Analytical figures of merit: from univariate to multi-way calibration

Alejandro C. Olivieri

Departamento de Química Analítica, Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Instituto de Química de Rosario (IQUIR-CONICET), Suipacha 531, Rosario, S2002LRK, Argentina

CONTENTS

1. INTRODUCTION....3
2. NOMENCLATURE....7
2.1. Sample constituents....7
2.2. Data arrays....8
3. DATA PROPERTIES, MODELS AND ALGORITHMS....10
3.1. Univariate and first-order data....10
3.2. Multi-way data....11
3.3. Multi-way models and algorithms....17
4. SENSITIVITY EXPRESSIONS BASED ON SIGNAL OR NET SIGNAL CHANGES....19
4.1. Univariate calibration....19
4.2. First-order calibration....19
4.3. Multi-way (higher-order) calibration....23
5. SENSITIVITY EXPRESSIONS BASED ON UNCERTAINTY PROPAGATION....25
5.1. The general sensitivity expression....25
5.2. Univariate calibration....30
5.3. First-order calibration....30
5.4. Multi-way (higher-order) calibration....31
5.4.1. Multi-linear algorithms....31
5.4.2. Multivariate curve resolution-alternating least-squares....33
5.4.3. Partial least-squares/residual multi-linearization....34
5.5. Other multi-way algorithms....34
5.6. Multi-way net analyte signal....35
6. OTHER FIGURES OF MERIT....36
6.1. Analytical sensitivity....37
6.2. Selectivity....37
6.3. Prediction uncertainty....40
6.4. Detection capabilities....43
7. AVAILABILITY OF SOFTWARE....46
8. COMPARISON OF FIGURES OF MERIT....46
9. CONCLUSIONS....49
ACKNOWLEDGMENTS....50
SUPPORTING INFORMATION AVAILABLE....50
SYMBOLS AND ACRONYMS....50
APPENDICES....55
A-1. First-order sensitivity....55
A-2. Multi-way sensitivity....58
REFERENCES....60
AUTHOR INFORMATION....71

1. INTRODUCTION

Figures of merit are numerical parameters which help to characterize the performance of a device or a system relative to alternative ones. In engineering, the former are often defined for particular materials or devices in order to determine their relative utility for certain applications. In commerce, they are usually employed as marketing tools to convince consumers to choose a particular brand. Their use in analytical calibration is comparable concerning the relative success of different methodologies.

The search for new ways to improve analytical figures of merit is an important driving force in modern analytical chemistry research, with the sensitivity occupying one of the prominent places among these figures.[1] Whether the purpose is the comparison of the performance of different experimental procedures, or the optimization of a given methodology under various experimental conditions, a consistent numerical sensitivity parameter is required in order to judge about the real improvement obtained from various experimental strategies. Analytical figures of merit are an integral part of official protocols of analysis, as documented in international standards.[2,3]

The sensitivity is a key element in the estimation of other figures of merit, such as: (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,[4] (2) selectivity, which helps to assess the possibility of analyte quantitation in the presence of interferences,[5] and (3) prediction uncertainty, limit of detection and limit of quantification, which are needed for assessing detection capabilities,[1] and are of prime importance in certain specific areas such as doping control in sports,[6,7] monitoring traces of contaminants in environmental samples,[8] etc.

The International Union of Pure and Applied Chemistry (IUPAC) has set sensitivity definitions in various calibration scenarios.[9-11] In the classical single-constituent or univariate calibration (involving a single instrumental measurement per sample), the sensitivity expression is well-known: it is defined as the change in the response of the instrument divided by the corresponding change in the stimulus (the concentration of the analyte of interest), i.e., the slope of the calibration curve.[9]

When multiple instrumental data are measured for a single sample, the calibration is known as multivariate. If the data can be arranged in vector form (e.g., spectra, chromatograms, electrochemical traces, etc.), they belong to the category of first-order (see below for nomenclature details on data and calibrations). A particularly successful form of first-order calibration, partial least-squares (PLS) regression, which is based on the so-called inverse regression model on latent variables, permits the quantitation of selected analytes in a sample without knowing the chemical identity of the interfering species.[12-14] The presence of the latter is adequately compensated by the calibration model, which is built from a training sample set where the interfering agents have been adequately incorporated. This is especially important for applications in fields such as industrial, food, environmental and life sciences, where the number and nature of interfering species is usually unknown.[15]

In first-order multivariate calibration the situation regarding the definition of sensitivity becomes more complex than for the univariate case.[16] In particular, the sensitivity is analyte-specific, meaning that a certain sensitivity parameter corresponds to each analyte of interest. Although this property may not seem natural, because traditionally the sensitivity characterizes the instrument, it is perfectly logical in the multivariate context, where an intense analyte signal may be useless under severe spectral overlapping with signals from other concomitant constituents.

Notwithstanding the difficulties, a useful generalization of the univariate definition has been developed for first-order multivariate calibration.[17,18] It is known as the LBOZ criterion (after Lorber,[17] Bergmann, von Oepen and Zinn[18]), and is based on an intuitive analogy between the true instrumental signal and the so-called net analyte signal (NAS) generated by a unit analyte concentration, as first proposed by Lorber.[17] The first-order NAS is defined in precise mathematical terms, and suitably interpreted as the portion of the overall signal which can be uniquely ascribed to a given analyte.[17,19] The subject of first-order multivariate figures of merit has been thoroughly reviewed in 2006,[11] and thus only the main concepts will be repeated here, where comparison with other calibration scenarios is appropriate.

Multi-way 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.[20] By processing these data, considerably more complex analytical problems can be solved,[21-31] and predictions are even possible in the presence of unexpected spectral interferences, i.e., sample constituents not considered in the calibration phase.[32] The latter will be called, in the remainder of this paper, simply as 'unexpected interferences'. Moreover, multi-way calibration often provides valuable physico-chemical information such as the pure-constituent signals. In some popular approaches to multi-way calibration, analytes and potential interfering agents are mathematically separated by retrieving pure-constituent profiles, followed by a pseudo-univariate calibration strategy for analyte quantitation (the prefix 'pseudo' distinguishes this calibration, constructed with multivariate signals that are the result of mathematical processing, from the classical one built with raw univariate signals). In this field, several different sensitivity expressions have been proposed, some of them based on extensions of the first-order NAS concept to further data modes.[33-36] However, there are difficulties with the NAS strategy, as there are various competing NAS

definitions, with no clear relationship among them.[37-39] What is even more worrying, the plainly extrapolated expressions to data arrays with higher number of modes appeared to lead to serious underestimation of true sensitivities.[40]

An alternative methodology for assessing the sensitivity in analytical calibration emerged in recent years, based on the analysis of how the uncertainty in instrumental signal propagates to the uncertainty in predicted concentration.[40-42] This approach has led to the development of closed-form expressions applicable to most multi-way data processing algorithms, and has been confirmed by extensive, additive noise Monte Carlo simulations.[40-42] 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 multi-way, and in the latter case for most multi-way algorithms. Whether the general expression fits into a broader scene incorporating an intuitively useful multi-way NAS concept is probably a matter of future debate. It is worth noticing that the multi-way sensitivity displays even more intriguing properties in comparison with the first-order counterpart: it is not only analyte-specific, but also strongly dependent on the test sample and on the data processing algorithm. This implies that the sensitivity can only be estimated for a particular group of test samples, all having similar qualitative chemical compositions. Likewise, the selected calibration algorithm greatly affects the analyte sensitivity, and hence the computational tools employed for processing the data should be regarded as an integral part of a multi-way analytical protocol.

In this report, the traditional definitions of sensitivity for zeroth- and first-order calibration are put in perspective with the new multi-way (higher-order) sensitivity expression. Once the sensitivity is computed, access is granted to the remaining figures of merit by analogy with the univariate counterparts. The review is organized as follows: first the established nomenclature of data arrays is introduced, together with a summary of some multi-way models

and calibration algorithms. Sensitivity expressions for univariate and first-order multivariate calibration are summarized, including the prospective extension to higher-order methodologies. Then a new approach to sensitivity is discussed, based on uncertainty propagation, which can be appropriately condensed into a single, general sensitivity equation. It is shown how the latter is able to reproduce the univariate and first-order equations, and also to yield consistent expressions for multi-way calibration, and in this case for the major data processing algorithms. Finally, the remaining figures of merit are discussed, with emphasis on their peculiarities regarding the multi-way calibration field.

2. NOMENCLATURE

2.1. Sample constituents

In univariate and first-order multivariate calibration, the composition of the calibration set of samples should be representative of future samples, both qualitatively and quantitatively. In contrast, in multi-way calibration the figures of merit which can be achieved by the various methodologies greatly depend on the presence or absence of the constituents of the various samples. It is thus essential in this context to appropriately classify sample constituents, a feature which derives from the sample-specificity of the multi-way figures of merit.

Calibration and validation sets of samples contain the so-called 'expected' constituents, because the analyst includes them in these sets to model their behaviour in future samples. The expected constituents have also been further divided into 'calibrated', those for which calibration concentrations are available, and 'uncalibrated', for which only the instrumental signals are measured.[35] In this review this latter distinction will not be made, because: (1) only the presence or absence from the calibration set is required for the application of the general sensitivity

equation, and (2) the expression 'uncalibrated' may be confused with 'unexpected' (see below), which implies a different concept.

The constituents present in unknown test samples, but not in the calibration or validation samples, are called 'unexpected'. Additional overlapping responses from the sample background may also be unexpected if not present during calibration. Unexpected constituents are also called potential interfering agents, with emphasis in 'potential', because in multi-way calibration their presence in a test sample may not lead to a systematic error in the analyte determination.[43] In contrast, in univariate and first-order calibration, unexpected constituents usually produce an interference.

Note, in the remainder of this review, the distinction of *constituent*, which is a real chemical compound present in a given sample, from *component*, which in general refers to a mathematical entity needed to model a data array, which may or may not directly represent the behavior of a specific chemical constituent.

2.2. Data arrays

For a consistent nomenclature of different data types, the concept of 'order' can be employed, as is widely done in analytical chemistry studies.[20] The order is a tensorial property of data measured *for a single sample*: scalars are zeroth-order tensors, and thus univariate calibration is also known as zeroth-order calibration. If spectra are measured (or other vectors per sample), the calibration becomes first-order (and multivariate instead of univariate). Increasing the number of data modes per sample leads to correspondingly complex data arrays, which give rise to higher-order multivariate calibration (Figure 1). The order is also linked to the popular expression 'second-order advantage', which is common among analytical chemists. The second-order

advantage refers to the possibility of quantitating an analyte in a mixture with potentially interfering constituents, even when calibrating with pure analyte standards. It is not restricted to second-order data, but to all data of *at least* second-order.[32] Interesting experimental applications in which this advantage has been exploited for a variety of samples can be found in recent reviews.[22-31]

An alternative nomenclature is based on the number of ways, which is equivalent to the number of modes of a data array for a group of samples.[32] Thus univariate and one-way calibration are synonymous, as are first-order and two-way calibration, second-order and three-way calibration, etc. Three-way systems and beyond are also known as multi-way. Figure 1 visually summarizes the data array nomenclatures up to three-way (second-order).

Popular examples of second-order data are excitation-emission fluorescence landscapes (Figure 2), which can be visually represented by plotting the fluorescence intensity as a function of excitation and emission wavelengths. When fluorescence excitation-emission matrix data (EEM data) are stacked together forming a three-way array (as in Figure 1), the latter object has three different modes: (1) the sample mode, (2) the excitation wavelength mode, and (3) the emission wavelength mode. The latter two will be referred as the instrumental data modes of the three-way data. Another popular form of second-order data is a chromatographic-spectral matrix, such as those collected on a liquid chromatograph with diode array detection (LC-DAD data) or fast-scanning fluorescence detection (LC-FSFD), or on a gas chromatograph with mass spectrometric detection (GC-MS data). Figure 3 shows a typical LC-FSFD landscape. Data matrices of this type can also be arranged into a three-way array whose modes are: (1) the sample mode, (2) the elution time mode, and (3) the spectral mode. Other three-way (second-order) data types are possible, as has been recently reviewed,[22-26] but the number of works devoted to them are considerably smaller than for EEM or LC-spectral data.

Third-order data and beyond can easily be seen as extensions of the objects shown in Figure 1. Two usual manners in which they can be collected are: (1) measuring EEM while they evolve as a function of reaction time,[44] or (2) acquiring two-dimensional chromatographic data (either LC-LC or GC-GC) with spectral (DAD or MS) detection.[45,46] In all of these cases, three instrumental modes occur for each sample (excitation wavelength, emission wavelength and reaction time in one case, and first column elution time, second column elution time and spectra in the second), while the sample mode is the remaining one for a four-way array obtained by joining data for a group of samples.

3. DATA PROPERTIES, MODELS AND ALGORITHMS

3.1. Univariate and first-order data

Univariate calibration is well-known as the cornerstone of classical single-constituent calibration in analytical chemistry. Details can be found in IUPAC official protocols.[9]

The origins of first-order multivariate calibration date back to 1960s. Today it is established as a robust and reliable methodology for the analysis of industrial materials, with a paradigmatic example of the marriage between near infrared spectroscopy and partial least-squares regression as a successful combination of instrumental and chemometric techniques.[15] PLS is today the *de facto* standard for most first-order applications. Excellent reviews and books exist on the matter.[12-14]

Multi-way calibration is relatively new in this regard: the first work describing a second-order calibration was published in 1978,[47] reporting the determination of polycyclic aromatic hydrocarbons in the presence of potential interfering agents. Then an impasse of ca. 15 years elapsed, until the subject was revived in the mid 1990's.[39] Although the number of multi-way

applications is exponentially growing, the methodology is still relatively unknown to the average analytical chemist. In contrast to univariate and first-order multivariate calibration, several multi-way algorithms compete with success in the multi-way terrain. Their application fields overlap to some extent, and this may make the selection of a specific data processing algorithm a complex task. Therefore, information is provided in the next sections on the different types of multi-way data the analyst may find, and their relationship with the underlying model of various data processing algorithms.

3.2. Multi-way data

It is advisable to have some insight into the properties of the measured multi-way data, referring to an underlying physical model which the data are suspected to follow. Knowledge of the model allows one to select a specific data processing tool, which in turn significantly affects the achieved figures of merit. This is another peculiar feature of multi-way calibration: the algorithm-specificity of these figures.

The simplest array in this regard is represented by second-order data measured for a single sample, which is also the basic ingredient of a three-way array. Two of the most popular experimental matrix data types, i.e., EEM and LC-spectral data, usually display a mathematical property called *bilinearity*. Assume an excitation-emission fluorescence data matrix **X** has been measured by scanning J different emission wavelengths and K different excitation wavelengths, or an LC-DAD matrix has been collected, consisting of J spectra measured at K elution times. In both cases, the data can be arranged into a data table or matrix with J rows and K columns, i.e., of size $J \times K$. If there are N responsive constituents in the sample, a generic element x_{ij} of these data matrices can be written as:

$$x_{ij} = \sum_{n=1}^{N} b_{jn} c_{kn} \quad (1)$$

where b_{jn} and c_{kn} define the specific properties at instrumental channels j and k for constituent n. An error term should be added to the right hand side of eq (1) for completeness; it was omitted in this paper in all pertinent expressions for clarity. The meaning of 'instrumental channel' depends on the experimental setup, i.e., excitation/emission/absorption wavelength, elution time, mass/charge ratio, etc. Notice that signal additivity of the N constituents is assumed in eq (1), as is usual in the above described experimental situations. Figure 4 illustrates the obtainment of a matrix element from the individual profiles.

Equation (1) is equivalent to:

$$\mathbf{X} = \mathbf{b}_1\mathbf{c}_1^{\mathrm{T}} + \ldots \mathbf{b}_N\mathbf{c}_N^{\mathrm{T}} = \mathbf{B}\,\mathbf{C}^{\mathrm{T}} \quad (2)$$

where the superscript 'T' indicates transposition, and $\mathbf{b}_n$ and $\mathbf{c}_n$ are vectors describing the profiles for constituent n in both data modes (Figure 4). Hence $\mathbf{X}$ is the sum of N terms, each of them linear in $\mathbf{b}_n$ and $\mathbf{c}_n$, which are called bilinear components. For these reasons $\mathbf{X}$ is known as a bilinear matrix. To be precise, any data matrix can be expressed as a product of two matrices, however if the number of bilinear components required to adequately model a data matrix to a reasonable degree (the mathematical rank) is small and ideally equal to the number of responsive chemical constituents, then the matrix is said to be bilinear, although a proper nomenclature would be low-rank bilinear. When a data matrix for a mixture of a few constituents cannot be expressed as a sum of a few bilinear terms, it is called non-bilinear. This occurs, for example, for two-dimensional mass spectrometry (MS-MS),[48] two-dimensional nuclear magnetic resonance correlated spectroscopy (^{2}D NMR COSY),[49] and total synchronous fluorescence (TSF) spectroscopy,[50] because a spectrum in one mode depends on its position in the second mode. In

general, bilinearity is lost when the phenomena occurring in the two instrumental modes are mutually dependent.

For the simplest multi-way data, i.e., a three-way array, a particularly appealing property is the trilinearity, which naturally follows as the next step after bilinearity. As noted above, a three-way data array is trilinear (more precisely low-rank trilinear) if it can be expressed as a sum of a few trilinear components when the mixture contains a few constituents. Since EEM data are prime examples of trilinearity, they will be used as example. Assume a number of excitation-emission fluorescence data matrices (I) has been measured. They can be stacked in the sample mode, creating a three-way array $\boldsymbol{X}$, of size $I \times J \times K$, whose generic element can be designated as x_{ijk}. If 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:

$$x_{ijk} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} \qquad (3)$$

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. Equation (3) is analogous to the one relating the fluorescence emission intensity to the usual chemical and instrumental parameters involved in this phenomenon,[51] except for a scaling factor and a change in symbols. It is customary to collect all a_{in} values into a vector $\mathbf{a}_n$, b_{jn} into a vector $\mathbf{b}_n$, and c_{kn} into a vector $\mathbf{c}_n$. The latter two vectors are usually normalized to unit length.

It is perhaps not directly apparent in eq (3), but trilinearity demands that: (1) individual data matrices are bilinear, i.e., **b** and **c** profiles do not depend on each other, and (2) **b** and **c** profiles do not depend on the sample index i, i.e., there should be unique **b** and **c** vectors describing the behaviour of each constituent in both instrumental modes in all samples. These

conditions are usually met by EEM spectroscopy, which provides primary examples of trilinear three-way data (except for the diffraction grating harmonics, which can generally be corrected).

In the case of data stemming from chromatography with spectral detection, the data matrices are individually bilinear; however, small changes in elution profiles for a given constituent from sample to sample usually occur. Hence a three-way array composed of these latter data matrices will not be, in general, trilinear. They would be if elution profiles were exactly reproducible from sample to sample. For this reason, the elution time mode is a potentially *trilinearity-breaking* mode.[52] Three-way data having a single trilinearity-breaking mode can be conveniently analyzed if the three-way array is unfolded along the elution time mode, i.e., if it is converted into a data matrix having all individual sample data matrices adjacent to each other in the direction of the elution time (Figure 5). The latter data matrix is called *augmented*, because it can be viewed as being built from the individual matrices by the process of augmentation (it can also be viewed as arising from the unfolding process which starts from the three-way array, see Figure 5). Since the individual data matrices are bilinear, an important property of a chromatographic-spectral matrix augmented in the time direction ($\mathbf{X}_{\text{aug}}$) is that it is also bilinear, i.e., it can be formulated as:

$$x_{\text{aug},pk} = \sum_{n=1}^{N} b_{\text{aug},pn}\, c_{kn} \qquad (4)$$

with the index p running from 1 to IJ, because the size of the augmented matrix is $IJ \times K$ (I = No. of samples, J = No. of elution times, K = No. of wavelengths). In eq (4), the spectral profile $\mathbf{c}_n$ (also called the non-augmented profile) is unique for each constituent and common to all samples, whereas $\mathbf{b}_{\text{aug},n}$ is the augmented time profile in the augmented elution time mode, and is composed of I successive time sub-profiles with J times each.

Table 1. Classification of second- and third-order data, and models/algorithms which can be applied to analyze them.

Data type		Example	Suitable algorithm
Three-way (second-order) data			
Trilinear		EEM	PARAFAC[a]
Non-trilinear with one breaking mode		LC-DAD LC-FSFD GC-MS	MCR-ALS
Other non-trilinear	Two trilinearity breaking modes	EEM with inner filter	PLS/RBL
	Non-bilinear individual matrices	MS-MS ^{2}D NMR COSY TSF	PLS/RBL[b]
Four-way (third-order) data			
Quadrilinear		EEM-time	PARAFAC[a]
Non-quadrilinear with two quadrilinearity breaking modes		LC-LC-DAD GC-GC-MS	MCR-ALS
Other non-quadrilinear		–	–

[a] Additional multi-linear decomposition variants are also possible.[55-58]

[b] PLS/RBL has only been applied to TSF data.[50]

Several other instrumental second-order data exist, although they are not employed as often for analytical calibration purposes. Among these other data, two additional non-trilinear data types may be found, for which: (1) individual data matrices are non-bilinear (see above for examples), and (2) individual data matrices have two trilinearity breaking modes, because both instrumental profiles vary from sample to sample, and augmentation into a bilinear matrix is not

possible, such as EEM fluorescence data in the presence of inner filter effects in both the excitation and emission modes.[53] As a consequence, it is sensible to classify three-way data in: (1) trilinear, (2) non-trilinear with a single trilinearity-breaking mode and unfoldable to a bilinear augmented matrix, and (3) other non-trilinear, as summarized in Table 1 including pertinent examples. It is interesting to note that ca. 90% of the published multi-way calibration works describe data belonging to the first two categories, distributed in almost equal shares.

In going to more data orders, a similar classification scheme is possible. In general, multi-way data can be arranged into a multi-way array, which is multi-linear if its elements obey an equation similar to eq (3); for four-way (third-order) data, for example:

$$x_{ijkl} = \sum_{n=1}^{N} a_{in} b_{jn} c_{kn} d_{ln} \qquad (5)$$

where the extra factor in comparison with eq (3) corresponds to the profile in the additional data mode. Multi-linearity requires profiles in all data modes which are independent of each other and independent of sample. A typical example involves the measurement of the time evolution of EEM data while following the kinetics of a reaction.[44]

If there are multi-linearity breaking instrumental modes in multi-way arrays, unfolding the array into a bilinear augmented matrix may be possible. This is typical of third-order chromatographic data such as LC-LC-DAD and GC-GC-MS, which display two potentially quadrilinearity-breaking modes (the two elution time modes). In this case, it is wise to unfold the four-way array into an augmented data matrix whose modes are: (1) the spectral mode (either DAD or MS), which is the non-augmented mode, and (2) a concatenation of both elution time modes into a single one, which is the augmented mode.[45,46] Additional four- and higher-way data types can be envisaged beyond those quoted in Table 1, as the different data modes might in

principle be multi-linearity breaking, and/or mutually interacting with each other. However, the number of experimental developments in this regard is still small.

3.3. Multi-way models and algorithms

There are many available algorithms for analytical calibration with multi-way data. Priority is given to those allowing for multiple calibration samples, because this leads to more robust and statistically efficient analytical results.[35] The most employed algorithms in this regard can be appropriately classified into three main groups according to a simple connection between their underlying models and the different data categories discussed in the previous section: (1) a multi-linear model, (2) a bilinear model for an augmented matrix, and (3) a flexible latent-variable model. Group (1) includes parallel factor analysis (PARAFAC)[54] and some variants,[55-58] group (2) multivariate curve resolution coupled to alternating least-squares (MCR-ALS)[59] particularly in the so-called extended version,[60] and group (3) unfolded and multi-way partial least-squares (U-PLS and N-PLS).[61,62] Additional, less employed algorithms for multi-way calibration are worth mentioning, such as multi-linear least-squares (MLLS),[63,64] a less flexible version of the PLS methodologies, and generalized rank annihilation (GRAM),[65] direct trilinear decomposition (DTLD),[66] and non-bilinear rank annihilation (NBRA).[49] The latter three algorithms are based on calibration with a single standard (either real or virtual).

A fundamental difference between PARAFAC and MCR-ALS is that the former often leads to unique solutions,[54] whereas MCR-ALS needs to apply a series of constraints (all of them based on natural physico-chemical assumptions) in order to reach a chemically reasonable solution.[59,60] The uniqueness property of PARAFAC regarding the decomposition of a multi-way array into a small number of trilinear components has a direct consequence in the achievement of

the second-order advantage, because it leads to pure signals and relative concentrations of all sample constituents, including the analyte of interest.

It may be noticed that both PARAFAC and MCR-ALS achieve the second-order advantage by simultaneously processing 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 post-calibration activity: the test sample is subjected to a procedure called residual multi-linearization (RML, including bi-, tri- and quadrilinearization, i.e., RBL,[63,67,68] RTL[64] and RQL[69]), which separates the portion of the signal which can be explained by calibration from the contribution of the potential interfering agents. This gives rise to the hybrid methodologies PLS/RML. Details on the operation of all these algorithms can be found in the Supporting Information and in relevant reviews.[22-26]

At the risk of some oversimplification, Table 1 shows a correspondence between data properties and algorithms. Multi-linear algorithms such as PARAFAC are the natural choice for multi-linear data, extended MCR-ALS is based on an augmented bilinear matrix and hence it is also natural to select it when the data are non-multi-linear but follows the augmented bilinear model, and finally latent-variable PLS/RML methodologies display a flexibility which should made them the algorithms of choice for other non-multilinear data. However, the application fields of these algorithms considerably overlap because of several facts: (1) MCR-ALS and PLS/RML can be applied to multi-linear data, (2) chromatographic profiles can in certain cases be aligned or synchronized,[70] so that their shapes and positions in the elution time axis become common to all samples, restoring multi-linearity, and (3) PARAFAC variants have been developed (i.e., PARAFAC2)[71] for coping with varying chromatographic profiles from sample to sample. These three events may make model and algorithm selection a more complex task for the

average chemist. However, it is likely that future developments will take into account sensitivity considerations as a helpful decision-making tool in this regard.

4. SENSITIVITY EXPRESSIONS BASED ON SIGNAL OR NET SIGNAL CHANGES

4.1. Univariate calibration

In univariate calibration, prediction of the analyte concentration (y) in a test sample from its signal (x) proceeds through the known expression:[9]

$$y = (x - n_0) / m_0 \quad (6)$$

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.[9]

4.2. First-order calibration

The concept of net analyte signal has been useful in assessing the sensitivity in first-order calibration, by extending the univariate definition to the change in NAS for a unit change in analyte concentration.[17] In order to fully understand the NAS concept and its consequences, it is highly useful to consider the simplest possible example, i.e., 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:

$$\mathbf{x} = y_1 \mathbf{s}_1 + y_2 \mathbf{s}_2 \quad (7)$$

where y_1 and y_2 are the constituent concentrations and $\mathbf{s}_1$ and $\mathbf{s}_2$ the pure constituent profiles at unit concentration. By employing eq (7) it is implicitly assumed 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 its analogues) apply. Figure 6A shows two typical pure constituent spectra at unit concentration, and Figure 6B the spectrum for a mixture at equal analyte concentrations, including the individual contributions of each constituent to the mixture spectrum. Focusing on analyte 1 as the compound of interest, the contribution from constituent 2 can be removed from eq (7) by left-multiplying both sides by an orthogonal projection matrix $[\mathbf{I} - \mathbf{s}_2 (\mathbf{s}_2^T \mathbf{s}_2)^{-1} \mathbf{s}_2^T]$, where $\mathbf{I}$ is an appropriately dimensioned unit matrix (orthogonal means in this context 'perpendicular' to a generalized plane in the multivariate space). Usually the result of $[(\mathbf{s}_2^T \mathbf{s}_2)^{-1} \mathbf{s}_2^T]$ is designated as $\mathbf{s}_2^+$, with the superscript '+' implying the generalized inverse operation. 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 are adequately retrieved by analysis of mixtures of pure constituents. Removal of the contribution of the concomitant No. 2 in the mixture by orthogonal projection is possible because $(\mathbf{I} - \mathbf{s}_2 \mathbf{s}_2^+) \times \mathbf{s}_2 = \mathbf{s}_2 - \mathbf{s}_2 = \mathbf{0}$. Equation (7) thus leads to:

$$(\mathbf{I} - \mathbf{s}_2 \mathbf{s}_2^+) \mathbf{x} = y_1 (\mathbf{I} - \mathbf{s}_2 \mathbf{s}_2^+) \mathbf{s}_1 \quad (8)$$

By performing this rather smart operation, a two-constituent problem has become a virtual single-constituent problem. Indeed, the left hand side of the latter equation defines the net analyte signal 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:

$$\mathbf{x}_1^* = y_1 \mathbf{s}_1^* \quad (9)$$

In sum, the NAS in the mixture and the NAS at unit concentration, both specific to analyte 1, are defined by projecting, orthogonal to the space spanned by the remaining sample constituent 2 ($\mathbf{s}_2$), the sample signal $\mathbf{x}$ and the pure analyte signal $\mathbf{s}_1$ respectively. Figure 7 pictorially illustrates the process of projecting the $\mathbf{x}$ vector orthogonally to the space spanned by constituent 2 (the plane in this figure is only intended as a graphical representation of the multi-dimensional surface corresponding to the vector $\mathbf{s}_2$). The relevant result to be gathered from eq (9) is that the NAS vector $\mathbf{x}_1^*$ is parallel to the NAS at unit concentration $\mathbf{s}_1^*$. Figure 8A shows these NAS vectors, which are seen to be rather abstract linear combinations of true profiles and thus lacking intuitive interpretation. 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 ($\| \mathbf{s}_1^* \|$) (Figure 8B).[72-74] This immediately leads to an intuitive definition of sensitivity for analyte 1 as follows (see Appendix A-1 for details):

$$\mathrm{SEN}_1 = \| \mathbf{x}_1^* \| / y_1 = \| \mathbf{s}_1^* \| = \left[\mathbf{s}_1^{\mathrm{T}} \left(\mathbf{I} - \mathbf{s}_2 \mathbf{s}_2^{+} \right) \mathbf{s}_1 \right]^{1/2} \qquad (10)$$

In conclusion, 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 precisely define the sensitivity towards 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 pseudo-inverse operations.[75] As discussed in Appendix A-1, eq (10) can be generalized to the *n*th constituent of interest in a multi-constituent sample in different forms. One useful form is expressed as a function of the pure profiles for all constituents, ubiquitous in CLS studies:

$$\mathrm{SEN}_n = \left[\boldsymbol{\delta}_n^{\mathrm{T}}\left(\mathbf{S}_{\mathrm{CLS}}^{\mathrm{T}}\mathbf{S}_{\mathrm{CLS}}\right)^{-1}\boldsymbol{\delta}_n\right]^{-1/2} \quad (11)$$

where $\boldsymbol{\delta}_n$ is an $N\times 1$ vector selecting the analyte of interest (see Appendix A-1), and the matrix $\mathbf{S}_{\mathrm{CLS}}$ contains N columns, each with the pure constituent profile $\mathbf{s}_n$ for the nth constituent.

Another useful generalization of the multivariate first-order sensitivity can be developed in terms of the so-called 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 predictive equation:

$$y_n = \boldsymbol{\beta}_{\mathrm{CLS},n}^{\mathrm{T}}\,\mathbf{x} \quad (12)$$

As a function of this vector, the sensitivity can be expressed as:

$$\mathrm{SEN}_n = \left(\boldsymbol{\beta}_{\mathrm{CLS},n}^{\mathrm{T}}\,\boldsymbol{\beta}_{\mathrm{CLS},n}\right)^{-1/2} \quad (13)$$

Equation (13) provides a useful link to first-order algorithms which do not rely on the estimation of pure constituent profiles. The latter ones are the so-called inverse models, such as inverse least-squares (ILS), principal component regression (PCR) and PLS.[37,76] In contrast to the direct approach of the classical Beer's law, inverse calibration models relate concentrations to signals, i.e.:

$$y_n = \boldsymbol{\beta}_n^{\mathrm{T}}\,\mathbf{x} \quad (14)$$

Inverse algorithms are able to provide a vector of regression coefficients $\boldsymbol{\beta}_n$ from a suitable set of calibration mixtures, and thus the analogue of eq (13) is a useful means of estimating the sensitivity for these methodologies. Appendix A-1 shows how eq (11) can be adapted to make it compatible with those for latent based methodologies, by replacement of $\boldsymbol{\delta}_n$ and $\mathbf{S}_{\mathrm{CLS}}$ with appropriate latent-variable mathematical objects. In this way, sensitivity expressions for both direct and inverse calibration first-order methodologies can be brought into a common form.

4.3. Multi-way (higher-order) calibration

One approach for estimating the SEN_n parameter in three-way (second-order) calibration is the calculation of the net analyte signal inspired in the useful first-order NAS philosophy, removing the contribution of constituents other than the analyte of interest using orthogonal projection matrices. One intriguing aspect of this multi-way NAS approach is the fact that, in principle, these projections can be carried out in different ways, leading to competing NAS definitions.[35]

A typical matrix signal (**X**) defined in two different instrumental modes for a simple binary mixture can be written as:

$$\mathbf{X} = y_1 \mathbf{M}_1 + y_2 \mathbf{M}_2 \quad (15)$$

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 **b** and **c**, the expression for **X** would be:

$$\mathbf{X} = y_1 \mathbf{b}_1 \mathbf{c}_1^{\mathrm{T}} + y_2 \mathbf{b}_2 \mathbf{c}_2^{\mathrm{T}} \quad (16)$$

where $\mathbf{b}_1$ and $\mathbf{b}_2$ the pure constituent profiles in the first data mode, and $\mathbf{c}_1$ and $\mathbf{c}_2$ those in the second data mode. For EEM data, **b** and **c** describe excitation and emission spectral profiles, while in LC-spectral data, they correspond to elution time and spectral profiles respectively.

Following the NAS approach, the contributing matrix signal for the constituent No. 2 may be removed from eq (16) by these simultaneous operations: left-multiplication with a projection matrix orthogonal to $\mathbf{b}_2$, and right-multiplication with an analogous matrix orthogonal to $\mathbf{c}_2$. Without going into the specific details, the relevant outcome is that this line of reasoning leads to one particular sensitivity expression known as HCD (acronym follows authors initials),[33] which is valid in a certain calibration scenario. Table 2 shows the specific expression and applicability.[35] Figure 9 shows the graphical result for a typical binary mixture where the constituents are described in both data modes by Gaussian-shaped profiles such as the pure

spectra shown in Figure 4. As was the case with the first-order NAS counterpart, the second-order NAS landscape for the HCD definition does not lead to an obvious physical interpretation.

Table 2. Different three-way (second-order) sensitivity definitions based on extensions of the NAS concept.

Expression[a]	Comments	Ref.
$SEN_n = m_n \{[(\mathbf{B}^T \mathbf{B})^{-1}]_{nn} [(\mathbf{C}^T \mathbf{C})^{-1}]_{nn}\}^{-1/2}$	HCD sensitivity, valid for one calibrated constituent in the presence of unexpected constituents	33
$SEN_n = m_n \{[(\mathbf{B}^T \mathbf{B}) * (\mathbf{C}^T \mathbf{C})]^{-1}\}_{nn}^{-1/2}$	MKL sensitivity, valid in the absence of unexpected constituents	34
$SEN_n = m_n \{[(\mathbf{B}_{exp}^T (\mathbf{I} - \mathbf{B}_{unx} \mathbf{B}_{unx}^+) \mathbf{B}_{exp}) * (\mathbf{C}_{exp}^T (\mathbf{I} - \mathbf{C}_{unx} \mathbf{C}_{unx}^+) \mathbf{C}_{exp})]^{-1}\}_{nn}^{-1/2}$	FO sensitivity, valid for any number of calibrated constituents in the presence of unexpected constituents	35

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

There is an alternative approach, which involves first unfolding the matrix **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 unfolded space is formally represented by the so-called Kronecker product $(\mathbf{c}_2 \otimes \mathbf{b}_2) = [c_{21} \mathbf{b}_2 \mid c_{22} \mathbf{b}_2 \mid \ldots]$.[77] This approach leads to a different second-order sensitivity definition, the MKL sensitivity (acronym follows authors initials),[34] valid in a different calibration situation in comparison with the HCD sensitivity. Table 2 provides the corresponding information. Notice that the original works on HCD and MKL sensitivity did

not employ NAS arguments for their derivation, but the results are identical to those provided by the above NAS-inspired procedures.[35]

Figure 9 shows the plot of the MKL NAS surface: this sensitivity is higher than the HCD one (the vertical scales of Figure 9 are arbitrary, but the numerical limits for the intensity axes are identical). From Table 2, this appears to be the expected outcome from two radically different calibration situations: (1) both constituents are calibrated (MKL) and (2) one is calibrated and the second one is a potential interfering agent (HCD), which decreases the sensitivity.

Both HCD and MKL equations were condensed into the more general FO definition (acronym follows authors initials),[35] conceived to take into account all possible calibration situations, including cases not covered by the former two expressions (Table 2). The derivation required a complicated series of steps, which combined removal of other sample constituents, partly in matrix form and partly in unfolded form. This multiplicity of definitions and procedures is puzzling and lacks the elegance of the first-order NAS-based sensitivity counterpart. More importantly, 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.[38] This situation prompted the finding of an alternative solution to the estimation of the multi-way sensitivity.

5. SENSITIVITY EXPRESSIONS BASED ON UNCERTAINTY PROPAGATION

5.1. The general sensitivity expression

An alternative and useful operative 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.[78,79] More sensitivity is achieved if low output noise is obtained for a given input noise (Figure 10). It then makes perfect sense to define the SEN_n parameter as the ratio of input to output noise:

$$SEN_n = \sigma_x / \sigma_y \quad (17)$$

where σ_x and σ_y are the uncertainties in signal and concentration respectively. This uncertainty propagation approach assumes that the input noise is independent and identically distributed, 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. Therefore the box labeled 'Calibration model' in Figure 10 refers to a precisely defined model in terms of a set of calibration samples with known reference concentrations, all of them carrying negligible uncertainty in both signals and concentrations. The scheme shown in Figure 10 can be mirrored by a Monte Carlo additive noise simulation for estimating sensitivities for any calibration model, whether univariate, multivariate or multi-way, as has been recently done.[40-42] This allowed operational values for the sensitivity in different calibration scenarios to be obtained, although they do not provide a closed-form sensitivity equation, which would be far more useful in this regard.

Recently, several expressions were derived using the concept of uncertainty propagation from a noisy test sample signal to the concentration predicted by a noiseless calibration model.[40-42] The developed equations allowed to estimate the sensitivity in most of the relevant multi-way calibration models, including PARAFAC, MCR-ALS and PLS/RML, with results

which are: (1) compatible with the second-order HCD, MKL and FO (when they apply, see Table 2), (2) in agreement with Monte Carlo additive noise simulations for data of various orders, and (3) extendable to data with increasing number of ways.

From this body of work, it is now 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. The main result is appropriately condensed into the following expression:

$$\mathrm{SEN}_n = \left\{ \mathbf{g}_n^{\mathrm{T}} \left[\mathbf{Z}_{\mathrm{exp}}^{\mathrm{T}} \left(\mathbf{I} - \mathbf{Z}_{\mathrm{unx}} \mathbf{Z}_{\mathrm{unx}}^{+} \right) \mathbf{Z}_{\mathrm{exp}} \right]^{-1} \mathbf{g}_n \right\}^{-1/2} \qquad (18)$$

The different factors appearing in eq (18) will be explained below in the context of each calibration scenario, but a qualitative description is appropriate at this point. 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 combines the latter information, making it specific for the nth analyte of interest. The final factor in eq (18) is the matrix $\left(\mathbf{I} - \mathbf{Z}_{\mathrm{unx}} \mathbf{Z}_{\mathrm{unx}}^{+} \right)$, which is the mathematical manifestation of the second-order advantage, and thus it only appears in higher-order (three-way and beyond) calibration methodologies. Its purpose is to correct the matrix of profiles for the expected constituents ($\mathbf{Z}_{\mathrm{exp}}$) for the overlapping effect of the profiles for the unexpected constituents (hence the subscript 'unx') or potential interfering agents. Specifically, the matrix $\left(\mathbf{I} - \mathbf{Z}_{\mathrm{unx}} \mathbf{Z}_{\mathrm{unx}}^{+} \right)$ depends on the profiles for the unexpected constituents which may occur in a given test sample, and only appears when achieving the second-order advantage, because only in this case is such information available. The profiles for the unexpected constituents may be: (1) true constituent profiles (or approximations to them) provided, for example, by MCR-ALS, PARAFAC and all its multi-

linear decomposition variants or (2) latent profiles (linear combinations or loadings) 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. Although the specific form of $\mathbf{Z}_{\text{unx}}$ differs from the simple intuitive expectations based on the direct extension of the NAS concept from first- to higher-order, a germ of eq (18) can already be anticipated from inspection of the FO expression shown in Table 2, which is not based on uncertainty propagation principles.

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 from zeroth-order to any order (see Table 3) implies that eq (18) is the most general expression available for estimating sensitivities. It is also worth noticing the properties of the multi-way sensitivity defined by eq (18): (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.

Table 3. General sensitivity expression and detailed parameters for the various calibration methodologies applicable to data of increasing order.

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

Model	Order	$\mathbf{g}_n$	$\mathbf{Z}_{\mathrm{exp}}$	$\mathbf{Z}_{\mathrm{unx}}$	Comments
Univariate	0	1	m_0	–	m_0 = slope of univariate graph
CLS	1	$\boldsymbol{\delta}_n$	$\mathbf{S}_{\mathrm{CLS}}$	–	$\boldsymbol{\delta}_n$ = Kronecker vector for analyte n $\mathbf{S}_{\mathrm{CLS}}$ = matrix of pure constituent profiles
ILS	1	$\mathbf{y}_{\mathrm{cal},n}$	$\mathbf{X}_{\mathrm{cal}}$	–	$\mathbf{y}_{\mathrm{cal},n}$ = vector of calibration analyte concentrations $\mathbf{X}_{\mathrm{cal}}$ = matrix of calibration signals
PCR	1	$\mathbf{v}_{\mathrm{PCR},n}$	$\mathbf{P}_{\mathrm{PCR}}$	–	$\mathbf{v}_{\mathrm{PCR},n}$ = vector of latent PCR coefficients $\mathbf{P}_{\mathrm{PCR}}$ = matrix of PCR calibration loadings
PLS	1	$\mathbf{q}_{\mathrm{PLS},n}$	$\mathbf{W}_{\mathrm{PLS}}$	–	$\mathbf{q}_{\mathrm{PLS},n} = (\mathbf{P}_{\mathrm{PLS}}^{\mathrm{T}} \mathbf{W}_{\mathrm{PLS}})^{-1} \mathbf{v}_{\mathrm{PLS},n}^{\mathrm{T}}$ $\mathbf{v}_{\mathrm{PLS},n}$ = vector of latent PLS coefficients $\mathbf{W}_{\mathrm{PLS}}$ = matrix of PLS calibration weights $\mathbf{P}_{\mathrm{PLS}}$ = matrix of PLS calibration loadings
MCR-ALS	2	$\boldsymbol{\delta}_n$	$\frac{m_n}{J^{1/2}} \mathbf{C}_{\mathrm{exp}}$	$\mathbf{C}_{\mathrm{unx}}$	J = No. of sensors of each sub-matrix in augmented mode m_n = slope of pseudo-univariate plot $\mathbf{C}_{\mathrm{exp}}$ = profiles in non-augmented mode for expected constituents in calibration $\mathbf{C}_{\mathrm{unx}}$ = profiles in non-augmented mode for unexpected constituents
PARAFAC	2, 3, ...	$\boldsymbol{\delta}_n$	See last column	See Table 4	Second-order: $\mathbf{Z}_{\mathrm{exp}} = m_n \mathbf{C}_{\mathrm{exp}} \odot \mathbf{B}_{\mathrm{exp}}$ Third-order: $\mathbf{Z}_{\mathrm{exp}} = m_n \mathbf{D}_{\mathrm{exp}} \odot \mathbf{C}_{\mathrm{exp}} \odot \mathbf{B}_{\mathrm{exp}}$ Fourth-order: $\mathbf{Z}_{\mathrm{exp}} = m_n \mathbf{E}_{\mathrm{exp}} \odot \mathbf{D}_{\mathrm{exp}} \odot \mathbf{C}_{\mathrm{exp}} \odot \mathbf{B}_{\mathrm{exp}}$ $\mathbf{B}_{\mathrm{exp}}$, $\mathbf{C}_{\mathrm{exp}}$, $\mathbf{D}_{\mathrm{exp}}$, and $\mathbf{E}_{\mathrm{exp}}$ are loading matrices in the various data modes for the expected constituents in calibration
U-PLS/RML[a]	2, 3, ...	$\mathbf{v}_{\mathrm{UPLS},n}$	$\mathbf{P}_{\mathrm{UPLS}}$	See Table 4	$\mathbf{P}_{\mathrm{UPLS}}$ = matrix of U-PLS calibration loadings $\mathbf{v}_{\mathrm{UPLS},n}$ = vector of latent PLS coefficients
N-PLS/RML[a]	2, 3, ...	$\mathbf{v}_{\mathrm{NPLS},n}$	$\mathbf{W}_{\mathrm{NPLS}}$	See Table 4	$\mathbf{W}_{\mathrm{NPLS}}$ = matrix of N-PLS calibration weights Second-order:[b] $\mathbf{W}_{\mathrm{NPLS}} = \mathbf{W}^K \odot \mathbf{W}^J$ Third-order:[b] $\mathbf{W}_{\mathrm{NPLS}} = \mathbf{W}^L \odot \mathbf{W}^K \odot \mathbf{W}^J$ Fourth-order:[b] $\mathbf{W}_{\mathrm{NPLS}} = \mathbf{W}^M \odot \mathbf{W}^L \odot \mathbf{W}^K \odot \mathbf{W}^J$ $\mathbf{v}_{\mathrm{NPLS},n}$ = vector of latent N-PLS coefficients

[a] RML = residual multi-linearization (includes RBL, RTL, RQL).
[b] J, K, L and M identify the different data modes.

5.2. Univariate calibration

In the classical one-way (zeroth-order) or univariate calibration, the relevant parameters from the general eq (18) are scalars (Z_{unx} does not exist, since no unexpected constituents are possible in this methodology), $g_n = 1$ and $Z_{exp} = m_0$, leading to $SEN_n = m_0$, the slope of the calibration graph. This agrees with the IUPAC definition, and of course with the simple and intuitive uncertainty analysis of eq (6): if calibration were precise, uncertainties in x will propagate to y through $\sigma_y = m_0^{-1} \sigma_x$, and thus SEN_n will be equal to m_0.

5.3. First-order calibration

In the first-order calibration world, the definitions of $\mathbf{Z}_{exp}$ and $\mathbf{g}_n$ depend on the specific data processing algorithm (Table 3). In any case, since no unexpected constituents should appear in the test samples, $\mathbf{Z}_{unx}$ does not exist and thus $\left(\mathbf{I} - \mathbf{Z}_{unx} \mathbf{Z}_{unx}^{+} \right) = \mathbf{I}$. It is apparent that the general eq (18) gives eq (11) for CLS, and analogous expressions for ILS, PCR and PLS [see Appendix A-1, eqs (A-10)-(A-12)], in full agreement with the NAS-based sensitivity approach.

Again, uncertainty propagation provides these results directly from the general predictive equation for analyte n:

$$y_n = \boldsymbol{\beta}_n^{\mathrm{T}} \mathbf{x} \tag{19}$$

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:

$$\sigma_y = \left(\boldsymbol{\beta}_n^{\mathrm{T}} \boldsymbol{\beta}_n \right)^{1/2} \sigma_x \tag{20}$$

From this latter expression, $SEN_n = \left(\boldsymbol{\beta}_n^{\mathrm{T}} \boldsymbol{\beta}_n \right)^{-1/2}$ immediately follows through the uncertainty propagation approach, in agreement with eq (A-10) of Appendix A-1. This sensitivity

parameter is clearly 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. One may argue that it is algorithm-specific, because different algorithms (CLS, ILS, PCR, PLS) will provide different regression vectors $\boldsymbol{\beta}_n$. However, algorithm-specificity explicitly refers to the wildly different $\mathbf{Z}_{unx}$ matrices provided by multi-way algorithms, which may cause the sensitivity to markedly differ from one algorithm to the other.

5.4. Multi-way (higher-order) calibration

5.4.1. Multi-linear algorithms

Multi-linear algorithms such as PARAFAC[54] and its variants based on the multi-linear model[55-58] provide approximations to pure constituent profiles, whether they belong to the category of expected or unexpected. 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 pseudo-univariate prediction graph (the latter is a plot of the scores or relative concentrations of a given analyte vs. its nominal calibration concentrations). What is peculiar in the case of these multi-way algorithms achieving the second-order advantage is that they are able to provide profiles for the potentially interfering constituents in a given test sample.

As a function of the relevant parameters for multi-linear multi-way calibration, the recently derived expression for the sensitivity in multi-linear models is:[40]

$$\mathrm{SEN}_n = m_n \,\|\, n\text{th row of } [(\mathbf{I} - \mathbf{Z}_{unx}\,\mathbf{Z}_{unx}^{+})\,\mathbf{Z}_{exp}]^{+} \,\|^{-1} \qquad (21)$$

The matrix $\mathbf{Z}_{exp}$ is defined in Table 3 as a function of the loading matrices for second-, third- and fourth-order as in eqs. (22), (23) and (24) respectively, while Table 4 shows the specific forms of the $\mathbf{Z}_{unx}$ matrix:

$$\mathbf{Z}_{exp} = m_n \, (\mathbf{C}_{exp} \odot \mathbf{B}_{exp}) \tag{22}$$

$$\mathbf{Z}_{exp} = m_n \, (\mathbf{D}_{exp} \odot \mathbf{C}_{exp} \odot \mathbf{B}_{exp}) \tag{23}$$

$$\mathbf{Z}_{exp} = m_n \, (\mathbf{E}_{exp} \odot \mathbf{D}_{exp} \odot \mathbf{C}_{exp} \odot \mathbf{B}_{exp}) \tag{24}$$

In the latter expressions, the symbol '⊙' indicates the Khathri-Rao product operator,[75] also known as the column-wise Kronecker product, because for matrices **A** and **B**, the *i*th. column of **A**⊙**B** follows from the *i*th. columns of **A** and **B** as $\mathbf{a}_i \otimes \mathbf{b}_i$. Therefore, the columns of $\mathbf{Z}_{exp}$ are proportional to the pure signals for each constituent in the calibration set, each unfolded into a vector and normalized to unit length.

Table 4. Content of the matrix representing the space spanned by the unexpected constituents in multi-way (higher-order) calibration.

Order	$\mathbf{Z}_{unx}$ [a]
2	$[\, \mathbf{c}_1 \otimes \mathbf{I}_b \mid \mathbf{I}_c \otimes \mathbf{b}_1 \mid \mathbf{c}_2 \otimes \mathbf{I}_b \mid \mathbf{I}_c \otimes \mathbf{b}_2 \mid ...]$
3	$[\mathbf{d}_1 \otimes \mathbf{c}_1 \otimes \mathbf{I}_b \mid \mathbf{d}_1 \otimes \mathbf{I}_c \otimes \mathbf{b}_1 \mid \mathbf{I}_d \otimes \mathbf{c}_1 \otimes \mathbf{b}_1 \mid \mathbf{d}_2 \otimes \mathbf{c}_2 \otimes \mathbf{I}_b \mid \mathbf{d}_2 \otimes \mathbf{I}_c \otimes \mathbf{b}_2 \mid \mathbf{I}_d \otimes \mathbf{c}_2 \otimes \mathbf{b}_2 \mid ...]$
4	$[\mathbf{e}_1 \otimes \mathbf{d}_1 \otimes \mathbf{c}_1 \otimes \mathbf{I}_b \mid \mathbf{e}_1 \otimes \mathbf{d}_1 \otimes \mathbf{I}_c \otimes \mathbf{b}_1 \mid \mathbf{e}_1 \otimes \mathbf{I}_d \otimes \mathbf{c}_1 \otimes \mathbf{b}_1 \mid \mathbf{I}_e \otimes \mathbf{d}_1 \otimes \mathbf{c}_1 \otimes \mathbf{b}_1 \mid \mathbf{e}_2 \otimes \mathbf{d}_2 \otimes \mathbf{c}_2 \otimes \mathbf{I}_b \mid \mathbf{e}_2 \otimes \mathbf{d}_2 \otimes \mathbf{I}_c \otimes \mathbf{b}_2 \mid \mathbf{e}_2 \otimes \mathbf{I}_d \otimes \mathbf{c}_2 \otimes \mathbf{b}_2 \mid \mathbf{I}_e \otimes \mathbf{d}_2 \otimes \mathbf{c}_2 \otimes \mathbf{b}_2 \mid ...]$

[a] The profiles $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{d}_1$, $\mathbf{d}_2$, $\mathbf{e}_1$, $\mathbf{e}_2$, … correspond to the unexpected constituents in the various data modes. $\mathbf{I}_b$, $\mathbf{I}_c$, $\mathbf{I}_d$ and $\mathbf{I}_e$ are appropriately dimensioned unit matrices, of size $J \times J$, $K \times K$, $L \times L$ and $M \times M$ respectively. The numbers 1, 2, ... run up to the total number of unexpected constituents.

In Appendix A-2 it is shown how this latter expression can be cast in the general format of eq (18). It only requires introduction of the vector $\mathbf{g}_n$, which is the previously discussed vector

$\boldsymbol{\delta}_n$ serving, as before, to select a particular constituent from the various calibrated constituents (Table 3).

It may be noticed that for three-way (second-order) calibration, eqs (18) and (21) appear to be different than the MKL, HCD and FO expressions (Table 2), however the latter numerical results are identical to those provided by eq (18), indicating that all previous approximations based on the net analyte signal are special cases of the general uncertainty propagation expression.[40]

5.4.2. Multivariate curve resolution-alternating least-squares

For the MCR-ALS algorithm applied in the so-called extended mode to a set of calibration and test data matrices forming an augmented matrix (see Supporting Information), the corresponding SEN_n expression has been recently derived:[41]

$$\mathrm{SEN}_n = m_n \left[J \left(\mathbf{C}^{\mathrm{T}} \mathbf{C} \right)^{-1}_{nn} \right]^{-1/2} \qquad (25)$$

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

The MCR-ALS sensitivity expression can also be shown (see Appendix A-2) to be adequately covered by the general eq (18). The connection is simple: $\mathbf{Z}_{\mathrm{exp}}$ and $\mathbf{Z}_{\mathrm{unx}}$ are equal to

$\mathbf{C}_{exp}$ and $\mathbf{C}_{unx}$ respectively, as indicated in Table 3, and the vector $\mathbf{g}_n$ is equal to the multi-linear selector $\boldsymbol{\delta}_n$.

5.4.3. Partial least-squares/residual multi-linearization

For multi-way 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 eq (18). In this case no pure constituent profiles are available, but combinations of the latter ones in abstract calibration loadings (see Supporting Information). For U-PLS calibration, for example, $\mathbf{Z}_{exp}$ is composed of columns which are the so-called calibration loadings contained in the matrix $\mathbf{P}_{UPLS}$, 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 (Table 3). The above discussion concerning the properties of the $\left(\mathbf{I}-\mathbf{Z}_{unx}\mathbf{Z}_{unx}^{+}\right)$ matrix is also pertinent in this case. The uncertainty propagation approach fully agrees with the expression for the U-PLS/RBL sensitivity which was previously derived from NAS considerations.[68] An analogous expression can be derived for N-PLS/RBL.[42]

5.5. Other multi-way algorithms

The general eq (18) has been applied to assess the sensitivity for several algorithms commonly employed for multi-way calibration. However, there are additional methodologies, such as the second-order GRAM and DTLD models,[65,66] which are somewhat less employed. It has been shown that GRAM always achieves the lowest HCD sensitivity (Table 2), even when various constituents are calibrated.[36] This is 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. Overall, it is an argument in favor of the latter calibration philosophy as opposed to one-sample calibration.

Other algorithms for which sensitivity studies are lacking are MLLS/RML,[63-68] the classical version of PLS/RML in what concerns the calibration phase, although an educated guess is that it would fit into the scheme of eq (18) under the PARAFAC umbrella, and PARAFAC2,[71] a variant of PARAFAC conceived to cope with non-multilinear multi-way data with one trilinearity-breaking mode, e.g., chromatographic-spectral second-order data, whose sensitivity properties have yet to be explored.

5.6. Multi-way net analyte signal

The above results may trigger a debate as to the existence of a multi-way net analyte signal bearing a link with analyte sensitivity. Interestingly, as shown in eqs (A-14) and (A-15) (see Appendix A-1), the left-hand side of eq (18) is the length of a vector, which can be considered as a multi-way net analyte signal vector (in unfolded format) at unit concentration for the analyte of interest, analogously to the useful concept employed in first-order calibration. Needless to say, the expression for the multi-way net analyte vector should reduce to the first-order calibration version for data with a single instrumental mode. In any case, the alleged multi-way NAS definition needs to be flexible in the interpretation of the expression 'space spanned by other sample constituents'. Careful inspection of the specific mathematical expressions for $\mathbf{Z}_{unx}$ in Table 4 indicates that the space spanned by this matrix is not a function of the individual spaces for the unexpected constituents in each of the data modes. Instead, the columns of $\mathbf{Z}_{unx}$ are expressed as combinations of profiles for a certain number of modes. As can be seen in Table 4, 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 No. 1 agent, the first block will look as follows:

$$\mathbf{Z}_{unx} = [\mathbf{d}_1 \otimes \mathbf{c}_1 \otimes \mathbf{I}_b \mid \mathbf{d}_1 \otimes \mathbf{I}_c \otimes \mathbf{b}_1 \mid \mathbf{I}_d \otimes \mathbf{c}_1 \otimes \mathbf{b}_1 \mid \ldots] \quad (26)$$

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. On the other hand, in five-way (fourth-order) calibration, the block of $\mathbf{Z}_{unx}$ corresponding to the first unexpected agent includes four columns:

$$\mathbf{Z}_{unx} = [\mathbf{e}_1 \otimes \mathbf{d}_1 \otimes \mathbf{c}_1 \otimes \mathbf{I}_b \mid \mathbf{e}_1 \otimes \mathbf{d}_1 \otimes \mathbf{I}_c \otimes \mathbf{b}_1 \mid \mathbf{e}_1 \otimes \mathbf{I}_d \otimes \mathbf{c}_1 \otimes \mathbf{b}_1 \mid \mathbf{I}_e \otimes \mathbf{d}_1 \otimes \mathbf{c}_1 \otimes \mathbf{b}_1 \mid \ldots] \quad (27)$$

where all symbols have analogous meanings to above. 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, $\mathbf{Z}_{unx}$ will display analogous forms to those found in Table 4. 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 the specific manner depicted in Table 4 for N = 2, 3 and 4. In sum, $\mathbf{Z}_{unx}$ conceivably represents the space spanned by the unexpected constituents, but in a non-classical way, although the systematic block characteristics of this matrix makes it easy to build it for any data order and number of interfering agents.

6. OTHER FIGURES OF MERIT

For the remaining figures of merit to be discussed in the present review, an analogy is made with the univariate counterparts, except for certain aspects which are specific of multivariate/multi-way calibration, as explained below.

6.1. 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 × 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:[4]

$$\gamma_n = SEN_n / \sigma_x \quad (28)$$

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 eqs (17) and (28) 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,[4] 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 eq (28).

6.2. 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.[5] This qualitative definition does not imply a specific procedure for the estimation of a numerical selectivity parameter, for which some controversy exists.[80]

Several requirements have been proposed for a consistent numerical selectivity:[80] (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 over-determined systems (having more sensors or wavelengths than components), and (6) generalization to multi-way 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:[11]

$$SEL_n = SEN_n \text{ (in a mixture)} / SEN_n \text{ (pure)} \quad (29)$$

In univariate calibration, SEL should be equal to 1 (100%, meaning full selectivity), because no interfering agents are allowed. In first-order classical least-squares calibration, eq (29) naturally follows as a consequence of the LBOZ criterion,[80] by setting the denominator as equal to $\| \mathbf{s}_n \|$, which is a measure of the pure analyte signal. However, for first-order latent based calibration models (in fact for latent models of any data order), no approximations are available to pure analyte profiles, 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 eq (29) in these cases,[19] i.e., the value of the overall $\| \mathbf{x} \|$ instead of $\| \mathbf{s}_n \|$, this makes the SEL_n parameter highly dependent on the unknown samples, even if the qualitative compositions of the latter are similar. In other words, two test samples A and B having the same number and type of constituents are expected to display the same selectivity towards a given analyte. However, if the overall signal increases two-fold in going from A to B because concomitants other than the analyte of interest are more concentrated in B than in A, then the first-order selectivity defined as a function of the overall signal would be twice as large in A than in B, which is not reasonable.

Therefore, 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.

The concept of selectivity has been generalized to multi-way analysis.[34,37,79] The defining eq (29) implies that the multi-way selectivity is accessible when the pure analyte signal is adequately retrieved by the processing algorithm. This is possible in the case of multi-linear (e.g., PARAFAC) analysis, for which the selectivity (SEL_n) is directly given by:

$$SEL_n = SEN_n / m_n \quad (30)$$

where m_n is the slope of the pseudo-univariate calibration graph. The degree by which SEN_n departs from m_n in eq (30) 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). The adequacy of the latter approach has been revealed in the chromatographic context, where a relationship between multi-way selectivity and classical separation metrics has been sought. Indeed, a direct relationship has been proposed to exist between the effective peak capacity of a chromatogram and the multi-way selectivity.[81]

Another multi-way framework in which selectivity can be defined is MCR-ALS, where the selectivity is:[41]

$$SEL_n = SEN_n J^{1/2} / m_n \quad (31)$$

Equation (31) also leads to continuous values in the range 0-1, depending on the relative degree of overlapping among the profile for the various sample constituents.

Finally, a different approach to selectivity is worth mentioning, based on quantifying the impact of individual interferences on analyte predictions, leading to a definition of pair-wise multivariate selectivity coefficients.[82-84] The idea is reminiscent of pair-wise selectivity measures,[85] traditionally applied to potentiometric ion selective electrodes,[86-88] which demands availability of specific potential interferents.

6.3. Prediction uncertainty

Since all analytical results should be accompanied by the corresponding uncertainty, the latter is an important figure of merit to be estimated and reported. Notice the compelling title of a publication in the field: *Measurement results without statements of reliability should not be taken seriously*.[89] Prediction uncertainties are also essential to assess detection capabilities, as shown in the next section.

Two basic proposals exist for estimating prediction standard errors in multivariate/multi-way analysis:[90] (1) resampling techniques such as jack-knife or bootstrap,[91] 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.[11,76]

The best approximation to concentration variance is the well-known three-term expression (valid for propagation of homoscedastic and uncorrelated noise): [11,76,92]

$$\sigma_y^2 = \mathrm{SEN}_n^{-2}\sigma_x^2 + h\ \mathrm{SEN}_n^{-2}\sigma_x^2 + h\ \sigma_{y\mathrm{cal}}^2 \qquad (32)$$

where σ_x^2 the variance in instrumental signals, h the sample leverage and $\sigma_{y\mathrm{cal}}^2$ the variance in calibration concentrations. The three terms in the right-hand side of eq (32) account for the propagation of uncertainties derived from (in the order in which they appear): (1) instrumental signals in the test sample data, (2) instrumental signals in the calibration data, and (3) calibration concentrations. The first and probably the most relevant of these contributions is transmitted directly *via* the inverse squared sensitivity, which is the most significant ingredient in eq (32). The second and third terms arise from calibration uncertainties, and are both scaled by the sample leverage h, a dimensionless parameter measuring the position of the sample relative to the calibration space. The leverage has a simple expression in univariate calibration,[9] and also in multi-way methodologies resorting to a pseudo-univariate calibration graph,[35] otherwise the

position of the test sample relative to the calibration space depends on the presence and level of other sample constituents. A general equation is able to appropriately cover all cases, however:

$$h = \mathbf{f}_{\text{test}}^{\text{T}}(\mathbf{F}_{\text{cal}}^{\text{T}}\mathbf{F}_{\text{cal}})^{-1}\mathbf{f}_{\text{test}} \quad (33)$$

where $\mathbf{F}_{\text{cal}}$ is a matrix (or vector) and $\mathbf{f}_{\text{test}}$ a vector, corresponding to the calibration set of samples and to the test sample respectively. Details on their specific forms in the different methodologies are given in Table 5. Recall that if data are mean-centered before calibration models are built, then $(1/I_{\text{cal}})$ (I_{cal} is the number calibration samples) should be added to the leverage in eq (33). This is also true when the data are modeled including an intercept, as in univariate calibration through eq (6), where h becomes the familiar expression:[9]

$$h = \frac{1}{I_{\text{cal}}} + \frac{(y - \bar{y}_{\text{cal}})^2}{\sum_{i=1}^{I\text{cal}} (y_i - \bar{y}_{\text{cal}})^2} \quad (34)$$

In eq (34), y is the predicted analyte concentration, y_i its nominal concentration in the ith. calibration sample, and $\bar{y}_{\text{cal}}$ the mean calibration concentration. It may be noticed that eq (32) is accurate for the univariate case[9] and for classical least-squares first-order calibration.[93] **Error! Bookmark not defined.**For the remaining calibration scenarios, the first term of eq (32) is accurate,[76] while the remaining two terms have been shown to be excellent approximations in most cases.[35,42]

Notice the relation between prediction uncertainty and sensitivity, which is direct if only the term propagating the noise in the test sample is considered. Inspection of eqs (30) and (32) indicates that there is also a relation between uncertainty in prediction and selectivity, which is one of the requisites for the latter to be consistent.

Table 5. Values of the leverage parameter in analytical methodologies based on data of various orders.

General expression[a] $h = \mathbf{f}_{test}^{T}(\mathbf{F}_{cal}^{T}\mathbf{F}_{cal})^{-1}\mathbf{f}_{test}$

Model	Order	$\mathbf{F}_{cal}$ (size)[b]		$\mathbf{f}_{test}$ (size)[c]		Ref.
Univariate	0	$\mathbf{y}_{cal}$	(I_{cal}×1)	y_n	(1×1)	9
CLS	1	$\mathbf{Y}_{cal}$	(I_{cal}×N)	$\mathbf{y}$	(N×1)	93
ILS	1	$\mathbf{X}_{cal}^{T}$	(I_{cal}×J)	$\mathbf{x}$	(J×1)	76
PCR	1	$\mathbf{T}_{cal}$	(I_{cal}×A)	$\mathbf{t}$	(A×1)	76
PLS	1	$\mathbf{T}_{cal}$	(I_{cal}×A)	$\mathbf{t}$	(A×1)	76
MCR-ALS	2	$\mathbf{y}_{cal,n}$	(I_{cal}×1)	y_n	(1×1)	[d]
PARAFAC	2, 3, ...	$\mathbf{y}_{cal,n}$	(I_{cal}×1)	y_n	(1×1)	35
U-PLS/RML	2, 3, ...	$\mathbf{T}_{cal}$	(I_{cal}×A)	$\mathbf{t}$	(A×1)	42
N-PLS/RML	2, 3, ...	$\mathbf{T}_{cal}$	(I_{cal}×A)	$\mathbf{t}$	(A×1)	42

[a] When data are mean-centered or the univariate/pseudo-univariate calibration includes an intercept, all the parameters quoted in the present table should be centered, and a term ($1/I_{cal}$) should be added to the leverage (I_{cal} = number of calibration samples).

[b] Calibration $\mathbf{F}_{cal}$: $\mathbf{y}_{cal}$ and $\mathbf{y}_{cal,n}$, vector of calibration concentrations for the analyte of interest, $\mathbf{Y}_{cal}$, matrix of calibration concentrations of all analytes, $\mathbf{T}_{cal}$, matrix of calibration scores, $\mathbf{X}_{cal}$, matrix of calibration signals, N = number of calibrated analytes, A = number of calibration latent variables, J = number of sensors or predictor variables.

[c] Test sample $\mathbf{f}_{test}$: y_n, predicted analyte concentration, $\mathbf{y}$, vector of predicted analyte concentrations, $\mathbf{t}$, vector of test sample scores, $\mathbf{x}$, vector of test sample signals.

[d] This leverage expression is an educated guess, because it has not been fully tested yet.

6.4. Detection capabilities

The limit of detection (LOD) for a given analyte is an important parameter to be reported as a figure of merit. The modern definition is due to Currie's pioneering work on hypothesis-based detection limit theory.[94] It can be qualitatively defined as the minimum analyte concentration which is detectable *with a certain degree of confidence* (notice the emphasis in the latter words). The precise definition of the LOD, officially recommended by IUPAC, however, is somewhat less simple. It first requires to define a critical concentration level (CL), which is the level for the detection decision, involving a certain risk of false detects (also called false positives, α-errors or Type I errors). The limit of detection is then defined as a concentration level for which the risk of false non-detects (false negatives, β-errors or Type II errors) has a probability β.[1,95-97] Both α and β are usually assigned reasonably small values, depending on the specific analytical application. Figure 11 illustrates the different concepts involved in the definition of the LOD.[98]

It is important to recognize the difference between minimum detectable concentration (CL) and minimum detectable concentration with a certain degree of confidence (LOD). Intriguing as it may seem, it is possible to detect the analyte when its concentration is actually below the limit of detection, because the detection decision is taken at CL and not at LOD (Figure 11). The critical level CL is sometimes confounded with the limit of detection LOD; however they only coincide for a 50% probability of β errors, which is clearly an unreasonable situation. As a further qualitative insight into these detection concepts, Figure 12 distinguishes three analyte concentration regions: (1) from zero to CL, where the analyst may declare that the analyte is absent (with a probability α of false detects), (2) above LOD, where the analyte may be declared present (with a probability β of false non-detects), and (3) a 'no man land' region

between CL and LOD, where not enough evidence exists for asserting the presence of the analyte.

Figure 11 allows one to intuitively reach an expression for the LOD. If 95% for confidence levels against both Type I and II errors is considered, and further assumed that both Gaussian curves in Figure 11 have similar widths, i.e., that $\sigma_{y,0} \approx \sigma_{y,\mathrm{LOD}}$ then the LOD is given by $(t_{\alpha,\nu} + t_{\beta,\nu})\,\sigma_{y,0}$, where $t_{\alpha,\nu}$ and $t_{\beta,\nu}$ are the *t*-coefficients for probabilities α and β, with ν the degrees of freedom. This makes the LOD proportional to the uncertainty in predicted concentration near a blank sample:[11,99,100]

$$\mathrm{LOD}_n = 3.3\,(\mathrm{SEN}_n^{-2}\sigma_x^2 + h_0\,\mathrm{SEN}_n^{-2}\sigma_x^2 + h_0\,\sigma_{y\mathrm{cal}}^2)^{1/2} \qquad (35)$$

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 equal to $(t_{\alpha,\nu} + t_{\beta,\nu})$ for $\alpha = 0.05$ and $\beta = 0.05$ and a large value of ν.[11,1,95] The factor in front of eq (35) may be corrected for other probabilities and degrees of freedom. Notice the assumptions underlying eq (35): (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,[101] 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 non-centrality parameter of a non-central *t* distribution instead of a sum of classical *t*-coefficients.[97] It is likely, however, that the values provided by eq (35) and more elaborate statistical approaches do not significantly differ.[101] In any case, for a thorough critique of the approaches based on prediction intervals and non-centrality parameters see the work of Voigtman.[102,103]

In univariate calibration, the subscript *n* may be dropped and the LOD characterizes the detection capability towards the analyte under study. In multivariate calibration, however, the situation is different. 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. Hence the LOD_n not only becomes analyte specific, but also sample specific. 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 pseudo-univariate calibration is employed).[36] This means that the detection capability towards a given analyte depends on various factors beyond the instrumental signals measured for a set of calibration samples. To overcome the sample-dependency issue, the usual criterion has been to report an average LOD_n value over a group of test samples of similar qualitative composition. This provides a reasonable estimate of the detection capability in a certain chemical environment, and helps to understand the effect of background and potential interfering agents on the analyte detection for complex samples.

The limit of quantitation (LOQ_n), in turn, is estimated as the concentration level for which the relative prediction error is 10%, and is easily set at a concentration value which is 10 times the associated prediction uncertainty:[11]

$$LOQ_n = 10\,(SEN_n^{-2}\sigma_x^2 + h_0\ SEN_n^{-2}\sigma_x^2 + h_0\ \sigma_{y\mathrm{cal}}^2)^{1/2} \qquad (36)$$

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

Another approach to estimating LOD_n and LOQ_n has been taken when multi-way calibration provides the analyte concentration through a pseudo-univariate linear calibration graph. This amounts to considering the latter as a true single-constituents calibration, and computing the detection capabilities directly from univariate analysis.[104-106] While this approach should in principle furnish similar detection capabilities as eqs (35) and (36), it does not allow

one to estimate neither the selectivity, because the pseudo-univariate graph is taken as a true univariate representation, nor the sensitivity, because the vertical scale of the pseudo-univariate graph is not based on original signals.

The reduction of multivariate calibration results to the univariate case for estimating detection capabilities has also been proposed for all possible methodologies, by analysis of the linear regression of predicted vs. nominal analyte concentrations.[107,108] Although this approach is appealing from the intuitive point of view, it is likely that detection and quantitation limits estimated in this way are only averages of those corresponding to samples with low analyte concentrations, but with varying levels of other constituents.

7. AVAILABILITY OF SOFTWARE

Currently few commonly available software packages have incorporated the latest developments in multivariate figures of merit. Examples of useful graphical interfaces which do offer these capabilities are MVC1, MVC2 and MVC3 for first-, second- and third-order calibration respectively, which are freely available at www.iquir-conicet.gov.ar/descargas/mvc1.rar, www.iquir-conicet.gov.ar/ descargas/mvc2.rar and www.iquir-conicet.gov.ar/descargas/mvc3.rar. However, the need of reporting these figures in multivariate/multi-way calibration works will drive multi-way program developers to include them in the near future.

8. COMPARISON OF FIGURES OF MERIT

It is already known that measuring and processing multivariate data leads to a sensitivity increase, derived from multiple redundant measurements and noise averaging.[21-26,109] The sensitivity increase can now be precisely computed using the general expression (18). This may

help in advanced planning, and in anticipating the sensitivity gain for complex multi-way experiments. A trade-off between sensitivity increase and experimental complexity is always desired;[39] eq (18) may help in deciding whether the extra experimental effort is worthwhile. Furthermore, eq (18) may be useful in selecting a data processing algorithm for a specific multi-way data set. It has already been suggested that unfolding a multi-way data array into arrays of lower modes before data processing may lead to a sensitivity loss, meaning that maintaining the original multi-way structure is always preferable.[109] The selection is only fair among algorithms whose internal models match the specific data properties at hand; otherwise, the estimated figures of merit would not be useful in this regard.

To illustrate the sensitivity gain when increasing the number of instrumental sensors and data orders, the relevant SEN_n parameter was computed for an analyte of interest in a simple set of simulated data for two constituents. Figure 13A shows the specific constituent profiles, consisting of highly overlapped Gaussian lines defined in a range of 50 different sensors in all instrumental data modes. From these profiles, SEN_n was calculated for the analyte (Figure 13A) for various data orders and algorithms, with the results shown in Figure 13B. In the case of first-order algorithms, ILS was assumed to involve the ten sensors most sensitive to the analyte (this methodology requires less sensors than calibration samples), whereas CLS, PCR and PLS employed full sensor data, with almost the same sensitivity for the latter three; hence they are grouped into a single bar in Figure 13B. It is apparent that the first-order sensitivity greatly increases in going from a small sensor set to full sensor data. As also indicated in Figure 13B, additional sensitivity is gained in going to second-order data (compare blue bars with red bars). Finally, third-order data (green bars in Figure 13B) and fourth-order data (black bars in Figure 13B) lead to even higher sensitivities. Some subtle differences among multi-way algorithms are discussed below for additional analytical scenarios.

It is also interesting to extend the analysis to multi-way algorithms, comparing these two relevant analytical situations: (1) both constituents are present in the calibration set, and (2) one of them is calibrated and the other one is a potential interferent. For this purpose, many different degrees of overlap of the constituent Gaussian profiles have been considered for second-, third-, and fourth-order data. A comparison of PARAFAC and MCR-ALS analyte selectivities [eqs (30) and (31) respectively] are shown in Figure 14, where interesting conclusions can be drawn (third- and fourth-order data are previously unfolded to matrices for MCR-ALS application). When both constituents are present in the calibration set (A, B and C), PARAFAC appears to be the best option. However, in the event one of the constituents is an analyte and the other one a potential interferent (D, E and F), MCR-ALS provides the best selectivity for second-order data. This trend is reversed for third- and fourth-order data, where PARAFAC outperforms MCR-ALS in most (although not all) cases. Of course all these results apply when the data are multi-linear, otherwise multi-way calibration based on PARAFAC analysis is not a good choice. The conclusion is that selection of potentially competing algorithms may be based on selectivity issues, which depend on the data order and also on the number and nature of constituents in the calibration set and in the unknown samples. In the end, selectivity studies may be useful in deciding the best multi-way calibration algorithm on purely analytical considerations.

The trends observed in Figures 13 and 14 will certainly derive in corresponding changes in detection capabilities, although the latter depend on other parameters beyond the sensitivity. Accordingly, they may improve, but probably not to the same degree by which SEN_n increases. An illustrative example concerns the decrease in LOD_n which can be achieved on increasing the sensitivity by increasing the number of data ways. Figure 15 shows a typical plot of LOD_n as a function of increasing SEN_n for different relative values of signal and concentration uncertainties. The LOD_n decreases for increasing sensitivity, and the effect is more significant for larger values

of σ_x with respect to σ_y, because this gives comparatively higher importance to the first two terms of eq (35), which are sensitivity-dependent. However, LOD_n tends to level off at a certain point, because of the contribution of the leverage-dependent term corresponding to the propagated calibration concentration uncertainty in eq (35), which is independent of the sensitivity. By only judging from the perspective of the sensitivity, any increase would be worth the experimental effort of increasing the data order. However, this may not be immediately translated into correspondingly lower limits of detection.

9. CONCLUSIONS

With the unified approach to sensitivity communicated in this report, the estimation of this important parameter in most calibration scenarios rests on a consistent ground, for data ranging from zeroth- to multi-way (higher-order). Moreover, the underlying uncertainty propagation approach provides a reliable platform for future sensitivity studies. Important lessons are to be learned from the study of the multi-way sensitivity. One is that the latter not only depends on the employed instrument and the measured data, but also on the data processing algorithm, the test sample under scrutiny, and whether the sample constituents are present in the calibration set or only in the test sample. Another important outcome from the uncertainty propagation approach is that increasing the number of data ways increases the sensitivity, but does not decrease the limit of detection by the same degree. All these considerations should become important when it comes to planning a specific multi-way calibration protocol, and point to an integrated view of the analytical process, which specifically includes the data processing algorithm.

ACKNOWLEDGMENTS

The author is indebted to Dr. Klaas Faber, Chemometry Consultancy, The Netherlands, for fruitful discussions during the last ten years, which ultimately led to the development of the latest expressions for the multi-way figures of merit. Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project No. PIP 1950), ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project No. PICT-2010-0084) are gratefully acknowledged for financial support.

SUPPORTING INFORMATION AVAILABLE

The document 'Supporting information multi way AFOMs.docx' is provided, containing a detailed description of the operation of the three multi-way calibration algorithms discussed in this review, i.e., PARAFAC, MCR-ALS and PLS/RML.

SYMBOLS AND ACRONYMS

$\|\|\ \|\|$	Length of a vector, norm or squared root of the sum of its squared elements
a_{in}	Element of vector $\mathbf{a}_n$
$\mathbf{a}_n$	Column vector with profiles in sample mode
$\mathbf{B}$, $\mathbf{C}$	Matrices with profiles in instrumental modes
$\mathbf{B}_{\text{aug}}$	Matrix with profiles in augmented mode
$\mathbf{b}_{\text{aug},n}$	Profile in the augmented mode
$b_{\text{aug},pn}$	Element of profile in the augmented mode
b_{jn}, c_{kn}, d_{ln}	Elements of vectors $\mathbf{b}_n$, $\mathbf{c}_n$, $\mathbf{d}_n$
$\mathbf{b}_n$, $\mathbf{c}_n$, $\mathbf{d}_n$	Column vector with profiles in instrumental modes

$\mathbf{C}_{exp}$, $\mathbf{B}_{exp}$	Matrices of profiles in the instrumental modes for expected constituents
CL	Critical limit
CLS	Classical least-squares
$\mathbf{C}_{unx}$, $\mathbf{B}_{unx}$	Matrices of profiles in the instrumental modes for unexpected constituents
DAD	Diode array detector
DTLD	Direct trilinear decomposition
EEM	Excitation-emission matrix
$\mathbf{F}_{cal}$	Calibration matrix used to compute leverage
FO	Faber and Olivieri
FSFD	Fast-scanning fluorescence detection
$\mathbf{f}_{test}$	Sample vector used to compute leverage
GC	Gas chromatography
$\mathbf{g}_n$	Information-selecting vector for analyte n
GRAM	Generalized rank annihilation method
h, h_0	Leverage of sample and blank
HCD	Ho, Christian and Davidson
i	Index for sample
I	Number of samples
$\mathbf{I}$, $\mathbf{I}_b$, $\mathbf{I}_c$	Unit matrices
ILS	Inverse least-squares
j, k	Indexes for data points in instrumental modes
J, K	Number of data points in instrumental modes
LBOZ	Lorber, Bergmann, von Oepen and Zinn
LC	Liquid chromatography

LOD_n	Limit of detection for analyte n
LOQ_n	Limit of quantitation for analyte n
m_0, m_n	Slope of linear univariate and pseudo-univariate calibrations
$\mathbf{M}_1$, $\mathbf{M}_2$	Unit concentration data matrices
MCR-ALS	Multivariate curve resolution-alternating least squares
MKL	Messick, Kalivas and Lang
MLLS	Multi-linear least-squares
MS	Mass spectrometry
n	Index for constituent
n_0	Intercept of linear calibration
N	Number of constituents
NAS	Net analyte signal
NBRA	Non-bilinear rank annihilation
N-PLS	Multi-way PLS
PARAFAC	Parallel factor analysis
PCR	Principal component regression
PLS	Partial least-squares
$\mathbf{P}_{UPLS}$	Matrix of unfolded PLS loadings
$\mathbf{P}_{PCR}$	Matrix of PCR loadings
$\mathbf{P}_{PLS}$	Matrix of PLS loadings
$\mathbf{q}_{PLS,n}$	PLS vector in latent space
RML	Residual multi-linearization (RBL, bi-, RTL, tri-, RQL, quadrilinearization)
$\mathbf{s}_1^*$	Net analyte signal for constituent 1 at unit concentration

$\mathbf{s}_1$, $\mathbf{s}_n$	Pure constituent column vectors (e.g., spectra)
$\mathbf{S}_{CLS}$	Matrix of pure constituent signals in CLS regression
SEL_n	Selectivity for analyte *n*
SEN_n	Sensitivity for analyte *n*
T	As a superscript, matrix transposition
$\mathbf{T}_{cal}$	Matrix of calibration scores in PCR and PLS
$\mathbf{t}$	Vector of sample scores in PCR and PLS
t	Student's *t*-statistics
TSF	Total synchronous fluorescence
U-PLS	Unfolded PLS
$\mathbf{v}_{UPLS,n}$	Vector of U-PLS regression coefficients in latent space
$\mathbf{v}_{PCR,n}$	Vector of PCR coefficients in latent space
$\mathbf{v}_{PLS,n}$	Vector of PLS coefficients in latent space
$\mathbf{W}_{PLS}$	Matrix of PLS weight loadings
$\mathbf{x}_1^*$	Net analyte signal for constituent 1 in a mixture
$\mathbf{x}$	Column vector (e.g., spectrum)
$\mathbf{X}$	Data matrix
$\mathbf{X}_{cal}$	Calibration data matrix
$\boldsymbol{X}$	Three- and four-way data array
x	Univariate signal
$\mathbf{X}_{aug}$	Augmented data matrix
x_{ij}, x_{ijk}, x_{ijkl}	Element of matrix, three-way and four-way array
y, y_1, y_n	Analyte concentration

$\mathbf{y}$	Vector of analyte concentrations
$\mathbf{y}_{cal}$	Vector of calibration analyte concentrations in univariate calibration
$\mathbf{y}_{cal,n}$	Vector of calibration analyte concentrations in multivariate calibration
$\mathbf{Y}_{cal}$	Matrix of calibration concentrations of all analytes
$\mathbf{Z}_{exp}$	Matrix of loadings for the expected constituents
$\mathbf{Z}_{unx}$	Matrix of loadings for the unexpected constituents
α, β	Probabilities
$\boldsymbol{\beta}_{CLS,n}$	Vector of regression coefficients for analyte n in CLS analysis
$\boldsymbol{\beta}_n$	Generalized vector of regression coefficients for analyte n
$\boldsymbol{\delta}_n$	Vector with all zeros except a 1 at the nth analyte index
γ_n	Analytical sensitivity for analyte n
ν	Degrees of freedom
σ_x	Uncertainty in signal
σ_y	Uncertainty in concentration

APPENDICES

A-1. First-order sensitivity

This Appendix shows the relationship between the various expressions for the sensitivity towards a particular analyte in first-order calibration, both direct (classical least-squares) and inverse (e.g., partial least-squares). For the simple example of a binary mixture of constituents, if $\mathbf{S}_{CLS}$ is a two-column matrix whose columns are the pure constituent profiles $\mathbf{s}_1$ and $\mathbf{s}_2$, i.e., $\mathbf{S}_{CLS} = [\mathbf{s}_1 \mid \mathbf{s}_2]$, then the pseudo-inverse $\mathbf{S}^+_{CLS}$ is the two-row matrix:[75]

$$\mathbf{S}^+_{CLS} = \begin{bmatrix} \left[\left(\mathbf{I} - \mathbf{s}_2\mathbf{s}_2^+\right)\mathbf{s}_1\right]^+ \\ \left[\left(\mathbf{I} - \mathbf{s}_1\mathbf{s}_1^+\right)\mathbf{s}_2\right]^+ \end{bmatrix} \tag{A-1}$$

Matrix multiplication of $\mathbf{S}^+_{CLS}$ by its transposed leads to:

$$\text{(1,1) element of } \mathbf{S}^+_{CLS}\left(\mathbf{S}^+_{CLS}\right)^T = \text{(1,1) element of } \left(\mathbf{S}^T_{CLS}\mathbf{S}_{CLS}\right)^{-1} = \left[\mathbf{s}_1^T\left(\mathbf{I} - \mathbf{s}_2\mathbf{s}_2^+\right)\mathbf{s}_1\right]^{-1} \tag{A-2}$$

Recall that the sensitivity towards analyte 1 is given by the length of the vector $\mathbf{s}_1^*$ in eq (10), which can be computed as:

$$\text{SEN}_1 = \| \mathbf{s}_1^* \| = \left(\mathbf{s}_1^{*T}\mathbf{s}_1^*\right)^{1/2} = \left[\mathbf{s}_1^T\left(\mathbf{I} - \mathbf{s}_2\mathbf{s}_2^+\right)\mathbf{s}_1\right]^{1/2} \tag{A-3}$$

where the last result follows since the orthogonal matrix $(\mathbf{I} - \mathbf{s}_2\,\mathbf{s}_2^+)$ is symmetric and idempotent, i.e., $(\mathbf{I} - \mathbf{s}_2\,\mathbf{s}_2^+)^T(\mathbf{I} - \mathbf{s}_2\,\mathbf{s}_2^+) = (\mathbf{I} - \mathbf{s}_2\,\mathbf{s}_2^+)(\mathbf{I} - \mathbf{s}_2\,\mathbf{s}_2^+) = (\mathbf{I} - \mathbf{s}_2\,\mathbf{s}_2^+)$. Therefore, eqs (A-2) and (A-3) show that the sensitivity for analyte 1 can also be obtained from the complete matrix of pure constituent spectra $\mathbf{S}_{CLS}$. The (1,1) element of $\left(\mathbf{S}^T_{CLS}\mathbf{S}_{CLS}\right)^{-1}$ in eq (A-2) can be formally expressed in this alternative form:

$$\text{SEN}_1 = \left[\boldsymbol{\delta}_1^T\left(\mathbf{S}^T_{CLS}\mathbf{S}_{CLS}\right)^{-1}\boldsymbol{\delta}_1\right]^{-1/2} \tag{A-4}$$

where the following vector $\boldsymbol{\delta}_1$ is introduced:

$$\boldsymbol{\delta}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{A-5}$$

This latter vector helps to select, from the two constituents in the mixture, the analyte of interest (1 in this case). A final useful generalization can be reached as a function of the vector of regression coefficients for analyte 1 in this CLS model. The latter is easily shown to be given by the transposed first row of the pseudo-inverse $\mathbf{S}^{+}_{\mathrm{CLS}}$, i.e.:

$$\boldsymbol{\beta}_{\mathrm{CLS},1} = (\text{first row of } \mathbf{S}^{+}_{\mathrm{CLS}})^{\mathrm{T}} = \left[\left(\mathbf{I} - \mathbf{s}_2\mathbf{s}_2^{+}\right)\mathbf{s}_1\right]^{+\mathrm{T}} \tag{A-6}$$

Comparing eqs (A-6) and (A-2), the sensitivity can be given in terms of the regression coefficient vector as:

$$\mathrm{SEN}_1 = \left(\boldsymbol{\beta}^{\mathrm{T}}_{\mathrm{CLS},1}\,\boldsymbol{\beta}_{\mathrm{CLS},1}\right)^{-1/2} \tag{A-7}$$

All the above expressions can be generalized to *N*-constituent mixtures, with *n* being the index for the analyte of interest. The net analyte signal for the *n*th analyte at unit concentration is given as:

$$\mathbf{s}^{*}_{n} = \left(\mathbf{I} - \mathbf{S}_{-n}\mathbf{S}^{+}_{-n}\right)\mathbf{s}_n \tag{A-8}$$

where $\mathbf{S}_{-n}$ is the matrix of profiles for all sample constituents except *n*. There are two other interesting expressions in the present context. One is the CLS general expression derived from eq (A-4):

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

where $\boldsymbol{\delta}_n$ is an $N\times 1$ vector selecting the analyte of interest (i.e., having all values of 0, except a 1 at the analyte index). In eq (A-9), the matrix $\mathbf{S}_{\mathrm{CLS}}$ contains *N* columns, one for each analyte, containing the pure constituent profile $\mathbf{s}_n$ for each constituent.

The second useful generalization is the analogue of eq (A-7):

$$\mathrm{SEN}_n = \left(\boldsymbol{\beta}_{\mathrm{CLS},n}^{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{CLS},n}\right)^{-1/2} \tag{A-10}$$

where $\boldsymbol{\beta}_{\mathrm{CLS},n}$ is the vector of regression coefficients for analyte n provided by the CLS multivariate model. Equation (A-10) also applies to inverse and latent-based first-order calibration methodologies. For inverse least-squares (ILS), for example, $\boldsymbol{\beta}_{\mathrm{ILS},n} = \mathbf{X}_{\mathrm{cal}}^{+}\mathbf{y}_{\mathrm{cal},n}$ (where $\mathbf{X}_{\mathrm{cal}}$ is the matrix of calibration signals and $\mathbf{y}_{\mathrm{cal},n}$ the vector of calibration analyte concentrations),[37] and thus:

$$\mathrm{SEN}_n = \left[\mathbf{y}_{\mathrm{cal},n}^{\mathrm{T}}\left(\mathbf{X}_{\mathrm{cal}}^{\mathrm{T}}\mathbf{X}_{\mathrm{cal}}\right)^{-1}\mathbf{y}_{\mathrm{cal},n}\right]^{-1/2} \tag{A-11}$$

For principal component regression (PCR), on the other hand, $\boldsymbol{\beta}_{\mathrm{PCR},n} = \mathbf{P}_{\mathrm{PCR}}{}^{+\mathrm{T}}\,\mathbf{v}_{\mathrm{PCR},n}$ (where $\mathbf{P}_{\mathrm{PCR}}$ is the matrix of calibration loadings and $\mathbf{v}_{\mathrm{PCR},n}$ the vector of regression coefficients for the analyte in latent space),[37] and:

$$\mathrm{SEN}_n = \left[\mathbf{v}_{\mathrm{PCR},n}^{\mathrm{T}}\left(\mathbf{P}_{\mathrm{PCR}}^{\mathrm{T}}\mathbf{P}_{\mathrm{PCR}}\right)^{-1}\mathbf{v}_{\mathrm{PCR},n}\right]^{-1/2} \tag{A-12}$$

Finally, in partial least-squares (PLS), the corresponding expression can be brought into a form compatible with PCR:

$$\mathrm{SEN}_n = \left[\mathbf{q}_{\mathrm{PLS},n}^{\mathrm{T}}\left(\mathbf{W}_{\mathrm{PLS}}^{\mathrm{T}}\mathbf{W}_{\mathrm{PLS}}\right)^{-1}\mathbf{q}_{\mathrm{PLS},n}\right]^{-1/2};\ \mathbf{q}_{\mathrm{PLS}n} = \left(\mathbf{P}_{\mathrm{PLS}}^{\mathrm{T}}\mathbf{W}_{\mathrm{PLS}}\right)^{-1}\mathbf{v}_{\mathrm{PLS}n}^{\mathrm{T}} \tag{A-13}$$

where $\mathbf{W}_{\mathrm{PLS}}$ and $\mathbf{P}_{\mathrm{PLS}}$ are the PLS weight loading and loading matrix respectively, and $\mathbf{v}_{\mathrm{PLS}}$ is the vector of PLS regression coefficients in latent space.

Underlying all the above expressions is the idea that the sensitivity is the length of the NAS vector at unit concentration. In terms of the vector of regression coefficients for any first-order calibration model, the following equations apply:[73]

$$\mathrm{SEN}_n = \| \mathbf{s}_n^* \| = \frac{1}{\| \boldsymbol{\beta}_n \|} \tag{A-14}$$

$$\mathbf{s}_n^* = \frac{\boldsymbol{\beta}_n}{\| \boldsymbol{\beta}_n \|^2} \tag{A-15}$$

A-2. Multi-way sensitivity

The originally derived PARAFAC sensitivity expression is:[40]

$$\text{SEN}_n = m_n \, \| \, n\text{th row of } [(\mathbf{I} - \mathbf{Z}_{\text{unx}} \mathbf{Z}_{\text{unx}}^{+}) \, \mathbf{Z}_{\text{exp}}]^{+} \, \|^{-1} \tag{A-16}$$

In the latter equation, the relevant matrix can also be written as $[\mathbf{Z}_{\text{exp}} - \mathbf{Z}_{\text{unx}} (\mathbf{Z}_{\text{unx}}^{+} \mathbf{Z}_{\text{exp}})]^{+}$ which is better from the computational standpoint, since often $(\mathbf{I} - \mathbf{Z}_{\text{unx}} \mathbf{Z}_{\text{unx}}^{+})$ is very large and may consume the available computer memory.

This latter expression can be easily shown to be identical to the general equation (18) by noting that: (1) $\| \mathbf{x} \|$ represents the length of vector $\mathbf{x}$, and can also be given by $(\mathbf{x}^{\text{T}} \mathbf{x})^{1/2}$, and (2) the nth row of a matrix can be selected through multiplication by the vector $\boldsymbol{\delta}_n$:

$$\| \, n\text{th row of } [(\mathbf{I} - \mathbf{Z}_{\text{unx}} \mathbf{Z}_{\text{unx}}^{+}) \, \mathbf{Z}_{\text{exp}}]^{+} \, \| =$$

$$= \{\boldsymbol{\delta}_n^{\text{T}} \, [\mathbf{Z}_{\text{exp}}^{\text{T}} \, (\mathbf{I} - \mathbf{Z}_{\text{unx}} \mathbf{Z}_{\text{unx}}^{+})^{\text{T}} \, (\mathbf{I} - \mathbf{Z}_{\text{unx}} \mathbf{Z}_{\text{unx}}^{+}) \, \mathbf{Z}_{\text{exp}}]^{-1} \, \boldsymbol{\delta}_n\}^{1/2} \tag{A-17}$$

Since $(\mathbf{I} - \mathbf{Z}_{\text{unx}} \mathbf{Z}_{\text{unx}}^{+})$ is symmetric and idempotent, eqs (A-16) and (A-17) lead to the result shown in Table 3:

$$\text{SEN}_n = m_n \, \{\boldsymbol{\delta}_n^{\text{T}} \, [\mathbf{Z}_{\text{exp}}^{\text{T}} \, (\mathbf{I} - \mathbf{Z}_{\text{unx}} \mathbf{Z}_{\text{unx}}^{+}) \, \mathbf{Z}_{\text{exp}}]^{-1} \, \boldsymbol{\delta}_n\}^{-1/2} \tag{A-18}$$

In the case of MCR-ALS, the original expression is:[41]

$$\text{SEN}_n = m_n \, [J \, (\mathbf{C}^{\text{T}} \mathbf{C})_{nn}^{-1}]^{-1/2} \tag{A-19}$$

In eq (A-19), J is the number of data points in each sub-matrix 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 column-wise. In the case of row-wise augmentation, J should be replaced by K in eq (A-19). On the other hand, the matrix $\mathbf{C}$ contains the profiles for all sample constituents in the

non-augmented data mode, and the shorthand notation $(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}_{nn}$ implies selecting the (n,n) diagonal element of the inverse of matrix $(\mathbf{C}^{\mathrm{T}}\mathbf{C})$. To adapt eq (A-19) to the present approach, the matrix $\mathbf{C}$ is divided in two blocks, one for the constituents present in calibration ($\mathbf{C}_{\mathrm{exp}}$) and another one for the unexpected constituents ($\mathbf{C}_{\mathrm{unx}}$):

$$\mathbf{C} = [\ \mathbf{C}_{\mathrm{exp}} \mid \mathbf{C}_{\mathrm{unx}}\ ] \tag{A-20}$$

It can further be shown that:

$$(\mathbf{C}^{\mathrm{T}}\ \mathbf{C})^{-1} = ([\ \mathbf{C}_{\mathrm{exp}} \mid \mathbf{C}_{\mathrm{unx}}\ ]^{\mathrm{T}}\ [\ \mathbf{C}_{\mathrm{exp}} \mid \mathbf{C}_{\mathrm{unx}}\ ])^{-1} = [\mathbf{C}_{\mathrm{exp}}{}^{\mathrm{T}}\ (\mathbf{I} - \mathbf{C}_{\mathrm{unx}}\ \mathbf{C}_{\mathrm{unx}}{}^{+})\ \mathbf{C}_{\mathrm{exp}}]^{-1} \tag{A-21}$$

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

$$(\mathbf{C}^{\mathrm{T}}\mathbf{C})^{-1}_{nn} = \boldsymbol{\delta}_n{}^{\mathrm{T}}\ [\mathbf{C}_{\mathrm{exp}}{}^{\mathrm{T}}\ (\mathbf{I} - \mathbf{C}_{\mathrm{unx}}\ \mathbf{C}_{\mathrm{unx}}{}^{+})\ \mathbf{C}_{\mathrm{exp}}]^{-1}\ \boldsymbol{\delta}_n \tag{A-22}$$

and finally the sensitivity is given by the equation shown in Table 3:

$$\mathrm{SEN}_n = \frac{m_n}{J^{1/2}} \left\{ \boldsymbol{\delta}_n^{\mathrm{T}} \left[\mathbf{C}_{\mathrm{exp}}^{\mathrm{T}} \left(\mathbf{I} - \mathbf{C}_{\mathrm{unx}} \mathbf{C}_{\mathrm{unx}}^{+} \right) \mathbf{C}_{\mathrm{exp}} \right]^{-1} \boldsymbol{\delta}_n \right\}^{-1/2} \tag{A-23}$$

REFERENCES

(1) Currie, L. A. *Pure Appl. Chem*. **1995**, *67*, 1699.

(2) ISO 5725:1996, *Accuracy (trueness and precision) of measurement methods and results*; International Organization of Standardization: Geneva, 1996.

(3) Document N° 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) Cuadros Rodríguez, L.; García Campaña, A. M.; Jiménez Linares, C.; Román Ceba, M. *Anal. Lett*. **1993**, *26*, 1243.

(5) Vessman, J.; Stefan, R. I.; van Staden, J. F.; Danzer, K.; Lindner, W.; Burns, D. T.; Fajgelj, A.; Müller, H. *Pure Appl. Chem*. **2001**, *73*, 1381.

(6) Faber, N. M.; Boqué, R. *Accredit. Qual. Assur.* **2006**, *11*, 536.

(7) Faber, N. M.; Vandeginste, B. G. M. *Accredit. Qual. Assur.* **2010**, *15*, 373.

(8) Gibbons, R. D.; Coleman, D. E. *Statistical methods for detection and quantification of environmental contamination*; John Wiley & Sons: New York, 2001.

(9) Danzer, K.; Currie, L. A. *Pure Appl. Chem*. **1998**, *70*, 993.

(10) Danzer, K.; Otto, M.; Currie, L. A. *Pure Appl. Chem*. **2004**, *76*, 1215.

(11) Olivieri, A. C.; Faber, N. M.; Ferré, J.; Boqué, R.; Kalivas, J. H.; Mark, H. *Pure Appl. Chem.* **2006**, *78*, 633.

(12) Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst*. **2001**, *58*, 109.

(13) Massart, D. L.; Vandeginste, B. G. M.; Buydens, L. M. C.; De Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of chemometrics and qualimetrics*; Elsevier: Amsterdam, 1997.

(14) Martens, H.; Næs, T. *Multivariate Calibration*, John Wiley, Chichester, 1989.

(15) Burns, D. A.; Ciurczak, E. W. *Handbook of near-infrared analysis*; 3rd Ed., *Practical Spectroscopy Series*; CRC Press: Boca Raton, USA, 2008; Vol. 35.

(16) Olivieri, A. C.; Faber, N. M. *Validation and error*; in Brown, S.; Tauler, R.; Walczak, B. Editors, *Comprehensive Chemometrics*; Elsevier: Amsterdam, 2009; Vol. 3, p. 91.

(17) Lorber, A. *Anal. Chem*. **1986**, *58*, 1167.

(18) Bergmann, G.; von Oepen, B.; Zinn, P. *Anal. Chem.* **1987**, *59*, 2522.

(19) Lorber, A.; Faber, K.; Kowalski, B. R. *Anal. Chem*. **1997**, *69*, 1620.

(20) Sanchez, E.; Kowalski, B. R. *J. Chemometr*. **1988**, *2*, 265.

(21) Smilde, A.; Bro, R.; Geladi, P. *Multi-way analysis: applications in the chemical sciences*; Wiley: Chichester, 2004.

(22) Olivieri, A. C. *Anal. Meth*. **2012**, *4*, 1876.

(23) Gómez, V.; Callao, M. P. *Anal. Chim. Acta* **2008**, *627*, 169.

(24) Bro, R. *Crit. Rev. Anal. Chem*. **2006**, *36*, 279.

(25) Amigo, J. M.; Skov, T.; Bro, R. *Chem. Rev*. **2010**, *110*, 4582.

(26) Faber, N. M.; Goicoechea, H. C.; Muñoz de la Peña, A.; Olivieri, A. C.; Poppi, R. J. *Trends Anal. Chem*. **2007**, *26*, 752.

(27) Skov, T.; Bro, R. *Anal. Bioanal. Chem*. **2008**, *390*, 281.

(28) Ortiz, M. C.; Sarabia, L. *J. Chromatogr. A* **2007**, *1158*, 94.

(29) Arancibia, J. A.; Damiani, P. C.; Escandar, G. M.; Ibañez, G. A.; Olivieri, A. C. *J. Chromatogr. B* **2012**, *910*, 22.

(30) Escandar, G. M.; Goicoechea, H. C.; Muñoz de la Peña, A.; Olivieri, A. C. *Anal. Chim. Acta* **2014**, *806*, 8.

(31) Wu, H. L.; Li, Y.; Yu, R. Q. *J. Chemometr.* (in press, DOI 10.1002/cem.2570).

(32) Booksh, K. S.; Kowalski, B. R. *Anal. Chem*. **1994**, 66, 782A.

(33) Ho, C. -N.; Christian, G. D.; Davidson, E. R. *Anal. Chem*. **1980**, *52*, 1071.

(34) Messick, N. J.; Kalivas, J. H.; Lang, P. M. *Anal. Chem*. **1996**, *68*, 1572.

(35) Olivieri, A. C.; Faber, N. M. *Chemom. Intell. Lab. Syst*. **2004**, *70*, 75.

(36) Olivieri, A. C.; Faber, N. M. *J. Chemometr.* **2005**, *19*, 583.

(37) Faber, K.; Lorber, A.; Kowalski, B. R. *J. Chemometr.* **1997**, *11*, 419.

(38) Olivieri, A. C. *Anal. Chem*. **2005**, *77*, 4936.

(39) Olivieri, A. C. *Anal. Chem*. **2008**, *80*, 5713.

(40) Olivieri, A. C.; Faber, N. M. *Anal. Chem.* **2012**, *84*, 186.

(41) Bauza, C.; Ibañez, G. A.; Tauler, R.; Olivieri, A. C. *Anal. Chem*. **2012**, *84*, 8697.

(42) Allegrini, F.; Olivieri, A. C. *Anal. Chem*. **2012**, *84*, 10823.

(43) Van der Linden, W. E. *Pure Appl. Chem*. **1989**, *61*, 91.

(44) Olivieri, A. C.; Arancibia, J. A.; Muñoz de la Peña, A.; Durán-Merás, I.; Espinosa Mansilla, A. *Anal. Chem*. **2004**, *76*, 5657.

(45) Bailey, H. P.; Rutan, S. C. *Chemom. Intell. Lab. Syst*. **2011**, *106*, 131.

(46) Parastar, H.; Radovic, J. R.; Jalali-Heravi, M.; Diez, S.; Bayona, J. M.; Tauler, R. *Anal. Chem*. **2011**, *83*, 9289.

(47) Ho, C. -N.; Christian, G. D.; Davidson, E. R. *Anal. Chem*. **1978**, *50*, 1108.

(48) Zampronio, C. G.; Gurden, S. P.; Moraes, L. A.; Eberlin, M. N.; Smilde, A. K.; Poppi, R. J. *Analyst* **2002**, *127*, 1054.

(49) Wilson B. E.; Lindberg W.; Kowalski B. R. *J. Am. Chem. Soc*. **1989**, *111*, 3797.

(50) Calimag-Williams, K.; Knobel, G.; Goicoechea, H. C.; Campiglia, A. D. *Anal. Chim. Acta* **2014**, *811*, 60.

(51) Lakowicz, J. R. *Principles of fluorescence spectroscopy*; 3rd. Ed., Springer: Berlin, 2006.

(52) de Juan, A; Tauler, R. *J. Chemometr.* **2001**, *15*, 749.

(53) Piccirilli, G. N.; Escandar, G. M. *Analyst* **2010**, *135*, 1299.

(54) Bro, R. *Chemom. Intell. Lab. Syst*. **1997**, *38*, 149.

(55) Chen, Z. P.; Wu, H. L.; Jiang, J. H.; Li, Y.; Yu, R. Q. *Chemom. Intell. Lab. Syst*. **2000**, *52*, 75.

(56) Xia, A. L.; Wu, H. L.; Fang, D. M.; Ding, Y. J.; Hu, L. Q.; Yu, R. Q. *J. Chemometr.* **2005**, *19*, 65.

(57) Xia, A. L.; Wu, H. L.; Li, S. F.; Zhu, S. H.; Hu, L. Q.; Yu, R. Q. *J. Chemometr.* **2007**, *21*, 133.

(58) Fu, H. Y.; Wu, H. L.; Yu, Y. J.; Yu, L. L.; Zhang, S. R.; Nie, J. F.; Li, S. F.; Yu, R. Q. *J. Chemometr.* **2011**, *25*, 408.

(59) Tauler, R. *Chemom. Intell. Lab. Syst*. **1995**, *30*, 133.

(60) Tauler, R.; Maeder, M.; de Juan, A. *Multiset Data Analysis: Extended Multivariate Curve Resolution*; in Brown, S.; Tauler, R.; Walczak, B., Editors, *Comprehensive Chemometrics*; Elsevier: Amsterdam, 2009; Vol. 2, p. 473.

(61) Wold, S.; Geladi, P.; Esbensen, K.; Øhman, J. *J. Chemometr.* **1987**, *1*, 41.

(62) Bro, R. *J. Chemometr.* **1996**, *10*, 47.

(63) Linder, M.; Sundberg, R. *J. Chemometr.* **2002**, *16*, 12.

(64) Arancibia, J. A.; Olivieri, A. C.; Bohoyo Gil, D.; Espinosa Mansilla, A.; Durán Merás, I.; Muñoz de la Peña, A. *Chemom. Intell. Lab. Syst.* **2006**, *80*, 77.

(65) Sanchez, E.; Kowalski, B. R. *Anal. Chem*. **1986**, *58*, 496.

(66) Sanchez, E.; Kowalski, B. R. *J. Chemometr*. **1990**, *4*, 29.

(67) Öhman, J.; Geladi, P.; Wold, S. *J. Chemometr.* **1990**, *4*, 79.

(68) Olivieri, A. C. *J. Chemometr.* **2005**, *19*, 253.

(69) Maggio, R. M.; Muñoz de la Peña, A.; Olivieri, A. C.; *Chemom. Intell. Lab. Syst*. **2011**, *109*, 178.

(70) Bloemberg, T. G.; Gerretzen, J.; Lunshof, A.; Wehrens, R.; Buydens, L. M. C. *Anal. Chim. Acta* **2013**, *781*, 14 and references therein.

(71) Kiers, H. A. L.; Ten Berge, J. M. F.; Bro, R. *J. Chemometr*. **1999**, *13*, 275.

(72) Faber, N. M. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 107.

(73) Ferré, J.; Brown, S. D.; Rius, F. X. *J. Chemometr.* **2001**, *15,* 537.

(74) Ferré, J.; Faber, N. M. *Chemom. Intell. Lab. Syst.* **2003**, *69*, 123.

(75) Rao, C. R.; Mitra, S. *Generalized Inverse of Matrices and its Applications*; Wiley: New York, 1971.

(76) Faber, K.; Kowalski, B. R. *J. Chemometr.* **1997**, *11*, 181.

(77) Horn, R. A.; Johnson, C. R. *Topics in matrix analysis*; Cambridge University Press: Cambridge, UK, 1991.

(78) Saltelli, A.; Ratto, M.; Tarantola, S.; Campolongo, F. *Chem. Rev*. **2005**, *105*, 2811.

(79) Faber, N. M.; Ferré, J.; Boqué, R.; Kalivas, J. H. *Chemom. Intell. Lab. Syst*. **2002**, *63*, 107.

(80) Faber, N. M.; Ferré, J.; Boqué, R.; Kalivas, J. H. *Trends Anal. Chem.* **2003**, *22*, 352.

(81) Cantwell, M. T.; Porter, S. E. G.; Rutan, S. C. *J. Chemometr.* **2007**, *21*, 335.

(82) Arnold, M. A.; Small, G. W.; Xiang, G.; Qui, J.; Murhammer, D. W. *Anal. Chem.* **2004**, *76*, 2583.

(83) Brown, C. D.; Ridder, T. D. *Appl. Spectrosc.* **2005**, *59*, 787.

(84) Ridder, T. D.; Brown, C. D.; Ver Steeg, B. J. *Appl. Spectrosc.* **2005**, *59*, 804.

(85) Thompson, M.; Ellison, S. L. R.; Wood, R. *Pure Appl. Chem*. **2002**, *74*, 835.

(86) Umezawa, Y.; Bühlmann, P.; Umezawa, K.; Tohda, K.; Amemiya, S. *Pure Appl. Chem*. **2000**, *72*, 1851.

(87) Umezawa, Y.; Umezawa, K.; Bühlmann, P.; Hamada, N.; Aoki, H.; Nakanishi, J.; Sato, M.; Xiao, K. P.; Nishimura, Y. *Pure Appl. Chem*. **2002**, *74*, 923.

(88) Umezawa, Y.; Bühlmann, P.; Umezawa, K.; Hamada, N. *Pure Appl. Chem*. **2002**, *74*, 995.

(89) De Bièvre, P. *Accred. Qual. Assur.* **1997**, *2*, 269.

(90) Geladi, P. *Chemom. Intell. Lab. Syst*. **2002**, *60*, 211.

(91) Riu, J.; Bro, R. *Chemom. Intell. Lab. Syst*. **2003**, *65*, 35.

(92) Serneels, S.; Faber, K.; Verdonck, T.; Van Espen, P. J. *Chemom. Intell. Lab. Syst.* **2011**, *108,* 93.

(93) Cabezón, M.; Olivieri, A. C. *Chem. Educator* **2006**, *11*, 394.

(94) Currie, L. A. *Anal. Chem*. **1968**, *40*, 586-593.

(95) Currie, L. A. *Anal.Chim. Acta* **1999**, *391*, 127.

(96) Hubaux, A.; Vos, G. *Anal. Chem.* **1970**, *42***,** 849.

(97) Clayton, C. A.; Hines, J. W.; Elkins, P. D. *Anal. Chem.* **1987**, *59*, 2506.

(98) Faber, N. M. *Accred. Qual. Assur.* **2008**, *13*, 277.

(99) Boqué, R.; Larrechi, M. S.; Rius F. X. *Chemom. Intell. Lab. Syst*. **1999**, *45*, 397.

(100) Boqué, R.; Ferré, J.; Faber, N. M.; Rius, F. X. *Anal. Chim. Acta* **2002**, *451*, 313.

(101) del Río Bocio, F. J.; Riu, J.; Boqué, R.; Rius, F. X. *J. Chemometr*. **2003**, *17*, 413.

(102) Voigtman, E. *Spectrochim. Acta* **2008***, 63*, 115.

(103) Voigtman, E. *Spectrochim. Acta* **2008***, 63*, 129.

(104) Saurina, J.; Leal, C.; Compañó, R.; Granados, M.; Dolors Prat, M.; Tauler, R. *Anal. Chim. Acta* **2001**, *432*, 241.

(105) Rodríguez-Cuesta, M. J.; Boqué, R.; Rius, F. X. *Anal. Chim. Acta* **2003**, *491*, 47.

(106) Rodríguez-Cuesta, M. J.; Boqué, R.; Rius, F. X.; Martínez Vidal, J. L.; Garrido Frenich, A. *Chemom. Intell. Lab. Syst*. **2005**, *77*, 251.

(107) Ortiz, M. C.; Sarabia, L. A.; Herrero, A.; Sánchez, M. S.; Sanza, M. B.; Rueda, M. E.; Giménez, D.; Meléndez, M. E. *Chemom. Intell. Lab. Syst*. **2003**, *69*, 21.

(108) Ortiz, M. C.; Sarabia, L. A.; Sánchez, M. S. *Anal. Chim. Acta* **2010**, *674*, 123.

(109) Liu, X.; Sidiropoulos, S. D. *IEEE Trans. Signal Processing* **2001**, *49*, 2074.

(110) Muñoz de la Peña, A.; Espinosa Mansilla, A.; González Gómez, D.; Olivieri, A. C.; Goicoechea, H. C. *Anal. Chem*. **2003**, *75*, 2640.

(111) Bortolato, S. A.; Arancibia, J. A.; Escandar, G. M. *Anal. Chem.* **2009**, *81*, 8074.

FIGURE CAPTIONS

Figure 1: Illustration of the various arrays which can be built with data of different order. Left: focus is on the data order, i.e., the complexity of the data for a single sample (yellow objects identify specific samples from a set of samples). Right: focus is on the number of ways of the mathematical object built with data for a set of samples.

Figure 2: Three-dimensional plot of the excitation-emission fluorescence matrix measured for an aqueous solution of three fluorescent fluoroquinolone antibiotics: norfloxacin (5.28 μg L^{-1}), enoxacin (63.40 μg L^{-1}) and ofloxacin (16.90 μg L^{-1}), showing the presence of a diffraction grating harmonics (H) and both Rayleigh (Rh) and Raman (Rn) scatterings, as indicated. Reprinted with permission from Reference 110. Copyright 2003 American Chemical Society.

Figure 3: Three-dimensional plot of a typical chromatographic-fluorescence spectral matrix of a sample containing twelve polycyclic aromatic hydrocarbons at the following concentrations (all in ng mL^{-1}): fluoranthene, 500, pyrene, 500, chrysene, 300, benz[*a*]anthracene, 100, benzo[*b*]fluoranthene, 100, benzo[*k*]fluoranthene, 20, benzo[*a*]pyrene, 50, dibenz[*a*,*h*]anthracene, 50, indeno[1,2,3-*cd*]pyrene, 100, benzo[*g*,*h*,*i*]perylene, 50, benzo[*e*]pyrene, 300 and benzo[*j*]fluoranthene, 300. Adapted with permission from Reference 111. Copyright 2009 American Chemical Society.

Figure 4: Contour plot of a typical data matrix (EEM or LC-DAD measurements, with intensity levels growing from blue to red contours). The profiles in one data mode are indicated as $\mathbf{b}_1$ and

$\mathbf{b}_2$ for each sample constituent, and as $\mathbf{c}_1$ and $\mathbf{c}_2$ in the other mode. A generic element x_{jk} is shown to be obtained as sum of products of profile elements.

Figure 5: Illustration of the process of building an augmented data matrix from chromatographic-spectral matrix data for a set of samples. Left: by unfolding a three-way data array in the direction of elution time. Right: by augmenting the individual data matrices in the elution time direction.

Figure 6: A) Typical overlapped spectra for two sample constituents, as indicated. B) Profile for a mixture of both constituents at equal concentration, and individual contributions from each constituent.

Figure 7: Geometric illustration of the net analyte signal concept. The spectrum of a sample ($\mathbf{x}$) is projected orthogonal to the space spanned by the other sample constituent ($\mathbf{s}_2$), giving the net analyte signal ($\mathbf{x}_1^*$). The projection is carried out through the orthogonal projection matrix ($\mathbf{I} - \mathbf{s}_2 \mathbf{s}_2^+$).

Figure 8: A) Net analyte signal for the constituent No. 1 of Figure 6 (solid blue line) and its net analyte signal at unit concentration (dashed blue line). B) Pseudo-univariate representation of the first-order net analyte signal in scalar form (i.e., the length of the net analyte signal vector) as a function of analyte concentration. The slope is the sensitivity.

Figure 9: Two different versions of a second-order NAS. A) Contour plots for the matrix data of two analytes (blue and green) and a mixture of them (red). B) Surface plot of the HCD NAS. C) Surface plot of the MKL NAS. Projections of the mixture signals needed to obtain both NAS versions are indicated.

Figure 10: Uncertainty propagation analysis in an analytical calibration system, showing how the input and output noise can be used to define the sensitivity.

Figure 11: The official (IUPAC) definition of limit of detection (LOD), showing two Gaussian bands centered at the blank and the LOD, and the critical decision limit (CL) which helps to decide whether the analyte is detected or not. The shaded areas correspond to the rate of false detects (blue) and false non-detects (red).

Figure 12: Expansion of the shaded areas in Figure 12, corresponding to the false detects (blue) and false non-detects (red). Intuitive explanations regarding the analyte presence are indicated on the bottom of each concentration region.

Figure 13: A) Representative profiles for two sample constituents, as overlapped Gaussian lines defined in a range of 50 data points in a given instrumental mode (the solid line represents the analyte and the dashed line an interferent). Both profiles are normalized to unit length. B) Relative sensitivities towards the analyte, achieved by different algorithms when calibration is performed using data of different orders. Red bars correspond to first-order data, blue bars to second-order data, green bars to third-order data and black bars to fourth-order data. All

sensitivity bars are relative to the one for ILS taken as 1 (notice the logarithmic vertical scale). Specific algorithms are as follows: 1, ILS, 2, CLS, PCR and PLS, 3, 5 and 7, PARAFAC (three-, four- and five-way respectively) and 4, 6 and 8, MCR-ALS (in 6 and 8, multi-way data were unfolded into matrices). ILS was implemented on the ten most sensitive sensors, while all the remaining algorithms employed full sensor data (see text).

Figure 14: Comparison of selectivity values for multi-way calibration using MCR-ALS [eq (31), vertical axis] and PARAFAC [eq (30), horizontal axis]. In A), B) and C), two constituents are calibrated and no potential interferents occur in test samples. In D), E) and F), one constituent is calibrated and one is a potential interferent. Plots A) and D) correspond to three-way (second-order) data, B) and E) to four-way (third-order) data and C) and F) to five-way (fourth-order data). One hundred different situations are shown, corresponding to constituent profiles in all data modes represented by overlapped Gaussian functions. The gray triangles indicate the regions where the selectivity is larger for PARAFAC than for MCR-ALS.

Figure 15: Changes in detection limit estimated from eq (35), as a function of sensitivity, and for different relative values of the signal uncertainty (σ_x) and the concentration uncertainty (σ_{ycal}). The following parameters have been used in eq (35): $h_0 = 0.5$, $\sigma_{ycal} = 1$ unit, and $\sigma_x = 1$, 2 3, 4 and 5 units, as indicated.

AUTHOR INFORMATION

Alejandro C. Olivieri was born in Rosario, Argentina, on July 28, 1958. He obtained his B.Sc. in Industrial Chemistry from the Catholic Faculty of Chemistry and Engineering in 1982, and his Ph.D. from the Faculty of Biochemical and Pharmaceutical Sciences, University of Rosario in 1986. He is a fellow of the National Research Council of Argentina (CONICET) and currently works in the Rosario Institute of Chemistry (IQUIR), Department of Analytical Chemistry, University of Rosario. He has published about 200 scientific papers in international journals, several books and book chapters and supervised nine Ph.D. Theses. He was John Simon Guggenheim Memorial Foundation fellow (2001-2002). In 2013 he received the Platinum Konex prize (Konex Foundation, Argentina) for his contributions to analytical chemistry in the last decade.

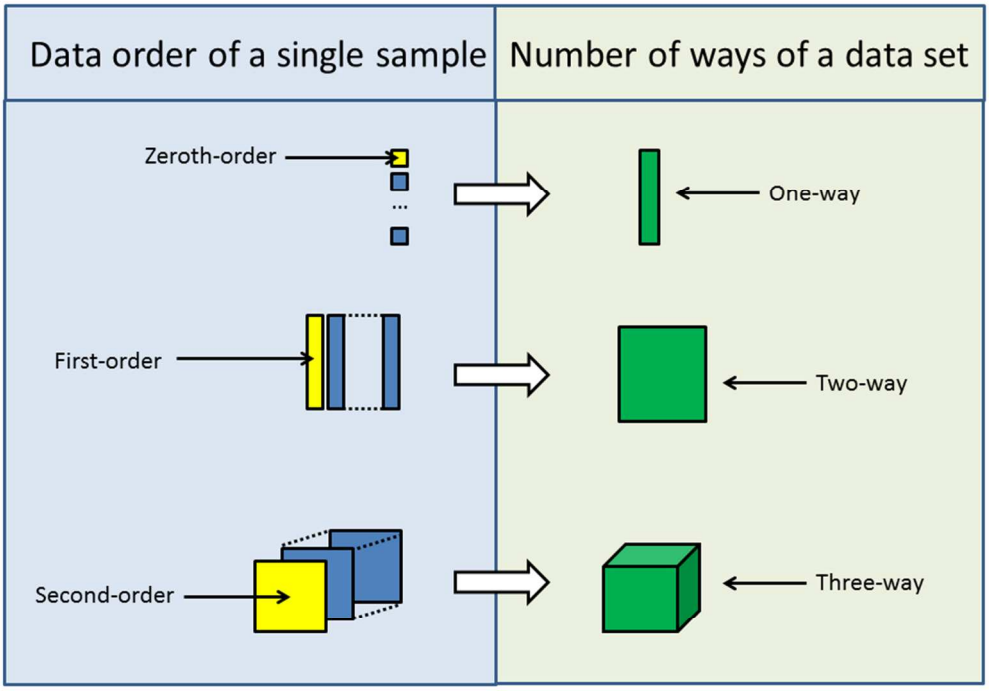

Figure 1. Illustration of the various arrays which can be built with data of different order. Left: focus is on the data order, i.e., the complexity of the data for a single sample (yellow objects identify specific samples from a set of samples). Right: focus is on the number of ways of the mathematical object built with data for a set of samples.
254x190mm (96 x 96 DPI)

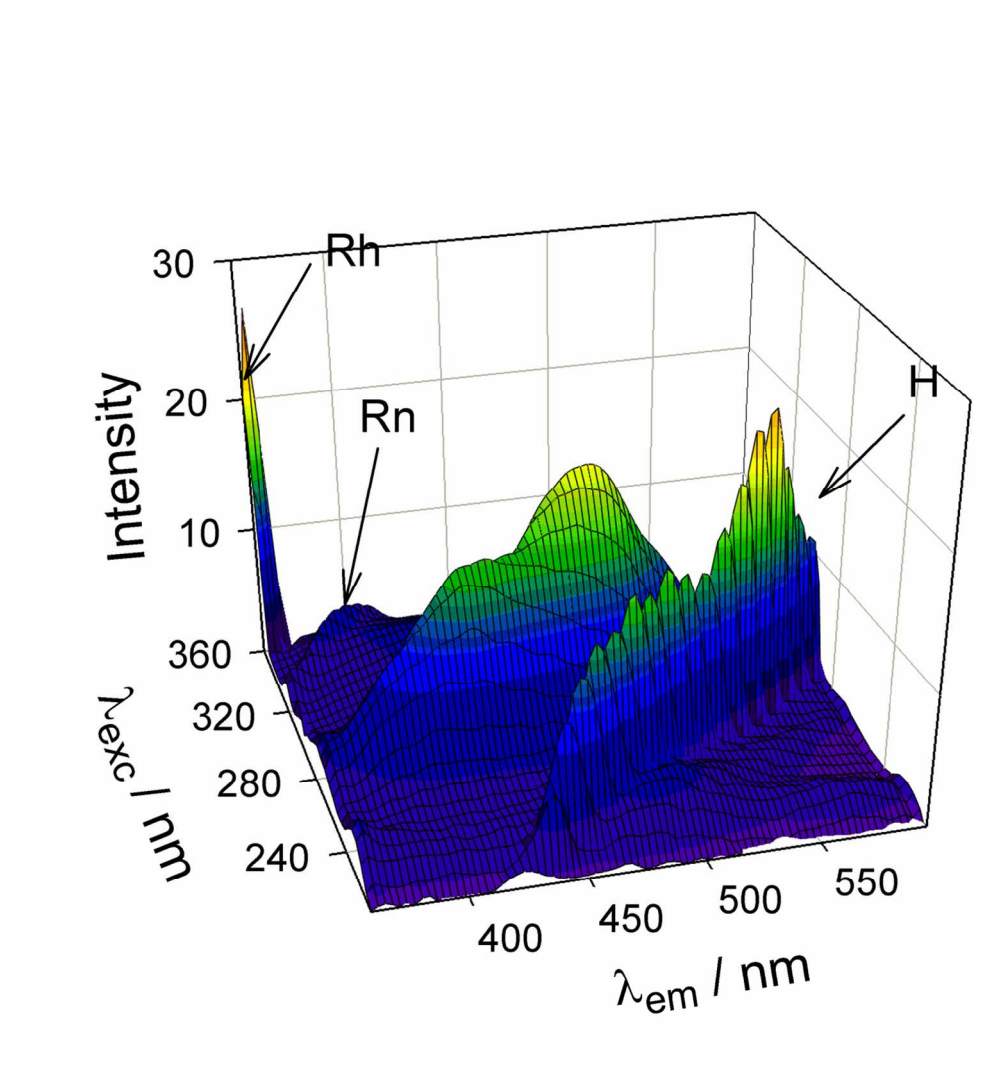

Figure 2. Three-dimensional plot of the excitation-emission fluorescence matrix measured for an aqueous solution of three fluorescent fluoroquinolone antibiotics: norfloxacin (5.28 μg L–1), enoxacin (63.40 μg L–1) and ofloxacin (16.90 μL–1), showing the presence of a diffraction grating harmonics (H) and both Rayleigh (Rh) and Raman (Rn) scatterings, as indicated. Reprinted with permission from ref 108. Copyright 2003 ACS.
117x126mm (300 x 300 DPI)

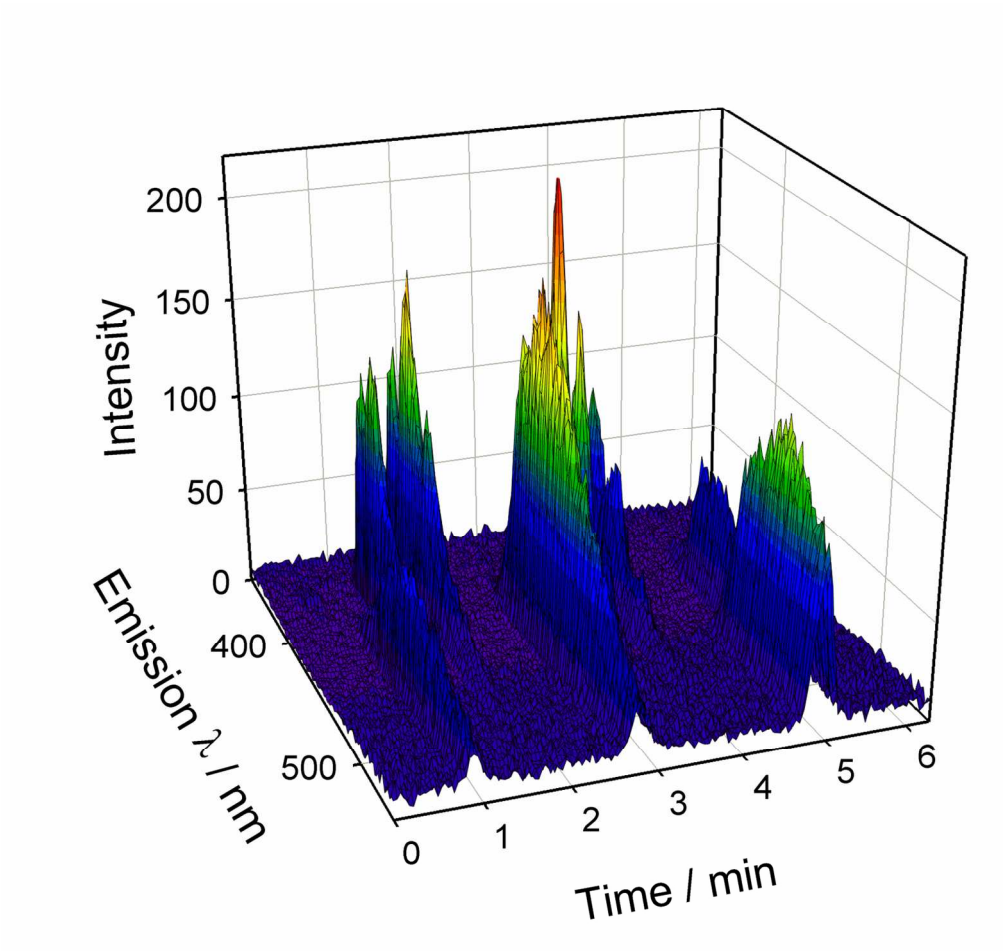

Figure 3. Three-dimensional plot of a typical chromatographic-fluorescence spectral matrix of a sample containing twelve polycyclic aromatic hydrocarbons at the following concentrations (all in ng mL−1): fluoranthene, 500, pyrene, 500, chrysene, 300, benz[a]anthracene, 100, benzo[b]fluoranthene, 100, benzo[k]fluoranthene, 20, benzo[a]pyrene, 50, dibenz[a,h]anthracene, 50, indeno[1,2,3-cd]pyrene, 100, benzo[g,h,i]perylene, 50, benzo[e]pyrene, 300 and benzo[j]fluoranthene, 300. Adapted with permission from Figure 1A of ref. 109. Copyright 2009 ACS.
139x134mm (300 x 300 DPI)

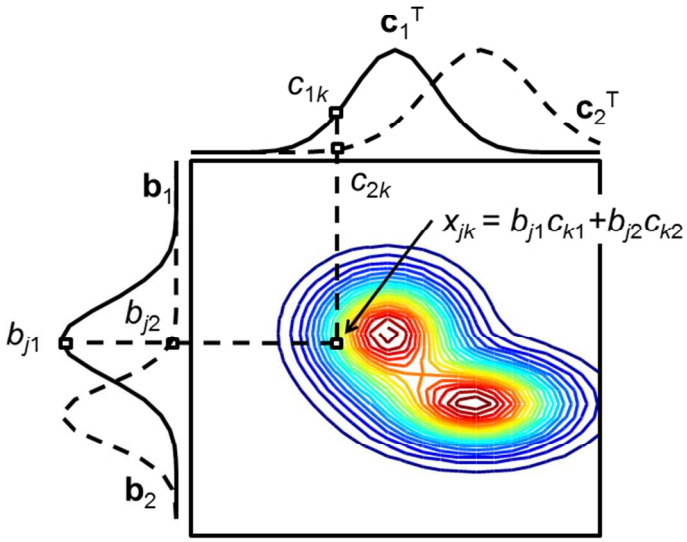

Figure 4. Contour plot of a typical data matrix (EEM or LC-DAD measurements, with intensity levels growing from blue to red contours). The profiles in one data mode are indicated as b1 and b2 for each sample constituent, and as c1 and c2 in the other mode. A generic element xjk is shown to be obtained as sum of products of profile elements.
254x190mm (96 x 96 DPI)

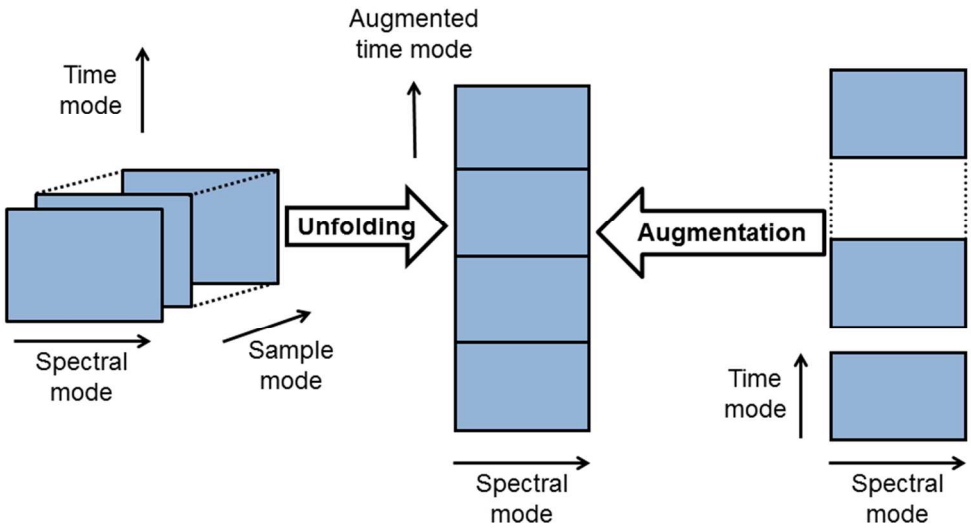

Figure 5. Illustration of the process of building an augmented data matrix from chromatographic-spectral matrix data for a set of samples. Left: by unfolding a three-way data array in the direction of elution time. Right: by augmenting the individual data matrices in the elution time direction.
254x190mm (96 x 96 DPI)

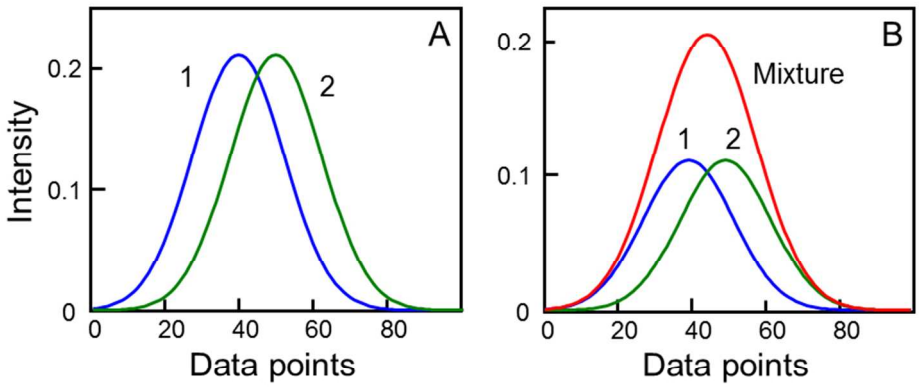

Figure 6. A) Typical overlapped spectra for two sample constituents, as indicated. B) Profile for a mixture of both constituents at equal concentration, and individual contributions from each constituent.
254x190mm (96 x 96 DPI)

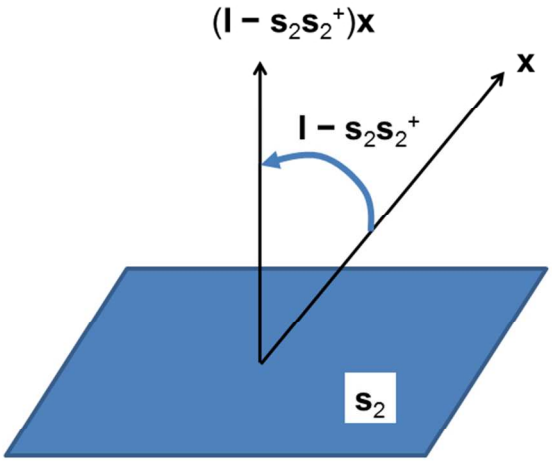

Figure 7. Geometric illustration of the net analyte signal concept. The spectrum of a sample (x) is projected orthogonal to the space spanned by the other sample constituent (s2), giving the net analyte signal (x1*). The projection is carried out through the orthogonal projection matrix (I – s2 s2+).
254x190mm (96 x 96 DPI)

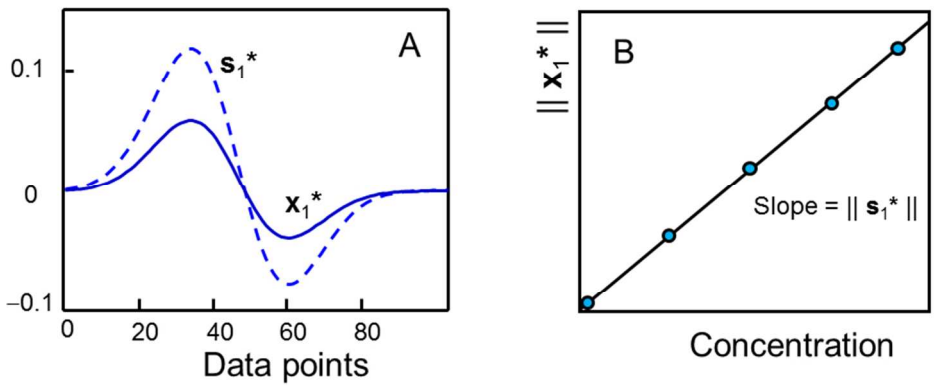

Figure 8. A) Net analyte signal for the constituent No. 1 of Figure 6 (solid blue line) and its net analyte signal at unit concentration (dashed blue line). B) Pseudo-univariate representation of the first-order net analyte signal in scalar form (i.e., the length of the net analyte signal vector) as a function of analyte concentration. The slope is the sensitivity.
254x190mm (96 x 96 DPI)

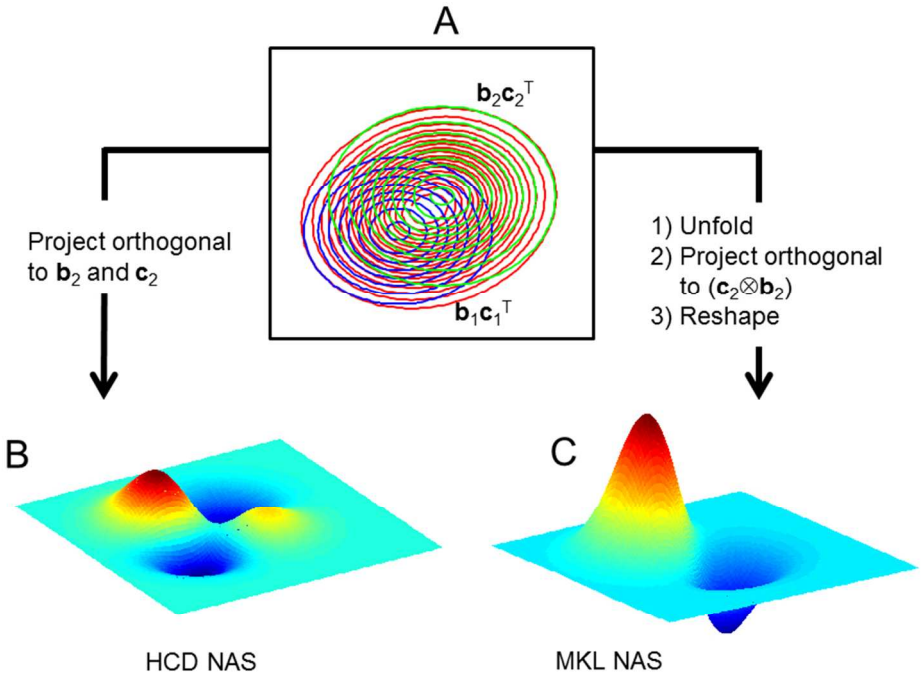

Figure 9. Two different versions of a second-order NAS. A) Contour plots for the matrix data of two analytes (blue and green) and a mixture of them (red). B) Surface plot of the HCD NAS. C) Surface plot of the MKL NAS. Projections of the mixture signals needed to obtain both NAS versions are indicated.
254x190mm (96 x 96 DPI)

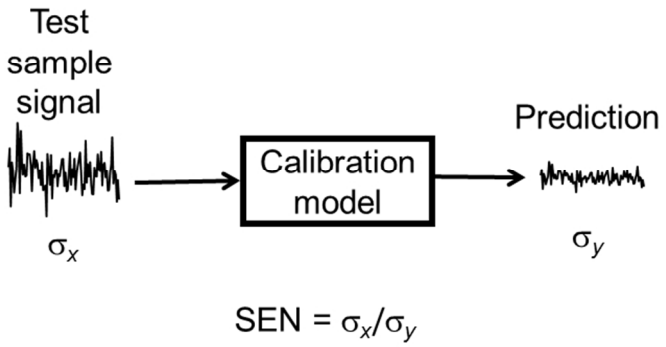

Figure 10. Uncertainty propagation analysis in an analytical calibration system, showing how the input and output noise can be used to define the sensitivity.
254x190mm (96 x 96 DPI)

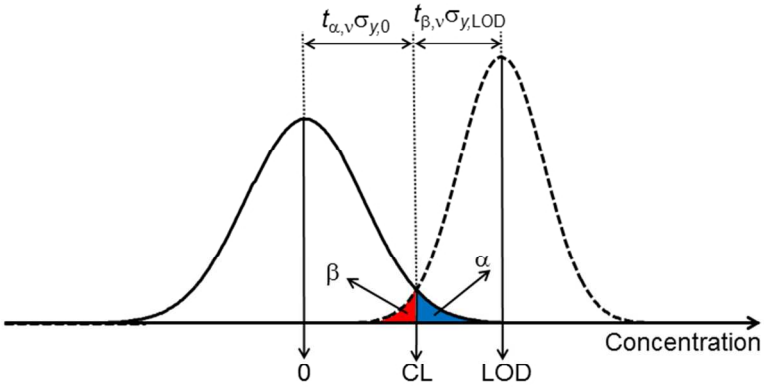

Figure 11. The official (IUPAC) definition of limit of detection (LOD), showing two Gaussian bands centered at the blank and the LOD, and the critical decision limit (CL) which helps to decide whether the analyte is detected or not. The shaded areas correspond to the rate of false detects (blue) and false non-detects (red). 254x190mm (96 x 96 DPI)

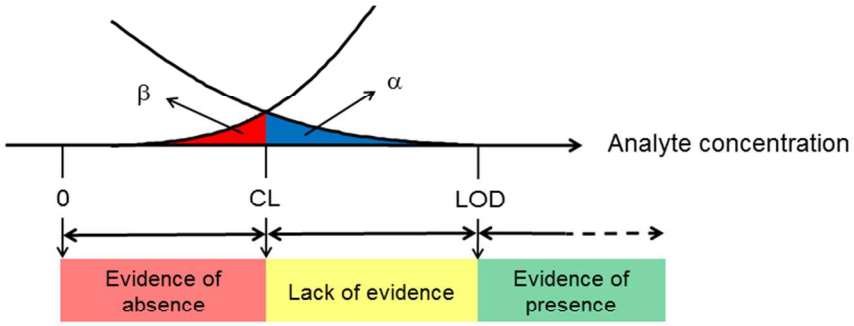

Figure 12. Expansion of the shaded areas in Figure 11, corresponding to the false detects (blue) and false non-detects (red). Intuitive explanations regarding the analyte presence are indicated on the bottom of each concentration region.
254x190mm (96 x 96 DPI)

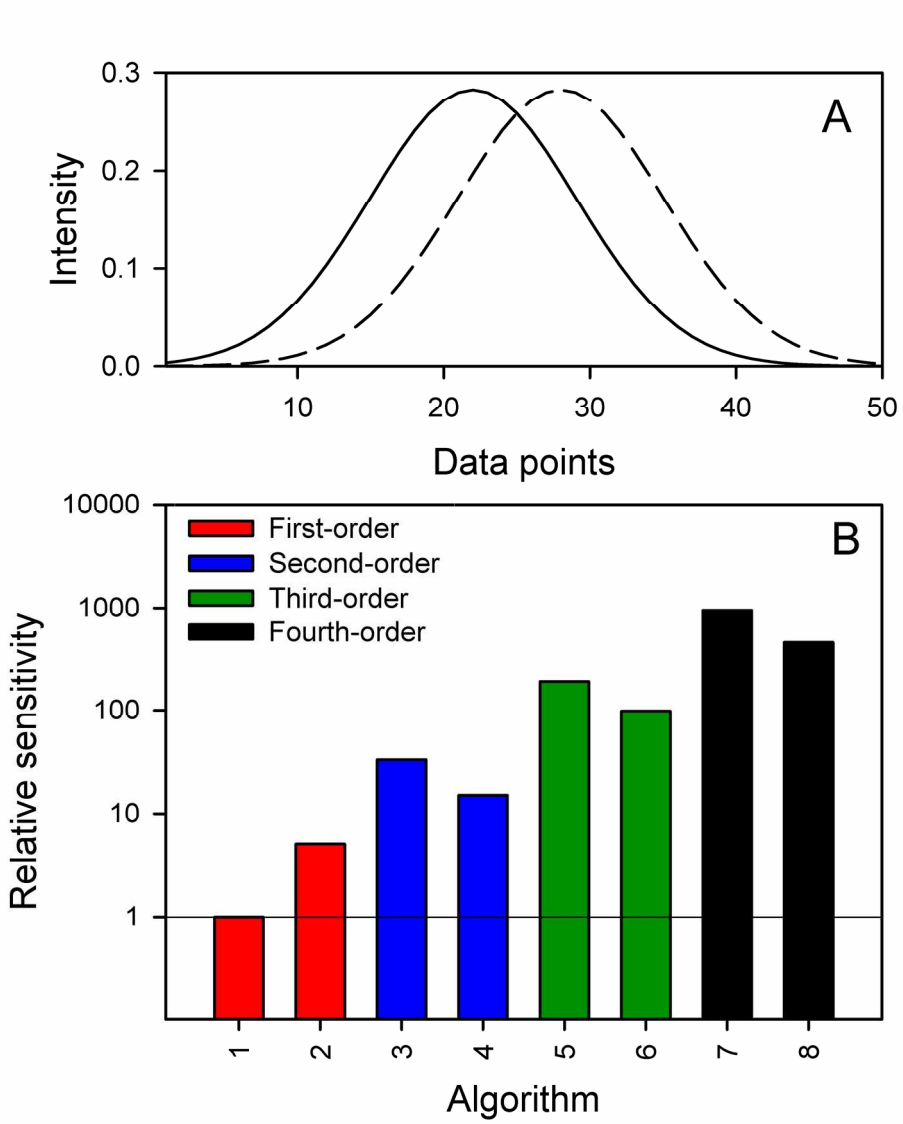

Figure 13: A) Representative profiles for two sample constituents, as overlapped Gaussian lines defined in a range of 50 data points in a given instrumental mode (the solid line represents the analyte and the dashed line an interferent). Both profiles are normalized to unit length. B) Relative sensitivities towards the analyte, achieved by different algorithms when calibration is performed using data of different orders. Red bars correspond to first-order data, blue bars to second-order data, green bars to third-order data and black bars to fourth-order data. All sensitivity bars are relative to the one for ILS taken as 1 (notice the logarithmic vertical scale). Specific algorithms are as follows: 1, ILS, 2, CLS, PCR and PLS, 3, 5 and 7, PARAFAC (three-, four- and five-way respectively) and 4, 6 and 8, MCR-ALS (in 6 and 8, multi-way data were unfolded into matrices). ILS was implemented on the ten most sensitive sensors, while all the remaining algorithms employed full sensor data (see text).
197x222mm (300 x 300 DPI)

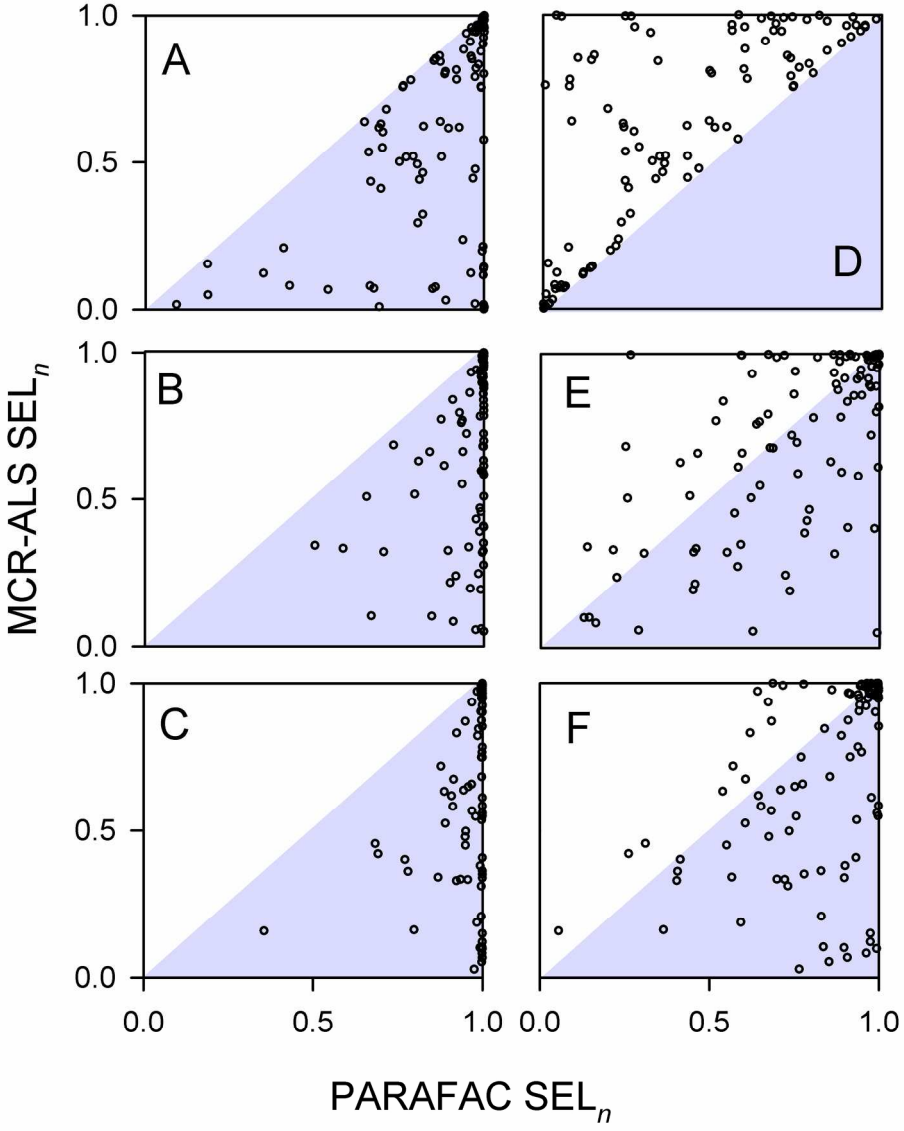

Figure 14: Comparison of selectivity values for multi-way calibration using MCR-ALS [eq (31), vertical axis] and PARAFAC [eq (30), horizontal axis]. In A), B) and C), two constituents are calibrated and no potential interferents occur in test samples. In D), E) and F), one constituent is calibrated and one is a potential interferent. Plots A) and D) correspond to three-way (second-order) data, B) and E) to four-way (third-order) data and C) and F) to five-way (fourth-order data). One hundred different situations are shown, corresponding to constituent profiles in all data modes represented by overlapped Gaussian functions. The gray triangles indicate the regions where the selectivity is larger for PARAFAC than for MCR-ALS.
196x253mm (300 x 300 DPI)

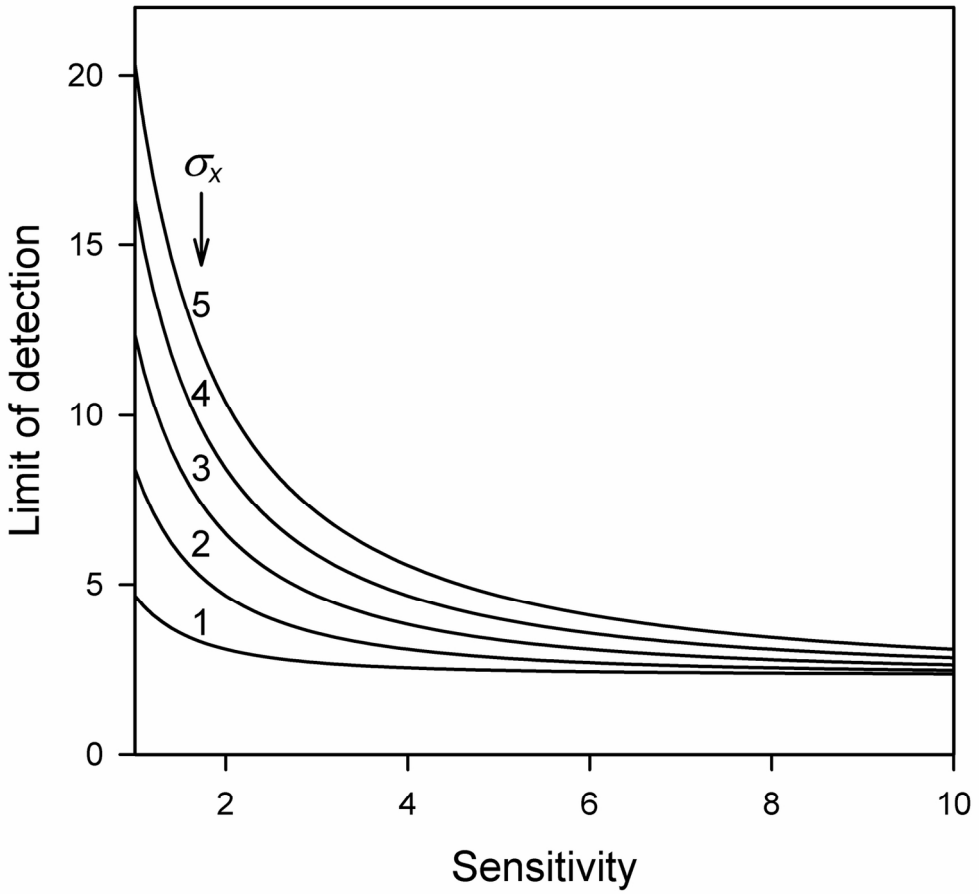

Figure 15: Changes in detection limit estimated from eq (35), as a function of sensitivity, and for different relative values of the signal uncertainty (σx) and the concentration uncertainty (σycal). The following parameters have been used in eq (35): h0 = 0.5, σycal = 1 unit, and σx = 1, 2 3, 4 and 5 units, as indicated.
148x143mm (300 x 300 DPI)