLD is a generalization of GRAM which can handle multiple samples [36,37]). In any case, PARAFAC and BLLS seem to be the most statistically efficient methodologies [38,39], and hence they are the main focus of the present work.

Each of the above mentioned methodologies achieves the second-order advantage by employing different strategies; hence, it may seem natural to expect different precision properties and sensitivities. There are basically two ways in which the second-order advantage can be obtained, schematically shown in Figure 1: either data for the unknown sample determine (together with calibration data) the regression coefficients leading to prediction (Figure 1(A)), or calibration is first performed using only calibration data, with the unknown sample aiding in the obtainment of unknown sample-specific regression coefficients in a subsequent step (Figure 1(B)). PARAFAC, SWATLD, GRAM and MCR-ALS operate according to Figure 1(A), whereas BLLS/RBL and PLS/RBL employ the scheme of Figure 1(B). In any case, the underlying philosophy implies that the unknown sample becomes part of the whole calibration process, an entirely new concept in analytical chemistry.

The existing closed-form expressions for estimating the sensitivity are based on Lorber's concept of net analyte signal [40], such as those developed by Messick, Kalivas and Lang (MKL) [1] and by Ho, Christian and Davidson (HCD) [2]. They seem to fit well to most second-order methodologies, although some conflicting aspects remain unclear, especially in what concerns the calibration of multiple analytes, and in the presence of unexpected unknown sample components [41]. A reliable estimate of sensitivity can be obtained by numerical Monte Carlo simulations, as has been abundantly described in the literature [10–12,14,17–19,41]. They are also helpful in related calibration fields such as computing leverages [42] and degrees of freedom when the model results from an extensive selection process [43].

A closed-form expression is desirable for estimating parameters such as the sensitivity, because, unlike Monte Carlo simulations, it provides insight into the variables affecting the parameter under investigation. In the present report, a general sensitivity expression is developed and shown to be applicable to cases not covered by either the MKL or the HCD approach. The general scheme contains, however, MKL and HCD values as special cases.

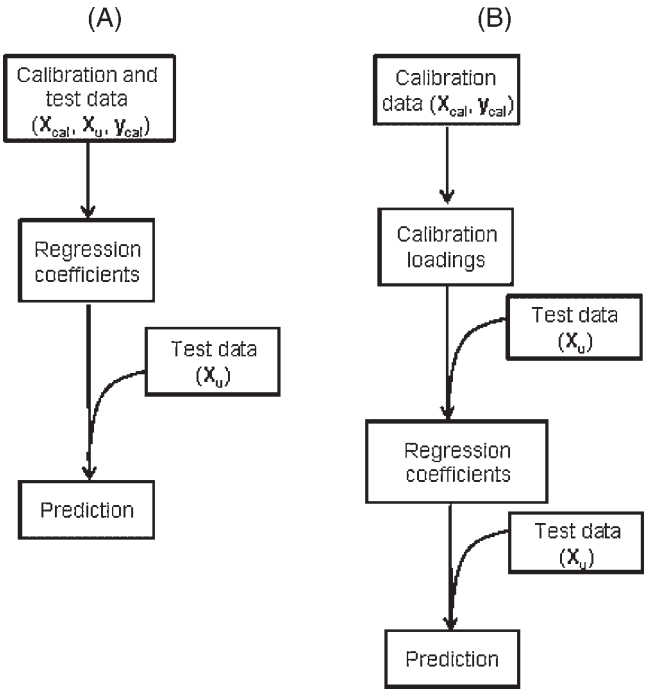

Figure 1. Two basic modes of obtaining the second-order advantage from higher-order data. A: Combining data from calibration and unknown samples before computing the regression coefficients. B: Calculating loadings from calibration data only, and then estimating regression coefficients after the unknown sample enters the scene. The information entered at each stage is noted: $\mathbf{X}_{cal}$ represents the calibration second-order signals, $\mathbf{X}_u$ the unknown sample signals and $\mathbf{y}_{cal}$ the analyte calibration concentrations.

2. THEORY

2.1. Terminology

It is important to define, in light of the forthcoming discussion, several sample component categories, with particular focus on components generating a signal that overlaps with the signal of the analyte of interest (or 'property' of interest), and can therefore be considered as potential interferents.

A distinction can be first made between components present in the calibration set of samples, and those which are only present in the unknown sample. The former ones can be called 'expected' components, because the analyst should include in the calibration set all components expected to be present in unknown samples, in order to have a sufficiently representative calibration set. However, truly unknown samples may contain additional components: these are called 'unexpected' ones. Note that the expected constituents can be further divided into 'calibrated' and 'uncalibrated': calibrated refers to components for which calibration concentrations are available (including, as a specific case, the analyte of interest), whereas uncalibrated refers to components for which only a common subspace that contains them is accessible. In inverse models, for example, not all calibration concentrations are available.

Additional phenomena producing overlapping signals, and also potentially interfering, are matrix effects, which may also belong to the category of expected (and also uncalibrated, if included into the signals produced by the calibration set) or unexpected (if only present in the unknown sample).

Notice that potential interferents will not always produce an interference, in the sense of generating a systematic error in the analyte determination [44]. Whether the interference will be actual or will only remain as potential, depends on the type of measured instrumental signals and on the employed calibration methodology. For first-order instrumental data, unexpected unknown sample components most likely constitute an interference. This may not be true, however, in the second-order domain involving the second-order advantage.

DOI: 10.1002/cem

Table I. Classification of sample components, with indications as to their presence in different sample types

	Expected components			
Type of sample	Calibrated analyte of interest	Other calibrated components	Other uncalibrated components	Unexpected components
Calibration	Yes	Yes	Yes	No
Validation	Possibly	Possibly	Possibly	No
Unknown	Possibly	Possibly	Possibly	Possibly

The terminology describing the different type of samples should also be mentioned. Samples can be divided in: (1) calibration or training samples, (2) test or validation samples, to guide choices during model construction or to monitor the predictive ability when the model is used for prediction and (3) unknown or prediction samples. For a test (validation) sample, the concentration value is available, unlike for a prediction sample (a 'true unknown').

Table I summarizes the nomenclature of the various sample and constituent types, including comments regarding the presence of the latter ones in each of the sample categories.

2.2. Net analyte signal

The philosophy behind the derivation of the closed-form expression discussed in the present paper is related to the concept of net analyte signal. This concept involves the decomposition of the total spectrum of a given sample ($\mathbf{x}$) in two orthogonal parts: a part that can be uniquely assigned to the analyte of interest (the net analyte signal, designated as $\mathbf{x}^*$), and the remaining part that contains the contribution from other components, which may be other expected or unexpected sample components (collectively included in $\mathbf{x}_{\text{other}}$):

$$\mathbf{x} = \mathbf{x}^* + \mathbf{x}_{\text{other}} = y_n\, \mathbf{s}_n^* + \mathbf{x}_{\text{other}} \tag{1}$$

where $\mathbf{x}_n^*$ and $\mathbf{s}_n^*$ are the net analyte signals corresponding to a given sample and to a sample having the nth analyte at unit concentration, respectively, and y_n is the analyte concentration.

In the case of Equation (1), vector-like net analyte signals are implied, but the latter can also be matrix signals:

$$\mathbf{X}_n^* = y_n\, \mathbf{S}_n^* \tag{2}$$

If a particular calibration model can be cast in either of the forms illustrated by Equations (1) and (2), then simple expressions for the sensitivity result, since the norm of the net analyte signal at unit concentration equals the sensitivity (S_n):

$$S_n = \left\|\mathbf{s}_n^*\right\| \text{ or } S_n = \left\|\mathbf{S}_n^*\right\| \tag{3}$$

The above considerations will be followed with regard to the BLLS model, with or without the second-order advantage. Since Monte Carlo simulations have shown that PARAFAC displays a similar precision as compared to BLLS, it is assumed that the developed expression will be suitable for both of these models.

2.3. Bilinear data models

We first note that the terminology herein employed is typically borrowed from analytical chemistry, but owing to their roots in linear algebra the concepts are of general applicability.

Many types of second-order data matrices for a pure component can be expressed, in the absence of noise, as an outer product of two vectors, that is $\mathbf{X} = g\mathbf{b}\mathbf{c}^{\mathrm{T}}$, where $\mathbf{b}$ and $\mathbf{c}$ denote the (normalized) profiles in both data dimensions and g is a scaling factor. These data have been termed *bilinear*, because they are linear in one variable when the other one is fixed and vice versa. Second-order bilinear calibration methods are of considerable interest since many instruments produce data that, ideally, follow the bilinear model. For a multi-component sample, the total matrix signal can be written as the sum of the various component contributions, that is in the form $\mathbf{X} = \sum_{n=1}^{N} g_n \mathbf{b}_n \mathbf{c}_n^{\mathrm{T}}$. The constituent profiles are usually collected into two matrices: $\mathbf{B}$ (size $J \times N$) and $\mathbf{C}$ (size $K \times N$), which contain the profiles for the N components present in the system (J and K are the number of channels in both data dimensions):

$$\mathbf{B} = [\mathbf{b}_1|\mathbf{b}_2|\ldots|\mathbf{b}_N] \tag{4}$$

$$\mathbf{C} = [\mathbf{c}_1|\mathbf{c}_2|\ldots|\mathbf{c}_N] \tag{5}$$

For reasons which will be clear below, the matrices $\mathbf{B}$ and $\mathbf{C}$ will be divided into sub-matrices containing the expected sample components that are present in the calibration set of samples (called $\mathbf{B}_{\text{exp}}$ and $\mathbf{C}_{\text{exp}}$, respectively), and sub-matrices with the unexpected component profiles which are only present in the unknowns ($\mathbf{B}_{\text{unx}}$ and $\mathbf{C}_{\text{unx}}$).

Both PARAFAC and BLLS specifically require the data to be bilinear, unlike PLS. However, if the second-order advantage is to be achieved by either BLLS or PLS using the RBL procedure, then the signal from the unexpected sample components should also have a bilinear structure.

2.4. MKL and HCD sensitivities

Two alternative definitions of sensitivity for second-order bilinear signals are available, both based on the concept of net analyte signal. They use the component profiles in both dimensions of the matrix data, as extracted by the different chemometric algorithms. The MKL definition for the sensitivity towards component n is:

$$S_{n,\text{MKL}} = s_n\{[(\mathbf{B}^{\mathrm{T}}\mathbf{B}) * (\mathbf{C}^{\mathrm{T}}\mathbf{C})]^{-1}\}_{nn}^{-1/2} \tag{6}$$

where '*' indicates the element-wise Hadamard matrix product, 'nn' implies the (n,n) element of a given matrix and s_n is the integrated total signal for component n at unit concentration.

When there are no unexpected sample components, the MKL sensitivity is exhibited by the PARAFAC method, and

DOI: 10.1002/cem

also by BLLS when using the so-called 'least-squares predictor' [16,32,41].

By contrast, the HCD sensitivity is defined as:

$$S_{n,\mathrm{HCD}} = s_n\{[(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}]_{nn}[(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}]_{nn}\}^{-1/2} \tag{7}$$

This sensitivity is lower than the MKL value, and has been shown to be connected with BLLS using the so-called 'naïve predictor', which is statistically less efficient than the least-squares predictor [16,32,41]. It is interesting to note that the naïve predictor stems from an unconstrained least-squares solution during the BLLS calibration phase, whereas the least-squares predictor arises from a specifically constrained solution [16]. This distinction is remindful of psychometrics works where 'non-constrained' and 'constrained' models are fitted [45].

When exploiting the second-order advantage, processing the data using both PARAFAC (see Reference [18] and [41]) and BLLS/RBL (see Reference [19]) leads to the HCD sensitivity, at least when a single analyte is calibrated, and the remaining components are unexpected constituents of the unknown sample.

2.5. BLLS without second-order advantage

It is useful to consider second-order bilinear instrumental signals for a three-component mixture, because in certain situations these multi-component systems show peculiar characteristics, which make them intractable by either the MKL or HCD approaches [41]. When all three components in an unknown sample are calibrated, the matrix signals are modelled by BLLS as:

$$\mathbf{X}_{\mathrm{u}} = y_1\mathbf{S}_1 + y_2\mathbf{S}_2 + y_3\mathbf{S}_3 + \mathbf{E} \tag{8}$$

where $\mathbf{X}_{\mathrm{u}}$ is the data matrix for the unknown sample (size $J \times K$), $\mathbf{S}_n$ is the matrix signal for component n at unit concentration ($n = 1, 2, 3$), y_n is the concentration of component n in the unknown sample and $\mathbf{E}$ is an error term of suitable size. Details on the implementation of BLLS can be found in the literature [32,33]. Notice that BLLS requires that all expected components be calibrated, and their sensitivities $\mathbf{S}_n$ be estimated from a set of calibration standards as the product of the profiles $\mathbf{b}_n$ and $\mathbf{c}_n$ in both dimensions and a scaling factor g_n:

$$\mathbf{S}_n = g_n\mathbf{b}_n\mathbf{c}_n^{\mathrm{T}} \tag{9}$$

When no unexpected components occur in the unknown sample, the matrices appearing in Equation (8) can all be vectorized for prediction of the analyte concentration, and the problem reduces to a classical least-squares (CLS) model using vectorized matrices:

$$\mathrm{vec}(\mathbf{X}_{\mathrm{u}}) = y_1\ \mathrm{vec}(\mathbf{S}_1) + y_2\ \mathrm{vec}(\mathbf{S}_2) + y_3\ \mathrm{vec}(\mathbf{S}_3) + \mathrm{vec}(\mathbf{E}) \tag{10}$$

where vec() indicates the column-wise vectorization operator. Equation (10) corresponds to the so-called least-squares predictor for estimating analyte concentrations in an unknown sample [32].

Notice that Equation (10) gives the total signal for the unknown sample, which can be easily decomposed into the contribution from the analyte (the first term on the right-hand side), and those from the remaining sample components. The signal from a specific analyte can be isolated from Equation (10) by projection of both sides onto the space orthogonal to the remaining components. For example, if the analyte of interest is 1, then the vectorized signals $\mathrm{vec}(\mathbf{S}_2)$ and $\mathrm{vec}(\mathbf{S}_3)$ can be used to construct an orthogonal projection matrix $\mathbf{P}_1$ of size $JK \times JK$:

$$\mathbf{P}_1 = \mathbf{I} - [\mathrm{vec}(\mathbf{S}_2)\,|\,\mathrm{vec}(\mathbf{S}_3)][\mathrm{vec}(\mathbf{S}_2)\,|\,\mathrm{vec}(\mathbf{S}_3)]^{+} \tag{11}$$

where $\mathbf{I}$ is an appropriately dimensioned identity matrix and '+' indicates the pseudo-inverse. The latter operation will succeed provided the $\mathrm{vec}(\mathbf{S}_n)$ vectors are not dependent (or nearly dependent), that is that the matrix $[\mathrm{vec}(\mathbf{S}_2)\,|\,\mathrm{vec}(\mathbf{S}_3)]$ is full column rank, which is a requirement for calibration with BLLS.

The result of multiplying Equation (10) by $\mathbf{P}_1$ is as follows:

$$\mathbf{P}_1\ \mathrm{vec}(\mathbf{X}_{\mathrm{u}}) = y_1\mathbf{P}_1\ \mathrm{vec}(\mathbf{S}_1) \tag{12}$$

since this last operation removes from Equation (10) the contribution of analytes 2 and 3. Hence, the vectorized net analyte signal at unit concentration for analyte 1 is equal to the $JK \times 1$ vector $\mathbf{s}_n^* = \mathbf{P}_1\ \mathrm{vec}(\mathbf{S}_1)$, and the sensitivity can be defined for this particular calibration scenario as:

$$S_1 = \|\mathbf{s}_n^*\| = \|\mathbf{P}_1\ \mathrm{vec}(\mathbf{S}_1)\| \tag{13}$$

An analogous equation can be found for the remaining two analytes. It may be noticed that the numerical value provided by Equation (13) is coincident with the MKL sensitivity, which has been developed as an extension of Lorber's method by working with vectorized matrices.

2.6. BLLS with second-order advantage obtained using RBL

Two different situations can be envisaged here: (1) a single analyte is calibrated and the remaining two components are unexpected constituents of the unknown sample signals, and (2) two analytes are calibrated, and the third component is a single unexpected constituent appearing in the unknown sample.

When a single analyte is calibrated, for example analyte 1, and components 2 and 3 are also present in the unknown sample, concentrations are estimated by first resorting to the RBL procedure, previous to CLS analysis based on Equation (10), in order to achieve the second-order advantage. In this case, vectorization cannot be applied to Equation (8), because the RBL procedure only works in matrix form. Hence, the removal of the contribution of components 2 and 3 from Equation (8) should be first done in matrix form, by applying two different left- and right-orthogonal projections:

$$\mathbf{P}_2\mathbf{X}_{\mathrm{u}}\mathbf{P}_3 = y_1\mathbf{P}_2\mathbf{S}_1\mathbf{P}_3 \tag{14}$$

where:

$$\mathbf{P}_2 = \mathbf{I} - [\mathbf{b}_2\,|\,\mathbf{b}_3][\mathbf{b}_2\,|\,\mathbf{b}_3]^{+} \tag{15}$$

$$\mathbf{P}_3 = \mathbf{I} - [\mathbf{c}_2\,|\,\mathbf{c}_3][\mathbf{c}_2\,|\,\mathbf{c}_3]^{+} \tag{16}$$

where $\mathbf{b}_{2,3}$ and $\mathbf{c}_{2,3}$ are the corresponding profiles for components 2 and 3 in both dimensions. The projection matrices $\mathbf{P}_2$ and $\mathbf{P}_3$ are of size $J \times J$ and $K \times K$, respectively. Equation (14) means that the net analyte signal of component 1 at unit concentration is in this particular case given by the $J \times K$ matrix $\mathbf{S}_n^* = \mathbf{P}_2\mathbf{S}_1\mathbf{P}_3$, and thus the sensitivity can be

DOI: 10.1002/cem

estimated in this case as follows:

$$S_1 = \left\|\mathbf{S}_n^*\right\| = \|\mathbf{P}_2\mathbf{S}_1\mathbf{P}_3\| \tag{17}$$

This last equation is equivalent to the HCD sensitivity, developed by extending Lorber's concept to each of the second-order dimensions separately.

A most interesting case appears in the event that both analytes 1 and 2 are calibrated, but component 3 is the unexpected one. Here, the sensitivity deviates from the values given by either the MKL or the HCD approach. Since the contribution of the unexpected component is modelled by RBL in order to gain the second-order advantage, a combination of the above operations is required to obtain the net analyte signal and the sensitivity in this case. In order to analyze the sensitivity towards analyte 1, component 3 is first removed in matrix form, because it is the one affected by the RBL process:

$$\mathbf{P}_4\mathbf{X}_\mathrm{u}\mathbf{P}_5 = y_1\mathbf{P}_4\mathbf{S}_1\mathbf{P}_5 + y_2\mathbf{P}_4\mathbf{S}_2\mathbf{P}_5 \tag{18}$$

where:

$$\mathbf{P}_4 = \mathbf{I} - \mathbf{b}_3\,\mathbf{b}_3^+ \tag{19}$$

$$\mathbf{P}_5 = \mathbf{I} - \mathbf{c}_3\,\mathbf{c}_3^+ \tag{20}$$

Once component 3 has been removed from the scene, vectorization of Equation (18) is possible because analytes 1 and 2 are calibrated:

$$\mathrm{vec}(\mathbf{P}_4\mathbf{X}_\mathrm{u}\mathbf{P}_5) = y_1\,\mathrm{vec}(\mathbf{P}_4\mathbf{S}_1\mathbf{P}_5) + y_2\,\mathrm{vec}(\mathbf{P}_4\mathbf{S}_2\mathbf{P}_5) \tag{21}$$

Now the removal of component 2 can be done by using the orthogonal projection matrix $\mathbf{P}_6$:

$$\mathbf{P}_6\,\mathrm{vec}(\mathbf{P}_4\mathbf{X}_\mathrm{u}\mathbf{P}_5) = y_1\mathbf{P}_6\,\mathrm{vec}(\mathbf{P}_4\mathbf{S}_1\mathbf{P}_5) \tag{22}$$

where:

$$\mathbf{P}_6 = \mathbf{I} - [\mathrm{vec}(\mathbf{P}_4\mathbf{S}_2\mathbf{P}_5)][\mathrm{vec}(\mathbf{P}_4\mathbf{S}_2\mathbf{P}_5)]^+ \tag{23}$$

The above combination of orthogonal projections makes the net analyte signal for analyte 1 at unit concentration equal to the $JK \times 1$ vector $\mathbf{s}_n^* = \mathbf{P}_6\,\mathrm{vec}(\mathbf{P}_4\mathbf{S}_1\mathbf{P}_5)$. Hence, the sensitivity is given by:

$$S_1 = \left\|\mathbf{s}_n^*\right\| = \|\mathbf{P}_6\,\mathrm{vec}(\mathbf{P}_4\mathbf{S}_1\mathbf{P}_5)\| \tag{24}$$

This last value is intermediate between those provided by the MKL and HCD definitions, because the result has been obtained by affecting the signal in part by the 'separate dimensions' approach to net analyte signal, and in part by the vectorized approach to net analyte signal.

2.7. The net analyte signal plot in different calibration situations

Since three different sensitivities occur in a three-component system such as the one discussed above, it follows that three different net analyte signal plots are possible. It is useful to consider the net analyte signal in matrix form, because this gives insight into the spectral regions in the dimensions of the original instrumental data. For this purpose, the vectorized net analyte signals can be reshaped into $J \times K$ matrices before plotting, except in the case of Equation (17), where the net analyte signal is directly given in matrix form. The corresponding three-dimensional surfaces and contour plots for these three different calibration situations are shown in Figure 2 for a specific example to be discussed below. Notice the distinct shapes of each of these signals.

The plots shown in Figure 2 have been calculated in the framework of the BLLS method, which requires that all expected components be calibrated. Methods such as PARAFAC and PLS/RBL, on the other hand, permit the occurrence of uncalibrated components in the calibration set of samples. Although it is not clear how the above approach can be applied to PARAFAC, the sensitivity results to be discussed below suggest a very similar behaviour of these two methodologies, pointing to similar net analyte signals. It is also interesting that the calculation of the net analyte signal has been recently discussed for the combination PLS/RBL [28]. Although the derivation is different than the one discussed above, and requires the estimation of calibration latent variables instead of spectral profiles, the results are almost identical to those shown in Figure 2.

3. RESULTS AND DISCUSSION

3.1. A general sensitivity expression

A general scheme which would cover all of the three above discussed cases is required. Recall that for a larger number of components, more calibration situations are possible, and correspondingly more sensitivity definitions exist. The general expression can be found by the following procedure. First consider the unexpected components, which are found by RBL, and define two projection matrices, orthogonal to the space spanned by all unexpected components in each mode:

$$\mathbf{P}_{\mathrm{b,unx}} = \mathbf{I} - \mathbf{B}_{\mathrm{unx}}\,\mathbf{B}_{\mathrm{unx}}{}^+ \tag{25}$$

$$\mathbf{P}_{\mathrm{c,unx}} = \mathbf{I} - \mathbf{C}_{\mathrm{unx}}\,\mathbf{C}_{\mathrm{unx}}{}^+ \tag{26}$$

where $\mathbf{B}_{\mathrm{unx}}$ and $\mathbf{C}_{\mathrm{unx}}$ contain the profiles for the unexpected components as columns. Notice that the matrices $\mathbf{B}_{\mathrm{unx}}$ and $\mathbf{C}_{\mathrm{unx}}$ can be built with columns representing the true spectral profiles for the unexpected components, or alternatively by columns representing the space spanned by them, for example, linear combinations obtained by singular value decomposition, as in the case of the application of the RBL procedure.

The orthogonal projection matrices shown in Equations (25) and (26) are employed to remove the contribution of the unexpected components from the matrix signal of the unknown sample:

$$\mathbf{X}_\mathrm{u} = \sum_{n=1}^{N} y_n\mathbf{S}_n \tag{27}$$

leading to an expression where the contribution from unexpected constituents has been removed:

$$\mathbf{P}_{\mathrm{b,unx}}\mathbf{X}_\mathrm{u}\,\mathbf{P}_{\mathrm{c,unx}} = \sum_{n=1}^{N}\mathbf{P}_{\mathrm{b,unx}}\mathbf{S}_n\mathbf{P}_{\mathrm{c,unx}} \tag{28}$$

Once this is done, the contribution of expected components other than the analyte of interest (No. 1 in this case) can be removed by vectorization before orthogonal projection:

$$\mathbf{P}_\mathrm{v}\,\mathrm{vec}(\mathbf{P}_{\mathrm{b,unx}}\,\mathbf{X}_\mathrm{u}\,\mathbf{P}_{\mathrm{c,unx}}) = y_1\,\mathbf{P}_\mathrm{v}\,\mathrm{vec}(\mathbf{P}_{\mathrm{b,unx}}\mathbf{S}_1\mathbf{P}_{\mathrm{c,unx}}) \tag{29}$$

DOI: 10.1002/cem

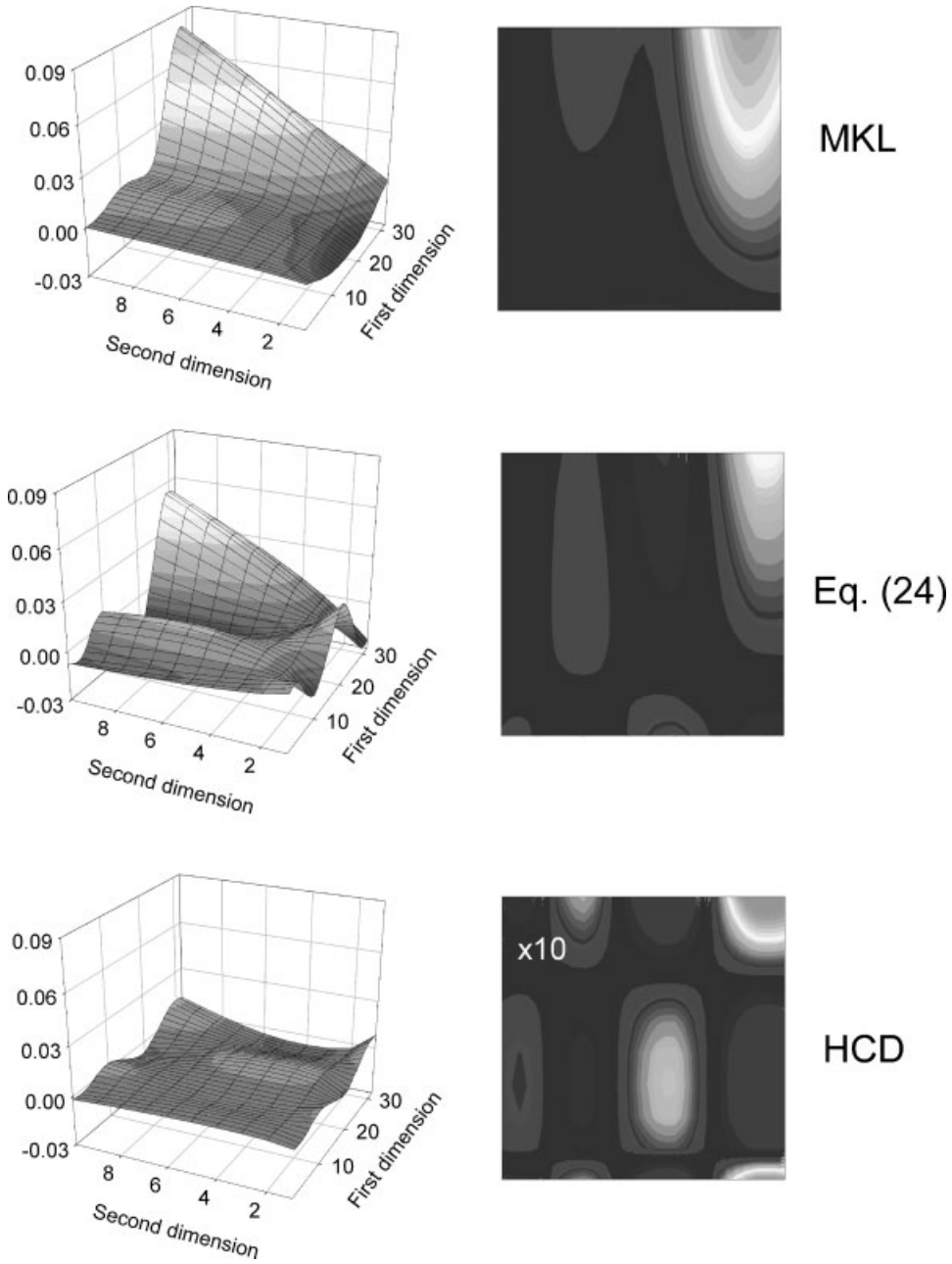

Figure 2. Three-dimensional surfaces (left) and contour plots (right) of the net analyte signal for analyte 1 in system 2 (see Table I), as computed for the three different approaches discussed in the present work (indicated in each plot). Those for the MKL and for the new approach have been obtained after reshaping into a matrix the vectorized net analyte signal, computed using Equations (13) and (24), respectively. For the HCD approach it is directly given as a matrix, cf. Equation (17). In the case of the contour plot obtained for the HCD approach, the values have been amplified by a factor of 10 to enhance details.

In this latter equation, the $JK \times JK$ projection matrix $\mathbf{P}_v$ is given by:

$$\mathbf{P}_v = \mathbf{I} - \mathbf{V}\mathbf{V}^{+} \tag{30}$$

$$\mathbf{V} = [\mathrm{vec}(\mathbf{P}_{b,unx}\,\mathbf{S}_2\,\mathbf{P}_{c,unx})|\ldots|\mathrm{vec}(\mathbf{P}_{b,unx}\,\mathbf{S}_{N\,exp}\,\mathbf{P}_{c,unx})] \tag{31}$$

where N_{exp} is the number of expected components. Notice that $\mathbf{V}$ does not contain the analyte of interest.

Vectorization leads to an MKL-type equation, which involves the Hadamard product of factors of the type $(\mathbf{Z}^T\mathbf{Z})$, where $\mathbf{Z}$ is a matrix of profiles. In the general expression, these relevant factors will have the following form:

$$(\mathbf{Z}^T\mathbf{Z}) = [(\mathbf{Z}^T_{exp}\,\mathbf{P}^T_{z,unx})(\mathbf{P}_{z,unx}\,\mathbf{Z}_{exp})] = (\mathbf{Z}^T_{exp}\,\mathbf{P}_{z,unx}\,\mathbf{Z}_{exp}) \tag{32}$$

since the orthogonal projection matrices are idempotent and symmetric. In this latter equation, $\mathbf{Z}_{exp}$ indicates the matrix of profiles for the expected components *only*, affected by the corresponding orthogonal projection $\mathbf{P}_{z,unx}$ in the dimension indicated by the subscript 'z'.

DOI: 10.1002/cem

Therefore, a general expression for computing the sensitivity in any calibration situation using second-order bilinear signals can be obtained as:

$$S_n = s_n \left\{ \left[\left(\mathbf{B}_{\text{exp}}{}^{\text{T}} \mathbf{P}_{\text{b,unx}} \mathbf{B}_{\text{exp}} \right) * \left(\mathbf{C}_{\text{exp}}{}^{\text{T}} \mathbf{P}_{\text{c,unx}} \mathbf{C}_{\text{exp}} \right) \right]^{-1} \right\}_{nn}^{-1/2} \tag{33}$$

where $\mathbf{B}_{\text{exp}}$ and $\mathbf{C}_{\text{exp}}$ are the matrices containing the profiles for all expected components in each dimension, and '*nn*' implies selecting the (*n*,*n*) element corresponding to the *n*th analyte of interest. Equation (33) should apply to PARAFAC and BLLS (in the latter case with the least-squares predictor), whether or not the second-order advantage is exploited, and for any specific calibration scenario involving multi-component samples. In the event the second-order advantage does not apply, $\mathbf{P}_{\text{b,unx}}$ and $\mathbf{P}_{\text{c,unx}}$ are both identity matrices.

3.2. An equivalent sensitivity expression for PLS/RBL

In Reference [28], the vector of regression coefficients $\boldsymbol{\beta}_n$ provided by the combined PLS/RBL model has been shown to be given by the following expression:

$$\boldsymbol{\beta}_n = (\mathbf{P}_{\text{c,unx}} \otimes \mathbf{P}_{\text{b,unx}})^{+} \boldsymbol{\beta}_{n,\text{cal}} \tag{34}$$

where $\boldsymbol{\beta}_{n,\text{cal}}$ is the corresponding vector of regression coefficients which would be computed in the absence of unexpected components, that is by applying the usual PLS model using vectorized calibration data. The matrices $\mathbf{P}_{\text{b,unx}}$ and $\mathbf{P}_{\text{c,unx}}$, on the other hand, have the same meaning as above. It follows from Equation (34) that the net analyte signal for the full PLS/RBL model, which includes the second-order advantage, can be estimated as:

$$\mathbf{s}_n^* = \boldsymbol{\beta}_n / \| \boldsymbol{\beta}_n \|^2 \tag{35}$$

leading to a sensitivity defined by Equation (3), that is $S_n = \|\boldsymbol{\beta}_n\|^{-1}$. Notice that when unexpected components are absent, $\boldsymbol{\beta}_n = \boldsymbol{\beta}_{n,\text{cal}}$, making Equation (35) analogous to the definition of net analyte signal in first-order multivariate calibration models [46,47] However, Equations (34 and 35) provide the correct picture by modifying the regression coefficients due to the effect of the unexpected components found by the RBL procedure.

It is important to note that Equation (35) could also be applied to PARAFAC and BLLS/RBL, yielding values which would be identical to those of Equation (33). In this case, the regression coefficients $\boldsymbol{\beta}_{n,\text{cal}}$ should be estimated by employing only calibration data. However, in the latter cases Equation (33) is preferable, because it provides more insight into the interplay of the relevant analyte profiles contained in the matrices $\mathbf{B}_{\text{exp}}$ and $\mathbf{C}_{\text{exp}}$, which are extracted by the models from the calibration data. Equation (35) is reserved for cases such as PLS, where the calibration process renders latent variables with no immediate physical meaning.

3.3. Comparison with Monte Carlo simulations

In Reference [41], extensive Monte Carlo numerical simulations have been carried out in order to estimate the sensitivity for particular second-order calibration scenarios. They were based on repeating the calibration/prediction process, introducing random noise in the signals corresponding to the unknown sample. The propagation of this noise to the estimation of regression coefficients and ultimately to the predicted analyte concentration in the unknown sample has been specifically investigated. Statistical analysis of the distribution of predicted concentration values rendered an estimation of the so-called variance inflation factor for a particular component (VIF_n) which quantifies the inflation of the instrumental noise when transmitted to a specific prediction:

$$\text{var}(y_n) = \text{VIF}_n \, \text{var}(\mathbf{X}_{\text{u}}) \tag{36}$$

The sensitivity can be gathered from the variance inflation factor by equating the latter one to the inverse squared sensitivity, that is: [41]

$$S_n = (\text{VIF}_n)^{-1/2} \tag{37}$$

When no unexpected components occur, the second-order advantage is not exploited, and Equation (33) gives values consistent with the MKL approach, as already established for PARAFAC and BLLS [17,33]. On the other extreme, when a single calibrated analyte occurs, and the remaining components act as unexpected constituents, Equation (33) gives a sensitivity which is identical to that provided by the HCD approach, as has already been shown for both PARAFAC and BLLS/RBL exploiting the second-order advantage [18].

Intermediate situations, not covered by either the MKL or HCD approaches, are also appropriately taken into account by Equation (33). This is best illustrated in Table II, where cases previously labeled as 'anomalous' are collected [41]. In the simulations leading to the values quoted in Table II, three components are present in the unknown sample, but only two of them are calibrated. Analytes 1 and 2 are thus part of

Table II. Comparison of sensitivities calculated by Equation (33) with MKL and HCD approaches, and Monte Carlo results for ternary systems[a]

	System					
	1		2		3	
	Analyte		Analyte		Analyte	
Method	1	2	1	2	1	2
Monte Carlo simulations						
PARAFAC	0.04	0.12	0.27	0.20	0.03	0.12
BLLS/RBL	0.04	0.11	0.26	0.19	0.03	0.12
PLS/RBL	0.04	0.11	0.27	0.20	0.03	0.12
GRAM	0.03	0.08	0.06	0.04	0.01	0.03
Closed-form expressions						
MKL	0.37	0.26	0.44	0.39	0.20	0.19
HCD	0.03	0.08	0.06	0.04	0.01	0.03
Equation (33)	0.04	0.12	0.27	0.19	0.03	0.12
Equation (35)	0.04	0.12	0.27	0.19	0.03	0.12

[a] The values reported in this Table have been taken from Reference [41], except the Monte Carlo values for the PLS/RBL method and those computed by Equation (35). They correspond to a situation where two analytes are calibrated, and a single unexpected component occurs in the unknown sample. Sensitivities are quoted for each analyte, in three different systems where the spectral overlapping is different (see Reference [41] for additional details).

DOI: 10.1002/cem

the calibration set of samples, and component 3 acts as an unexpected constituent. Three different systems with distinct overlapping in the component profiles have been analyzed [41]. In each case, the concentrations of both analytes were predicted using PARAFAC, BLLS/RBL and GRAM, and statistically analyzed with Monte Carlo noise addition numerical simulations. The results, in terms of sensitivity for each system and analyte, are given in Table II. For comparison, the Monte Carlo sensitivity calculated for PLS/RBL has also been included, and seen to be analogous to those for PARAFAC and BLLS.

Table II shows the sensitivity values computed with the MKL and HCD approaches and with Equation (33). As can be seen, neither of the previous theoretical approaches are able to explain the observed Monte Carlo sensitivity, as already discussed. However, Equation (33) provides values which are in close agreement to those obtained by Monte Carlo simulation, and to those obtained by Equation (35) in the case of the PLS/RBL model.

The progression of sensitivity values in a specific case, for example, analyte 1 in system 2 (Table II), can be pictorially appreciated through the plot of the net analyte signal for the three sensitivity schemes, that is MKL, Equation (24) and HCD. Figure 2 shows the corresponding three-dimensional surfaces and contour plots for the net analyte signal in matrix form, for the three different calibration situations which are possible in this three-component system. It is apparent that the sensitivity, given as the norm of the net analyte signal, decreases from top to bottom in Figure 2, accompanying the values quoted in Table II. Figures such as this one might also be helpful in assessing spectral regions within the bi-dimensional sensor plane where a given analyte is most responsive. In any case, it is interesting to note that these responsive spectral regions are not only analyte-specific, but also calibration-specific, that is they depend on what components are included in the calibration set of samples and which ones act as unexpected ones.

3.4. Summary of expressions

Table III shows a summary of the known expressions for computing the sensitivity in several chemometric methodologies applicable to second-order bilinear data, including the corresponding literature references for both the method and the sensitivity approach.

First notice that several sensitivities are given as the inverse of the norm of the regression coefficients provided by the model. Although this is a general definition, applicable to all models, it is intended to mean, in the context of Table III, that no special relationship exists with the approach discussed in this work, which involves spectral profiles estimated by the model. The PLS models quoted in Table III, on the other hand, have a latent variable structure, like principal component regression (PCR), and more generally, continuum regression [48] and cyclic subspace regression [49] which both contain PLS and PCR as special cases. However, when the spectral identity of the sample components is known, then the unfolded PLS model shows the MKL sensitivity, and the PLS/RBL model follows its specific Equation (35), which gives values consistent with the presently discussed Equation (33). Finally, nPLS appears to be unrelated to the MKL approach, although its sensitivity is known to be *larger* than the MKL one [14].

It follows that the MKL sensitivity is shown by most methods when all sample components are present in the calibration set of samples, that is they belong to the class of expected components. It is in the presence of unexpected components, which require the second-order advantage, when differences appear among the various models. Whereas GRAM always assumes the lowest possible (HCD) sensitivity, methodologies, such as PARAFAC, BLLS/RBL and PLS/RBL are able to gain a higher sensitivity. This is an important outcome from the presently discussed results. Another important conclusion is that the sensitivities for PARAFAC and the pair BLLS-PLS/RBL appear to be equal, even when they exploit the second-order advantage using the seemingly different strategies outlined in Figures 1(A), (B), respectively.

Finally, note that MCR-ALS has not been included in Table III. Previous studies have employed an analogy with univariate calibration for estimating figures of merit [20]. This ignores the effect of the unknown sample on the regression parameters, and hence may be overoptimistic. Preliminary results obtained by applying Monte Carlo methods to MCR-ALS indicate similar precision properties as compared with PARAFAC of BLLS when the second-order advantage is not exploited, that is a sensitivity compatible with the MKL approach. However, the results when achieving the second-order advantage do not seem to

Table III. Second-order sensitivity for different algorithms

Method	Reference[a]	Second-order advantage	Sensitivity definition	Reference[b]
Unfold PLS	24	No	$\|\|\boldsymbol{\beta}_n\|\|^{-1}$	16
nPLS	26	No	$\|\|\boldsymbol{\beta}_n\|\|^{-1}$	16
GRAM	35	Yes/No	HCD[c]	6
PARAFAC	29	No	MKL[d]	17
PARAFAC	29	Yes	Equation (33)	This work,18
BLLS/naïve predictor	32,33	No	HCD[c]	33
BLLS/LS predictor	32,33	No	MKL[d]	33
BLLS/RBL	32,41	Yes	Equation (33)	This work,19
PLS/RBL	28	Yes	$\|\|\boldsymbol{\beta}_n\|\|^{-1}$	28

[a] Reference to method.
[b] Reference to sensitivity definition.
[c] Equal to Equation (33) when the analyte of interest is only present in the calibration set.
[d] Equal to Equation (33) when no unexpected components occur.

DOI: 10.1002/cem

have a definite relationship with the above approaches. This would certainly require additional research.

4. CONCLUDING REMARKS

An important conclusion to be gathered from the present study is that second-order methods, such as PARAFAC, BLLS/RBL and PLS/RBL make a clear distinction between: (1) the calibrated analyte of interest, (2) other components present in the calibration set of samples and (3) unexpected components, only present in unknown samples. In this way, they exhibit a sensitivity parameter which is dependent on the composition of the calibration set, and are able to achieve the maximum possible sensitivity for each particular calibration situation. In contrast, methods such as GRAM only achieve extreme sensitivity values: a maximum value when all components are present in the calibration set, and the lowest possible value when unexpected sample components occur.

Since the sensitivity is usually ascribed to the instrument delivering the data, it may come as a surprise that the same second-order data set, after being processed by different chemometric methods, shows significantly distinct sensitivities towards the target analyte. This result can be viewed from Valcárcel's perspective: [50] 'The best selectivity (and sensitivity) levels can be obtained by applying chemometrics in the various physico-chemical methods for discrimination of analytes'. In the second-order domain, it follows that each chemometric methodology makes a specific contribution to the sensitivity of the complete analytical protocol, calling for a more integrated vision of data processing techniques and instrumental data. Other interesting surprises may await chemometricians in the higher-order multivariate world.

Acknowledgements

Financial support from the Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project No. PIP 5303) and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project No. PICT05-25825) is gratefully acknowledged. The authors thank the reviewers for their comments, which helped to improve the presentation of this work.

REFERENCES

1. Messick NJ, Kalivas JH, Lang PM. Selectivity and related measures for nth-order data. *Anal. Chem*. 1996; **68**: 1572–1579.
2. Ho C-N, Christian GD, Davidson ER. Application of the method of rank annihilation to fluorescent multi-component mixtures of polynuclear aromatic hydrocarbons. *Anal. Chem*. 1980; **52**: 1071–1079.
3. Wang Y, Borgen OS, Kowalski BR, Gu M, Turecek F. Advances in second-order calibration. *J. Chemometrics* 1993; **7**: 117–130.
4. Appellof CJ, Davidson ER. Three-dimensional rank annihilation for multi-component determinations. *Anal. Chim. Acta* 1983; **146**: 9–14.
5. Faber NM, Buydens LMC, Kateman G. Generalized rank annihilation method. II: bias and variance in the estimated eigenvalues. *J. Chemometrics* 1994; **8**: 147–154.
6. Faber NM. The price paid for the second-order advantage when using the generalized rank annihilation method (GRAM). *J. Chemometrics* 2001; **15**: 743–748.
7. Faber K, Lorber A, Kowalski BR. Analytical figures of merit for tensorial calibration. *J. Chemometrics* 1997; **11**: 419–461.
8. Faber K, Kowalski BR. Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares. *J. Chemometrics* 1997; **11**: 181–238.
9. Liu X, Sidiropoulos N. Cramer-Rao lower bounds for low-rank decomposition of multidimensional arrays. *IEEE Trans. Signal Process*. 2001; **49**: 2074–2086.
10. Olivieri AC. A simple approach to uncertainty propagation in preprocessed multivariate calibration. *J. Chemometrics* 2002; **16**: 207–217.
11. Faber NM, Boqué R, Ferré J. Iteratively reweighted generalized rank annihilation method. 1. Improved handling of prediction bias. *Chemom. Intell. Lab. Syst*. 2001; **55**: 67–90.
12. Faber NM, Boqué R, Ferré J. Iteratively reweighted generalized rank annihilation method: 2. Least squares property and variance expressions. *Chemom. Intell. Lab. Syst*. 2001; **55**: 91–100.
13. Boqué R, Ferré J, Faber NM, Rius FX. Limit of detection estimator for second-order bilinear calibration. *Anal. Chim. Acta* 2002; **451**: 313–321.
14. Faber NM, Bro R. Standard error of prediction for multiway PLS: 1. Background and a simulation study. *Chemom. Intell. Lab. Syst*. 2002; **61**: 133–149.
15. Riu J, Bro R. Jack-knife technique for outlier detection and estimation of standard errors in PARAFAC models. *Chemom. Intell. Lab. Syst*. 2003; **65**: 35–49.
16. Faber NM, Ferré J, Boqué R, Kalivas JH. Second-order bilinear calibration: the effects of vectorizing the data matrices of the calibration set. *Chemom. Intell. Lab. Syst*. 2002; **63**: 107–116.
17. Olivieri AC, Faber NM. Standard error of prediction in parallel factor (PARAFAC) analysis of three-way data. *Chemom. Intell. Lab. Syst*. 2004; **70**: 75–82.
18. Olivieri AC. Sample-specific standard prediction errors in three-way parallel factor analysis (PARAFAC) exploiting the second-order advantage. *J. Chemometrics* 2004; **18**: 363–371.
19. Haimovich A, Orselli R, Escandar G, Olivieri AC. Sensitivity and prediction error for spectroscopic bilinear least-squares exploiting the second-order advantage. Theoretical and experimental study. *Chemom. Intell. Lab. Syst*. 2006; **80**: 99–108.
20. Saurina J, Leal C, Compañó R, Gramados M, Dolors Prat M, Tauler R. Determination of triphenyltin in sea-water by excitation—emission matrix fluorescence and multivariate curve resolution. *Anal. Chim. Acta* 2001; **432**: 241–251.
21. Olivieri AC, Faber NM, Ferré J, Boqué R, Kalivas JH, Mark H. Uncertainty estimation and figures of merit for multivariate calibration. *Pure & Appl. Chem*. 2006; **78**: 633–661.
22. Baffi G, Martin E, Morris J. Prediction intervals for nonlinear projection to latent structures regression models. *Chemom. Intell. Lab. Syst*. 2002; **61**: 151–165.
23. Booksh KS, Kowalski BR. Theory of Analytical Chemistry. *Anal. Chem*. 1994; **66**: 782A–791A.
24. Wold S, Geladi P, Esbensen K, Øhman J. Multiway principal components and PLS analysis. *J. Chemometrics* 1987; **1**: 41–56.
25. Kiers HAL. Towards a standardized notation and terminology in multiway analysis. *J. Chemometrics* 2000; **14**: 105–122.
26. Bro R. Multiway calibration. Multilinear PLS *J. Chemometrics* 1996; **10**: 47–61.

DOI: 10.1002/cem

27. Öhman J, Geladi P, Wold S. Residual bilinearization. Part I. Theory and algorithms. *J. Chemometrics* 1990; **4**: 79–90.
28. Olivieri AC. On a versatile second-order multivariate calibration method based on partial least-squares and residual bilinearization. Second-order advantage and precision properties. *J. Chemometrics* 2005; **19**: 253–265.
29. Bro R. PARAFAC. Tutorial and applications. *Chemom. Intell. Lab. Syst*. 1997; **38**: 149–171.
30. Chen Z-P, Wu H-L, Jiang J-H, Li Y, Yu R-Q. A novel trilinear decomposition algorithm for second-order linear calibration. *Chemom. Intell. Lab. Syst*. 2000; **52**: 75–86.
31. Tauler R. Multivariate curve resolution applied to second order data. *Chemom. Intell. Lab. Syst*. 1995; **30**: 133–146.
32. Linder M, Sundberg R. Second-order calibration: bilinear least squares regression and a simple alternative. *Chemom. Intell. Lab. Syst*. 1998; **42**: 159–178.
33. Linder M, Sundberg R. Precision of prediction in second-order calibration, with focus in bilinear regression method. *J. Chemometrics* 2002; **16**: 12–27.
34. Goicoechea HC, Olivieri AC. A new robust bilinear least-squares method for the analysis of spectral-pH matrix data. *Appl. Spectrosc*. 2002; **59**: 926–933.
35. Sanchez E, Kowalski BR. Generalized rank annihilation factor analysis. *Anal. Chem*. 1986; **58**: 496–499.
36. Burdick DS, Tu XM, McGown LB, Millican DW. Resolution of multicomponent fluorescent mixtures by analysis of the excitation-emission-frequency array. *J. Chemometrics* 1990; **4**: 15–28.
37. Sanchez E, Kowalski BR. Tensorial resolution: a direct trilinear decomposition. *J. Chemometrics* 1990; **4**: 29–45.
38. Faber NM, Bro R, Hopke PK. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review. *Chemom. Intell. Lab. Syst*. 2003; **65**: 119–137.
39. Faber NM. Towards a rehabilitation of the generalized rank annihilation method (GRAM). *Anal. Bioanal. Chem*. 2002; **372**: 683–687.
40. Lorber A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem*. 1986; **58**: 1167–1172.
41. Olivieri AC. Computing sensitivity and selectivity in parallel factor analysis and related multi-way techniques: the need for further developments in net analyte signal theory. *Anal. Chem*. 2005; **77**: 4936–4946.
42. Baumann K, Stiefl N. Validation tools for variable subset regression. *J. Comput. Aid. Mol. Des*. 2004; **18**: 549–562.
43. Ye JC. Asymptotic global confidence regions in parametric shape estimation problems. *IEEE Trans. Inform. Theory* 2000; **46**: 1881–1895.
44. Van der Linden WE. Definition and classification of interferences in analytical procedures. *Pure & Appl. Chem*. 1989; **61**: 91–95.
45. Takane Y. Matrices with special reference to applications in psychometrics. *Linear Algebra Appl*. 2004; **388C**: 341–361.
46. Bro R, Andersen CM. Theory of net analyte signal vectors in inverse regression. *J. Chemometrics* 2003; **17**: 646–652.
47. Ferré J, Faber NM. Net analyte signal calculation for multivariate calibration. *Chemom. Intell. Lab. Syst*. 2003; **69**: 123–136.
48. Björkström A, Sundberg R. A generalized view on continuum regression. *Scand. J. Statist*. 1999; **26**: 17–30.
49. Lang PM, Brenchley JM, Nieves RG, Kalivas JH. Cyclic subspace regression. *J. Multivar. Anal*. 1998; **65**: 58–70.
50. Valcárcel M, Gómez-Hens A, Rubio S. Selectivity in analytical chemistry revisited. *Trends Anal. Chem*. 2001; **20**: 386–393.

DOI: 10.1002/cem