

MINIREVIEW

CrossMark
click for updatesCite this: *Anal. Methods*, 2017, 9, 739Received 24th October 2016
Accepted 22nd December 2016

DOI: 10.1039/c6ay02916g

www.rsc.org/methods

Recent advances in analytical figures of merit:
heteroscedasticity strikes back

Franco Allegrini and Alejandro C. Olivieri*

With a growing number of analytical methodologies incorporating multivariate calibration models, the development of suitable estimators of their analytical figures of merit is playing an important role in modern analytical chemistry. The main goal of this mini-review is to briefly present the recent trends regarding reliable and interpretable estimators of traditional figures of merit such as sensitivity, prediction uncertainty and detection limit. Heteroscedasticity in instrumental measurements is driving some of these recent developments and influencing the definition of new estimators.

Introduction

Since the first applications of multivariate analysis to chemical data, analytical chemistry has been one of the most benefited disciplines.¹ The major impact was on the development of quantitative first-order (*e.g.*, spectral) models in analytical calibration. As a clear proof, numerous analytical methodologies have been designed and optimized by the combination of spectroscopic data and multivariate algorithms. Such was the expansion that principal component regression (PCR) and

partial least-squares regression (PLS) are nowadays the battle horses of many quality control procedures, especially those which are based on near infrared spectroscopy (NIRS) measurements.² On the other hand, when the mathematical relation between the instrumental response and the target variable to be predicted is significantly nonlinear, the use of artificial neural networks (ANN) has become the main choice.³

The introduction of multivariate analytical models was accompanied by the immediate need for defining reliable estimators for analytical figures of merit (AFOMs), in analogy with univariate calibration. This opened a new field of study combining elements from metrology and chemometrics, guided by analytical chemistry criteria. AFOMs can be defined as statistical indicators giving a quantitative measure about the quality or performance of a particular analytical methodology.

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. E-mail: olivieri@iquir-conicet.gov.ar



Franco Allegrini was born in Venado Tuerto, Santa Fe, Argentina (10/30/1986). He obtained his B. Sc. (2011) and his Ph. D (2015) from the National University of Rosario. He is currently a Post-doc at the Rosario Institute of Chemistry (IQUIR) with a fellowship from the National Research Council (CONICET), under the supervision of Prof. Olivieri. He has published 9 papers in interna-

tional journals and a book chapter, including collaborations with different research groups around the world. His main research topic is chemometrics oriented to multivariate calibration in Analytical Chemistry.



Alejandro C. Olivieri was born in Rosario, Argentina (07/28/1958). He obtained his B.Sc. from the Catholic University (1982), and his Ph.D. from the National University of Rosario (1986). He did Postdoctoral Research at the University of Illinois, Urbana-Champaign, USA. In 1990, he returned to the University of Rosario, and joined the National Research Council (CONICET). He founded

a research group in chemometrics in analytical chemistry, publishing about 200 papers in international journals, books and book chapters, and supervised nine Ph.D. theses. He received several national and international awards, including the John Simon Guggenheim Memorial Foundation fellowship (2001–2002).

Although the root mean square error of prediction (RMSEP) and the relative error of prediction (REP) are widely employed to compare different methods, other figures of merit, such as the sensitivity and the sample-dependent prediction uncertainty are still ignored in many publications where multivariate calibration is applied.⁴ The effort to define suitable estimators for these figures is understandable in the light of the need to validate chemometric procedures.^{5,6} The most useful AFOMs for the analytical community are the sensitivity, the limit of detection and the prediction error.⁷ The latter can be calculated as a global parameter (*i.e.*, the root mean square error of prediction, RMSEP) or as a sample-specific standard error. In recent publications, Kalivas *et al.* showed that it is possible to combine AFOMs in order to assist parameter selection in models such as PLS and ridge regression.^{8,9}

Although sometimes neglected, an important assumption is that first-order regression methods are basically least-squares methodologies.¹⁰ This implies that the projections employed to represent the data in a latent variable space will work optimally only if the errors are independent and identically distributed (iid). The notion of iid is associated with that of homoscedastic noise in classical univariate calibration. However, multivariate iid noise is not a general rule, and can only be seen as an approximation.¹⁰

A partial solution to the previous dilemma is the use of mathematical pre-processing methods. However, this is neither general nor optimal, and only allows for a better approximation to the iid situation which is valid in some specific cases. In response to this, Wentzell *et al.* went a step further both from a conceptual and theoretical point of view, and developed an innovative technique known as maximum likelihood PCR (ML-PCR). By incorporating the overall error properties in the model, ML-PCR improves the performance of traditional PCR in systems where deviations of the iid condition are notorious. However, as recognized by the author in a recent publication, disseminating these methodologies has been difficult due to the complexity associated with the different code variants, the potentially demanding computer memory requirements, and the slow algorithmic convergence.¹¹

Despite the existence of ML methodologies, many current routine applications are still well satisfied by traditional multivariate methods, especially if the errors do not appreciably deviate from the iid assumption.¹⁰ Moreover, the use of suboptimal calibration models does not mean that AFOMs to judge them are also suboptimal. This simple observation was the starting point of a new research line which aims to integrate the properties of multivariate errors into the estimation of AFOMs.¹²

With this background in mind, we report recent advances in the development of estimators for three important AFOMs: sensitivity, sample specific prediction uncertainty and detection limit, and provide some perspectives in this growing research field. Heteroscedasticity strikes back.

The error structure

In univariate calibration, errors are normally classified as homoscedastic or heteroscedastic, corresponding to uniform or

non-uniform variance respectively. In multivariate calibration, these terms are normally replaced by the more general concepts of iid and non-iid noise. The latter, however, includes not only heteroscedastic but also correlated noise. To maintain the analogy between the univariate and multivariate nomenclature, heteroscedastic and non-iid are used as synonymous.

A more general concept in the multivariate context is the error structure, which is characterized by the error covariance matrix (Σ_x). In a simple and illustrative way, this matrix can be seen as a surface taking different shapes, depending on the type of instrumental noise affecting the measurement. The most important difficulty in considering non-iid noise types in multivariate analytical methods is the fact that the covariance matrix is not known in advance. For this reason, the development of methodologies for unbiased estimation of the error covariance matrix has a fundamental importance. This is the reason why this research topic is gaining popularity among chemometricians.^{10,13}

The “iid paradigm”

The error-in-variable (EIV) model is the starting point to develop, from a theoretical basis, the expressions to calculate prediction uncertainty in PCR/PLS. EIV allows one to postulate the following three-term expression to calculate prediction uncertainty:

$$\sigma_y^2 = \text{SEN}^{-2}\sigma_x^2 + h\text{SEN}^{-2}\sigma_x^2 + h\sigma_{\text{ycal}}^2 \quad (1)$$

where σ_y^2 is the prediction variance, σ_x^2 is the variance of the error in the signal, σ_{ycal}^2 is the variance of the errors in calibration concentrations, and SEN is the sensitivity, calculated as $1/(\mathbf{b}^T\mathbf{b})^{1/2}$, with \mathbf{b} being the vector of regression coefficients of the model under evaluation (size $n \times 1$, where n is the number of sensors). In eqn (1), the first term takes into account the uncertainty stemming from the signal of the unknown sample, and represents the most important contribution. The other two terms depend on the position of the sample to be predicted in the calibration space, *i.e.*, the leverage. This latter dimensionless parameter is calculated from the calibration and test scores as: $h = \mathbf{t}^T(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{t}$, where \mathbf{t} is the vector of scores of the sample to be predicted, and \mathbf{T} is the matrix of calibration scores.

The limit of detection is a function of the standard deviation of a blank, *i.e.*, a sample with zero concentration of the analyte of interest. Because of this, finding an expression to calculate the detection limit implies an extrapolation of eqn (1) to zero concentration, which is equivalent to obtaining the value of the leverage at zero concentration. The latter can be calculated through an orthogonal projection of the scores to the hyperplane defined by the score values for which the analyte concentration is equal to zero. By using a mathematical derivation applying basic geometric principles, it was demonstrated that the multivariate limit of detection is in fact given as a range of values, from a minimum (LOD_{min}) to a maximum (LOD_{max}):¹⁴

$$\text{LOD}_{\text{min}} = 3.3(\text{SEN}^{-2}\sigma_x^2 + h_{0\text{min}}\text{SEN}^{-2}\sigma_x^2 + h_{0\text{min}}\sigma_{\text{ycal}}^2) \quad (2)$$

$$\text{LOD}_{\text{max}} = 3.3(\text{SEN}^{-2}\sigma_x^2 + h_{0\text{max}}\text{SEN}^{-2}\sigma_x^2 + h_{0\text{max}}\sigma_{\text{ycal}}^2) \quad (3)$$

where $h_{0\min}$ and $h_{0\max}$ are the minimum and maximum values of the zero-concentration leverage. Details on the estimation of the latter parameters are provided in ref. 14. In any case, it is interesting to note that the expression to calculate $h_{0\min}$ is the same as the one used to calculate the univariate leverage in a traditional linear regression.¹⁴

Validating an estimator

The expressions allowing the calculation of AFOMs are known as estimators. For each particular AFOM, many estimators might be proposed based on theoretical and empirical bases. In any case, it is essential to verify if a proposed estimator is reliable. A simple and effective alternative to perform this verification in the case of uncertainty and sensitivity calculation is the use of noise addition simulations (NASIMs). The main steps of this procedure can be summarized as follows: simulated spectra are first generated using Gaussian profiles and random concentrations for a previously defined number of compounds in two sample sets, one for calibration and the other one for the test. The degree of overlapping of the analyte peak with those for the interferents determines the sensitivity and selectivity towards the analyte of interest in each system. A particular type of noise (iid, proportional or correlated) is then added, considering four different situations: (1) noise is only added to calibration signals, (2) noise is only added to test signals, (3) noise is only added to calibration concentrations and (4) noise is added to both signals and concentrations. Once the latter step is completed, the data are submitted to the specific multivariate model for calibration and prediction. The prediction results are saved at each iteration cycle, and the loop of noise addition and prediction is repeated a large number of times (normally between 1000 and 10 000) (see Fig. 1 for a summary). This methodology has the following advantages:

(1) it can be easily adapted to study a wide number of multivariate and even multiway models under different noise structures and different cases of noise propagation and (2) it operates iteratively, where each iteration is independent from the rest, allowing an easy parallelization to reduce calculation times.

The previously described NASIM strategy can be used to validate theoretical expressions for figures of merit such as the sample dependent prediction uncertainty or the sensitivity. In the case of the limit of detection, however, since its definition involves probabilities of false positives and false negatives, the validation becomes more complex. Finding a simple and reliable way to validate detection limit estimators by means of computational calculations will require additional research efforts.

The “non-iid paradigm”

ML-PCR can be interpreted as a generalized version of PCR where the measurement error structure provides a directional guide for the projection of the spectrum of the sample to be predicted onto the calibration subspace. In mathematical terms, ML-PCR uses an oblique projection instead of an orthogonal projection.

If an orthogonal projection is used to project non-iid noise into the latent variable subspace, as it occurs in PCR and PLS, noise will propagate through the model in a different way than iid, leading to suboptimal estimations by eqn (1). This implies that the prediction uncertainty will be larger as the system under analysis presents a larger deviation from the iid situation. In this latter case, the use of eqn (1) is inadequate because it is not consistent with the type of error that is being propagated to the final prediction. In response to the latter difficulty, recent efforts were oriented to the integration of heteroscedastic noise

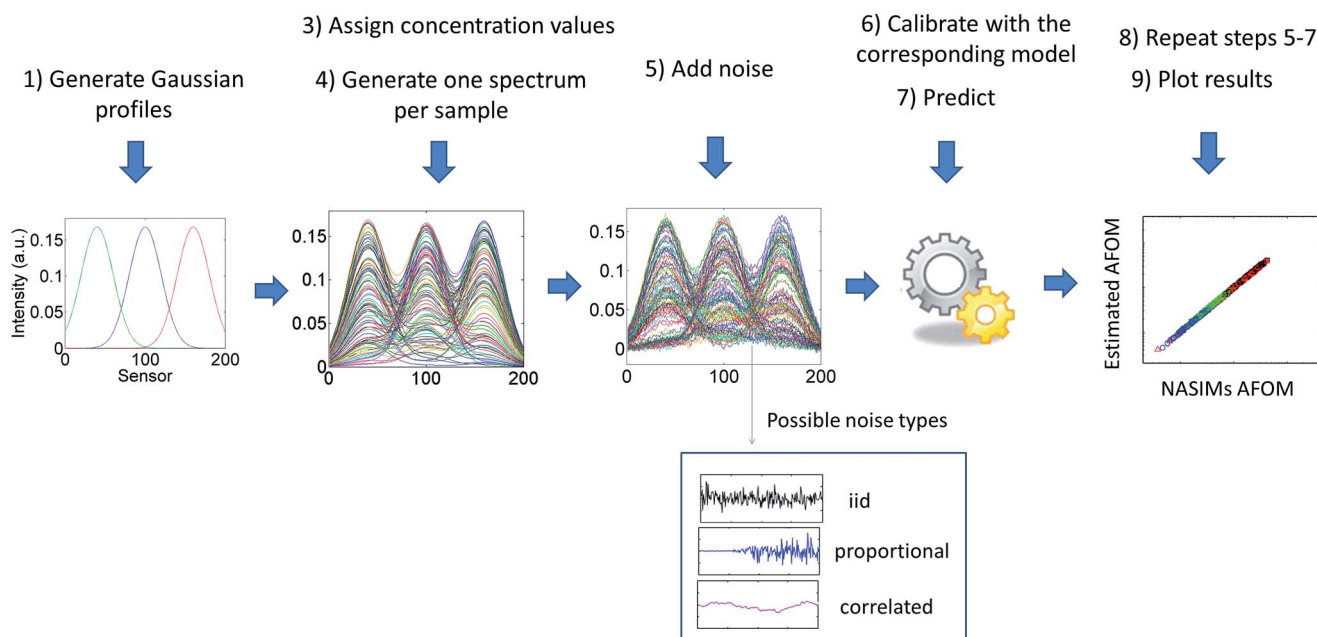


Fig. 1 General scheme for the validation of AFOM estimators.

with prediction uncertainty on the basis of error propagation theory.^{12,15} The results of these efforts can be condensed in the following general expression:

$$\sigma_y^2 = \mathbf{b}^T \Sigma_x \mathbf{b} + h \mathbf{b}^T \Sigma_{x,\text{eff}} \mathbf{b} + h \sigma_{\text{y,cal}}^2 \quad (4)$$

where $\Sigma_{x,\text{eff}}$ is an effective matrix calculated as a weighted sum of the matrices of each individual calibration sample. The presence of the error covariance matrix in eqn (4) can be understood because the latter has been derived by applying error propagation to the prediction expression $y = \mathbf{b}^T \mathbf{x}$. This requires to consider the covariance terms among different sensors, as well as the variance at each sensor. All the latter elements are included in the error covariance matrix Σ_x (for details see ref. 12). An important consequence of the latter analysis is that the particular error structure of the system under analysis, characterized by Σ_x , has a fundamental incidence in the estimation of the final uncertainty. Finally, it is important to remark that, in practice, there are two main alternatives to estimate the error covariance matrix: (1) pooling the error covariance estimates from replicates belonging to different samples or (2) finding a model capable of providing a reliable estimate of the error covariance.^{10,12}

Generalized analytical sensitivity (GAS)

The operational definition of sensitivity in multivariate and multi-way calibration involves the ratio of input to output noise:

$$\text{SEN} = \sigma_x / \sigma_y \quad (5)$$

This approach is based on two assumptions:⁷ (1) the input noise is iid and (2) a small perturbing noise is employed to probe how the latter is propagated to prediction. Contrasting these assumptions with those involved in eqn (4), the question that emerges is if it is possible to extend the definition of sensitivity to incorporate noise properties. In response to this, a recent proposal was to define a generalized analytical sensitivity (GAS) as the inverse of the first term of eqn (4).¹⁶ GAS follows the trend of RMSEP for different kinds of first-order models, allowing reliable method comparison under all types of error structures. Whether this new AFOM is incorporated into the arsenal of performance indicators will be a matter of future experimental work.

Nonlinear models

As shown in a recent publication, it is possible to calculate sensitivity coefficients in radial basis function neural networks (RBF-NN), much as in first-order linear models.¹⁷ The sensitivity coefficients are similar to the linear multivariate regression coefficients, but have no predictive properties. Since the RBF-NN prediction equation is linear in the network training weights, it is possible to apply uncertainty propagation along the same line of reasoning as that used to arrive to eqn (1) and (4). As a consequence, an analogous three-term expression is obtained, with the following specific properties: (1) the sensitivity is estimated using sensitivity coefficients which are sample-specific, and (2) the leverage is obtained from the training design matrix (instead

of the calibration score matrix) and the design vector of the unknown sample (instead of the sample score vector).¹⁷ In this way, neural network calibrations may include complete reports including sensitivity (as a range of values), prediction uncertainties and limit of detection.

For other non-linear multivariate methodologies, such as back propagation neural networks (BP-NN) and support vector machines (SVM), no expressions are currently available to estimate figures of merit. However, the procedure developed for RBF networks could in principle be adapted to these additional models.¹⁷ Work is in progress in our laboratory in this direction.

Multi-way figures of merit

A general expression to estimate sensitivity for different models and different data orders has been derived, validated by simulations, and applied to analyze experimental systems.^{7,18} In the case of multi-way algorithms, the expression shows a number of interesting features. The sensitivity not only depends on the predictive model, but also on the test sample under analysis. This is due to the fact that the general equation, applied to multi-way models, includes an orthogonal projection onto the space defined by the profiles of the interferences, and these are specific for a particular test sample. Incidentally, this is directly related to the second-order advantage property presented by some multi-way strategies such as parallel factor analysis (PARAFAC), multivariate curve resolution coupled to alternating least-squares (MCR-ALS) and PLS coupled to residual multilinearization (PLS-RML). Furthermore, for the unfolded version of the latter model (U-PLS/RML), AFOMs were recently derived in the presence of non-iid noise.¹⁹

A complete report

Thanks to the latest developments in the estimation of AFOMs described in the previous sections, it is now possible to go deeper into the comparison between different kinds of first-

Table 1 Figures of merit for the first-order multivariate determination of phenanthrene in a mixture of polycyclic aromatic hydrocarbons, using fluorescence emission spectra

Figure of merit	iid assumption		
	PCR	ML-PCR	ANN-RBF
RMSE ($\mu\text{g g}^{-1}$)	2.9×10^{-4}	1.8×10^{-4}	3.3×10^{-4}
SEN (fluorescence $\text{g } \mu\text{g}^{-1}$)	1013	523	1001
LOD _{min} ($\mu\text{g g}^{-1}$)	2.3×10^{-4}	4.4×10^{-4}	2.4×10^{-4}
LOD _{max} ($\mu\text{g g}^{-1}$)	2.4×10^{-4}	4.5×10^{-4}	2.9×10^{-4}
Figure of merit	Non-iid assumption		
	PCR	ML-PCR	ANN-RBF
GAS ($\text{g } \mu\text{g}^{-1}$)	3932	7706	3768
LOD _{min} ($\mu\text{g g}^{-1}$)	8.6×10^{-4}	4.4×10^{-4}	0.0010
LOD _{max} ($\mu\text{g g}^{-1}$)	9.2×10^{-4}	4.5×10^{-4}	0.0013

order predictive models. They can be evaluated not only in terms of the RMSE, but also of the information extracted from specific figures of merit such as the sensitivity and the limit of detection. As an example, Table 1 presents a complete AFOM report for the multivariate quantification of phenanthrene in the presence of other polycyclic aromatic hydrocarbons (PAHs), using full fluorescence emission spectra. As previously reported, in this analytical system the error structure presents a significant deviation from the iid condition.¹² As a consequence, when the error structure is taken into account, predictions improve and AFOMs become more consistent. If iid noise is assumed, prediction errors do not have a logical correlation with sensitivities and detection limits: the lowest RMSE corresponds to the lowest value of SEN and the largest LOD. On the other hand, under non-iid assumptions, the relationship among AFOMs becomes consistent: lower prediction errors correspond to higher sensitivities and lower detection limits.

Conclusion

This review discusses the latest advances in the estimation of analytical figures of merit. The style and focus were not directed to mathematical derivations, but to the conceptual line that led to the emergence of new expressions. Future research efforts in the field will be definitely oriented to the unbiased and efficient estimation of error covariance matrices, which are needed to define correct figures of merit for multivariate and multi-way analytical systems. Another important perspective regarding the use and implementation of new figures by the analytical community is the development of a toolbox to evaluate AFOMs for different calibration scenarios, taking into account the latest advances mentioned during this review.

Acknowledgements

Universidad Nacional de Rosario, CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas, Project No. PIP 0163), and ANPCyT (Agencia Nacional de Promoción Científica y Tecnológica, Project No. PICT-2013-0136) are gratefully acknowledged for financial support. F. A. thanks CONICET for a post-doctoral fellowship.

References

- 1 R. Tauler, F. Marini, B. Walczak, L. Buydens, R. G. Brereton, W. Buchberger and P. J. Worsfold, *Trends Anal. Chem.*, 2014, **56**, 9.
- 2 M. Manley, *Chem. Soc. Rev.*, 2014, **43**, 8200.
- 3 W. Li, L. Norgaard and M. Morup, *Anal. Chim. Acta*, 2014, **813**, 1.
- 4 A. C. Olivieri, S. Bortolato and F. Allegrini, Figures of merit in multiway calibration, in Data Handling in Science and Technology, *Fundamentals and Analytical Applications of Multiway Calibration*, ed. A. Muñoz de la Peña, H.C. Goicoechea, G. M. Escandar and A. C. Olivieri, Elsevier, Amsterdam, 2015, vol. 29, ch. 13.
- 5 F. Westad and F. Marini, *Anal. Chim. Acta*, 2015, **893**, 14.
- 6 A. C. Olivieri, *Anal. Chim. Acta*, 2015, **868**, 10.
- 7 A. C. Olivieri, *Chem. Rev.*, 2014, **114**, 5358.
- 8 J. H. Kalivas, K. Héberger and E. Andries, *Anal. Chim. Acta*, 2015, **869**, 21.
- 9 A. J. Tencate, J. H. Kalivas and A. J. White, *Anal. Chim. Acta*, 2016, **921**, 28.
- 10 P. D. Wentzell, *J. Braz. Chem. Soc.*, 2014, **25**, 183.
- 11 P. D. Wentzell, *The Errors of My Ways: Maximum Likelihood PCA Seventeen Years after Bruce, in 40 Years of Chemometrics – from Bruce Kowalski to the Future*, ed. Lavine, et al., American Chemical Society, Washington, DC, 2015, ACS Symposium Series, ch. 3.
- 12 F. Allegrini, P. D. Wentzell and A. C. Olivieri, *Anal. Chim. Acta*, 2016, **903**, 51.
- 13 T. Feital, D. M. Prata and J. C. Pinto, *Can. J. Chem. Eng.*, 2014, **92**, 2228.
- 14 F. Allegrini and A. C. Olivieri, *Anal. Chem.*, 2014, **86**, 7858.
- 15 E. Fernandez-Ahumada, J. M. Roger and B. Palagos, *Anal. Chim. Acta*, 2012, **721**, 28.
- 16 W. Fragoso, F. Allegrini and A. C. Olivieri, *Anal. Chim. Acta*, 2016, **933**, 43.
- 17 F. Allegrini and A. C. Olivieri, *Anal. Chem.*, 2016, **88**, 7807.
- 18 A. Eftekhari and H. Parastar, *J. Chromatogr. A*, 2016, **1466**, 155.
- 19 F. Allegrini and A. C. Olivieri, *Chemom. Intell. Lab. Syst.*, 2016, **158**, 200.