Analytica Chimica Acta 868 (2015) 10–22

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Tutorial

Practical guidelines for reporting results in single- and multi-component analytical calibration: A tutorial

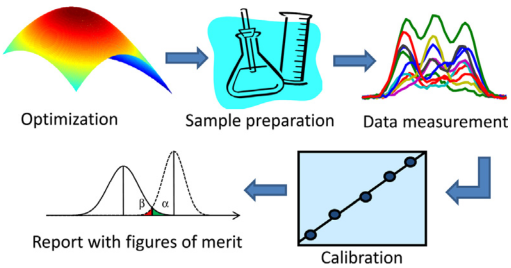
Alejandro C. Olivieri *

Instituto de Química Rosario (CONICET-UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, 2000 Rosario, Argentina

HIGHLIGHTS

- Practical guidelines for reporting analytical results are provided.
- Single- and multi-component calibrations are covered.
- Common mistakes and misconceptions are reviewed.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 29 October 2014
Received in revised form 8 January 2015
Accepted 12 January 2015
Available online 15 January 2015

Keywords:
Analytical calibration
Reporting results
Univariate calibration
First-order multivariate calibration

ABSTRACT

Practical guidelines for reporting analytical calibration results are provided. General topics, such as the number of reported significant figures and the optimization of analytical procedures, affect all calibration scenarios. In the specific case of single-component or univariate calibration, relevant issues discussed in the present Tutorial include: (1) how linearity can be assessed, (2) how to correctly estimate the limits of detection and quantitation, (2) when and how standard addition should be employed, (3) how to apply recovery studies for evaluating accuracy and precision, and (4) how average prediction errors can be compared for different analytical methodologies. For multi-component calibration procedures based on multivariate data, pertinent subjects here included are the choice of algorithms, the estimation of analytical figures of merit (detection capabilities, sensitivity, selectivity), the use of non-linear models, the consideration of the model regression coefficients for variable selection, and the application of certain mathematical pre-processing procedures such as smoothing.

Contents

* Tel.: +54 341 4372704; fax: +54 341 4372704.
E-mail address: olivieri@iquir-conicet.gov.ar (A.C. Olivieri).

http://dx.doi.org/10.1016/j.aca.2015.01.017
0003-2670/© 2015 Elsevier B.V. All rights reserved.

Alejandro 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 currently works in the Department of Analytical Chemistry of the latter Faculty, and is a fellow of the National Research Council of Argentina (CONICET). 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) and won the Platinum Konex prize in 2014 for his contributions to *Analytical Chemistry* in Argentina over the past 10 years.

1. Introduction

Univariate calibration is synonymous of single-component calibration, and is well-known as the cornerstone of many analytical chemistry procedures. The latter calibration protocol can be safely employed when the instrumental signal is selective enough, or the interferents have been separated from the analyte in the test sample. The separation can be physical (e.g., chromatography) or chemical (e.g., complexation to mask a species) [1]. When this is not the case, alternatives are based on multivariate calibration procedures, which are now firmly established. They can compensate for the presence of interferents by including them in the calibration phase, as when first-order data are measured (e.g., spectra) [2], or simply by purely mathematical means, as when achieving the second-order advantage from multi-way data [3].

In this context, it is rather paradoxical that analytical chemists regularly apply official regulations as regards analytical protocols, but they do not follow the same recommendations when reporting their results. After many years of reviewing manuscripts for mainstream analytical journals, including *Analytica Chimica Acta*, a list has been compiled of common mistakes and misconceptions which should be avoided when processing and reporting calibration data. This applies to both uni- and multivariate data, in the latter case with particular focus on first-order data, and specifically using the most popular partial least-squares (PLS) regression model [2].

The present report intends to provide a practical guide for improving the presentation of calibration results. It has been divided, for clarity, in three main sections: the first one contains general recommendations, applicable to all forms of calibration procedures, the second one specifically applies to univariate calibration, and the final one to first-order multivariate calibration. Although the latter is by far the most popular form of calibration with multiple data per sample, some advice will occasionally be directed to multi-way calibration. The sections devoted to general aspects and to univariate calibration are perhaps more elaborated than the one for multivariate calibration, in part because the former scenarios are more popular and also to keep a reasonable length of the tutorial.

2. General

2.1. Significant figures

An analytically oriented paper should pay proper attention to the significant figures employed in reporting results, i.e., unreasonably large numbers of significant figures should be avoided [4]. This is not purely cosmetic; authors should pay attention to this important fact, otherwise, a wrong impression would be produced in readers. It is also the recommended action by international convention, and should be honored by all chemists alike.

In general, all results should be reported with a number of significant figures compatible with the standard error associated to the result. Uncertainties should be reported with one or at most two significant figures. A good rule of thumb is to use two significant figures when the first one is 1, or when the first is 2 and the second is smaller than 5; in the remaining cases a single significant figure should be reported. For example, a predicted concentration should not be reported as 13.89 mg L^{-1} with a standard deviation of 2.85 mg L^{-1}, but as 14 mg L^{-1} with a standard error of 3 mg L^{-1}. A concentration value reported as 13.89 mg L^{-1} would give the reader the wrong impression that the uncertainty in the prediction of this concentration is on the order of 0.01 mg L^{-1}, while in fact it is two orders of magnitude larger! Even if the standard error is not provided, reporting the concentration as 14 mg L^{-1} conveys the implicit message that the uncertainty in such determination is on the order of the mg L^{-1}, which is correct.

Likewise, if the slope of a calibration graph is 2158.2 AU L mg^{-1} (AU = arbitrary signal units), with a standard error of 32 AU L mg^{-1}, the sensitivity for the analyte determination, which is equal to the slope, should not be reported as 2158.2 AU L mg^{-1}. The correct report should be 2.16×10^3 AU L mg^{-1}, because the uncertainty affects the third significant digit.

Additional parameters derived from uncertainties should be reported with one or at most two significant figures. For example, a limit of detection (a figure of merit which depends on the uncertainties in signals and concentrations, see below) should not be reported as 0.1187 mg L^{-1}, but as 0.12 mg L^{-1}.

Suppose the sensitivity of a method has been found to be 6.48×10^2 AU L mg^{-1}, and the instrumental noise level is 1.5 AU. Then the analytical sensitivity, which is the ratio of sensitivity to noise [5], should be reported as 4.3×10^2 L mg^{-1}, and not as 432 L mg^{-1}. This is because the number of significant figures for the noise (two in this case) controls the number of significant figures of the analytical sensitivity.

Other figures of merit to be reported with at most two significant figures are the limit of decision (LD), the limit of quantitation (LOQ), the average prediction error or the root mean square error (RMSE) and the relative error of prediction (REP). The latter two parameters can be estimated according to:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} (c_{\mathrm{nom},n} - c_{\mathrm{pred},n})^2}{N}} \quad (1)$$

$$\mathrm{REP} = 100 \frac{\mathrm{RMSE}}{\bar{c}_{\mathrm{cal}}} \quad (2)$$

where $c_{\mathrm{nom},n}$ and $c_{\mathrm{pred},n}$ are the nominal and predicted concentrations for the test sample n,N is the number of test samples, and $\bar{c}_{\mathrm{cal}}$ is the mean calibration concentration.

Finally, recoveries expressed in % should be provided with a number of significant figures compatible with those for the reported analyte concentrations.

2.2. Optimization of analytical methods

Usually the optimization of analytical methods is conducted by changing one variable at a time. This is not the recommended optimization procedure in analytical chemistry and other chemistry fields, since: (1) it requires considerably more experimental points than other rational surface response optimization (SRO) procedures, (2) it only provides local optima, in comparison with global optima furnished by SRO, (3) it does not take into consideration possible interactions among affecting factors, and, perhaps more importantly, (4) it leads to sub-optimal results, and hence the final result may not be the best one for the purpose the authors are pursuing [6].

It is worth repeating the following rather sad excerpt from the conclusion of a tutorial on multivariate design and optimization of experiments [7]: "... *When browsing through the papers published in Analytica Chimica Acta in 2009 (from volume 631 to volume 645, plus the papers available in the "Articles in Press" section on June 3, 2009), I found 165 of them having the general title or a section title containing the words "optimization" or "development", or "improvement", or "effect of". Only in 11 papers (i.e., one out of 15 ...) a multivariate approach has been followed, while in the great majority of them the "optimization" was performed one variable at a time, sometimes with the titles of the subsections proudly remarking it ("3.1. Effect of pH", "3.2. Effect of temperature", "3.3. Effect of flow", and so on.)* ..." An inspection of the analytical literature from 2009 to date reveals that the approach of changing one variable at a time for attempting optimization is still in use.

Authors are encouraged to follow the well-established and reliable multivariate optimization procedures described in Refs. [6,7] and references therein. As an appropriate example, consider a typical optimization procedure employed in the determination of the anti-allergic epinastine in human sera by capillary electrophoresis [8]. It was apparent that two important analytical parameters (the time of analysis and the resolution between the peaks for the analyte and an internal standard) depended on various experimental factors. The aim was to minimize the time and to reach a target value of 2 for the resolution. The factors were the concentration and pH of the buffer, the injection time, the injection voltage and the separation voltage. When the number of factors to be optimized is rather large, such as in the present example, it is advisable to first conduct a screening phase of experiments. The analytical responses are measured for a small number of runs, designed to explore the relative significance of the various factors. This calls for an experimental design of many factors at a small number of levels, such as the Plackett–Burman design [6,7], which only requires twelve experiments. As a result of the statistical analysis of the responses for the screening experiments, four factors were found to be important: the concentration and pH of the buffer, the injection voltage and the separation voltage [8]. In the next optimization phase, a design was employed with more levels per factor, in order to be able to explore the (possibly non-linear) surface response. In this case a central composite design [6,7] (30 experiments with 5 levels per factor) was used. The responses were modeled as a function of the factors using cubic polynomials. When two responses are simultaneously studied, full optimization is not generally possible, but desirable results can be obtained by combining the responses into a desirability function to be optimized [9]. The approach allowed to estimate the desirable analytical responses and the corresponding factor values [8]. This illustrates the complete screening and optimization process which is advisable when designing complex analytical experiments, instead of modifying variables one at a time.

3. Single-component calibration

3.1. Analyte determination

The set of concentrations designed for calibration should include the blank. The sample with zero analyte concentration allows one to gain better insight into the region of low analyte concentrations and detection capabilities [1].

The mathematical expression employed to fit the data to a linear model should include an intercept. The latter accounts for the blank signal, even if the blank signal is suspected to be zero. In fact it is never exactly zero, because of either the existence of a small blank signal, or because of the universal presence of instrumental noise.

Analyte concentrations in the calibration set should be included as replicate samples, and not as single samples. This allows to obtain more robust regression results, and to assess the linearity of the calibration graph (see below), as well as other statistical parameters [1].

3.2. Linearity

The correlation coefficient (R) of a calibration graph is usually employed for assessing its linearity, regularly by visual inspection of its closeness to 1. However, the International Union of Pure and Applied Chemistry (IUPAC) discourages the correlation coefficient as an indicator of linearity in the relationship between concentration and signal. This is literally expressed in Ref. [1]: " ... *the correlation coefficient, which is a measure of relationship of two random variables, has no meaning in calibration* ... ". A less stringent view is offered in Ref. [10], where the correlation coefficient is said to provide a measure of the degree of linear association between concentration and signal. However, the linearity is suggested to be checked using the test to be discussed below.

To test for linearity, authors should report the experimental F value corresponding to the ratio of residual variance to squared pure error, and the tabulated critical F for comparison. Specifically, the experimental F ratio is given by:

$$F_{\mathrm{exp}} = \left(\frac{s_{y/x}}{s_y}\right)^2 \quad (3)$$

where $s_{y/x}$ is the residual standard deviation and s_y is the so-called pure error (a measure of the instrumental noise). These parameters can be estimated from the calibration data as:

$$s_{y/x} = \sqrt{\frac{\sum_{i=1}^{I}(y_i - \hat{y}_l)^2}{I-2}} \quad (4)$$

$$s_y = \sqrt{\frac{\sum_{l=1}^{L}\sum_{q=1}^{Q} y_{lq} - y_l^{\,2}}{I-Q}} \quad (5)$$

In the latter expressions, y_i and $\hat{y}_i$ are the experimental and estimated response values for sample i, y_{lq} is the calibration response for replicate q at level l, $\bar{y}_l$ is the mean response at level l, and I, L and Q are the total number of calibration samples, levels and replicates at each level, respectively.

The statistical hypotheses are thus H_0 (the data are linear) and the alternative H_a (the data are non-linear), and the null hypothesis would be rejected at significance level α if F_{exp} exceeds the critical value at level α, $F(\alpha, I-2, I-L)$ (I is the number of calibration samples and L the number of concentration levels). This test is the best linearity indicator, as recommended by IUPAC [1], and amounts to statistically check whether the residual variance is larger than the squared pure error derived from the study of replicate samples.

It may be noticed that an alternative assessment of the linearity has also been discussed by resorting to the analysis of variance (ANOVA) of the calibration data [10]. In this latter methodology, comparison is made of the so-called lack-of-fit variance to the squared pure error through an F test. While the pure error is defined as in Eq. (5) above, the lack-of-fit differs from $s_{y/x}$ and leads to different experimental and critical values of F [10].

It would be advisable to read the excellent brief reports by the Analytical Methods Committee of the Royal Society of Chemistry [11]. Can a calibration data set be fitted to a linear regression analysis, giving a correlation coefficient close to unity and still be non-linear? Table 1 provides an example, where it is easy to grasp the hazards of employing correlation coefficients for assessing linearity.

3.3. Limit of detection

The univariate limit of detection ($\mathrm{LOD_u}$) is usually estimated using the old definition, now abandoned by IUPAC, based on the analyte concentration which gives a signal at least three times larger than the standard deviation of the blank signal. This $\mathrm{LOD_u}$ value is usually an underestimation [12,13].

The modern IUPAC recommendation first requires to define a level for the detection decision (LD), involving a certain risk of false detects (also called false positives, α-errors or Type I errors). As illustrated in Fig. 1, the green-shaded area represents a portion of

Table 1
A calibration data set for which the linearity test is not fulfilled but the correlation coefficient is close to unity.[a]

Calibration sample	Concentration	Signal	
		Replicate 1	Replicate 2
1	64	138	142
2	128	280	282
3	192	423	425
4	256	565	567
5	320	720	725
6	384	870	872
Calibration parameters[b]			
Slope (standard deviation)	2.286 (0.014)		
Intercept (standard deviation)	−11 (3)		
Linearity assessment			
Calibration sample/replicate	Estimated concentration	Residual error	Squared residual error
1/1	134.9	−3.1	9.4
2/1	281.3	1.3	1.6
3/1	427.6	4.6	21.0
4/1	573.9	8.9	79.5
5/1	720.2	0.2	0.1
6/1	866.6	−3.4	11.8
1/2	134.9	−7.1	50.0
2/2	281.3	−0.7	0.6
3/2	427.6	2.6	6.7
4/2	573.9	6.9	47.8
5/2	720.2	−4.8	22.6
6/2	866.6	−5.4	29.5
Sum of squared errors		280.5	
Residual standard deviation ($s_{y/x}$)		5.3	
Residual variance [$(s_{y/x})^2$]		28	
Pure error (s_y)		2.2	
Squared pure error [$(s_y)^2$]		4.8	
F_{exp}		5.9	
Critical $F(0.05,10,6)$		4.1	
R		0.9998	

[a] Concentrations and signals are given in arbitrary signal units; the data have been adapted from Ref. [11].
[b] Standard deviation in parenthesis. F_{exp} is the ratio of residual variance to squared pure error, critical $F(0.05,10,6)$ is the critical value of F with $(I-2)=10$ and $(I-L)=6$ degrees of freedom at 95% confidence level, where I is the number of calibration samples (12) and L the number of concentration levels (6) and R is the correlation coefficient.

the Gaussian distribution of concentration values having a probability α of declaring 'analyte absent' while in fact it is present. This is the meaning of false detect or false positive. 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. This corresponds to the red-shaded area in Fig. 1, where the analyte may be declared present while it is in fact absent. Both α and β are usually assigned reasonably small values, depending on the specific analytical application. Fig. 1 allows to understand the expression for the univariate LOD_u:

$$\text{LOD}_\text{u} = t(\alpha,\nu)\sigma_{c,0} + t(\beta,\nu)\sigma_{c,\text{LOD}} = \frac{3.3 s_{y/x}}{A}\sqrt{1 + h_0 + \frac{1}{I}} \tag{6}$$

where $t(\alpha,\nu)$ and $t(\beta,\nu)$ are Student coefficients with ν degrees of freedom and α and β probabilities, respectively, $\sigma_{c,0}$ and $\sigma_{c,\text{LOD}}$ are the concentration standard errors for the blank and LOD_u levels, A is the slope of the univariate calibration graph, I is the number of calibration samples and $s_{y/x}$ is the residual standard deviation. Assuming $\sigma_{c,0} = \sigma_{c,\text{LOD}}$, 95% confidence level ($\alpha = \beta = 0.05$) and a large number of degrees of freedom, the right-hand side of Eq. (6) is obtained, where h_0 is the leverage for the blank sample:

$$h_0 = \frac{\bar{c}_\text{cal}^2}{\sum_{i=1}^{I} (c_\text{i} - \bar{c}_\text{cal})^2} \tag{7}$$

where $\bar{c}_\text{cal}$ is the mean calibration concentration and c_i is each of the calibration concentration values. Similar concepts apply to the limit of quantitation (LOQ_u):

$$\text{LOQ}_\text{u} = \frac{10 s_{y/x}}{A}\sqrt{1 + h_0 + \frac{1}{I}} \tag{8}$$

where the factor 10 ensures a maximum relative prediction uncertainty of 10%.

Table 2 shows typical results for a calibration graph constructed with four duplicate concentration levels. In this particular example, the LOD_u computed with the old IUPAC definition is almost half the value estimated with the modern approach. Usually, the old definition significantly underestimates the detection limit.

A freely downloadable software written in MATLAB [14] is available at www.iquir-conicet.gov.ar/descargas/univar.rar, which performs univariate calibration, applies the linearity F test and provides various analytical figures of merit according to the above commented criteria.

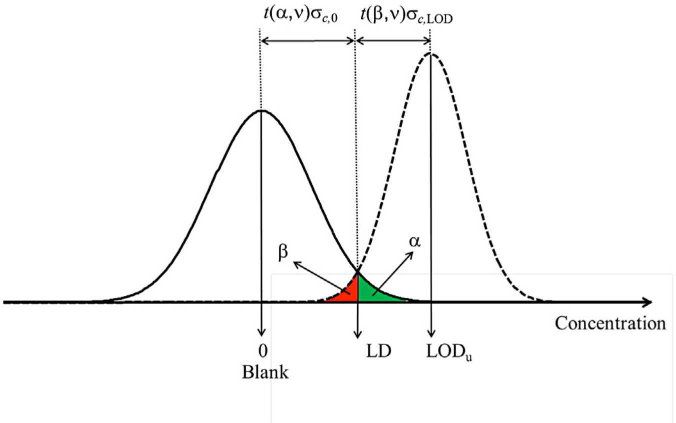

$t(\alpha,\nu)\sigma_{c,0}$
$t(\beta,\nu)\sigma_{c,\text{LOD}}$
β
α
Concentration
0
Blank
LD
LOD_u

Fig. 1. Illustration of fhe official (IUPAC) definition of the univariate decision limit (LD) and limit of detection (LOD_u). Two Gaussian bands are centered at the blank and at the LOD_u, respectively. The LD helps to decide whether the analyte is detected or not with a rate α of false detects, whereas the LOD_u implies detection with a rate α of false detects and a rate β of false non-detects. The shaded areas correspond to the rate of false detects (green) and false non-detects (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Calibration results for typical univariate data.

Concentration	Signal	
	Replicate 1	Replicate 2
0.00	0.06	0.08
1.00	1.44	1.6
3.00	4.15	4.2
5.00	6.61	6.54
Calibration parameters		
Slope (A)	1.30	
Residual standard deviation ($s_{y/x}$)	0.12	
Blank leverage (h_0)	0.17	
Number of samples (I)	8	
LOD_u (old IUPAC definition)	0.2	
LOD_u (new IUPAC definition)	0.4	

It may be noticed that the European Commission has adopted similar recommendations [15], including the capability of detection (CCβ), which is the smallest concentration of the substance that may be detected, identified and/or quantified in a sample with an error probability of β. This can be interpreted as the minimum analyte concentration that can be discriminated from the blank, controlling the risks of false positives and false negatives. The definition of CCβ is analogous to the LOD_u above.

3.4. Standard addition

When standard addition is employed for analyzing samples, an appropriate justification should be supplied. It is not enough to say that the samples are complex, or that they come from a biological origin, because this does not mean, *per se*, that standard addition is required for analyte quantitation [16].

Univariate standard addition should be employed only when: (1) the slope of the response-concentration relationship differs from the pure analyte to the analyte embedded in a certain background, and (2) the background is not responsive. This can be checked by statistically comparing the slopes with a certain confidence level and number of degrees of freedom [10,17], and not by visual comparison. Then, only in the case the slopes significantly differ (see Refs. [10,17]), standard addition should be employed, because the latter analytical mode is highly time consuming and expensive in comparison with classical calibration.

The comparison of two slopes A_1 and A_2 is based on the hypotheses H_0 ($A_1 = A_2$) and the alternative H_a ($A_1 \neq A_2$), rejecting the null hypothesis at significance level α if t_exp exceeds the critical value at level α, $t(\alpha, N_1 + N_2 - 4)$ (N_1 and N_2 are the number of concentration values used to estimate each slope). The experimental t value is estimated as [10,17]:

$$t_\text{exp} = \frac{A_1 - A_2}{\sqrt{s_p^2\left[\frac{1}{\sum_{n1=1}^{N1}(c_{n1} - \bar{C}_1)^2} + \frac{1}{\sum_{n2=1}^{N2}(c_{n2} - \bar{C}_2)^2}\right]}} \tag{9}$$

where each method is evaluated using N_1 and N_2 concentration values c_{n1} and c_{n2}, whose averages are $\bar{c}_1$ and $\bar{c}_2$ respectively, and s_p^2 is the pooled variance:

$$s_p^2 = \frac{s_{y/x1}^2(N_1 - 2) + s_{y/x2}^2(N_2 - 2)}{N_1 + N_2 - 4} \tag{10}$$

where $s^2_{y/x1}$ and $s^2_{y/x2}$ are the residual variances of each calibration graph. It should be noticed that Eq. (10) is only valid when the variances of each regression line are comparable, a fact which can be checked with a suitable *F* test [10,17].

An example is provided by the analysis of the antibiotic ciprofloxacin in water and human serum, using six different concentrations in the range 0.00–0.50 mg L^{-1} [18]. Table 3 shows the replicate signals, slopes and statistical parameters for the comparison. Since t_{exp} is larger than the tabulated value, the conclusion is that the slopes are different at 95% confidence level, justifying the use of standard addition.

3.5. Comparison of two analytical methods

The root mean square error (RMSE) is usually employed as an indicator of whether a given analytical methodology provides better predictive ability. However, the comparison of RMSE values should not be based on visual inspection. A suitable statistical test should be applied to assess whether two RMSE values are statistically different, such as the randomization test described in Ref. [19]. Accordingly, no conclusions should be drawn on the basis of RMSE values being larger or smaller, until a proper test has been applied. A short MATLAB code for applying the randomization test is provided in Ref. [19], and an adapted version is now given in Table 4.

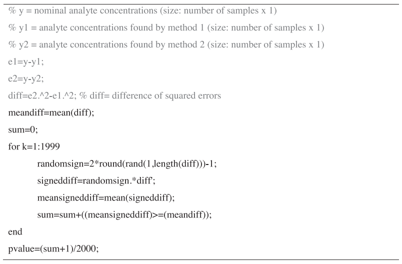
To illustrate the philosophy of the randomization test, a simple example is provided in Table 5 (top). For a group of five test samples, concentrations are estimated with three different methods, giving rise to three RMSE values: RMSE1 = 1.0, RMSE2 = 2.0 and RMSE3 = 1.2. Both RMSE2 and RMSE3 are larger than RMSE1, but the question is whether this is statistically relevant. In the middle of Table 5, the operation of the randomization test is shown for the comparison of RMSE2 with RMSE1. First the difference between squared errors is computed, as well as the mean difference, equal to 3 units in Table 5. Then the sign of each difference is randomly inverted, as seen in the subsequent columns as three pertinent examples of the randomization operation. The mean of each of the new column of differences is compared to the original mean, and a statistics is registered of the relative number of times these new differences are larger than the original one. This is the *p*-value associated to the test. As seen in Table 5 (middle), in the three example cases the new differences are smaller than the original ones. In fact, repeating the randomization process a large number of times leads to the conclusion that the value of *p* is ≪0.05, indicating that RMSE2 is significantly larger than RMSE1. Here the hypotheses are H_0 ($RMSE_2 = RMSE_1$) and H_a ($RMSE_2 > RMSE_1$), and the test directly provides the probability *p* associated to H_0, suggesting rejection of the null hypothesis.

Table 3
Comparison of slopes of two calibration graphs, previous to developing a standard addition method.

Concentration mg L^{-1}	Signal in water		Signal in serum	
	Replicate 1	Replicate 2	Replicate 1	Replicate 2
0.00	0.8	0.0	9.6	10.8
0.10	23.4	25.8	25.8	25.1
0.20	38.8	42.5	35.7	38.1
0.30	56.5	58.3	47.9	53.0
0.40	75.6	73.7	63.9	65.8
0.50	93.0	93.0	81.7	82.7
Calibration parameters				
	In water			In serum
Slope	180			141
Residual variance	5.8			4.4
Comparison of slopes				
s^2_p	5.1			
t_{exp}	5.1			
t(0.025,8)[a]	2.3			

[a] t(0.025,8) is the one-tail *t*-coefficient at 95% confidence level with 6 + 6–4 = 8 degrees of freedom (6 is the number of concentrations levels).

Table 4
MATLAB code for implementing the randomization test.[a]



```
% y = nominal analyte concentrations (size: number of samples x 1)
% y1 = analyte concentrations found by method 1 (size: number of samples x 1)
% y2 = analyte concentrations found by method 2 (size: number of samples x 1)
e1=y-y1;
e2=y-y2;
diff=e2.^2-e1.^2; % diff= difference of squared errors
meandiff=mean(diff);
sum=0;
for k=1:1999
        randomsign=2*round(rand(1,length(diff)))-1;
        signeddiff=randomsign.*diff';
        meansigneddiff=mean(signeddiff);
        sum=sum+((meansigneddiff)>=(meandiff));
end
pvalue=(sum+1)/2000;
```

[a] This routine tests whether errors by method 2 are larger than errors by method 1.

In the bottom part of Table 5 an analogous comparison is made between RMSE3 and RMSE1. In this case, the means of the new differences (arising after randomly inverting the signs of the original difference) are sometimes smaller and sometimes larger than the original value of 0.4. Application of the test a large number of times leads to the conclusion that *p* = 0.5, indicating that RMSE3 and RMSE1 are not statistically different, i.e., the null hypothesis H_0 is accepted.

Table 6 shows a second example involving the results of the randomization test of the predictions of two analytical methods in a set of test samples. Random errors equally affect the predicted concentrations, with a standard deviation of 1 unit. Method 1 is assumed to be unbiased, while method 2 is increasingly biased.

Table 5
Illustration of the randomization test for the comparison of the average errors from three different analytical methods.

Test sample	c_{nom}	c_1	c_2	c_3
1	10	11	12	10
2	20	19	22	22
3	30	29	28	31
4	40	41	42	39
5	50	49	52	51
RMSE		1.0	2.0	1.2
Comparison of RMSE2 with RMSE1				
Test sample	$(c_2-c_{\text{nom}})^2-(c_1-c_{\text{nom}})^2$	Randomize signs		
1	3	3	−3	3
2	3	−3	3	−3
3	3	3	−3	3
4	3	−3	3	3
5	3	3	3	3
Mean value	3	0.6	0.6	1.8
Comparison with 3		<	<	<
p value		≪0.05		
Comparison of RMSE3 with RMSE1				
Test sample	$(c_3-c_{\text{nom}})^2-(c_1-c_{\text{nom}})^2$	Randomize signs		
1	−1	−1	1	1
2	3	−3	−3	3
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
Mean value	0.4	−0.8	−0.4	0.8
Comparison with 0.4		<	<	>
*p*value		0.5		

Table 6
Comparison of RMSE values for two different methods, in terms of the % of times that p is <0.05.[a]

Bias1	RMSE1	Bias2	RMSE2	Number of test samples		
				10	50	100
0	1.0	0	1.0	10%	10%	4%
0	1.0	0.5	1.1	10%	15%	30%
0	1.0	1	1.4	25%	80%	100%
0	1.0	1.5	1.8	50%	100%	100%

[a] Random errors in concentrations have a standard deviation of 1 unit.

When the bias in method 2 is zero, the relative number of cases for which differences are found in both methods is small, independently of the number of test samples (Table 6). However, as the bias in method 2 increases, the RMSE2 increases, and the number of cases for which significant differences are found also increases. However, notice in Table 6 that for a bias of 1.5 units, when the number of test specimens is small (10), there is still a 50% chance of not finding significant differences between both methods, even when RMSE2 is almost twice RMSE1.

These results highlight the need of employing suitable tests to prove that a certain RMSE is smaller (or larger) than the one provided by an alternative analytical method. Visually checking that RMSE > RMSE1 does not guarantee that the observed difference is statistically meaningful.

3.6. Recovery studies

When discussing recovery results, it is usually stated that they are satisfactory, only by visual inspection of the predicted values or the closeness of recoveries to 100%. Nevertheless, statistical tests should be applied to assess whether a recovery is not statistically different than 100% [20], reporting both experimental and critical t values [20]. It should be noticed, however, that these tests assume certain conditions that they data should fulfill, such as constant variance [21].

Table 7 shows a typical problem of assessing the recoveries of an analyte from a group of pharmaceutical samples of similar composition, whose concentrations may be assumed to have a similar variance. Usually, tables such as Table 7 are left unprocessed, resorting only to visual inspection of the recoveries for justifying good analytical predictions. A suitable statistical analysis involves a hypothesis test of whether the average recovery is significantly different from 100% or not. The hypotheses are H_0 $(\overline{R}_{\mathrm{exp}} = 100\%)$ and the alternative H_a $(\overline{R}_{\mathrm{exp}} \neq 100\%)$; the null hypothesis is rejected at significance level α if t_{exp} exceeds the critical value at level α, $t(\alpha, N-1)$ (N is the number of test samples). The experimental t_{exp} value is estimated from:

$$t_{\mathrm{exp}} = |100 - \frac{\overline{R}_{\mathrm{exp}}|\sqrt{N}}{S_{\mathrm{R}}} \tag{11}$$

where $\overline{R}_{\mathrm{exp}}$ is the average experimental recovery and S_{R} the standard deviation of the recoveries. Comparison is usually made at 95% confidence level, as detailed in Table 7. According to the results quoted in the latter table, the method is accurate, since $t_{\mathrm{exp}} < t(0.025, N-1)$.

Table 7
Analyte recoveries for a group of samples, and statistical test of whether the mean recovery is significantly different than 100% or not.

Sample	Nominal mg^{-1}	Found mg^{-1}	Recovery/%
1	50	49.5	99.0
2	50	50.2	100.4
3	100	100.3	100.3
4	100	98.4	98.4
5	150	149.3	99.5
Recovery analysis			
Mean recovery $\overline{R}_{\mathrm{exp}}$		99.5	
Standard deviation of recoveries (S_R)		0.9	
Number of samples (N)		5	
t_{exp}		1.2	
t(0.025,4)		2.8	

When the analyte concentration range in the test samples is rather wide and constant variance cannot be assumed, linear regression of predicted vs. nominal analyte concentrations is recommended [22]. The analysis of these results should not be based on the study of whether the ideal values of unit slope and zero intercept are individually included within their confidence ranges around the means. The recommended test is the so-called elliptical joint confidence region (EJCR) test, which implies drawing the EJCR for the slope and intercept of the above linear plot, and checking whether the ideal point (slope = 1, intercept = 0) is included in the ellipse [22]. It should be noticed that for non-constant prediction variance, a regression technique should be employed accounting for the fact that there are non-constant errors, such as weighted least-squares (WLS) or bilinear least-squares (BLS), and not ordinary least-squares (OLS), as is usually done [23,24]. Otherwise, dramatically different conclusions might be obtained [24].

The specific expression describing the EJCR is:

$$N(y-B)^2 + 2(x-A)(y-B)\sum_{n=1}^{N} c_{xn} + (x-A)^2 \sum_{n=1}^{N} c_{xn}^2 \\ = 2s_{y/x}^2 F_{2,N-2} \tag{12}$$

where A and B are the estimated slope and intercept for the regression of predicted vs. nominal analyte concentrations, N is the number of test samples, c_{xn} the concentration of the nth sample used as reference and placed in the x axis of the regression analysis, $s_{y/x}$ is the residual standard deviation of this specific linear regression (not to be confused with the one corresponding to the calibration graph), and $F(0.05, 2, N-2)$ is the critical value of the

Table 8
Illustration of the assessment of the accuracy of an analytical method using the EJCR test, and the separate confidence regions for the slope and intercept.

Sample	Reference value	Predicted value	Standard deviation
1	0.00	0.06	0.03
2	0.05	0.13	0.05
3	0.11	0.10	0.09
4	0.16	0.07	0.08
5	0.21	0.25	0.04
6	0.26	0.22	0.10
7	0.32	0.23	0.08
8	0.37	0.37	0.05
9	0.42	0.43	0.04
10	0.47	0.50	0.02
11	0.53	0.54	0.03
12	0.58	0.55	0.08
13	0.63	0.61	0.05
14	0.68	0.67	0.05
15	0.74	0.74	0.02
16	0.79	0.77	0.04
17	0.84	0.80	0.12
18	0.89	0.82	0.05
19	0.95	1.00	0.13
20	1.00	0.97	0.19
Recovery analysis			
WLS EJCR			(1,0) point not included in EJCR
OLS EJCR			(1,0) point included in EJCR
OLS slope	0.96 ± 0.07		Slope = 1 included in interval
OLS intercept	0.01 ± 0.04		Intercept = 0 included in interval

parameter F with 2 and $N-2$ degrees of freedom and a 95% confidence level.

An appropriate example is illustrated in Table 8, which collects data for a number of samples, including the reference concentration in standards, the mean of triplicate sample analysis by a method under scrutiny, and the corresponding standard deviations. Regression using WLS, which takes into account the variance at each concentration, provides the elliptical region shown in Fig. 2 (red line), indicating that the method is not accurate. However, OLS regression points otherwise (blue line in Fig. 2). Incidentally, the consideration of the individual confidence intervals for the OLS slope and intercept (black rectangle in Fig. 2) leads to the same conclusion as the OLS elliptical region (Table 8). The EJCR test can be implemented using the MATLAB code provided in www.iquir-conicet.gov.ar/descargas/ejcr.rarwww.iquir-conicet.gov.ar/descargas/ejcr.rar.

4. Multi-component calibration

4.1. First-order algorithms

For first-order multivariate calibration, e.g., near infrared (NIR) spectroscopic studies, PLS seems to be preferred as the *de facto* standard [25], although principal component regression (PCR) has been shown to provide equivalent results to PLS in terms of prediction ability [26]. As stated in the Abstract of the latter paper, " . . . *In all cases, except when artificial constraints were placed on the number of latent variables retained, no significant differences were reported in the prediction errors reported by PCR and PLS. PLS almost always required fewer latent variables than PCR, but this did not appear to influence predictive ability*."

Usually first-order calibration data contain fewer samples than variables, because instruments measure hundreds or thousands of variables per sample. In this case one can in principle apply: (1) multiple linear regression (MLR), which requires the selection of a suitable number of wavelengths which should be smaller than the number of samples [27], or (2) PLS/PCR, which involves compression of full spectral data into a few latent variables [28]. The subject 'MLR vs. PLS/PCR' has given rise to considerable debate in the past.

It should be noticed that PLS and PCR show a number of advantages over MLR: (1) noise reduction, because of the averaging of correlated measurements, (2) chemical information in the scores and loadings, (3) better handling of spectral correlations using latent variables, and perhaps more importantly, (4) diagnostic information in the full spectral residuals. The latter allow to flag samples as outliers if they do not follow the model, a property which has come to be known as the first-order advantage. With MLR it would be impossible to obtain such beautiful results in NIR spectroscopy as the provision of early evidence of non-conformity and contamination of intact foodstuff at the entrance of a feed mill [29].

Variable selection is mandatory in MLR due to the need of having a full-rank calibration data matrix. In PLS/PCR, however, variable selection is in principle not needed, although it may be beneficial in a subtle way. Wavelength selection may provide better quality information to the model, i.e., variables which are more informative as regards the analyte or property of interest [30].

An appropriately simple example may illustrate these concepts. Fig. 3 shows the spectra of four pure components at unit concentration, with component No. 1 being the analyte of interest. The latter shows three active spectral peaks, each of them partially overlapped with the remaining three constituents. With these spectra, a 50-sample calibration and a 100-sample test set were built, both with random component concentrations in the range 0–1 units. For PLS and PCR calibration, the full 50-wavelength spectra were employed with 4 latent variables. For MLR, only a few wavelengths were selected, based on the well-known and efficient successive projection algorithm (SPA) [31]. Gaussian random noise of different size was introduced in all concentrations and signals, and predictions using these algorithms are compared in Table 9. As can be seen, as the relative impact of noise in signal and concentration increases, the average prediction errors increase, as expected. However, the average PLS/PCR prediction error is always lower than that for MLR (the randomization test for comparison of average prediction errors gives $p < 0.05$ in all cases, Section 3.5, meaning that PLS/PCR predictions are significantly better than MLR ones). The sensitivity of both models can be computed as the inverse of the length of the regression coefficients [13]. For PLS/PCR, the analyte sensitivity is computed as 1.8 units (signal × concentration^{-1}), significantly larger than that for MLR, which varied in the range 0.2–0.8 units, depending on the number

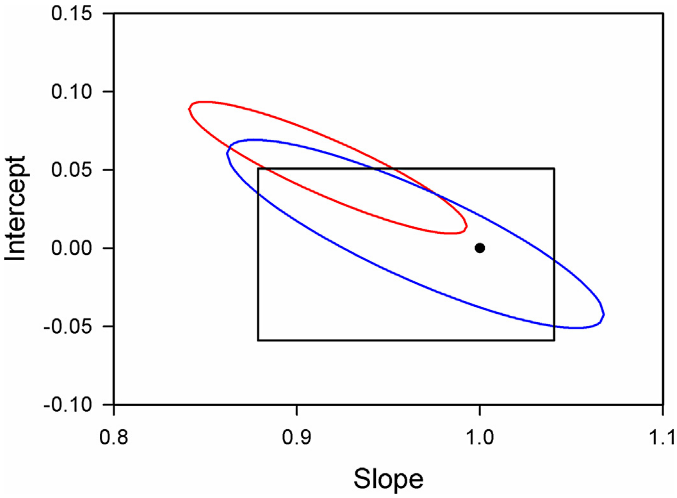


Fig. 2. Different regions in the slope-intercept plane: blue ellipse, EJCR for the slope and intercept estimated by OLS regression, red ellipse, EJCR estimated by WLS regression, black rectangle, region limited by the individual confidence intervals for the slope and intercept. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

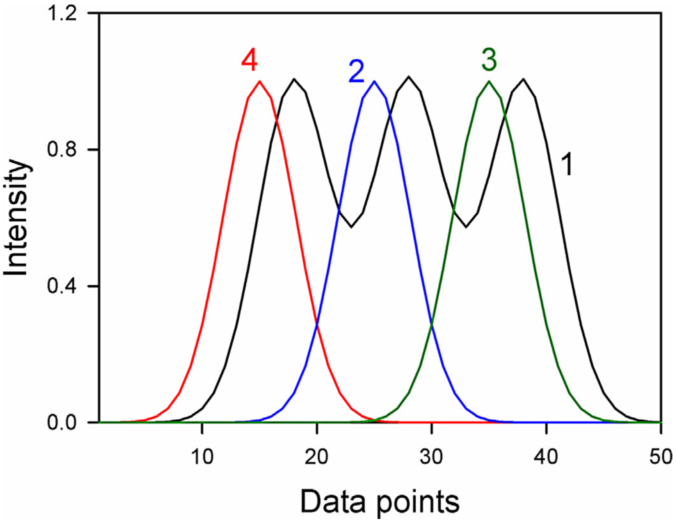


Fig. 3. Simulated overlapped pure spectra for four components. The analyte of interest (No. 1) corresponds to the solid black line, while the remaining blue, green and red solid lines describe the spectra for the remaining sample components No. 2, 3 and 4, as indicated. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 9
Comparison of PLS, PCR and MLR predictive results on a simulated data set, as a function of uncertainty in both signals and concentrations (see text).

Uncertainty in concentrations[a]	Uncertainty in signal[a]	RMSE			LOD_m	
		PLS	PCR	MLR[b]	PLS[c]	MLR
0.1	0.1	0.0026	0.0026	0.0032 (8)	0.004–0.006	0.01–0.02
0.1	1	0.016	0.016	0.041 (6)	0.04–0.05	0.13–0.23
1	0.1	0.014	0.014	0.018 (8)	0.008–0.02	0.04–0.07
1	1	0.028	0.028	0.041 (7)	0.04–0.06	0.16–0.28

[a] Uncertainties are expressed as % of the maximum calibration concentration or signal.
[b] Number of SPA-selected wavelengths between parenthesis.
[c] The limit of detection is only provided for PLS.

of selected wavelengths. This would lead to correspondingly lower detection and quantitation limits for PLS in comparison with MLR, which is confirmed in Table 9, where detection limits are reported using a recent approach (see below). Incidentally, Table 9 shows that PLS and PCR prediction results are of the same quality.

It is also worth noticing that PLS is sometimes employed in the so-called PLS-2 version, which permits the simultaneous calibration of all analytes of interest, at the expense of using the same calibration parameters and spectral regions for all analytes. The more common PLS-1 version, on the other hand, requires one to calibrate a model for each analyte at a time. This allows one to optimize calibration parameters, wavelength regions, etc. in a specific way for each analyte. PLS-1 appears to be the model of choice, with specific advantages over PLS-2 which have been analyzed in detail [32]. The PLS-2 version is recommended only for highly correlated data in the concentration block, which is not the case when the concentration calibrations are carefully designed to have low inner correlations [25].

4.2. Regression coefficients

Various wavelength selection procedures rely on the use (either direct or indirect) of regression coefficients: regions with significant values of the regression vector (either positive or negative) are suggested to be included in the model, while spectral windows where the regression vector is noisy or low-intensity are discarded [33]. Several modifications of this simple strategy are known, including: (1) uninformative variable elimination (UVE) based on the addition of noise [34], and (2) variable importance in projection (VIP) [35].

However, the use of regression coefficients for selecting wavelengths may be dangerous [36]. The selected ranges may only accidentally coincide with known absorption values by the analyte, and thus regression coefficients may misguide the search of useful regions. The conclusions heavily depend on the assumed relation between significant PLS regression coefficients and the relative importance of a given feature or variable [37–39].

There are reasons why care must be taken for the interpretation of regression vectors. One is the contravariance constraint: regression vectors are orthogonal or nearly orthogonal to the space spanned by the interferents, which naturally leads them to show negative peaks, making chemical interpretation difficult [37]. The second one is their dependence on the samples in the calibration and on the signal to noise ratio. In many cases, the largest regression coefficients have no correspondence to the largest bands in the analyte spectra [36].

An example is shown in Fig. 4A, which describes ternary mixtures where the analyte of interest shows a spectrum represented by the black solid line, embedded in mixtures with other two analytes (red and blue solid lines in Fig. 4A). The building of a PLS model leads to regression coefficients (green solid line in Fig. 4B) whose largest (negative) peaks coincide with the response peaks for the interferents and not with those for the analyte itself. Only the two secondary peaks of significantly lower absolute intensity are closer to the analyte peak, although they lie on its sides and not on the maximum. In such cases, wrong information would be obtained by judging analyte responses from regression coefficients.

Alternative methodologies are available, not based on PLS regression coefficients, for variable selection [30].

4.3. Linear vs. non-linear models

Sometimes PCR and PLS are compared in calibration performance with sophisticated non-linear models based on the neural network philosophy [40], such as least-squares support vector machines [41], perceptron networks [41], radial basis functions

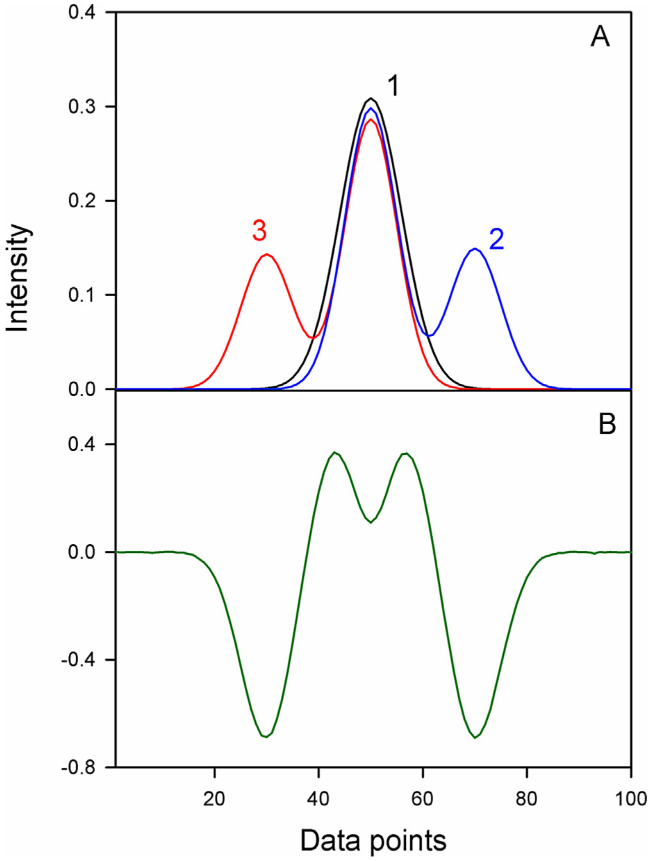


Fig. 4. (A) Simulated overlapped pure spectra for three components. The analyte of interest (No. 1) corresponds to the solid black line, while the remaining blue and red solid lines describe the spectra for the remaining sample components. (B) Vector of regression coefficients obtained by PLS processing from calibration data for mixtures of the three components shown in (A). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

[41], kernel-PLS, etc. [42]. This might appear as self-contradictory: either an adequate underlying model is linear, and hence PCR/PLS should be the calibration tool of choice, or the model is not linear, in which case it might be justified to move to a non-linear calibration model. However, the performance of these models should be carefully checked through suitable statistical procedures. It is necessary to check whether the RMSE for a sufficiently large group of test samples is indeed smaller (statistically significantly, e.g., by using the randomization test commented above). Marginal improvements in predictive ability, i.e., a non-significant decrease in RMSE, should not be regarded as a proof that a non-linear neural network is required to model the data.

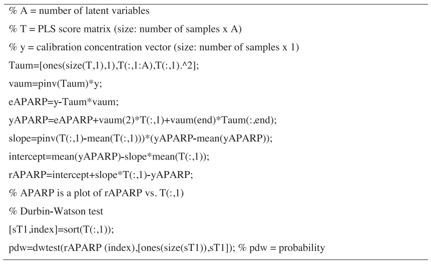
In any case, the lack of linearity can be checked in the multivariate case with suitable statistical tests [43,44]. The results of these tests, combined with a statistically significant improvement in prediction error by applying non-linear models, might constitute a proof that the system behaves in a non-linear manner. One particularly appealing technique is based on the augmented partial residual plots (APARP) which can be constructed for PLS/PCR models [44]. The following is the recommended protocol: (1) model the multivariate data with A latent variables, (2) regress analyte concentrations against an augmented model including the A scores and the squared values of the first score, (3) plot the portion of the concentration data modeled by the first score and the squared first score vs. the first score, (4) compute the residuals of the linear regression of the latter plot, and finally (5) check for the presence of correlations in the residuals of the latter regression using a suitable statistical test [44]. If significant trends are found in these APARP residuals, the data can be considered non-linear. Table 10 shows a short MATLAB code which can be used to apply this methodology. Correlations in the residuals are checked using the Durbin-Watson (DW) test [45]. MATLAB directly estimates the probability associated to the DW statistics:

$$\mathrm{DW} = \frac{\sum_{i=1}^{I-1}(r_{i+1} - r_i)^2}{\sum_{i=1}^{I} r_i^2} \tag{13}$$

using the built-on ‘dwtest·m’ routine, where r_i is the ith APARP residual and I the number of calibration samples. In this case the hypotheses are H_0 (data are linear) and H_a (data are non-linear); the test directly gives the probability p associated to H_0. If the null hypothesis is rejected, the data are declared as non-linear and vice-versa.

Fig. 5 compares the APARP results for the linear example described in Section 4.1 when processed by PLS, and a similarly simulated example where the signal-concentration relationship is non-linear. The probabilities associated to the DW test indicate absence of non-linearities in the APARP of Fig. 5B ($p = 0.8$) and significant non-linearity in Fig. 5D ($p \ll 0.05$). In any case, residual trends for the non-linear case are apparent (Fig. 5D).

Table 10
MATLAB code for implementing the APARP test.



```
% A = number of latent variables
% T = PLS score matrix (size: number of samples x A)
% y = calibration concentration vector (size: number of samples x 1)
Taum=[ones(size(T,1),1),T(:,1:A),T(:,1).^2];
vaum=pinv(Taum)*y;
eAPARP=y-Taum*vaum;
yAPARP=eAPARP+vaum(2)*T(:,1)+vaum(end)*Taum(:,end);
slope=pinv(T(:,1)-mean(T(:,1)))*(yAPARP-mean(yAPARP));
intercept=mean(yAPARP)-slope*mean(T(:,1));
rAPARP=intercept+slope*T(:,1)-yAPARP;
% APARP is a plot of rAPARP vs. T(:,1)
% Durbin-Watson test
[sT1,index]=sort(T(:,1));
pdw=dwtest(rAPARP (index),[ones(size(sT1)),sT1]); % pdw = probability
```

4.4. Mathematical pre-processing

Mathematical pre-processing techniques exist for removing variations in spectra from run to run, which are unrelated to analyte concentration changes [46,47]. The removal of these unwanted effects, e.g., dispersion in near infrared (NIR) spectra of solid or semi-solid materials, usually leads to more parsimonious partial least-squares (PLS) models. The latter require less latent variables than those based on raw data, and often produce better statistical indicators. Usually, however, these tools are applied on a trial-and-error basis, although rational approaches to selecting the best pre-processing have been proposed [48].

Pre-processing techniques applied before PLS calibration should be justified if they lead to considerably simpler and more parsimonious models than regular PLS. For example, orthogonal signal correction (OSC) and other variants [49] can be employed to simplify the models when significant sources of spectral variation due to dispersion in NIR spectroscopy or other phenomena. Proper justification includes a significant decrease in prediction error and number of calibration latent variables. In the study of liquid samples with no dispersion effects, it is preferable to avoid OSC or related procedures, aimed at decreasing the calibration PLS factors by one or two, but leading to insignificant improvement in prediction ability [50].

However, mathematical pre-processing is not always beneficial, as one may naively expect. In some applications, it is essential to leave the dispersion component of the NIR spectra, since they carry information on physical, rather than on chemical properties of the studied materials. For example, when measuring wood density from NIR data, removing the dispersion effects by scattering correction makes the data less sensitive to the target property [51].

Sometimes prediction is shown to improve after spectral smoothing. However, this might be a dangerous activity [52]. The effect of smoothing is most times negligible or only marginal in terms of calibration performance, but sometimes it is detrimental. The reason for the latter result is that smoothing introduces correlations into the noise structure, and regular PCR/PLS (as most multivariate techniques) may lead to worst predictions in comparison with raw data processing if they do not take into account the effect of correlated noise [52].

4.5. Limits of detection and quantitation

In PLS studies, the detection limit is sometimes estimated using a univariate approach, regressing predicted vs. nominal analyte concentrations for the calibration set, and computing $\mathrm{LOD_u}$ from Eq. (6) [53]. However, this procedure is debatable, since it provides a single PLS detection limit, whereas other authors have suggested that sample-specific LOD values exist, which depend on the level of other background constituents [54].

Recently, the multivariate detection limit for PLS has been suggested to be available in the form of a range of values, whose lower and upper limits are given by [55]:

$$\mathrm{LOD_{min}} = 3.3\sqrt{\frac{var(x)(1 + h_{0\mathrm{min}})}{\parallel \beta \|^2} + h_{0\mathrm{min}} var(c_{\mathrm{cal}})} \tag{14}$$

$$\mathrm{LOD_{min}} = 3.3\sqrt{\frac{var(x)(1 + h_{0\mathrm{max}})}{\parallel \beta \|^2} + h_{0\mathrm{max}} var(c_{\mathrm{cal}})} \tag{15}$$

where β is the vector of regression coefficients, indicates the norm or vector length, var(x) is the variance in the instrumental signal,

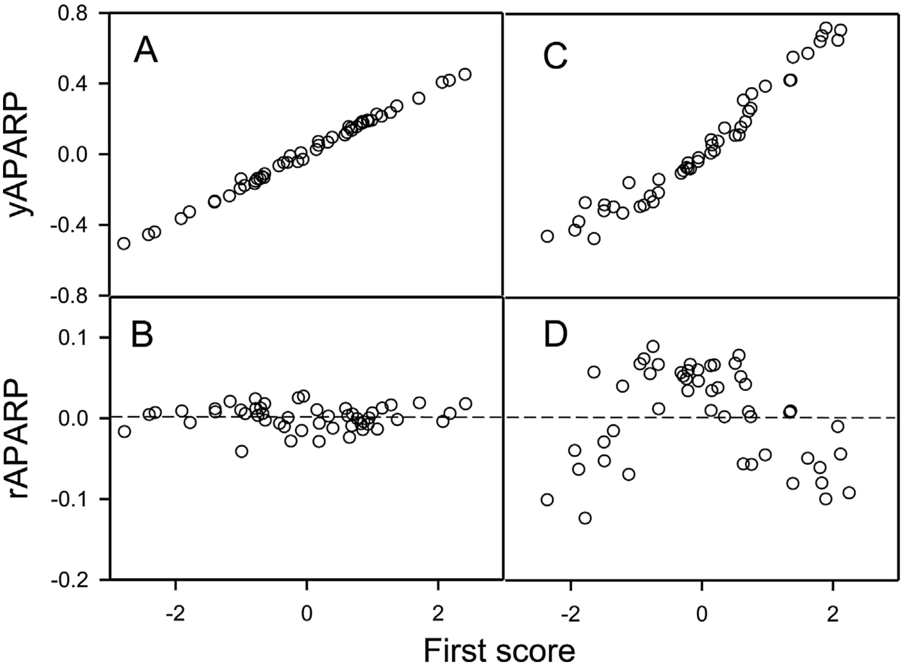


Fig. 5. Detection of non-linearities in two simulated examples: (A) and (B) plots correspond to linear data, and (C) and (D) plots to non-linear data. Both data sets involve 4 calibrated components and 50 calibration samples. Plots (A) and (C) show the modelled part by the first score and its squared values (yAPARP in Table 10) as a function of the first scores (T(:,1) in Table 10). Plots (B) and (D) show the residuals (rAPARP in Table 10) of the linear regression of plots (A) and (C) as a function of the first scores. The probabilities associated to the Durbin-Watson statistics of the residuals are 0.8 for the linear case (plot B) and ≪0.05 for the non-linear case (plot D).

var(c_{cal}) is the variance in calibration concentrations, and $h_{0\mathrm{min}}$ and $h_{0\mathrm{max}}$ are the minimum and maximum values of the leverage at the blank level. The interpretation of the factor 3.3 in Eqs. (14) and (15) is the same as that given for univariate calibration. The value of $h_{0\mathrm{min}}$ in Eq. (14) is identical to h_0 in Eq. (7), while $h_{0\mathrm{max}}$ is:

$$h_{0\mathrm{max}} = \max\left(h_i + h_{0\mathrm{min}}\left[1 - \left(\frac{c_i - \bar{c}_{\mathrm{cal}}}{\bar{c}_{\mathrm{cal}}}\right)^2\right]\right) \qquad (16)$$

where h_i and c_i are the leverage and (uncentered) analyte concentration of a generic calibration sample (mean-centering is assumed in building the PLS model). The relationship between the univariate approach, which provides a single detection limit, and the range of detection limits given by the new approach has been discussed in Ref. [55].

A MATLAB software for first-order calibration is available at www.iquir-conicet/descargas/mvc1.rar [56], which includes the estimation of figures of merit according to the latest findings.

4.6. Selectivity

According to IUPAC, selectivity can be defined as the extent to which a method can be used to determine individual analytes in mixtures or matrices without interferences from other constituents of like behavior [57]. In practical terms, it can be evaluated as the ratio between sensitivity and unit-concentration pure analyte signal. In the context of first-order multivariate calibration, the sensitivity can be defined as the net analyte signal at unit concentration, and estimated as the inverse of the norm of the regression vector [13].

The above selectivity definition is debatable, because it cannot be applied to inverse latent-structured methods such as PLS, where no approximations are available to pure analyte spectra [13]. An alternative is to replace the pure analyte signal by the signal for a test sample as denominator in the selectivity expression, although this makes the selectivity sample-dependent: two test samples A and B having the same number of chemical constituents should display the same analyte selectivity. However, if the signal for sample B is larger than that for sample A because constituents other than the analyte are more concentrated in B than in A, the selectivity would be larger in A than in B, which is unreasonable. Therefore, it may only be sensible to define the selectivity when the pure analyte signal is either known from separate experiments, or is adequately retrieved by the data processing algorithm.

4.7. Multivariate standard addition

Standard addition is designed to circumvent the effect of a background on the analyte response leading to a change in sensitivity, i.e., a change in the slope of the univariate signal-concentration relationship. The generalized standard addition method (GSAM) [58], is the first-order multivariate counterpart of univariate standard addition, and is realized by measuring first-order data for various overlapping analytes embedded in a sample background. Generalized standard addition not only demands knowledge of the number and identity of the analytes, but also that standards of each of them are available, in order to be added in perfectly known amounts to each sample. In any case, the limitations of this method regarding the background effects are analogous to those for the univariate standard addition mode.

A background signal arising from responsive non-analytes constitutes an interference in univariate analysis, and cannot be corrected by means of standard addition. This is typical of most biological samples, where the second-order advantage is required for successful quantitation. The presence of a responsive background, which affects the analyte response in a sample (for example, through analyte-background interactions such as complex formation or protein binding) requires at least second-order standard addition for analyte quantitation [59].

4.8. First- vs. second-order data

Whenever there is the possibility of measuring and processing second-order data, they should be preferred over first-order data, because of various reasons, including the second-order advantage, i.e., quantitating analytes in test mixtures in the presence of interferents which have not been included in the calibration phase [3]. This means that if the same instrument employed for registering first-order data is also able to provide second-order data at no extra cost, the latter should be the option of choice.

For example, first-order synchronous fluorescence spectra are sometimes employed to perform first-order multivariate calibration, on the basis that they are more selective than either excitation or emission spectra [60]. Some efforts have also been made in optimizing the wavelength offset for synchronous fluorescence optimal results. However, modern spectrofluorimeters are capable of measuring second-order fluorescence excitation-emission matrices (EEM) at no extra cost and very rapidly; these second-order data are immensely more powerful than first-order synchronous fluorescence data. This is because EEM data: (1) do not require optimization of the wavelength offset, simply because they use the complete data matrix, (2) they allow one to achieve the second-order advantage, and (3) they provide additional selectivity and sensitivity to that obtained by first-order measurements [61]. This should be convincing enough to move from first-order synchronous spectra to second-order fluorescence matrices.

A similar situation is found when registering liquid chromatograms at a single detection wavelength, either UV-vis absorption or fluorescence emission, and the detector is able to measure multi-wavelength spectra. The latter measuring mode leads to elution time-spectral data matrices, which may allow for background and interference corrections without sample clean-up [62].

5. Conclusions

A series of practical guidelines has been provided for the processing and reporting of both univariate and multivariate calibration data, in accordance with international standards and official protocols. Following a set of reporting rules contributes to the use of a common analytical language, and aids in reaching mutual understanding among analytical users.

Acknowledgements

Financial support from Universidad Nacional de Rosario (Project BIO342), Consejo Nacional de Investigaciones Científicas y Técnicas (Project PIP-163) and Agencia Nacional de Promoción Científica y Tecnológica (Project PICT 2013-0136) is acknowledged.

References

[1] K. Danzer, L.A. Currie, Guidelines for calibration in analytical chemistry. Part 1. Fundamentals and single component calibration, Pure Appl. Chem. 70 (1998) 993–1014.
[2] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers- Verbeke, Handbook of Chemometrics and Qualimetrics, Part A, Elsevier, Amsterdam, 1997 (Chapter 36).
[3] A.C. Olivieri, G.M. Escandar, Practical Three-way Calibration, Elsevier, Waltham, MA, USA, 2014.
[4] ASTM E29, American Society for Testing and Materials, Standard Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications, January (1) (1967).
[5] L. Cuadros Rodríguez, A.M. García Campaña, C. Jiménez Linares, M. Román Ceba, Estimation of performance characteristics of an analytical method using the data set of the calibration experiment, Anal. Lett. 26 (1993) 1243–1258.
[6] S. Brown, R. Tauler, B. Walczak (Eds.), Comprehensive Chemometrics, 1, Elsevier, Amsterdam, 2009 (Chapters 1.09–1.20).
[7] R. Leardi, Experimental design in chemistry: a tutorial, Anal. Chim. Acta 652 (2009) 161–172.
[8] L. Vera-Candioti, A.C. Olivieri, H.C. Goicoechea, Simultaneous multiresponse optimization applied to epinastine determination in human serum by using capillary electrophoresis, Anal. Chim. Acta 595 (2007) 310–318.
[9] G. Derringer, R. Suich, Simultaneous optimization of several response variables, J. Qual. Technol. 12 (1980) 214–219.
[10] M.C. Ortiz, M.S. Sánchez, L.A. Sarabia, Quality of analytical measurements: univariate regression, 1, 127–169 (Chapter 1.05).
[11] Analytical Methods Committee, Is my calibration linear? Analyst 119 (1994) 2363–2366.
[12] L.A. Currie, Recommendations in evaluation of analytical methods including detection and quantification capabilities, Pure Appl. Chem. 67 (1995) 1699–1723.
[13] A.C. Olivieri, Analytical figures of merit: from univariate to multiway calibration, Chem. Rev. 114 (2014) 5358–5378.
[14] MATLAB, The Mathworks, Natick, Massachussets.
[15] European Commission, Decision 2002/657/EC, implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results, OJ L221, 17.8.2002, 8–36.
[16] R.C. Castells, M.A. Castillo, Systematic errors: detection and correction by means of standard calibration, Youden calibration and standard additions method in conjunction with a method response model, Anal. Chim. Acta 423 (2000) 179–185.
[17] J.M. Andrade, M.G. Estévez-Pérez, Statistical comparison of the slopes of two regression lines: a tutorial, Anal. Chim. Acta 838 (2014) 1–12.
[18] V.A. Lozano, R. Tauler, G.A. Ibañez, A.C. Olivieri, Standard addition analysis of fluoroquinolones in human serum in the presence of the interferent salicylate using lanthanide-sensitized excitation-time decay luminescence data and multivariate curve resolution, Talanta 77 (2009) 1715–1723.
[19] H. van der Voet, Comparing the predictive accuracy of models using a simple randomization test, Chemom. Intell. Lab. Syst. 25 (1994) 313–323.
[20] J.N. Miller, J.C. Miller, Statistics Chemometrics for Analytical Chemistry, fourth ed., Prentice-Hall, Harlow, UK, 2000.
[21] A.G. González, M.A. Herrador, A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles, Trends Anal. Chem. 26 (2007) 227–238.
[22] A.G. González, M.A. Herrador, A.G. Asuero, Intra-laboratory testing of method accuracy from recovery assays, Talanta 48 (1999) 729–736.
[23] J. Riu, F.X. Rius, Assessing the accuracy of analytical methods using linear regression with errors in both axes, Anal. Chem. 68 (1996) 1851–1857.
[24] V. Franco, V.E. Mantovani, H.C. Goicoechea, A.C. Olivieri, Teaching chemometrics with a bioprocess: analytical methods comparison using bivariate linear regression, Chem. Edu. 7 (2002) 265–269.
[25] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[26] P.D. Wentzell, L. Vega Montoto, Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures, Chemom. Intell. Lab. Syst. 65 (2003) 257–279.
[27] G.-C. Zhang, Z. Li, X.-M. Yan, C.-G. Cheng, P. Zhou, G.-L. Lin, C.-J. Zhou, N. Liu, X.-R. Han, Rapid analysis of apple leaf nitrogen using near infrared spectroscopy and multiple linear regression, Commun. Soil Sci. Plant Anal. 43 (2012) 1768–1772.
[28] D.M. Haaland, E.V. Thomas, Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information, Anal. Chem. 60 (1988) 1193–1202.
[29] J.A. Fernández-Pierna, O. Abbas, B. Lecler, P. Hogrel, P. Dardenne, V. Baeten, NIR fingerprint screening for early control of non-conformity at feed mills, Food Chem. (2014) , doi:http://dx.doi.org/10.1016/j.foodchem.2014.09.105 (in press).
[30] T. Mehmood, K.H. Liland, L. Snipen, S. Sæbø, A review of variable selection methods in partial least squares regression, Chemom. Intell. Lab. Syst. 118 (2012) 62–69.
[31] M.C. Ugulino Araújo, T.C. Bezerra Saldanha, R. Kawakami Harrop Galvão, T. Yoneyama, H. Caldas Chame, V. Visani, The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, Chemom. Intell. Lab. Syst. 57 (2001) 65–73.
[32] R. Manne, Analysis of two PLS algorithms for multivariate calibration, Chemom. Intell. Lab. Syst. 2 (1987) 187–197.
[33] A. Garrido-Frenich, D. Jouan-Rimbaud, D.L. Massart, S. Kuttatharmmakul, M. Martínez- Galera, J.L. Martínez-Vidal, Wavelength selection method for multicomponent spectrophotometric determinations using partial least squares, Analyst 120 (1995) 2787–2792.
[34] V. Centner, D.L. Massart, O.E. de Noord, S. de Jong, B.M. Vandeginste, C. Sterna, Elimination of uninformative variables in multivariate calibration, Anal. Chem. 68 (1996) 3851–3858.
[35] I.-G. Chong, C.-H. Jun, Performance of some variable selection methods when multicollinearity is present, Chemom. Intell. Lab. Syst. 78 (2005) 103–112.
[36] C.D. Brown, R.L. Green, Critical factors limiting the interpretation of regression vectors in multivariate calibration, Trends Anal. Chem. 28 (2009) 506–514.
[37] M.B. Seasholtz, B.R. Kowalski, Qualitative information for multivariate calibration models, Appl. Spectrosc. 44 (1990) 1337–1348.
[38] A.J. Burnham, J.F. MacGregor, R. Viveros, Interpretation of regression coefficients under a latent variable regression model, J. Chemometr. 15 (2001) 265–284.
[39] O.M. Kvalheim, T.V. Karstaag, Interpretation of latent- variable regression models, Chemom. Intell. Lab. Syst. 7 (1989) 39–51.
[40] F. Despagne, D.L. Massart, Neural networks in multivariate calibration, Analyst 123 (1998) 157R–178R.

[41] S. Haykin, Neural Networks. A Comprehensive Foundation, second ed., Prentice-Hall, Upper Saddle River, NJ, USA, 1999.

[42] T. Czekaj, W. Wu, B. Walczak, About kernel latent variable approaches and SVM, J. Chemometr. 19 (2005) 341–354.

[43] S.V.C. de Souza, R.G. Junqueira, A procedure to assess linearity by ordinary least squares method, Anal. Chim. Acta 552 (2005) 25–35.

[44] V. Centner, O.E. de Noord, D.L. Massart, Detection of nonlinearity in multivariate calibration, Anal. Chim. Acta 376 (1998) 153–168.

[45] J. Durbin, G.S. Watson, Testing for serial correlation in least squares regression I, Biometrika 37 (1959) 409–428.

[46] P. Geladi, D. MacDougall, H. Martens, Linearization and scatter-correction for near-infrared reflectance spectra of meat, Appl. Spectrosc. 39 (1985) 491–500.

[47] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra, Appl. Spectrosc. 43 (1989) 772–777.

[48] C.D. Brown, Rational Approaches to Data Preprocessing in Multivariate Calibration, Ph.D. Thesis, Dalhousie University, Halifax, Nova Scotia, Canada, 2000.

[49] S. Wold, H. Antti, F. Lindgren, J. Öhman, Orthogonal signal correction of near-infrared spectra, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.

[50] T. Fearn, On orthogonal signal correction, Chemom. Intell. Lab. Syst. 50 (2000) 47–52.

[51] T.A. Lestander, J. Lindeberg, D. Eriksson, U. Bergsten, Prediction of Pinus sylvestris clearwood properties using NIR spectroscopy and biorthogonal partial least squares regression, Can. J. For. Res. 38 (2008) 2052–2062.

[52] C.D. Brown, P. Wentzell, Hazards of digital smoothing filters as a preprocessing tool in multivariate calibration, J. Chemometr. 13 (1999) 133–152.

[53] M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, Tutorial on evaluation of type I and type II errors in chemical analyses: from the analytical detection to authentication of products and process control, Anal. Chim. Acta 674 (2010) 123–142.

[54] R. Boque, M.S. Larrechi, F.X. Rius, Multivariate detection limits with fixed probabilities of error, Chemom. Intell. Lab. Syst. 45 (1999) 397–408.

[55] F. Allegrini, A.C. Olivieri, IUPAC-consistent approach to the limit of detection in partial least-squares calibration, Anal. Chem. 86 (2014) 7858–7866.

[56] A.C. Olivieri, H.C. Goicoechea, F.A. Iñón, MVC1: an integrated Matlab toolbox for firstorder multivariate calibration, Chemom. Intell. Lab. Syst. 73 (2004) 189–197.

[57] J. Vessman, R.I. Stefan, J.F. van Staden, K. Danzer, W. Lindner, D.T. Burns, A. Fajgelj, H. Müller, Selectivity in analytical chemistry (IUPAC recommendations 2001), Pure Appl. Chem. 73 (2001) 1381–1386.

[58] B.E.H. Saxberg, B.R. Kowalski, Generalized standard addition method, Anal. Chem. 51 (1979) 1031–1038.

[59] V.A. Lozano, R. Tauler, G.A. Ibañez, A.C. Olivieri, Standard addition analysis of fluoroquinolones in human serum in the presence of the interferent salicylate using lanthanide-sensitized excitation-time decay luminescence data and multivariate curve resolution, Talanta 77 (2009) 1715–1723.

[60] S. Rubio, A. Gomez-Hens, M. Valcarcel, Analytical applications of synchronous fluorescence spectroscopy, Talanta 33 (1986) 633–640.

[61] A.C. Olivieri, Analytical advantages of multivariate data processing. One, two, three, infinity? Anal. Chem. 80 (2008) 5713–5720.

[62] V. Boeris, J.A. Arancibia, A.C. Olivieri, Determination of five pesticides in juice, fruit and vegetable samples by means of liquid chromatography combined with multivariate curve resolution, Anal. Chim. Acta 814 (2014) 23–30.