Chapter 13

Figures of Merit in Multiway Calibration

Alejandro C. Olivieri[1], **Santiago Bortolato and Franco Allegrini**
Instituto de Química Rosario (CONICET-UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina
[1]*Corresponding author: e-mail: aolivier@fbioyf.unr.edu.ar; olivieri@iquir-conicet.gov.ar*

Chapter Outline

Data Handling in Science and Technology, Vol. 29. http://dx.doi.org/10.1016/B978-0-444-63527-3.00013-8

1 INTRODUCTION

1.1 Figures of Merit in Modern Analytical Chemistry: Why Are They Important?

As previously presented in this book, measuring and processing multiway data provides analytical chemists with a number of advantages, such as (1) improved sensitivity, derived from noise averaging multiple measurements of redundant data, (2) increased selectivity, because each new data mode provides an additional degree of partial selectivity, and (3) modeling the analyte contribution and its quantitative determination in the presence of unknown interferences, absent in calibration samples (second-order advantage) [1]. Regarding items 1 and 2, a question which immediately emerges is how figures of merit like sensitivity, selectivity, and even the limit of detection (LOD) should be estimated when dealing with multivariate and multiway data? As analytical chemistry is the science of chemical measurements, finding a reliable way to judge them properly is not a minor issue, and this is the reason why this chapter is focused on trying to give a response to this question.

In modern analytical chemistry research, the search for new estimators to improve analytical figures of merit is an important driving force, with the sensitivity (SEN) and the LOD occupying prominent places among these figures. Finding an expression to calculate consistent estimators for these figures has a relevant influence on different activities, such as (1) comparison of the performance of different experimental procedures, (2) optimization of a given methodology under various experimental conditions, and (3) development of official protocols of validation and analysis, as documented in international standards [2,3].

The LOD has gained significant popularity as a descriptor of the quality of a method in applied analytical chemistry. This is because of two reasons: (1) it is expressed in concentration units, allowing to easily comparing in a direct way different methodologies and (2) it is needed for assessing detection capabilities which are of fundamental importance in certain specific areas: doping control in sports, monitoring traces of contaminants in environmental samples, etc. However, at the core of LOD definition lies the sensitivity, which is also a key parameter in the estimation of other figures of merit: (1) analytical sensitivity, which is important for the comparison of methodologies based on widely different signals, because it is independent of the instrument and technique applied, (2) selectivity, which helps to assess the possibility of analyte quantitation in the presence of interferences, and (3) prediction uncertainty, which gives an idea of the precision of a certain prediction.

1.2 From Univariate to Multivariate and Multiway Calibration

1.2.1 Sensitivity

According to the International Union of Pure and Applied Chemistry (IUPAC), in the classical single-constituent or univariate calibration, the sensitivity is defined as "the change in the response of the instrument divided by the corresponding stimulus (the concentration of the analyte of interest)," i.e., the slope of the calibration curve [4]. In first-order multivariate calibration, unfortunately, the situation regarding the definition of sensitivity becomes more complex [5]. The difficulty arising in this case is the fact that an intense signal may be useless under severe spectral overlapping with signals from other concomitant constituents. This leads to an important property of multivariate sensitivity: analyte specificity, which means that a certain sensitivity parameter corresponds to each analyte of interest.

As explained in other chapters, multiway calibration involves the measurement of data matrices per sample (or data arrays with three or more modes) for analyte calibration purposes and constitutes a powerful generalization of multivariate calibration [6]. In this field, several different sensitivity expressions have been proposed, some of them based on the extension of the first-order net analyte signal (NAS) concept to further data modes [7–10]. The NAS is the portion of the total signal which can be uniquely ascribed to the analyte of interest, and hence, the slope of the pseudounivariate NAS concentration graph, or the NAS at unit concentration, appears to be a reasonable definition of sensitivity. However, this kind of intuitive extension has caused several difficulties, as there are various competing NAS definitions, with no clear relationships among them [11–13]. Moreover, extrapolation to higher-order calibration led to serious underestimation of sensitivities.

As an alternative to the NAS approach, a general expression emerged in recent years, based on the analysis on how the uncertainty in instrumental signal propagates to the uncertainty in predicted concentrations [14–16]. As will be explained in more detail below, thanks to the developments in this direction, it is now possible to cast all the available sensitivity expressions into a general mathematical equation encompassing all possible degrees of data complexity, from univariate to multiway, and in the latter case for most multiway algorithms [17].

The general expression is in accordance with the fact that the multiway sensitivity displays even more intriguing properties in comparison with the univariate and first-order counterparts: it is not only analyte specific but also strongly dependent on the test sample and on the data processing algorithm. This implies that each test sample can have its own qualitative chemical composition, leading to a specific value of sensitivity. Likewise, the

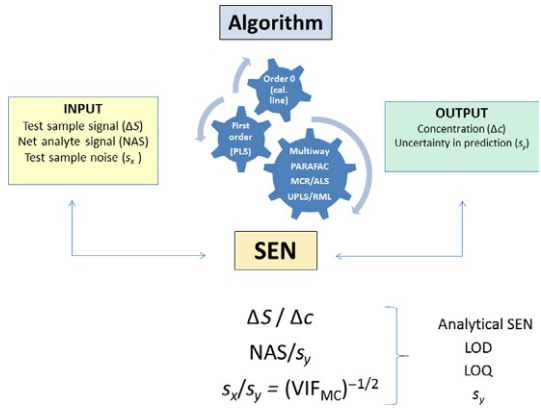

FIGURE 1 Schematic representation of the different possible ways to define sensitivity according to the data order and the corresponding algorithms.

computational tools employed by each algorithm will affect the analyte sensitivity, and hence, they should be regarded as an integral part of a multiway analytical protocol. Figure 1 summarizes the different ways in which sensitivity, generally understood as the relation between an input and an output, can be defined depending on the order of the data under analysis and the figures of merit resulting from the former.

1.2.2 Limit of Detection

Regarding the LOD, the transition between univariate and multivariate calibration requires a special attention, as in the case of SEN. The terms associated with detection capabilities have been present in the scientific literature for at least the past 100 years. Numerous terms, definition, and calculation approaches have been presented during the latter time period. A recent review with a historical overview about this topic has been recently published [18].

Currently, IUPAC adopts the definition given by the International Standardization Organization (document ISO 11843) [19] for the capability (or limit) of detection as "the lowest quantity of substance that can be distinguished from the absence of that substance (a blank value) within a stated confidence limit" [20–22]. This implies that the LOD is the minimum quantity detectable with a preset probability of false positives (false detects, α- or Type I errors) and false negatives (false nondetects, β- or Type II errors) [20–22]. When the analytical signal is univariate and analyte specific, the estimator is well defined and the LOD can be directly estimated from the univariate calibration line. The recommended detection rule is based on Neyman–Pearson test that considers false detects and false nondetects for the null hypothesis "there is no analyte" and the alternative hypothesis "there is analyte" [20].

However, when dealing with multivariate calibration, as is the case of the well-known partial least-squares (PLS) regression analysis, some aspects which remain outside the field of application of the ISO norm need to be considered [23]. In fact, there is still no generally accepted LOD estimator for PLS studies. Nevertheless, there is a high interest in the topic [17], undoubtedly tied to the inclusion of PLS regression in many commercial instruments, particularly those based on near-infrared spectral measurements [24], in addition to the continuous emergence of new and more sensitive analytical techniques, and the release of regulations on human or environmental exposure to low levels of chemical health hazards.

The main difficulty to estimate a multivariate LOD is that instrumental signals are not specific for a particular analyte. For this reason, the composition of the background of the sample, when extrapolating to zero analyte concentration, plays a fundamental role. As will be developed further in this chapter, a reasonable proposal based on an LOD estimator which adopts the form of a detection interval was recently investigated to try to overcome the previously stated difficulties [25].

1.2.3 Other Figures of Merit

As postulated in Section 1.1, the sensitivity constitutes the core from which other figures of merit can be defined. Among them, the LOD is one of the most well known and widely used. However, in some cases, it might be useful to define other figures in order to emphasize certain features and differences in the data, the samples, or the methodology under analysis. Examples of this kind of figures of merit are the analytical sensitivity, the uncertainty in prediction, and the limit of quantification. A more detailed description of the corresponding estimators will be presented in Section 5.

2 A BRIEF INSIGHT INTO DATA PROPERTIES, MODELS, AND ALGORITHMS

Knowledge of the properties and the structure of measured multiway data allow selecting a corresponding model and a data processing tool. As previously stated, this will significantly affect the achieved figures of merit, leading to another peculiar feature of multiway calibration: the algorithm specificity of these figures. This is why a brief reference to the main features of multiway data will be made at this point.

2.1 Bilinear and Trilinear Models

As discussed in other chapters, the simplest data array that can be found is a matrix for a single sample, leading to second-order data: if there are N

responsive constituents in the sample, a generic element x_{ij} of these data matrices can be written as:

$$x_{ij} = \sum_{n=1}^{N} b_{jn} c_{kn} + e_{ij} \tag{1}$$

where b_{jn} and c_{kn} define the specific properties at instrumental channels j and k for constituent n, and e_{ij} is an error term. Any data matrix can be expressed as the product of two matrices; however, if the mathematical rank, i.e., the number of bilinear components required to model the data, is small (ideally equal to the number of chemical constituents), the matrix is considered to be bilinear (strictly speaking, low-rank bilinear). When the matrix cannot be expressed as a sum of a few bilinear terms, it is considered as nonbilinear. In general, bilinearity is lost when the phenomena occurring in the two instrumental modes are mutually dependent [26].

Second-order data are the basic ingredient of a three-way array, which is the simplest multiway data. A three-way data array is low-rank trilinear if it can be expressed as a sum of a few trilinear components when the mixture contains a few constituents. Excitation–emission fluorescence (EEFM) data are typical examples where trilinearity applies. If a number of EEFM (I) are stacked in the sample mode, creating a three-way array $\mathbf{X}$, and the samples are mixtures of N fluorescent constituents, a specific signal x_{ijk} at sample i, emission wavelength j, and excitation wavelength k can be written as:

$$x_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} + e_{ijk} \tag{2}$$

where a_{in} is proportional to the concentration of constituent n in sample i, b_{jn} to the emission quantum yield at wavelength j, and c_{kn} to the absorption coefficient at excitation wavelength k.

2.2 Models to Deal with Trilinearity Deviations

Although it is not directly apparent in Equation (2), for practical purposes it is useful to outline the trilinearity demands: (1) individual data matrices should be bilinear, i.e., **b** and **c** profiles should not depend on each other and (2) **b** and **c** profiles should not depend on the sample, i.e., there should be unique **b** and **c** vectors in both instrumental modes in all samples. In chromatographic-spectral matrix data, elution profiles are not always exactly reproducible from sample to sample. Because of this, a three-way array composed of these latter data matrices will not be, in general, trilinear. Nevertheless, since individual data matrices are bilinear, an augmentation along the elution time mode can be performed. This leads to a chromatographic-spectral matrix augmented in the time direction ($\mathbf{X}_{\text{aug}}$), which is also bilinear and can be formulated as:

$$x_{\text{aug},pk} = \sum_{n=1}^{N} b_{\text{aug},pn} c_{kn} + e_{\text{aug},pk} \tag{3}$$

where $x_{\text{aug},pk}$, $b_{\text{aug},pn}$ and c_{kn} are elements of the $\mathbf{X}_{\text{aug}}$ matrix, and of the $\mathbf{b}_{\text{aug}}$ and $\mathbf{c}$ profiles respectively, with the index p runs from 1 to IJ, because the size of the augmented matrix is $IJ \times K$ (I = number of samples, J = number of elution times, K = number of wavelengths or other spectral sensors). The spectral profile $\mathbf{c}_n$ (in the nonaugmented mode) is unique for each constituent and common to all samples, whereas $\mathbf{b}_{aug,n}$ is the augmented time profile in the augmented elution time mode and is composed of I successive time subprofiles with J times each.

As a consequence of the above discussion, it is possible to classify three-way data as (1) trilinear, (2) nontrilinear with a single trilinearity breaking mode and unfoldable to a bilinear augmented matrix, and (3) other nontrilinear. This latter category refers to data structures in which two trilinearity breaking modes can be identified or cases in which the individual matrices are nonbilinear, as discussed in previous chapters.

Regarding the available algorithms to deal with the different data types previously described, they can be classified into three main groups according to a simple connection between their underlying models and the different data categories discussed above: (1) a multilinear model, (2) a bilinear model for an augmented matrix, and (3) a latent variable model. Group 1 includes parallel factor analysis (PARAFAC) [27], group 2 multivariate curve resolution coupled to alternating least-squares (MCR-ALS) [28] particularly in the extended version [29], and group 3 unfolded and multiway partial least-squares (U-PLS and N-PLS) [30,31].

As already explained in other chapters, both PARAFAC and MCR-ALS achieve the second-order advantage by simultaneously processing the multiple calibration samples and unknowns because their internal algorithmic models are able to decompose the contribution of the potential interfering agents and the analytes to the total signal. However in the case of the PLS-based methodologies, the achievement of the second-order advantage is a postcalibration activity based on a procedure called residual multilinearization (RML) [32–36].

As a final note for this section, it is important to consider that sometimes the border between these algorithms in terms of application fields is not entirely clear, and there may be a considerable overlapping. However, it is likely that future developments will take into account sensitivity and LOD considerations as a helpful decision-making tool in this regard.

2.3 From Homoscedastic to Correlated and Heteroscedastic Noise

While many chemometric tools have been designed to extract information from multivariate chemical measurements, one issue that has been somewhat

ignored is the role of multivariate measurement errors in this process. However, in recent years, a new branch of chemometrics has gained place: the development of methodologies to try to uncover, understand, and characterize the error structure of a given data set [37–39].

Most of the models and algorithms to deal with multivariate and multiway calibration have been conceived under the simplistic assumption that the error term, as presented in Equations (1–3), is identical and independently distributed (*iid*). In many cases, especially when measurement errors are small or the assumptions are approximately valid, traditional chemometric tools can be applied with excellent results, but in other cases, consideration of the measurement error structure can mean the difference between the success and the failure of the data analysis [40,41]. The same discussion could be extended to the estimation of figures of merit. Although the sensitivity obtained by the uncertainty propagation, the approach (see below) is defined assuming that the input noise is *iid*, when the uncertainty in predictions is to be estimated in the context of correlated and heteroscedastic errors, the scenario is different, and some extra considerations are needed. This is why further research will be required to integrate the information about the error structure to the current available estimators of multivariate uncertainty.

3 SENSITIVITY EXPRESSIONS BASED ON NET SIGNAL CHANGES

3.1 Univariate Calibration

In this case, prediction of the analyte concentration (y) in a test sample from its signal (x) proceeds through the known expression [4]:

$$y = (x - n_0)/m_0 \tag{4}$$

where m_0 and n_0 are the slope and intercept, respectively, of the zeroth-order linear calibration graph. The slope m_0 is the sensitivity since it measures the change in signal for a unit change in concentration.

3.2 First-Order Calibration

The concept of NAS, graphically represented in Figure 2, was helpful on the evaluation of the sensitivity in first-order calibration, by extending the univariate definition to the change in NAS for a unit change in analyte concentration [42]. In order to fully understand the NAS concept and its consequences, it is highly useful to consider, as a simple example, a binary mixture where two constituents occur, with the vector signal, e.g., a spectrum for a test sample measured at a number of sensors and given by:

$$\mathbf{x} = y_1\mathbf{s}_1 + y_2\mathbf{s}_2 \tag{5}$$

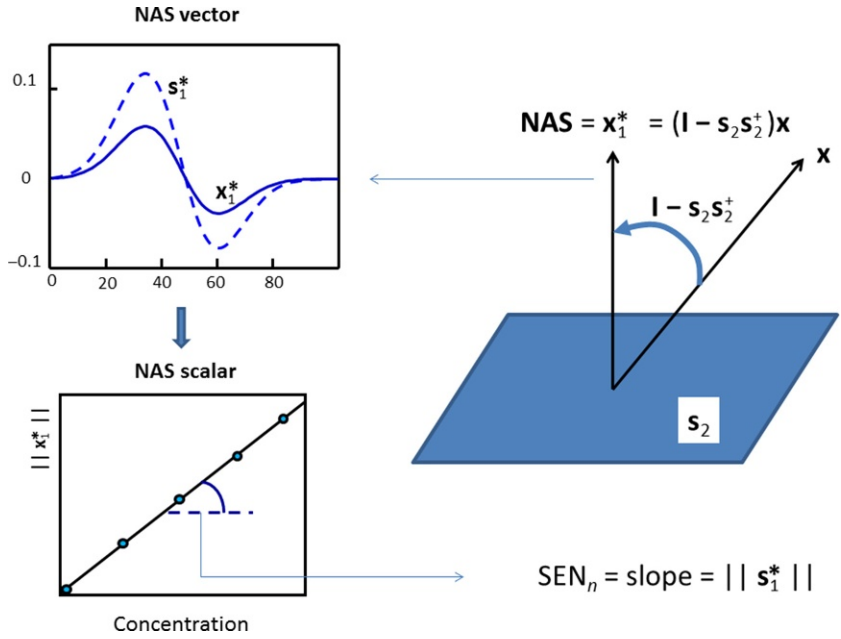

FIGURE 2 Graphical representation of the NAS concept. *Adapted with permission from Ref. [17]. Copyright 2014 American Chemical Society.*

where y_1 and y_2 are the constituent concentrations, and $\mathbf{s}_1$ and $\mathbf{s}_2$ are the pure constituent profiles at unit concentration. Equation (5) assumes that: (1) the studied signal is additive (i.e., the total signal is the sum of the individual contributions from both sample constituents) and (2) the constituent signals are proportional to their concentrations, meaning that Beer's law (or similar ones) applies. Focusing on analyte 1 as the constituent of interest, the contribution from constituent 2 can be removed from Equation (5) by left-multiplying both sides by an orthogonal projection matrix $\left[\mathbf{I}-\mathbf{s}_2\left(\mathbf{s}_2^{\mathrm{T}}\mathbf{s}_2\right)^{-1}\mathbf{s}_2^{\mathrm{T}}\right]$, where $\mathbf{I}$ is an appropriately dimensioned unit matrix, because $\left(\mathbf{I}-\mathbf{s}_2\mathbf{s}_2^{+}\right)\times\mathbf{s}_2=\mathbf{s}_2\times\mathbf{s}_2=0$, with $\left[\mathbf{s}_2\left(\mathbf{s}_2^{\mathrm{T}}\mathbf{s}_2\right)^{-1}\mathbf{s}_2^{\mathrm{T}}\right]$ designated as $\mathbf{s}_2^{+}$ (the superscript "+" indicates the generalized inverse). Notice that knowledge of $\mathbf{s}_1$and $\mathbf{s}_2$ is assumed, which is only possible in the context of first-order methodologies such as classical least-squares (CLS) analysis, where the pure spectra are either supplied to the model from separate measurements on pure constituents or adequately retrieved by analysis of mixtures of pure constituents. Thus, Equation (5) leads to Equation (6), and a two-constituent problem has become a virtual single-constituent problem:

$$\left(\mathbf{I}-\mathbf{s}_2\mathbf{s}_2^{+}\right)\mathbf{x}=y_1\left(\mathbf{I}-\mathbf{s}_2\mathbf{s}_2^{+}\right)\mathbf{s}_1 \tag{6}$$

Indeed, the left-hand side of the latter equation defines the NAS for constituent 1 in the mixture ($\mathbf{x}_1^*$), as being proportional to its NAS at unit concentration ($\mathbf{s}_1^*$), the proportionality constant being the analyte concentration y_1:

$$\mathbf{x}_1^* = y_1 \mathbf{s}_1^* \tag{7}$$

The relevant result to be gathered from Equation (7) is that the NAS vector $\mathbf{x}_1^*$ is parallel to the NAS at unit concentration $\mathbf{s}_1^*$. The real usefulness of the NAS lies in the fact that a plot of the length of the NAS vector ($||\mathbf{x}_1^*||$), also called the scalar NAS, and given as the square root of the sum of the squared elements of the $\mathbf{x}_1^*$ vector) as a function of analyte concentration is linear, the slope being the length of the NAS vector at unit concentration ($||\mathbf{s}_1^*||$). This result immediately leads to an intuitive definition of sensitivity for analyte 1 [43–45]:

$$\mathrm{SEN}_1 = ||\mathbf{x}_1^*||/y_1 = ||\mathbf{s}_1^*|| = \left[\mathbf{s}_1^{\mathrm{T}}\left(\mathbf{I} - \mathbf{s}_2\mathbf{s}_2^{+}\right)\mathbf{s}_1\right]^{1/2} \tag{8}$$

In summary, if both vectorial signals at unit concentration for the pure constituents of a mixture are known or can be estimated from the analysis of mixtures of pure constituents, simple matrix manipulation allows one to define precisely the sensitivity toward a given constituent. A useful relationship between the sensitivity based on this NAS approach and the complete matrix of pure constituent signals can be found by invoking the theory of block pseudoinverse operations [46]: Equation(8) can be generalized to the *n*th constituent of interest in a multiconstituent sample in different forms. One useful form is expressed as a function of the pure profiles for all constituents, ubiquitous in CLS studies:

$$\mathrm{SEN}_n = \left[\delta_n^{\mathrm{T}}\left(\mathbf{S}_{\mathrm{CLS}}^{\mathrm{T}}\mathbf{S}_{\mathrm{CLS}}\right)^{-1}\boldsymbol{\delta}_n\right]^{-1/2} \tag{9}$$

where $\boldsymbol{\delta}_n$ is an $N \times 1$ vector selecting the analyte of interest, and the matrix $\mathbf{S}_{\mathrm{CLS}}$ contains N columns, each with the pure constituent profile $\mathbf{s}_n$ for the *n*th constituent.

A different useful generalization of the multivariate first-order sensitivity can be developed in terms of the vector of regression coefficients, which is specific for a given analyte in a mixture ($\boldsymbol{\beta}_{\mathrm{CLS},n}$). This vector provides the analyte concentration from the below predictive equation

$$y_n = \boldsymbol{\beta}_{\mathrm{CLS},n}^{\mathrm{T}}\mathbf{x} \tag{10}$$

From Equation (10), the sensitivity can be expressed as:

$$\mathrm{SEN}_n = \left(\boldsymbol{\beta}_{\mathrm{CLS},n}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{CLS},n}\right)^{-1/2} \tag{11}$$

Interestingly, Equation (11) provides a useful link to estimate the SEN for first-order algorithms based on inverse models, such as inverse least-squares

(ILS), principal component regression (PCR), and PLS [11]. In contrast to the direct approach of the classical Beer's law, inverse calibration models relate concentrations to signals $\left(y_n = \boldsymbol{\beta}_n^{\mathrm{T}}\mathbf{x}\right)$, and these are able to provide a vector of regression coefficients $\boldsymbol{\beta}_n$ from a suitable set of calibration mixtures, and thus, the analogue of Equation (11) is a useful means of estimating the sensitivity for these methodologies. In this way, sensitivity expressions for both direct and inverse calibration first-order methodologies can be brought into a common form.

3.3 Multiway (Higher-Order) Calibration

Inspired in the useful first-order NAS philosophy, it is possible to estimate the SEN parameter in three-way (second-order) calibration, calculate the NAS, and remove the contribution of constituents other than the analyte of interest using orthogonal projection matrices. One intriguing aspect of this multiway NAS approach is the fact that, in principle, these projections can be carried out in different ways, leading to competing NAS definitions [9].

A matrix signal $\mathbf{X}$ defined in two different instrumental modes for a simple binary mixture can be written as:

$$\mathbf{X} = y_1\mathbf{M}_1 + y_2\mathbf{M}_2 \tag{12}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ are matrix signals at unit concentration for each analyte, and, as before, signal additivity and signal concentration linearity are assumed. If the signals are bilinear and the profiles in both data modes are designated as $\mathbf{b}$ and $\mathbf{c}$, the expression for $\mathbf{X}$ would be:

$$\mathbf{X} = y_1\mathbf{b}_1\mathbf{c}_1^{\mathrm{T}} + y_2\mathbf{b}_2\mathbf{c}_2^{\mathrm{T}} \tag{13}$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ are the pure constituent profiles in the first data mode, and $\mathbf{c}_1$ and $\mathbf{c}_2$ those in the second data mode.

According to the NAS approach, the contributing matrix signal for constituent 2 may be removed from Equation (13) by these simultaneous operations: left-multiplication with a projection matrix orthogonal to $\mathbf{b}_2$ and right-multiplication with an analogous matrix orthogonal to $\mathbf{c}_2$. This outcome leads to one particular sensitivity expression known as HCD (the acronym follows the authors' final initials) [7], which is valid in a certain calibration scenario (see Table 1).

There is an alternative procedure, which involves first unfolding the matrix $\mathbf{X}$ into a vector and then removing the contribution of constituent 2 with a single removing matrix, orthogonal to the unfolded space spanned by constituent 2. This approach leads to a different second-order sensitivity definition, the MKL sensitivity [8], which is valid in a different calibration situation in comparison with the HCD sensitivity (see Table 1). Although the original works on HCD and MKL sensitivity did not employ NAS

TABLE 1 Different Three-Way Sensitivity Definitions Based on Extensions of the NAS Concept

HCD sensitivity	
Authors and ref.	C.N. Ho, G.D. Christian, E.R. Davidson [7]
Comments	Valid for one calibrated constituent in the presence of unexpected constituents
Expression[a]	$\mathrm{SEN}_n = m_n\{[(\mathbf{B}^{\mathrm{T}}\mathbf{B})^{-1}]_{nn}\,[(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}]_{nn}\}^{-1/2}$
MKL sensitivity	
Authors and ref.	N.J. Messick, J.H. Kalivas, P.M. Lang [8]
Comments	Valid in the absence of unexpected constituents
Expression[a]	$\mathrm{SEN}_n = m_n\left\{\left[(\mathbf{B}^{\mathrm{T}}\mathbf{B}) * (\mathbf{C}^{\mathrm{T}}\mathbf{C})\right]^{-1}\right\}_{nn}^{-1/2}$
FO sensitivity	
Authors and ref.	A.C. Olivieri, N.M. Faber [9]
Comments	Valid for any number of calibrated constituents in the presence of unexpected constituents
Expression[a]	$\mathrm{SEN}_n = m_n\left\{\left[\left(\mathbf{B}_{\mathrm{exp}}^{\mathrm{T}}(\mathbf{I}-\mathbf{B}_{\mathrm{unx}}\mathbf{B}_{\mathrm{unx}}^{+})\mathbf{B}_{\mathrm{exp}}\right) * \left(\mathbf{C}_{\mathrm{exp}}^{\mathrm{T}}(\mathbf{I}-\mathbf{C}_{\mathrm{unx}}\mathbf{C}_{\mathrm{unx}}^{+})\mathbf{C}_{\mathrm{exp}}\right)\right]^{-1}\right\} m_n^{-1/2}$

[a]*The symbol "*" indicates the Hadamard matrix product, and the subscript "*nn*" denotes the (n,n) diagonal element of the matrix. The parameter* m *is the total signal of the analyte considered at unit concentration, while the matrices* ***B*** *and* ***C*** *collect the loadings (profiles for the sample constituents in both data modes, normalized to unit length). The subscripts "exp" and "unx" for the FO expression refer to expected and unexpected, respectively.*

arguments for their derivation, the results are identical to those provided by the above NAS-inspired procedures.

Both HCD and MKL equations were condensed into the more general FO definition [9], conceived to take into account all possible calibration situations, including cases not covered by the former two expressions (see Table 1). The derivation required a complicated series of steps, which combined removal of other sample constituents, partly in matrix form and partly in unfolded form. However, the approach could not be straightforwardly extended to four-way (third-order) calibration, where it is apparent that even more alternative NAS definitions may exist. This situation prompted the finding of an alternative solution to the estimation of the multiway sensitivity.

4 SENSITIVITY EXPRESSIONS BASED ON UNCERTAINTY PROPAGATION

4.1 The General Sensitivity Expression

A different definition of sensitivity can be given in terms of uncertainty propagation: the sensitivity parameter SEN_n is considered to measure the degree of output noise from a system for a given input noise. More sensitivity is achieved if low output noise is obtained for a given input noise. In this way, the SEN_n parameter can be defined as the ratio of input to output noise:

$$\mathrm{SEN}_n = \sigma_x / \sigma_y \tag{14}$$

where σ_x and σ_y are the uncertainties in signal and concentration, respectively. This uncertainty propagation approach assumes that the input noise is *iid* and employs a small, perturbing noise value to interrogate how the latter is propagated to prediction. However, it does not imply specific assumptions regarding the properties of the real experimental noise. When calibration is precise, the main source of uncertainty in the predicted concentration is the one stemming from the test sample signals, and the ratio of these uncertainties is a good measure of the SEN_n.

Equation (14) can be used to apply a Monte Carlo additive noise simulation for estimating sensitivities for any calibration model, whether univariate, multivariate, or multiway, as has been recently done [14–16]. This allowed to obtain operational values for the sensitivity in different calibration scenarios, although it does not provide a closed-form sensitivity equation. Equation (14) was also invoked to derive expression for estimating the sensitivity in most of the relevant multiway calibration models, including PARAFAC, MCR-ALS, and PLS/RML, with results which are (1) compatible with the second-order HCD, MKL, and FO (see Table 1 for application scenarios), (2) in agreement with Monte Carlo additive noise simulations, and (3) extendable to data with increasing number of ways. According to these results, it was possible to write an expression for casting all sensitivity equations into a single unified scheme, covering from zeroth-order (univariate calibration) to calibration models based on data of any order and ways [17]. The main result is appropriately condensed into the following expression:

$$\mathrm{SEN}_n = \left\{ \mathbf{g}_n^{\mathrm{T}} \left[\mathbf{Z}_{\mathrm{exp}}^{\mathrm{T}} \left(\mathbf{I} - \mathbf{Z}_{\mathrm{unx}} \mathbf{Z}_{\mathrm{unx}}^{+} \right) \mathbf{Z}_{\mathrm{exp}} \right]^{-1} \mathbf{g}_n \right\}^{-1/2} \tag{15}$$

The different factors appearing in Equation (15) are summarized in Table 2 for the most commonly used models in each calibration order. Both the matrix $\mathbf{Z}_{\mathrm{exp}}$ (the subscript "exp" stands for expected) and the analyte-specific vector $\mathbf{g}_n$ correspond to the calibration phase. The matrix $\mathbf{Z}_{\mathrm{exp}}$ collects profiles (either in pure form or as linear combinations) for the expected constituents present in the calibration set, while $\mathbf{g}_n$ adequately selects or

TABLE 2 General Sensitivity Formula and Corresponding Parameters for Different Data Orders and Algorithms

$$\mathrm{SEN}_n = \left\{ \mathbf{g}_n^{\mathrm{T}} \left[\mathbf{Z}_{\mathrm{exp}}^{\mathrm{T}} \left(\mathbf{I} - \mathbf{Z}_{\mathrm{unx}} \mathbf{Z}_{\mathrm{unx}}^{+} \right) \mathbf{Z}_{\mathrm{exp}} \right]^{-1} \mathbf{g}_n \right\}^{-1/2}$$

		Analyte selector	*Calibration components*	*Unexpected components*		
Model	*Order*	$g_{\mathbf{n}}$	Z_{exp}	Z_{unx}	*Parameters description*	*References*
Univariate calibration						
Univariate	0	1	m_0	–	m_0 = slope of univariate graph	[4]
Multivariate calibration						
CLS	1	$\boldsymbol{\delta}_n$	$\mathbf{S}_{\mathrm{CLS}}$	–	$\boldsymbol{\delta}_n$: $N \times 1$ vector selecting the analyte of interest (all values of 0, except a 1 at analyte index) $\mathbf{S}_{\mathrm{CLS}}$: matrix of pure constituent profiles	[4]
ILS		$\mathbf{y}_{\mathrm{cal},n}$	$\mathbf{X}_{\mathrm{cal}}$		$\mathbf{y}_{\mathrm{cal},n}$: vector of calibration analyte concentrations $\mathbf{X}_{\mathrm{cal}}$: matrix of calibration signals	
PCR		$\mathbf{v}_{\mathrm{PCR},n}$	$\mathbf{P}_{\mathrm{PCR}}$		$\mathbf{v}_{\mathrm{PCR},n}$: vector of latent PLS coefficients $\mathbf{P}_{\mathrm{PCR}}$: matrix of PCR calibration loadings	
PLS		$\mathbf{q}_{\mathrm{PLS},n}$	$\mathbf{W}_{\mathrm{PLS}}$		$\mathbf{q}_{\mathrm{PLS},n}$: $\left(\mathbf{P}_{\mathrm{PLS}}^{\mathrm{T}} \mathbf{W}_{\mathrm{PLS}}\right)^{-1} \mathbf{v}_{\mathrm{PLS},n}^{\mathrm{T}}$ $\mathbf{v}_{\mathrm{PLS},n}$: vector of latent PLS coefficients $\mathbf{W}_{\mathrm{PLS}}$: matrix of PLS calibration weights $\mathbf{P}_{\mathrm{PLS}}$: matrix of PLS calibration loadings	[4,11]

Multiway calibration						
MCR-ALS	2	$\boldsymbol{\delta}_n$	$(m_n/J^{1/2})\mathbf{C}_{exp}$	$\mathbf{C}_{unx}$	J: no. of sensors of each submatrix in augmented mode m_n: slope of pseudounivariate plot $\mathbf{C}_{exp}$: profiles in nonaugmented mode for expected constituents $\mathbf{C}_{unx}$: profiles in nonaugmented mode for unexpected constituents	[15,17]
PARAFAC	2,3,…	$\boldsymbol{\delta}_n$	See parameters description	Second-order: Equation (16); third-order: Equation (26); fourth-order: same pattern as Equations (16) and (26) adding the extra data mode	Second-order: $\mathbf{Z}_{exp} = m_n \mathbf{C}_{exp} \odot \mathbf{B}_{exp}$[a] Third-order: $\mathbf{Z}_{exp} = m_n\, \mathbf{D}_{exp} \odot \mathbf{C}_{exp} \odot \mathbf{B}_{exp}$ Fourth-order: $\mathbf{Z}_{exp} = m_n\, \mathbf{E}_{exp} \odot \mathbf{D}_{exp} \odot \mathbf{C}_{exp} \odot \mathbf{B}_{exp}$	[14,17]
U-PLS/RML		$\mathbf{v}_{UPLS,n}$	$\mathbf{P}_{UPLS}$		$\mathbf{P}_{UPLS}$: matrix of U-PLS calibration loadings $\mathbf{v}_{UPLS,n}$: vector of latent PLS coefficients	[16,17]

[a] *The symbol ⊙ refers to the Khatri–Rao product. For two matrices* $\mathbf{A}$ *and* $\mathbf{B}$*, the ith. column of* $\mathbf{A} \odot \mathbf{B}$ *follows from the ith. columns of* $\mathbf{A}$ *and* $\mathbf{B}$ *as* $\mathrm{vec}(\mathbf{b}_i^T \mathbf{a}_i)$.

combines the latter information, making it specific for the nth analyte of interest. The final factor in Equation (15) is the matrix $\left(\mathbf{I}-\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)$, which is the mathematical manifestation of the second-order advantage, and thus, it only appears in higher-order (three-way and beyond) calibration methodologies (see Table 2). $\mathbf{Z}_{\text{unx}}$ is a block matrix built from the extracted profiles of the unexpected constituents, and for second-order data it can be expressed in a general way as:

$$\mathbf{Z}_{\text{unx}} = [\mathbf{c}_1 \otimes \mathbf{I}_b | \mathbf{I}_c \otimes \mathbf{b}_1 | \mathbf{c}_2 \otimes \mathbf{I}_b | \mathbf{I}_c \otimes \mathbf{b}_2 | \cdots] \tag{16}$$

The purpose of $\left(\mathbf{I}-\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)$ is to correct the matrix of profiles for the expected constituents ($\mathbf{Z}_{\text{exp}}$), for the overlapping effect of the profiles for the unexpected constituents (hence the subscript "unx"), or potential interfering agents. Specifically, the matrix $\left(\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)$ only appears when achieving the second-order advantage because only in this case is such information available. This is why in Table 2, only the entries corresponding to multiway algorithms are filled with respect to this part of Equation (16). The profiles for the unexpected constituents may be: (1) true constituent profiles (or approximations to them) provided, for example, by MCR-ALS, PARAFAC, and other all multilinear decompositions or (2) latent profiles (linear combinations or loadings), as retrieved by RML. What is relevant is that $\left(\mathbf{I}-\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)$ defines a projection orthogonal to the space spanned by the unexpected constituents, because $\mathbf{Z}_{\text{unx}}$ only contains information relative to the signals for the latter agents.

The fact that closed expressions for $\mathbf{Z}_{\text{exp}}$, $\mathbf{g}_n$, and $\left(\mathbf{I}-\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)$ can be written for all calibration methodologies (see Table 2) from zeroth- to any order indicates that Equation (15) is a completely general expression for estimating sensitivities. It is also worth noticing the properties of the multiway sensitivity defined by Equation (15): (1) it is analyte specific because the factor $\mathbf{g}_n$ depends on the analyte of interest; (2) it is sample specific because the composition of each test sample is unique as regards the unexpected constituents, generating a unique $\mathbf{Z}_{\text{unx}}$ matrix; and (3) it is algorithm specific because each data processing methodology provides a set of specific $\mathbf{Z}_{\text{exp}}$, $\mathbf{g}_n$, and $\left(\mathbf{I}-\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)$ factors.

4.2 Univariate Calibration

For the case of univariate calibration, the terms of the general Equation (15) are scalars, except $\mathbf{Z}_{\text{unx}}$, since no unexpected constituents are possible in this methodology and, therefore, does not defined: $\mathbf{g}_n = 1$, and $\mathbf{Z}_{\text{exp}} = m_0$, leading to $\text{SEN}_n = m_0$, the slope of the calibration graph. This agrees with the IUPAC definition and also with the simple and intuitive uncertainty analysis of Equation (4): if the calibration were precise, uncertainties in x will propagate to y through $\boldsymbol{\sigma}_y = m_0^{-1}\boldsymbol{\sigma}_x$, and thus SEN_n will be equal to m_0 (see Table 2).

4.3 First-Order Calibration

The definitions of $\mathbf{Z}_{\text{exp}}$ depend on the specific data processing algorithm used in first-order calibration, but since no unexpected constituents should appear in the test samples, $\mathbf{Z}_{\text{unx}}$ does not exist in none of the cases, and thus $\left(\mathbf{I}-\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)=\mathbf{I}$. Therefore, the general Equation (15) becomes Equation (9) for CLS and analogous expressions for ILS, PCR, and PLS [17], in full agreement with the NAS-based sensitivity approach.

Uncertainty propagation allows one to achieve the same results, directly from the general predictive equation for analyte n:

$$y_n = \boldsymbol{\beta}_n^{\text{T}}\mathbf{x} \tag{17}$$

where $\boldsymbol{\beta}_n$ is the vector of regression coefficient for any first-order methodology. If only $\mathbf{x}$ carries uncertainty, it follows that the uncertainty in concentration is given by:

$$\sigma_y = \left(\boldsymbol{\beta}_n^{\text{T}}\boldsymbol{\beta}_n\right)^{1/2}\sigma_x \tag{18}$$

From Equation (18), $\text{SEN}_n = \left(\boldsymbol{\beta}_n^{\text{T}}\boldsymbol{\beta}_n\right)^{-1/2}$ immediately follows through the uncertainty propagation approach. This sensitivity parameter is analyte specific but does not depend on the composition of the test sample, because the vector of regression coefficients stems from the processing of the calibration data only.

4.4 Multiway (Higher-Order) Calibration

4.4.1 Multilinear Algorithms

Multilinear algorithms such as PARAFAC [27] and its variants based on the multilinear model [47–49] provide approximations to pure constituent profiles, whether they belong to the category of expected (calibrated) or unexpected (potentially interferent), thus achieving the second-order advantage. Each constituent is characterized by instrumental profiles describing their behavior in the different data modes. In the usual setting, these profile vectors are normalized to unit length, and thus, the scaling factor with respect to analyte concentration is left to the slope (m_n) of the pseudounivariate prediction graph (the latter is a plot of the scores or relative concentrations of a given analyte vs. its nominal calibration concentrations). As a function of the relevant parameters for multilinear multiway calibration, the recently derived expression for the sensitivity in multilinear models is [50]

$$\text{SEN}_n = m_n\left\|n\text{th row of }\left[\mathbf{Z}_{\text{exp}}^{\text{T}}\left(\mathbf{I}-\mathbf{Z}_{\text{unx}}\mathbf{Z}_{\text{unx}}^{+}\right)\mathbf{Z}_{\text{exp}}\right]^{+}\right\|^{-1} \tag{19}$$

The matrix $\mathbf{Z}_{\text{exp}}$ is defined as a function of a loading matrices for second- (two data modes), third- (three data modes), and fourth-order (four data

modes), while the $\mathbf{Z}_{unx}$ is defined as a function of the loading profiles of the unexpected constituents in the various data modes [17]. It may be noticed that for three-way (second-order) calibration, Equations (15) and (19) appear to be different than the MKL, HCD, and FO expressions (see Table 1); however, the latter numerical results are identical to those provided by Equation (15), indicating that all previous approximations based on the NAS are special cases of the general uncertainty propagation expression [14].

4.4.2 Multivariate Curve Resolution-Alternating Least-Squares

The corresponding SEN_n expression for the MCR-ALS algorithm applied in the extended mode has been recently derived [15].

$$SEN_n = m_n \left[J \left(\mathbf{C}^T \mathbf{C} \right)^{-1}_{nn} \right]^{-1/2} \tag{20}$$

In Equation (20), J is the number of data points in each submatrix in the augmented mode, and m_n is the slope of the MCR-ALS pseudounivariate graph (built in a similar manner to PARAFAC, i.e., plotting analyte scores vs. nominal calibration concentrations). Assuming successful decomposition of the augmented matrix $\mathbf{X}_{aug}$ into two matrices ($\mathbf{B}_{aug}$ and $\mathbf{C}$), containing the constituent profiles in the augmented mode and in the nonaugmented mode, respectively, the sensitivity depends on the nonaugmented profiles $\mathbf{C}$, which can be further separated into $\mathbf{C}_{exp}$ and $\mathbf{C}_{unx}$, containing the profiles for the expected (present in calibration) and unexpected constituents, respectively.

The MCR-ALS sensitivity expression can also be shown to be adequately covered by the general Equation (15). In Equation (20), J is the number of data points in each submatrix in the augmented mode. Since each data matrix is assumed to be of size $J \times K$, this also assumes that augmentation has been performed columnwise. In the case of row-wise augmentation, J should be replaced by K in Equation (20). On the other hand, the matrix $\mathbf{C}$ contains the profiles for all sample constituents in the nonaugmented data mode, and the shorthand notation $\left(\mathbf{C}^T \mathbf{C} \right)^{-1}_{nn}$ implies selecting the (n,n) diagonal element of the inverse of matrix $(\mathbf{C}^T \mathbf{C})$. To adapt Equation (20) to the present approach, the matrix $\mathbf{C}$ is divided into two blocks, one for the constituents present in calibration ($\mathbf{C}_{exp}$) and another one for the unexpected constituents ($\mathbf{C}_{unx}$):

$$\mathbf{C} = \left[\mathbf{C}_{exp} | \mathbf{C}_{unx} \right] \tag{21}$$

It can further be shown that:

$$\begin{aligned} \left(\mathbf{C}^T \mathbf{C} \right)^{-1} &= \left(\left[\mathbf{C}_{exp} | \mathbf{C}_{unx} \right]^T \left[\mathbf{C}_{exp} | \mathbf{C}_{unx} \right] \right)^{-1} \\ &= \left[\mathbf{C}^T_{exp} \left(\mathbf{I} - \mathbf{C}_{unx} \mathbf{C}^+_{unx} \right) \mathbf{C}_{exp} \right]^{-1} \end{aligned} \tag{22}$$

Then, the (n,n) diagonal element of the latter matrix can be found using the vector $\boldsymbol{\delta}_n$ as selector:

$$\left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\right)^{-1}_{nn} = \boldsymbol{\delta}_n^{\mathrm{T}}\left[\mathbf{C}_{\mathrm{exp}}^{\mathrm{T}}\left(\mathbf{I}-\mathbf{C}_{\mathrm{unx}}\mathbf{C}_{\mathrm{unx}}^{+}\right)\mathbf{C}_{\mathrm{exp}}\right]^{-1}\boldsymbol{\delta}_n \tag{23}$$

and the sensitivity is given by the equation:

$$\mathrm{SEN}_n = m_n J^{-1/2}\left\{\boldsymbol{\delta}_n^{\mathrm{T}}\left[\mathbf{C}_{\mathrm{exp}}^{\mathrm{T}}\left(\mathbf{I}-\mathbf{C}_{\mathrm{unx}}\mathbf{C}_{\mathrm{unx}}^{+}\right)\mathbf{C}_{\mathrm{exp}}\right]^{-1}\boldsymbol{\delta}_n\right\}^{-1/2} \tag{24}$$

Finally, $\mathbf{Z}_{\mathrm{exp}}$ and $\mathbf{Z}_{\mathrm{unx}}$ from Equation (15) are equal to $\mathbf{C}_{\mathrm{exp}}$ and $\mathbf{C}_{\mathrm{unx}}$, respectively, and the vector $\mathbf{g}_n$ is equal to the multilinear selector $\boldsymbol{\delta}_n$, thereby achieving an expression equivalent to general Equation (15).

4.4.3 Partial Least-Squares/Residual Multilinearization

For multiway algorithms with a latent-based calibration, such as PLS/RML, the corresponding sensitivity expression has already been developed in the same format as the general equation (15). For U-PLS calibration, for example, $\mathbf{Z}_{\mathrm{exp}}$ is composed of columns which are the calibration loadings, which is understandable, since they represent the behavior of the calibrated constituents in signal space. Here, the vector $\mathbf{g}_n$ does not act as selector of a particular analyte loading, but appropriately combines the loadings in a manner, which specifically reflects the behavior of the analyte of interest. It is equal to the vector of analyte-specific regression coefficients, defined in the space of the latent variables. The above discussion concerning the properties of the $\left(\mathbf{I}-\mathbf{Z}_{\mathrm{unx}}\mathbf{Z}_{\mathrm{unx}}^{+}\right)$ matrix is also pertinent in this case. The uncertainty propagation approach fully agrees with the expression for the U-PLS/residual bilinearization (RBL) sensitivity, which was previously derived from NAS considerations [51]. An analogous expression can be derived for N-PLS/RBL [16].

4.5 Other Multiway Algorithms

The general Equation (15) has been applied to assess the sensitivity for several algorithms commonly employed for multiway calibration. However, there are additional methodologies, based on eigenvector–eigenvalue analysis [52,53], which are somewhat less employed. In addition, the latter ones always achieve the lowest HCD sensitivity, even when various constituents are calibrated [51], probably due to the very limited information provided to the model for the single calibration sample, in contrast to methodologies relying on multiple calibration samples.

Other algorithms for which sensitivity studies are lacking are multilinear least-squares/RML (MLLS/RML) [32,34,35,52,53] and PARAFAC2, a variant of PARAFAC conceived to cope with nonmultilinear multiway data with one trilinearity breaking mode, e.g., chromatographic-spectral second-order

data, whose sensitivity properties have yet to be explored. Recently, a new version of PARAFAC (Augmented PARAFAC), specifically designed for four-way nonmultilinear data, was used to resolve different third-order chromatographic systems [54]. In the latter work, sensitivities were calculated using expressions based on the multilinear approach, i.e., as if the system were processable by quadrilinear four-way PARAFAC [14]. Thus, the results should in principle be rather overoptimistic. However, it is now possible to postulate an expression for sensitivity that fits better to the three-way augmented PARAFAC model: as for MCR-ALS, Equation (15) becomes similar to Equation (23), but in this case, it is possible to use the information of the two pure profiles **B** and **C** (see FO sensitivity in Table 1):

$$\mathrm{SEN}_n = m_n J^{-1/2} \left\{ \boldsymbol{\delta}_n^{\mathrm{T}} \left[\left(\mathbf{B}_{\mathrm{exp}}^{\mathrm{T}} \left(\mathbf{I} - \mathbf{B}_{\mathrm{unx}} \mathbf{B}_{\mathrm{unx}}^{+} \right) \mathbf{B}_{\mathrm{exp}} \right) \left(\mathbf{C}_{\mathrm{exp}}^{\mathrm{T}} \left(\mathbf{I} - \mathbf{C}_{\mathrm{unx}} \mathbf{C}_{\mathrm{unx}}^{+} \right) \mathbf{C}_{\mathrm{exp}} \right) \right]^{-1} \boldsymbol{\delta}_n \right\}^{-1/2} \tag{25}$$

In Equation (25), all symbols have the same meaning that the MCR-ALS context, but the matrices **B** and **C** contain the profiles for all sample constituents in the two nonaugmented data modes, an option which is not possible in the expression of MCR-ALS sensitivity. Figure 3 shows the sensitivities for simulated four-way data ternary systems, consisting in two analytes and one interferent, estimated through Monte Carlo/augmented PARAFAC and Monte Carlo/MCR-ALS calculations, using Equation (24) for MCR-ALS and Equation (25) for Augmented PARAFAC. Interestingly, the sensitivity expression for augmented PARAFAC leads to better results than for MCR-ALS, meaning that Equation (25) is well suited for the proposed objective.

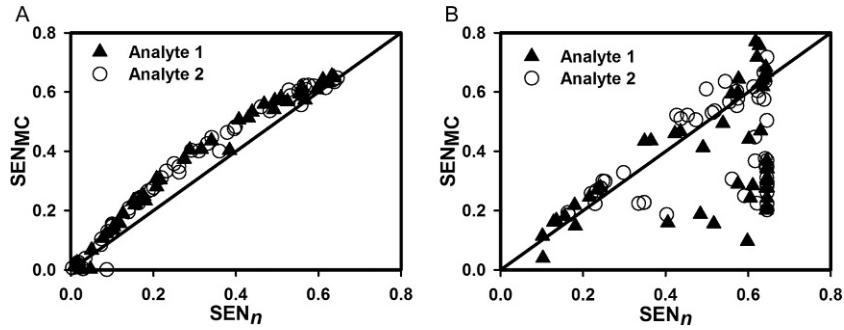

FIGURE 3 Four-way sensitivity expressions versus Monte Carlo values for 100 cases of a ternary system: (A) Augmented PARAFAC (Equation 25) and (B) MCR-ALS (Equation 24).

4.6 Further Multiway Systems

Generalization of $\mathbf{Z}_{unx}$ in Equation (15) to further data modes is possible by noting, for example, that in four-way (third-order) calibration, the spaces spanned by $\mathbf{Z}_{unx}$ are the three possible combinations of pairs of modes: for a given interfering constituent, $\mathbf{Z}_{unx}$ contains blocks of columns for each unexpected constituent, e.g., for the unexpected agent 1, the first block will look as follows [17]:

$$\mathbf{Z}_{unx} = [\mathbf{d}_1 \otimes \mathbf{c}_1 \otimes \mathbf{I}_b | \mathbf{d}_1 \otimes \mathbf{I}_c \otimes \mathbf{b}_1 | \mathbf{I}_d \otimes \mathbf{c}_1 \otimes \mathbf{b}_1 | \cdots] \tag{26}$$

where $\mathbf{I}_b$, $\mathbf{I}_c$, and $\mathbf{I}_d$ are $J \times J$, $K \times K$, and $L \times L$ identity matrices. This $\mathbf{Z}_{unx}$ matrix is easily constructed for any number of unexpected constituents.

If a data mode is added (five-way calibration), four different combinations of triads of profiles in each of the possible sets of three modes will provide a $\mathbf{Z}_{unx}$ matrix for each unexpected agent. For more data modes, specifically, for $(N+1)$-way (Nth-order) calibration, the blocks of $\mathbf{Z}_{unx}$ for each unexpected agent will include all possible combination of profiles in $(N-1)$ modes. In sum, $\mathbf{Z}_{unx}$ conceivably represents the space spanned by the unexpected constituents, but in a nonclassical way, although the systematic block characteristics of this matrix makes it easy to build it for any data order and number of interfering agents.

5 OTHER FIGURES OF MERIT

5.1 Analytical Sensitivity

One potential problem with the interpretation of the plain sensitivity is that it depends on the specific type of signal employed for developing a calibration methodology. The value of SEN_n has units of (signal $\times$ concentration^{-1}), and therefore, sensitivities derived from spectral and electrochemical measurements cannot be compared on an equal basis. For these reasons, the analytical sensitivity (γ) has been proposed as a better indicator for comparison purposes, as the ratio between sensitivity and instrumental noise [55]:

$$\gamma_n = \mathrm{SEN}_n / \sigma_x \tag{27}$$

The parameter γ_n has units of (concentration)$^{-1}$, is independent of the measured signal, and can be employed to compare different methodologies. Comparison of Equations (14) and (27) implies that $\gamma_n = \sigma_y^{-1}$, and thus the analytical sensitivity has been interpreted as the inverse of the minimum concentration difference which can be appreciated across the linear analytical range, although this appears to be a rather qualitative statement, less rigorous than the detection capabilities to be described below. In any case, having estimated the sensitivity, a measure of the instrumental noise level allows one to compute the analytical sensitivity through Equation (27).

5.2 Selectivity

According to IUPAC, selectivity is the extent to which a method can be used to determine particular analytes in mixtures or matrices without interferences from other constituents of similar behavior [56]. However, this qualitative definition does not imply a specific procedure for the estimation of a numerical selectivity parameter [50].

Several requirements have been proposed for a consistent numerical selectivity [50]: (1) a change in the calibration data should be reflected in changes in selectivity, (2) changes in individual analyte selectivities should produce corresponding changes in the selectivity and the amount of these changes should be comparable in size, (3) values such as infinity should not be obtained, (4) a relation between selectivity and prediction uncertainty is desirable, (5) numerical results should be possible for overdetermined systems (having more sensors or wavelengths than components), and (6) generalization to multiway data should be straightforward.

The simplest way in which a selectivity parameter can be defined for most calibration scenarios, complying with the above requirements, is as the dimensionless ratio between two analyte sensitivity values: the sensitivity in a mixture and the sensitivity when all other sample constituents are absent [57].

$$\mathrm{SEL}_n = \mathrm{SEN}_n(\text{in a mixture})/\mathrm{SEN}_n(\text{pure}) \tag{28}$$

In univariate calibration, SEL should be equal to 1 (100%, meaning full selectivity), because no interfering agents are allowed. In first-order CLS calibration, Equation (28) naturally follows as Equation (28) by setting the denominator as a measure of the pure analyte signal [50]. However, for latent variables-based calibration models, no approximations to pure analyte profiles are available, and hence, the selectivity cannot be precisely defined. Although there have been proposals to use the total signal for a given test sample as denominator in Equation (28) in these cases, i.e., this makes the selectivity highly dependent on the unknown samples. Consequently, it may only be sensible to define the selectivity when the pure analyte signal is either adequately retrieved by the processing algorithm or known from separate experiments.

For the multiway analysis world, the defining Equation (28) implies that the multiway selectivity is accessible when the pure analyte signal is adequately retrieved by the processing algorithm [8,11,58], as in the case of multilinear decomposition analysis, for which the selectivity (SEL_n) is directly given by:

$$\mathrm{SEL}_n = \mathrm{SEN}_n/m_n \tag{29}$$

where m_n is the slope of the pseudounivariate calibration graph. The degree by which SEN_n departs from m_n in Equation (29) is adequately measured by the level of overlapping among the profiles for the various constituents. Since

$SEN_n < m_n$, the value of SEL_n continuously varies between 0 (null selectivity) and 1 (100%, full selectivity).

In the case of MCR-ALS, the selectivity is [15] as follows:

$$SEL_n = SEN_n J^{1/2}/m_n \tag{30}$$

As was the case with Equation (29), Equation (30) leads to continuous values in the range 0–1, depending on the relative degree of overlapping among the profile for the various sample constituents.

The latter expressions are not the only ones employed for the selectivity calculation in multiway calibration [59–61], which shows that there is some controversy about the numerical concept of selectivity.

5.3 Prediction Uncertainty

Prediction uncertainty is based on two proposals for estimating standard errors in multivariate/multiway analysis [62]: (1) resampling techniques such as jack-knife or bootstrap [63] and (2) error propagation, which is preferable because it leads to closed-form expressions and permits better insight into the relative impact of various uncertainty sources on the prediction error [57,64].

The best approximation to concentration variance is the well-known three-term expression (valid for propagation of homoscedastic and uncorrelated noise) [57,64,65]:

$$\sigma_y^2 = SEN_n^{-2}\sigma_x^2 + hSEN_n^{-2}\sigma_x^2 + h\sigma_{y_{cal}}^2 \tag{31}$$

where σ_x^2 is the variance in instrumental signals, h is the sample leverage, and $\sigma_{y_{cal}}^2$ is the variance in calibration concentrations. The three terms in the right-hand side of Equation (31) derive, by uncertainty propagation, from the following three sources: (1) instrumental signals for the test sample, (2) instrumental signals for the calibration samples, and (3) calibration concentrations. The first and probably the most relevant of these contributions is directly propagated and is inversely proportional to the squared sensitivity. The second and third terms stem from calibration uncertainties and are scaled by the sample leverage h, a dimensionless parameter placing the sample relative to the calibration space. A simple expression exists for h in univariate calibration and in pseudounivariate multiway calibration, whereas in other situations the value of h depends on the presence and levels of additional sample constituents. A general equation is able to appropriately cover all cases, however:

$$h = \mathbf{f}_{test}^{T}\left(\mathbf{F}_{cal}^{T}\mathbf{F}_{cal}\right)^{-1}\mathbf{f}_{test} \tag{32}$$

where $\mathbf{F}_{cal}$ is a matrix (or vector) corresponding to the calibration set of samples, and $\mathbf{f}_{test}$ is a vector for the test sample. It is important to notice that $1/I_{cal}$

(I_{cal} is the number calibration samples) should be added to Equation (32) for mean-centering data, and when the model includes an intercept, as in univariate calibration through Equation (4), where h becomes the familiar expression:

$$h = \frac{1}{I_{cal}} + \frac{(y - \bar{y}_{cal})^2}{\sum_{i=1}^{I_{cal}} (y - \bar{y}_{cal})^2} \tag{33}$$

where y is the predicted analyte concentration, y_i is its nominal concentration in the ith calibration sample, and $\bar{y}_{cal}$ is the mean calibration concentration. It may be noticed that Equation (31) is accurate for the univariate case [4] and for CLS first-order calibration [66]. For the remaining calibrations, the first term of Equation (31) is accurate [64], while the remaining two terms are appropriate approximations [9,16].

5.4 Detection Capabilities

The modern definition of the LOD is due to Currie's pioneering work on hypothesis-based detection limit theory [67] and is an important figure of merit to be reported because it defines the minimum analyte concentration that is detectable with a certain degree of confidence. However, the definition of the LOD officially recommended by IUPAC is somewhat less simpler than the latter idea: it first requires one to define a critical concentration level (CL), which is the level for the detection decision, involving a certain risk of false detects. The LOD is then defined as a CL for which the risk of false nondetects has a probability β [68–71]. Following the above cited works, and according to the expressions presented herein, an LOD expression can be achieved, as proportional to the uncertainty in predicted concentration near a blank sample [17,72,73]:

$$\mathrm{LOD}_n = 3.3\left(\mathrm{SEN}_n^{-2}\sigma_x^2 + h_0\mathrm{SEN}_n^{-2}\sigma_x^2 + h_0\sigma_{y_{cal}}^2\right) \tag{34}$$

where the subscript n identifies a particular analyte of interest, h_0 is the leverage for the blank sample, and the factor 3.3 is the sum of t-coefficients accounting for types I and II errors at 95% confidence level [57]. The factor in front of Equation (34) may be corrected for other probabilities and degrees of freedom. Notice the assumptions underlying Equation (34): (1) the LOD_n is close enough to the blank so that the leverage at the LOD level is equal to the blank leverage h_0, otherwise, complex corrections are required [74] and (2) the distance from the blank to the LOD is given as a sum of two confidence intervals; a more rigorous treatment suggests the use of a noncentrality parameter of a noncentral t distribution instead of a sum of classical t-coefficients [71]. It is likely, however, that the values provided by Equation (34) and more elaborate statistical approaches do not significantly differ [75,76].

In univariate calibration, the subscript n may be dropped and the LOD characterizes the detection capability toward the analyte under study. However, for first-order methodologies, SEN_n is analyte specific, as explained above, but the leverage h_0 is also sample specific, meaning that different blank samples (samples where the analyte is absent, but contain varying proportions of the remaining constituents) have different associated values of h_0. Therefore, the LOD_n not only becomes analyte specific but also sample specific. Even more, in higher-order calibration, as already discussed, the value of SEN_n is analyte, sample, and algorithm specific (incidentally, the leverage h_0 is not sample specific when pseudounivariate calibration is employed) [10]. For this reasons, the detection capability toward a given analyte depends on various factors beyond the instrumental signals measured for a set of calibration samples. Following this concept, and in order to overcome the sample dependency issue, a recent approach has been proposed based on the interpretation of the LOD as an interval [25]. Although this proposal has been specifically developed for first-order PLS calibration, it could be potentially extended to other algorithms as long as they are consistent with Equation (34). The main ideas underlying this definition are (1) each test sample has a specific LOD value, (2) the universe of test samples is represented by the calibration set of samples, (3) the leverages for the calibration samples can be extrapolated to zero analyte concentration, and (4) a range of LOD values can be estimated for the PLS model as a whole.

As shown in Figure 4, the lower and upper limits of the LOD interval correspond to the calibration samples with the lowest and largest extrapolated leverages to zero analyte concentration and can be calculated using the following expressions [25]:

$$\mathrm{LOD}_{\min} = 3.3\left[\mathrm{SEN}^{-2}\mathrm{var}(x) + h_{0\min}\mathrm{SEN}^{-2}\mathrm{var}(x) + h_{0\min}\mathrm{var}(y_{\mathrm{cal}})\right]^{1/2} \quad (35)$$

$$\mathrm{LOD}_{\max} = 3.3\left[\mathrm{SEN}^{-2}\mathrm{var}(x) + h_{0\max}\mathrm{SEN}^{-2}\mathrm{var}(x) + h_{0\max}\mathrm{var}(y_{\mathrm{cal}})\right]^{1/2} \quad (36)$$

where

$$h_{0\min} = \frac{\bar{y}_{\mathrm{cal}}^2}{\sum_{i=1}^{I} y_i^2} \quad (37)$$

$$h_{0\max} = \max\left(h_{0\mathrm{cal}}\right) \quad (38)$$

It is interesting to note that the leverage in Equation (34) corresponds to the value obtained in univariate calibration with a given calibration set, provided other sample components are absent. In Equation (35), on the other hand, $h_{0\mathrm{cal}}$ refers to the leverages for the projections of all calibration samples onto H_0, which is the zero concentration hyperplane in a score space having a

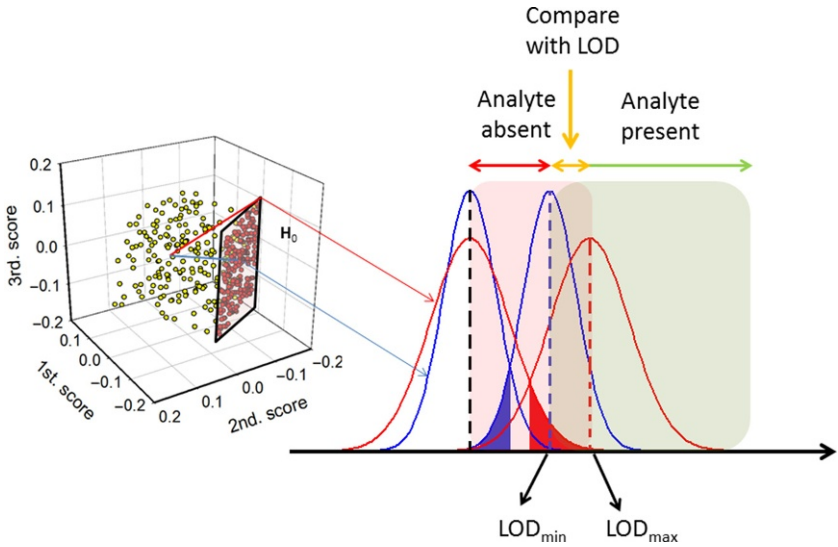

FIGURE 4 Schematic representation of the limit of detection (LOD) interval approach to calculate the LOD in PLS calibration. Red (gray in the print version) solid lines, leverage and corresponding standard deviation used to calculate LOD_{min}; blue (dark gray in the print version) solid lines, leverage and corresponding standard deviation used to calculate LOD_{max}. *Adapted with permission from Ref. [25]. Copyright 2014 American Chemical Society.*

number of dimensions equal to the number of calibration latent variables (Figure 4), and h_{0cal} can be calculated as:

$$h_{0cal} = h_{cal} + h_{0min}\left[1 - \left(\frac{y_{cal}}{\bar{y}_{cal}}\right)^2\right] \tag{39}$$

In the concentration range $LOD_{min} < y < LOD_{max}$, a reasonable way to decide if the analyte is detected or not is to estimate a specific LOD value for the corresponding test sample, approximating its real leverage h to the leverage h_0, which would correspond to its background components (i.e., in the absence of analyte). This is equivalent to taking the sample as if it were a blank, which is conceivable because its analyte concentration is most probably very low. The obtained LOD value can then be employed to check whether the predicted concentration is below (analyte absent) or above (analyte present) the sample-specific LOD.

Another common approach to calculate limits of detection in multivariate calibration bears some similarities with the univariate approach, and this is why it is sometimes called "pseudounivariate LOD" [77]. In this strategy, the analyte concentrations estimated for the calibration set of samples by the PLS model are plotted against their nominal or reference concentrations. The result is a pseudounivariate calibration graph in which the vertical scale

is the estimated analyte concentration instead of either instrumental or latent variables. The graph is processed as in univariate calibration, assuming that the detection limit is insensitive to linear transformations applied to the signal. This leads to an LOD_{pu} value, estimated from the classical univariate Equation (4):

$$LOD_{pu} = 3.3 s_{pu}^{-1} \left[(1 + h_{0min} + 1/I) \, var_{pu} \right]^{1/2} \tag{40}$$

where s_{pu} is the slope of the pseudounivariate line and var_{pu} is the variance of the regression residuals.

The parameter LOD_{pu} has the advantage of being a single figure of merit characterizing the overall PLS calibration model. However, one critical point about this estimator is the absence of a term accounting for calibration concentration uncertainties. This difference shows up if LOD_{pu} and LOD_{min}/LOD_{max} are compared by Monte Carlo noise addition studies, which indicate that the LOD_{pu} distribution is centered at the lower limit LOD_{min} of the presently proposed LOD interval, provided the noise in calibration concentrations is negligible compared to the level of noise in instrumental signals. In contrast, when concentration uncertainties compete with the instrumental noise in relative size, the mutual relationship among LOD_{pu}, LOD_{min}, and LOD_{max} is less clear.

This can be explained on the basis of how the errors in calibration concentrations $var(y_{cal})$ are incorporated into the LOD definitions. In the estimation of both LOD_{min} and LOD_{max}, the latter contribution is scaled by the leverage, but in LOD_{pu}, it is directly incorporated into the first, test-sample-dependent term of the LOD expression. In the latter case, the "signal" is replaced by the estimated concentration. This means that although the term that takes into account the errors in calibration concentrations is not present in an explicit way, concentrations errors are directly propagated to the standard error in predicted concentrations. In any case, it is important to remark that the conceptual approach to LOD_{pu} is radically different than the presently proposed range of LOD values, which is clearly consistent with the latest advances in error-in-variables theory [11], and leads to a better insight into PLS detection capabilities, because it helps to understand the effect of background and potential interfering agents on the analyte detection for complex samples [25].

The limit of quantitation (LOQ_n), in turn, is estimated as the CL for which the relative prediction error is 10% and is easily set at a concentration value which is 10 times the associated prediction uncertainty [57]:

$$LOQ_n = 10 \left(SEN_n^{-2} \sigma_x^2 + h_0 SEN_n^{-2} \sigma_x^2 + h_0 \sigma_{y_{cal}}^2 \right) \tag{41}$$

Analogous considerations to those for LOD_n regarding the analyte and sample dependence of the LOQ_n apply.

6 COMPARISON OF FIGURES OF MERIT

The measuring and processing of multivariate data leads to a sensitivity increase, derived from multiple redundant measurements and noise averaging. The sensitivity increase can now be precisely computed using the general expression (15). This may help in advanced planning and in anticipating the sensitivity gain for complex multiway experiments. To illustrate the sensitivity gain when increasing the number of instrumental sensors and data orders, a real example will be presented, which shows the usefulness of the general expression to calculate sensitivity and the resulting figures of merit. In this experimental fourth-order data set, excitation–emission fluorescence matrices were measured as a function of time and pH to quantify the amount of the fluorescent pesticide carbaryl, which hydrolyzes in alkaline media to fluorescent 1-naphthol. The calibration samples only contain the analyte carbaryl, but the test samples contain, in addition to the analyte, another fluorescent pesticide as an interfering agent (thiabendazole or fuberidazole). Hence, the second-order advantage is required for successful analyte determination [16].

Given the features of the data, it is possible to explore all the different possibilities provided by second-, third-, and fourth-order calibration. To do this, the corresponding subarrays can be generated by fixing the time, the pH (third-order data), or both pH and time simultaneously (second-order data). When the second-order data were analyzed using U-PLS/RBL, the use of cross-validation rendered an optimum number of calibration latent variables equal to 1. It is important to notice that two chemical components occur in calibration. However, since they are mutually correlated because one component is hydrolyzed to yield the second one, a single U-PLS latent variable is understandable. The same situation stands for higher orders. As can be seen in Figure 5 for two typical test samples with different interferents, the RBL procedure was able to return the corresponding profiles, achieving the second-order advantage. The figures of merit are shown in Table 3 and will be compared below with those corresponding to third- and fourth-order data analysis. When third-order data are generated by fixing the pH value to 10, residual trilinearization (RTL), as in the case of RBL, allowed to model the corresponding interfering agent in each sample. The complete fourth-order data were finally submitted to U-PLS/RQL, with similar qualitative results in comparison with above analysis, but with an additional profile in the pH mode for the interfering agent profiles.

The figures of merit for the analyte carbaryl are reported in Table 3 for comparison with previous methodologies. In comparing the results for second-, third-, and fourth-order data for the studied experimental system, increasing sensitivities and analytical sensitivities are apparent on increasing the data order (Figure 6). A steady improvement in the average concentration error indicators is also observed, as well as in uncertainty in predicted concentrations and detecting capabilities (LOD and LOQ). However, in agreement

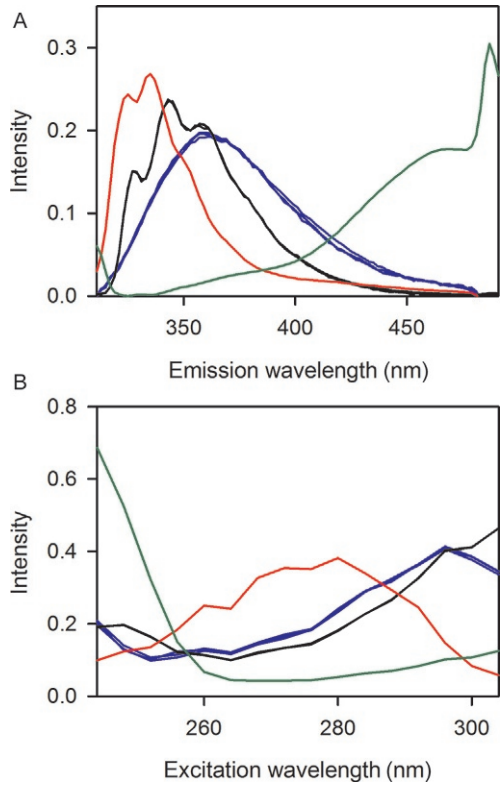

FIGURE 5 Excitation (A) and emission (B) profiles for the various components of a typical experimental example. Green (gray in the print version) and red (gray in the print version) lines correspond to the experimental spectra for the analyte carbaryl and its hydrolysis product 1-naphthol, respectively. Blue (black in the print version) and black lines (three superimposed similar traces) indicate the profiles for the interfering agents fuberidazole and thiabendazole, as retrieved from test samples containing interferents, by U-PLS/RML analysis of second-, third-, and fourth-order data. *Reprinted with permission from Ref. [16]. Copyright 2012 American Chemical Society.*

with the simulations presented in the previous section, the improvement in concentration uncertainty, LOD, and LOQ is not directly proportional to the gain in sensitivity. Again, this is expected on inspection of Equation (31), where only two of the three terms are directly affected by the sensitivity parameter.

From this example can be clearly seen that, with the expressions now available, it is possible to obtain a reliable quantitative measure of the improvement that could result from the addition of an extra data mode. This gives the analyst a powerful criterion to decide whether it is necessary or not to add extra information to a particular analysis.

TABLE 3 Figures of Merit for the Experimental Example Using U-PLS/RBL, U-PLS/RTL, and U-PLS/RQL

Figures of merit[a]			
SEN/AFU (L μg^{-1})	1.3	5.5	12
γ (L μg^{-1})	0.7	3.1	6.7
LOD ($\mu g\ L^{-1}$)	5.3	2	1.5
LOQ ($\mu g\ L^{-1}$)	16	6	4.5

[a] $[var(y_{cal})]^{1/2} = 1\ \mu g\ L^{-1}$, $[var(x)]^{1/2} = 2$ AFU (arbitrary fluorescence units).

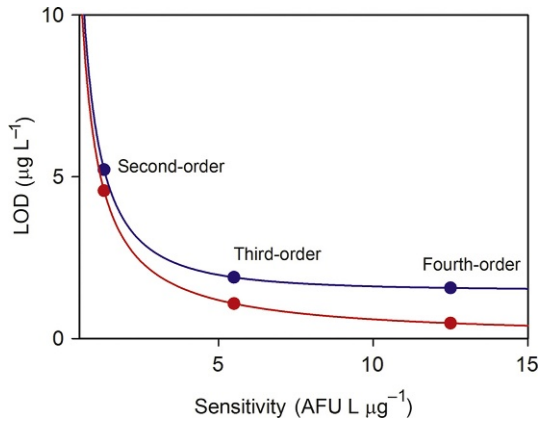

FIGURE 6 Variation of the limit of detection (LOD) with the sensitivity for the determination of the analyte carbaryl in the experimental example. Red (gray in the print version) circles, LOD values as obtained with U-PLS/RML for the different experimental data orders using as an approximation only the first term of Equation (34). Blue (black in the print version) circles, LOD values from the complete IUPAC's recommended expression (Equation 34), inserting the corresponding values of *h*. *Reprinted with permission from Ref. [16]. Copyright 2012 American Chemical Society.*

7 CONCLUSIONS

In this chapter, it has been clearly shown that the task of calculating figures of merit in multivariate and multiway calibration is not as simple as an intuitive and direct extension of univariate expressions and requires special attention. An insight into the data structure clearly shows that it becomes necessary to study and rationalize the definition of estimators, depending not only on the sample being analyzed but also on the structure of the algorithm used to

process the corresponding data. In this sense, the derivation of a general formula to calculate sensitivity, applicable to any data order and for the most important processing tools, constitutes an important step to a better understanding of the information needed to develop a reliable validation methodology.

The definition of the LOD is also an important issue, since this figure of merit brings together two important analytical concepts: the sensitivity and the precision in the analytical determination. Regarding this figure of merit, some efforts to define a reliable estimator in first-order calibration have been presented, but further studies should be made to extend it to more complex multiway data.

Finally, one of the main perspectives of this chapter is related to uncertainty estimation. The extension of the proposed expressions to cases in which the error structure is not *iid* is a topic of fundamental importance, since this latter assumption is not always completely adequate. This implies a deep insight into the different sources of instrumental noise that could affect a measurement, and how the noise is propagated through the structure of each calibration algorithm to give a reliable uncertainty value in the estimated concentration.

ACKNOWLEDGMENTS

Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project No. PICT-2013-0136) are gratefully acknowledged for financial support. F.A. thanks CONICET for a fellowship.

REFERENCES

[1] Booksh KS, Kowalski BR. Theory of analytical chemistry. Anal Chem 1994;66:782A–791A.
[2] ISO 5725:1996. Accuracy (trueness and precision) of measurement methods and results. Geneva, Switzerland: International Organization of Standardization; 1996.
[3] Document No. SANCO/12495/2011. Method validation and quality control procedures for pesticide residues analysis in food and feed, Directorate General for Health and Consumer Affairs (SANCO), European Commission; 2012.
[4] Danzer K, Currie LA. Guidelines for calibration in analytical chemistry. Part 1. Fundamentals and single component calibration. Pure Appl Chem 1998;70:993–1014.
[5] Olivieri AC, Faber NM. Validation and error. In: Brown S, Tauler R, Walczak B, editors. Comprehensive chemometrics. vol. 3. Amsterdam: Elsevier; 2009. p. 91.
[6] Sanchez E, Kowalski BR. Tensorial calibration: I. First-order calibration. J Chemometr 1988;2:247–63.
[7] Ho CN, Christian GD, Davidson ER. Application of the method of rank annihilation to fluorescent multicomponent mixtures of polynuclear aromatic hydrocarbons. Anal Chem 1980;52:1071–9.
[8] Messick NJ, Kalivas JH, Lang PM. Selectivity and related measures for *n*th-order data. Anal Chem 1996;68:1572–9.

[9] 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.
[10] Olivieri AC, Faber NM. A closed-form expression for computing the sensitivity in second-order bilinear calibration. J Chemometr 2005;19:583–92.
[11] Faber K, Lorber A, Kowalski BR. Analytical figures of merit for tensorial calibration. J Chemometr 1997;11:419–61.
[12] 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;77:4936–46.
[13] Olivieri AC. Analytical advantages of multivariate data processing. One, two, three, infinity? Anal Chem 2008;80:5713–20.
[14] Olivieri AC, Faber K. New developments for the sensitivity estimation in four-way calibration with the quadrilinear parallel factor model. Anal Chem 2012;84:186–93.
[15] Bauza MC, Ibañez GA, Tauler R, Olivieri AC. Sensitivity equation for quantitative analysis with multivariate curve resolution-alternating least-squares: theoretical and experimental approach. Anal Chem 2012;84:8697–706.
[16] Allegrini F, Olivieri AC. Analytical figures of merit for partial least-squares coupled to residual multilinearization. Anal Chem 2012;84:10823–30.
[17] Olivieri AC. Analytical figures of merit: from univariate to multiway calibration. Chem Rev 2014;114:5358–78.
[18] Belter M, Sajnóg A, Barałkiewicz D. Over a century of detection and quantification capabilities in analytical chemistry—historical overview and trends. Talanta 2014;129:606–16.
[19] International Organization for Standardization (ISO). Capability of detectionReport No. ISO 11843–1. Genève, Switzerland: ISO; 1997.
[20] International Organization for Standardization (ISO). Capability of detectionReport No. ISO 11843-2. Genève, Switzerland: ISO; 2000.
[21] McNaught AD, Wilkinson A, IUPAC. Compendium of chemical terminology. 2nd ed. Oxford: Blackwell Scientific Publications; 1997.
[22] Inczédy J, Lengyel T, Ure AM, Gelencsér A, Hulanicki A. IUPAC analytical chemistry division, compendium of analytical nomenclature. 3rd ed. Oxford: Blackwell; 1998.
[23] Ortiz MC, Sarabia LA, Herrero A, Sánchez MS, Sanz MB, Rueda ME, et al. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. Chemom Intell Lab Syst 2003;69:21–33.
[24] Burns DA, Ciurczak EW. Handbook of near-infrared analysis. 3rd ed. Practical spectroscopy series. , vol. 35. Boca Raton, FL: CRC Press; 2008.
[25] Allegrini F, Olivieri AC. IUPAC-consistent approach to the limit of detection in partial least-squares calibration. Anal Chem 2014;86:7858–66.
[26] Escandar GM, Olivieri AC. Practical three-way calibration. 1st ed. Waltham, MA: Elsevier; 2014.
[27] Bro R. PARAFAC: tutorial & applications. Chemom Intell Lab Syst 1997;38:149–71.
[28] Tauler R. Multivariate curve resolution applied to second-order data. Chemom Intell Lab Syst 1995;30:133–46.
[29] Tauler R, Maeder M, de Juan A. Multiset data analysis: extended multivariate curve resolution. In: Brown S, Tauler R, Walczak B, editors. Comprehensive chemometrics. vol. 2. Amsterdam: Elsevier; 2009. p. 473.
[30] Wold S, Esbensen K, Geladi P. Principal component analysis. Chemom Intell Lab Syst 1987;2:37–52.
[31] Bro R. Multi-way calibration. Multi-linear PLS. J Chemometr 1996;10:47–62.

[32] Linder M, Sundberg R. Precision of prediction in second order calibration, with focus on bilinear regression methods. J Chemometr 2002;16:12–27.
[33] Olivieri AC. A combined artificial neural network/residual bilinearization approach for obtaining the second-order advantage from three-way non-linear data. J Chemometr 2005;19:615–24.
[34] Arancibia JA, Olivieri AC, Gil DB, Mansilla AE, Durán-Merás I, de la Peña AM. Trilinear least-squares and unfolded-PLS coupled to residual trilinearization: new chemometric tools for the analysis of four-way instrumental data. Chemom Intell Lab Syst 2006;80:77–86.
[35] Ohman J, Geladi P, Wold S. Residual bilinearization. Part 1: theory and algorithms. J Chemometr 1990;4:79–90.
[36] Maggio RM, de la Peña AM, Olivieri AC. Unfolded partial least-squares with residual quadrilinearization: a new multivariate algorithm for processing five-way data achieving the second-order advantage. Application to fourth-order excitation-emission-kinetic-pH fluorescence analytical data. Chemom Intell Lab Syst 2011;109:178–85.
[37] Leger MN, Vega-Montoto L, Wentzell PD. Methods for systematic investigation of measurement error covariance vatrices. Chemom Intell Lab Syst 2005;77:181–205.
[38] Wentzell PD. Measurement errors in multivariate chemical data. J Braz Chem Soc 2014;25:183–96.
[39] Wentzell PD, Tarasuk AC. Characterization of heteroscedastic measurement noise in the absence of replicates. Anal Chim Acta 2014;847:16–28.
[40] Wentzell PD, Andrews DT, Hamilton DC, Faber K, Kowalski BR. Maximum likelihood principal component analysis. J Chemometr 1997;11:339–66.
[41] Wentzell PD, Andrews DT, Kowalski BR. Maximum likelihood multivariate calibration. Anal Chem 1997;69:2299–311.
[42] Lorber A. Error propagation and figures of merit for quantification by solving matrix equations. Anal Chem 1986;58:1167–72.
[43] Faber NM. Exact presentation of multivariate calibration model as univariate calibration graph. Chemom Intell Lab Syst 2000;50:107–14.
[44] Ferré J, Brown SD, Rius FX. Improved calculation of the net analyte signal in inverse multivariate calibration. J Chemometr 2001;15:537–53.
[45] Ferré J, Faber NM. Net analyte signal calculation for multivariate calibration. Chemom Intell Lab Syst 2003;69:123–36.
[46] Rao CR, Mitra S. Generalized inverse of matrices and its applications. New York: Wiley; 1971.
[47] Chen ZP, Wu HL, Jiang JH, Li Y, Yu RQ. A novel trilinear decomposition algorithm for second-order linear calibration. Chemom Intell Lab Syst 2000;52:75–86.
[48] Xia AL, Wu HL, Fang DM, Ding YJ, Hu LQ, Yu RQ. Alternating penalty trilinear decomposition algorithm for second-order calibration with application to interference-free analysis of excitation–emission matrix fluorescence data. J Chemometr 2005;19:65–76.
[49] Xia AL, Wu HL, Li SF, Zhu SH, Hu LQ, Yu RQ. Alternating penalty quadrilinear decomposition algorithm for an analysis of four-way data arrays. J Chemometr 2007;21:133–44.
[50] Faber NM, Ferré J, Boqué R, Kalivas JH. Quantifying selectivity in spectrophotometric multicomponent analysis. Trends Anal Chem 2003;22:352–61.
[51] 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 Chemometr 2005;19:253–65.
[52] Sanchez E, Kowalski BR. Generalized rank annihilation factor analysis. Anal Chem 1986;58:496–9.

[53] Sanchez E, Kowalski BR. Tensorial resolution: a direct trilinear decomposition. J Chemometr 1990;4:29–45.
[54] Bortolato SA, Lozano VA, de la Peña AM, Olivieri AC. Novel augmented parallel factor model for four-way calibration of high-performance liquid chromatography-fluorescence excitation-emission data. Chemom Intell Lab Syst 2015;141:1–11.
[55] Cuadros Rodríguez L, García Campaña AM, Jiménez Linares C, Román Ceba M. Estimation of performance characteristics of an analytical method using the data set of the calibration experiment. Anal Lett 1993;26:1243–58.
[56] Vessman J, Stefan RI, van Staden JF, Danzer K, Lindner W, Burns DT, et al. Selectivity in analytical chemistry (IUPAC Recommendations 2001). Pure Appl Chem 2001;73:1381–6.
[57] Olivieri AC, Faber NM, Ferré J, Boqué R, Kalivas JH, Mark H. Uncertainty estimation in spectroscopic multivariate calibration. Pure Appl Chem 2006;78:633–61.
[58] Faber NM, Ferré J, Boqué R, Kalivas JH. Second-order bilinear calibration: the effects of vectorising the data matrices of the calibration set. Chemom Intell Lab Syst 2002;63:107–16.
[59] Arnold MA, Small GW, Xiang G, Qui J, Murhammer DW. Pure component selectivity analysis of multivariate calibration models from near-infrared spectra. Anal Chem 2004;76:2583–90.
[60] Brown CD, Ridder TD. Framework for multivariate selectivity analysis, Part I: theoretical and practical merits. Appl Spectrosc 2005;59:787–803.
[61] Ridder TD, Brown CD, Ver Steeg BJ. Framework for multivariate selectivity analysis, Part II: experimental applications. Appl Spectrosc 2005;59:804–15.
[62] Geladi P. Some recent trends in the calibration literature. Chemom Intell Lab Syst 2002;60:211–24.
[63] 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.
[64] Faber K, Kowalski BR. Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares. J Chemometr 1997;11:181–238.
[65] Serneels S, Faber K, Verdonck T, Van Espen PJ. Case specific prediction intervals for tri-PLS1: the full local linearisation. Chemom Intell Lab Syst 2011;108:93–9.
[66] Cabezón M, Olivieri AC. Precision in multi-wavelength spectroscopic analysis using classical least-squares regression. Chem Educ 2006;11:394–401.
[67] Currie LA. Limits for qualitative detection and quantitative determination. Application to radiochemistry. Anal Chem 1968;40:586–93.
[68] Currie LA. Nomenclature in evaluation of analytical methods including detection and quantification capabilities. Pure Appl Chem 1995;67:1699–723.
[69] Currie LA. Detection and quantification limits: origins and historical overview. Anal Chim Acta 1999;391:127–34.
[70] Hubaux A, Vos G. Decision and detection limits for linear calibration curves. Anal Chem 1970;42:849–55.
[71] Clayton CA, Hines JW, Elkins PD. Detection limits with specified assurance probabilities. Anal Chem 1987;59:2506–14.
[72] Boqué R, Larrechi MS, Rius FX. Multivariate detection limits with fixed probabilities of error. Chemom Intell Lab Syst 1999;45:397–408.
[73] Boqué R, Ferré J, Faber NM, Rius FX. Limit of detection estimator for bilinear second-order calibration. Anal Chim Acta 2002;421:313–21.

[74] del Río Bocio FJ, Riu J, Boqué R, Rius FX. Limits of detection in linear regression with errors in the concentration. J Chemometr 2003;17:413–21.
[75] Voigtman E. Limits of detection and decision. Part 1. Spectrochim Acta 2008;63:115–28.
[76] Voigtman E. Limits of detection and decision. Part 2. Spectrochim Acta 2008;63:129–41.
[77] Ortiz MC, Sarabia LA, Herrero A, Sánchez MS, Sanz MB, Rueda ME, et al. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. Chemom Intell Lab Syst 2003;69:21–33.