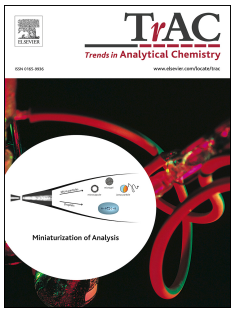
# Journal Pre-proof




Assessment of goodness-of-fit for the main analytical calibration models: Guidelines and case studies

Francisco Raposo, Damià Barceló

PII: S0165-9936(21)00196-5

DOI: https://doi.org/10.1016/j.trac.2021.116373

Reference: TRAC 116373

To appear in: *Trends in Analytical Chemistry*

Received Date: 25 February 2021

Revised Date: 14 June 2021

Accepted Date: 14 June 2021

Please cite this article as: F. Raposo, D. Barceló, Assessment of goodness-of-fit for the main analytical calibration models: Guidelines and case studies, *Trends in Analytical Chemistry*, https://doi.org/10.1016/j.trac.2021.116373.

# ANALYTICAL CALIBRATION

## Goodness-of-Fit

**Graphical Plots**

- *General x-y plot*
- *Residuals plot*

**Statistical Significance Tests**

- *Linearity and Homoscedasticity Tests*

**Numerical Parameters**

- *$R^2$*
- *x-Residuals (%RE)*

# Assessment of goodness-of-fit for the main analytical calibration models: Guidelines and case studies

Francisco Raposo[1]*; Damià Barceló [2,3]

[1] Instituto de la Grasa, Consejo Superior de Investigaciones Científicas (IG-CSIC), Campus Universitario Pablo de Olavide, Carretera de Utrera km 1, Edificio 46 41013 Seville, Spain

[2]Water and Soil Quality Research Group, Department of Environmental Chemistry, IDAEA-CSIC C/ Jordi Girona 18-26, Barcelona, 08034, Spain

[3]Catalan Institute for Water Research (ICRA), C/Emili Grahit 101, Girona, 17003, Spain

* To whom correspondence should be addressed: (Tel.: +34 954611550; E-mail: fraposo@ig.csic.es)

## ABSTRACT

This critical review paper will discuss the main analytical calibration models as well as the guidelines for their practical use. The main models used to fit a multiple-point calibration dataset are: 1) linear unweighted or ordinary least squares regression (OLSR); 2) quadratic unweighted least squares regression (QLSR); 3) linear weighted least squares regression (WLSR). Unfortunately, there is no standard procedure in analytical chemistry for objectively testing the goodness-of-fit of calibration models. Different proposals were reported in the literature. However, none is more commonly used, and probably not more controversial than $R^2$. In this document, a three step simple calibration diagnosis has been proposed. It is based on a combination of different procedures such as graphical plots, statistical significance tests and numerical parameters. Experimental conditions and design of calibration procedures are very relevant for appropriate selection. Finally, some information on the choice of the different models will be reported in four case studies.

## KEYWORDS

Back-calculated concentration; Calibration; Determination coefficient; Goodness-of-fit; Least squares regression; Linear regression, Quadratic regression; Relative error; Weighted regression

## ABBREVIATIONS

$b_0$: intercept; $b_{0w}$: weighted intercept; $b_1$: slope; $b_{1w}$: weighted slope; $b_2$: quadratic coefficient; C.I: confidence interval; C.V: coefficient of variation; $F_{exp/cri}$: experimental/ critical values of Fisher-Snedecor test; GoF: goodness-of-fit; LOQ: limit of quantification; LSR: least squares regression; OLSR: linear unweighted or ordinary least squares regression; QLSR: quadratic unweighted least squares regression; *%RE*: percentage of relative error; *s*: standard deviation; SQT: significance of quadratic term; $s_{y/x}$: residual variance; $t_{exp/cri}$: experimental/ critical values of t-test; *x*: independent variable (concentration or amount); $x_{BC}$: back-calculated or experimental concentration (or amount); $x_{NC}$: nominal or theoretical concentration (or amount); *y*: dependent variable (instrument signal or response); WLSR: linear weighted least squares regression

## CONTENTS

## 1. Introduction to analytical calibration for method development and validation

The calibration models used with analytical methods are at the core of analytical science due to their importance for quantification and must be considered carefully during development and validation of methods [1]. Calibration procedures are carried out by regression analysis, which is a statistical inference method that estimates the relationship between a dependent variable and one (univariate) or more (multivariate) independent variable(s) or predictor(s). From the methodological calibration point of view, in a regression analysis the instrument signal or response (*y*) is assumed to be a dependent variable, while the concentration or amount (*x*) is assumed to be an independent variable [2]. Each regression model can be characterized in terms of method, order and fitting technique. The method of least squares regression (LSR) is a standard statistical approach in regression analysis and is considered the most common statistical method used for analytical calibration. However, for LSR different polynomial functions can be used according to:

- The selection of the model order. The polynomial function can be of first-order, giving a straight-line relationship. Alternatively, a second-order or quadratic polynomial can be used in the case of a curved relationship [3, 4].
- The selection of the model fitting technique. Two options are available: weighted and unweighted. The model chosen depends on whether there is variation of absolute precision from instrument response data [5, 6].

Consequently, from a theoretical viewpoint, different models can be designated to fit a calibration dataset. Selection of a calibration model is important because the choice of calibration function defines the mathematical equation that relates both variables and is therefore used to calculate the concentration or amount of analyte(s) in unknown samples [7]. Even though calibration models are used in analytical science, a reduction of accuracy in some routine analytical measurements can be caused by the inappropriate selection of the specific LSR calibration model [8]. There are three main LSR models which can be used to fit a multiple-point calibration dataset such as: 1) linear unweighted or ordinary (OLSR); 2) quadratic unweighted (QLSR); 3) linear weighted (WLSR). It is important to note that the quadratic weighted model and some other non-linear models can also be found reported in the literature [9, 10] but they are used less frequently than the models discussed in this paper. The most common model reported in all types of laboratories is OLSR because it is simple, easy to use and easy to interpret. However, it must be pointed out that OLSR is not always the most appropriate fitting model to use with many calibration methods routinely used in laboratories. On the contrary, it should be avoided in some specific areas such as bioanalysis, environmental or food analysis, where the calibration model must be applied to a broad concentration range [8, 11].

As different LSR models may yield different results, the reliability of the quantitative performance of the analytical method may be highly dependent on the calibration model selected. Unfortunately, there is no standard procedure in the analytical chemistry field for objectively testing the fit of calibration curves and how to proceed when in doubt regarding the best model to be selected. A model “good fit” might be a model that describes a high proportion of the variability in the experimental data and is able to predict new observations with high certainty [12]. Regrettably, recommendations as to how to make these choices are frequently vague and do not address all necessary decisions [13].

This manuscript has three main objectives. Firstly, to describe the main analytical calibration models according to their performance. Secondly, to demonstrate how to choose an appropriate calibration function based on the goodness-of-fit concept

(GoF). Thirdly, to highlight the practical applications of this approach through some case studies.

## 2. Main analytical calibration models obtained by LSR method

The statistical method of LSR is a standard approach in regression analysis for which the objective is to find the best fit through the experimental data that minimizes the value of the residual sum of squares [14]. The statistical term residual is used as a measure of the discrepancy between experimental and estimate values provided by the selected regression model for each observation [15]. This study was limited to three candidate LSR models: OLSR, QLSR and WLSR. These regression models for multiple-point analytical calibration datasets are the most frequently used as a basis for method development and validation.

### 2.1. OLSR model

This, due to its simplicity, is the most common regression approach for using the LS method to fit a calibration model to bivariate data points ($x$, $y$) due to its simplicity. This model makes some underlying assumptions about the nature of errors linked to instrument response values such as normality, linearity, constant variance and independence [9, 15, 16]. Unfortunately, in most cases analysts routinely apply calibration models with no preliminary considerations about the fulfilment of the above mentioned assumptions [8]. For each assumption that does not hold true, there is a chance that the reliability of the statistical results is compromised. As a consequence, OLSR model may produce suboptimal quantitation of estimated results since one or more of the above-listed assumptions are often violated [11]. For example, the assumption of linearity cannot be supported in the case of many calibration curves, particularly at high concentration ranges. In addition, the assumption of constant variance is seldom true for most instrumental analytical methods based on the regression of instrument response versus concentration or amount as variables [11, 17]. Given that it is quite probable that some analytical methods do not provide a linear relationship or that the variance is not constant, approaches other than OLSR, such as non-linear or weighted approaches should be used, respectively. Additionally, there are some situations in which the new calibration model should be established from data with the extra complexity of being non-linear and without constant variance [18].

### 2.2. QLSR model

Application of the OLSR model means that the assumption of a linear relationship between both variables should be true. Some instrumental analytical techniques produce calibration graphs that may be linear over several orders of magnitude [19]. whereas some are essentially non-linear even at low concentration range [16]. Furthermore, detectors frequently display a linear response over a restricted concentration range, but become non-linear at higher concentrations [17]. That means, a calibration graph is generally linear at low analyte concentration but shows negative deviations or curvature at higher analyte concentration [19].

There are several reasons why the application of a non-linear calibration model is overlooked when a calibration dataset shows a minor degree of curvature. Firstly, non-linear calibration models are challenging because the mathematics behind the model can be difficult to understand, especially when the data are noisy [20]. Secondly, non-linear calibration models are more complicated because many different polynomial equations have to be considered [16]. A quadratic model is

frequently reported to fit calibration curves [21]. However, when a curvature is present, in some cases it is difficult to cope with a single quadratic regression that could be used over the entire range [22].

The difference between the equations provided by linear and quadratic models is evident:

*(1) Linear model:* $y = b_1 \cdot x + b_0$

*(2) Quadratic model:* $y = b_2 \cdot x^2 + b_1 \cdot x + b_0$

where x is the concentration value (or amount) and $b_0$, $b_1$ and $b_2$ are coefficients to define the intercept, slope and quadratic term, respectively. However, in essence, the only difference between the quadratic and the linear equation is provided by the inclusion of a second-order term used to describe the deviation from linearity. Hence, a quadratic model should not be regarded as overly complex.

## 2.3. WLSR model

Application of the OLSR model means that the assumption that precision remains constant over the entire working range should be fulfilled. In analytical science it is very often reported that absolute precision (in the form of standard deviation or variance) changes throughout the concentration range. This behaviour is called heteroscedasticity [22]. This is the normal trend in most bioanalytical assays, where a wide concentration range is covered, and therefore the variance is expected to increase with concentration [23]. This fact has a lot of influence in the selection of the regression model because deviations at a high concentration range will affect the calibration model more than deviations at a low concentration range [7]. This is due to the data with the biggest variance making the largest contribution to the residual sum of squares values and dominating the selection of the regression coefficients for the model equation. Thus, the use of OLSR as an unweighted (or more appropriately equally weighted) model for a calibration curve with heteroscedastic data will lead to compromised accuracy due to the coefficient estimates being less reliable [24]. Therefore, when the variance of any calibration dataset is not homogeneous, a WLSR model should be used to produce the calibration curve [25]. It is important to highlight that a WLSR model could yield significant differences, when compared to using an OLSR model, particularly when calibration datasets at low analyte concentrations are evaluated [26].

Weighting is specifically designed to compensate for the limitation of LSR models when applied to heteroscedastic experimental data and the contribution of such data to the residual sum of squares. More importance (weight) is applied to more consistent experimental data (i.e. data with low variance), and less importance (weight) is applied to less consistent experimental data (i.e. data with higher variance). Therefore, an optimal weighted fitting technique is necessary to balance the regression model in a mode in which the error is uniformly distributed throughout the whole calibration range [24].

Another important subject is the selection of weighting factors for appropriately resolving the heteroscedastic data generated. It has been frequently described as a weighting procedure based on the use of the inverse of variance response [23]. Alternatively, the inverse of variance has been replaced by either the inverse of concentration (or amount) and the intensity of instrumental signal. Therefore, weighting factors as functions of $1/x^{0.5}$, $1/x$, $1/x^2$ and $1/y^{0.5}$, $1/y$, $1/y^2$ have been

proposed to establish regression models [27, 28]. The selection of weighting factors based on $1/x^2$ has been suggested in the literature as being most appropriate [27-29]

## 3. Goodness-of-fit for analytical calibration models

Multiple-point calibration curves are commonly used in the analytical chemistry field for quantitative analysis, that means using the calibration model equation to estimate the concentration values of analyte(s) in unknown samples [1, 2, 7]. The need to ensure equally good fits at both low and high concentration ranges has already been mentioned. However, it is quite improbable that the calibration model will exactly match the instrument response versus concentration function over the entire working range. Furthermore, when the application of various calibration models is possible, a reliable tool is necessary to allow the different models to be compared.

The selection of an appropriate calibration model for a defined analytical procedure is essential. FDA method validation guidelines state~~s~~ that: "standard curve fitting is determined by applying the simplest model that satisfactorily describes the concentration versus response association using appropriate weighting and statistical tests for GoF" [30]. So, an appropriate method to evaluate GoF is important.

Different methodologies are described in the scientific literature to evaluate GoF of calibration models [31,32]. They can be classified into three main types: graphical plots, statistical tests and numerical parameters. Unfortunately, estimation of GoF is often a source of disagreement among the analytical chemistry community because the different methods frequently produce contradictory results. As a consequence, there is no consensus about which standard procedure and criteria are required to evaluate how well the selected calibration model fits the calibration data.

### 3.1. Graphical plots

Two types of plots are used: general *x-y* trend and residual errors.

#### 3.1.1. General *x-y* plot

Plotting the original dataset is often carried out to organize data, so the overall relationship between both variables (instrument response versus concentration) can be visualized. Accordingly, this graphical or scatter plot can be useful to:

- Evaluate the general trend i.e does the data show a linear or non-linear relationship? If there is some degree of curvature, this will probably be visually apparent.
- Highlight data points which may be outliers.

It is important to note that visual inspection of the regression plot is not sufficient to fully investigate the validity of the proposed calibration model.

#### 3.1.2. Residuals plot

A conventional residuals plot is a graph that displays the residuals values on the vertical *y*-axis and the independent variable on the horizontal *x*-axis. It is very useful in validating a calibration model by helping with the visual interpretation of the magnitude to which the model accounts for deviation in the dataset. Using residual plots can be vital in helping to choose the correct model by means of assessing whether the observed error (residuals) is consistent with the random error [33]. Ideally, residual values for OLSR model should be [15, 34]:

- randomly scattered around zero for the complete range of fitted values.
- centered on zero (indicating that the model's estimates are right on average rather than frequently too high or low).

- normally distributed with the same degree of scattering for all fitted values.

On the other hand, if a residual plot shows a non-random structure, it is a clear sign that the model fits the data poorly [15]. Some datasets which are not good candidates for regression, include:

- Datasets with U-shaped or inverted U-shaped curvature, suggesting a QLSR model would yield a better fit.
- Heteroscedastic datasets, indicated by the presence of a "funnel or trumpet-like" pattern. I these cases a WLSR model may yield a better field.

Residual plots are very effective, but they cannot be considered as an irrefutable tool for properly investigating the validity of an OLSR model. This is due to the lack of statistical significance in relation to identifying deviations from assumptions. Additionally, some degree of practice may be necessary for understanding these plots [31].

### 3.2. Statistical significance tests

Statistical significance is likely to be found when a relationship concerning two or more variables in an analysis is not purely unintentional, but it is in fact caused by another factor.

#### 3.2.1. Statistical testing of the linearity assumption

GoF in the form of linearity statistics is useful for increasing the statistical rigour when it is not clear if the association between variables in the selected calibration model is not clear if it is linear (OLSR) or curved (QLSR). For this reason, two main statistical tests are suggested. These are Mandel´s and significance of quadratic term (SQT) tests.

##### 3.2.1.1. Mandel´s test

This is an *F*-test which works by comparing the residual variance values resulting from the linear and the quadratic models, with a null hypothesis that there is not a significant difference between both models.

$$(3)\ F_{Mandel} = [(N-2) \cdot (s_{y/x})^2_{LIN} - (N-3) \cdot (s_{y/x})^2_{QUA})] / (s_{y/x})^2_{QUA}$$

where N is the overall number of calibration standards necessary to calculate the degree of freedom and $s_{y/x}$ is the residual variance. Values of $F_{exp}$ are compared against $F_{crit}$, or alternatively the complementary *p*-value corresponding to $F_{exp}$, to evaluate if the QLSR model can offer a significantly better fit than the OLSR model [35, 36].

##### 3.2.1.2. SQT test

This is a *t*-test where the experimental value is calculated as a ratio between the quadratic coefficient and its corresponding standard deviation in the form:

$$(4)\ t_{exp} = b_2 / s_{b2}$$

There are three ways to evaluate the SQT considering a null hypothesis that the $b_2$ coefficient is not significant [37, 38]:

- Comparing $t_{exp}$ versus $t_{crit}$.
- Assessing the complementary *p*-value corresponding to $t_{exp}$.
- Assessing when the $b_2$ coefficient is statistically different from zero, that means the confidence interval where the $b_2$ coefficient includes zero.

#### 3.2.2. Statistical testing of the homoscedasticity assumption

GoF in the form of variance homogeneity is useful in statistical reasoning when there is doubt if the linear association between variables in the selected calibration model

is unweighted (OLSR) or weighted (WLSR). Testing for homogeneity of variances requires the measurement of replicates at constant values of the predictor variable [39]. There are numerous statistical tests reported in the literature as suitable to check the existence of homogeneity/heterogeneity relating to variances (or standard deviation). Typically, a simple Fisher-Snedecor *or F*-test between variances using the highest and lowest experimental data point has been suggested in some documents [27, 40, 41]. In addition, some of the most convenient tests were previously reviewed [42]. In this manuscript, five different statistical tests: Bartlett, Cochran, Fisher-Snedecor, Hartley and Levene were recommended for further evaluation of instrument response behaviour.

### 3.3. Numerical parameters

The numerical parameters for model validation tend to be more narrowly focused on a particular aspect of the calibration dataset. They often try to simplify the analytical information into a single number. However, some cautiousness is necessary in the analytical chemistry field because it is easy to generate a number without a real understanding of its meaning. In this document, two numerical parameters have been selected for review, the commonly used determination coefficient ($R^2$) and the less common but challenging *x*-residuals.

#### 3.3.1. $R^2$ value

This numerical parameter is usually defined as the ratio between the regression sum of squares and the total sum of squares. Alternatively, it is the proportion of the variation in the *y*-variable that is accounted for by the variation in the *x*-variable.

In the analytical field the use of $R^2$ *is* quite common, it has been used almost universally by chemical laboratories as a measurement of how well a set of data fits a calibration model [43, 44]. It was previously explained that in the LSR method the best fit is selected to minimize the squares of the residual values [14]. However, an $R^2$ value is not useful in indicating the errors relating to an individual measurement [17]. This coefficient cannot be considered as reliable enough to be used consistently for quantifying the GoF over the entire calibration curve [45], because $R^2$ values are based on absolute residual values. The effect of absolute bias will be more important to high values while for relative bias is more important at low values. Even if an $R^2$ value of 0.999 instinctively seems excellent, the circumstance that absolute error is used to determine $R^2$ can cause misunderstanding of this value [46].

#### 3.3.2. Residuals with regard to concentration or amount (*x*-residuals)

It was previously reported that, due to the limitations of the LSR method, large errors at the lower end of the concentration range can coincide with acceptable $R^2$ values [17]. Analyte concentrations used to establish the calibration relationship over the working range are supposed to have small errors. From the preceding comment, it is clear that selection of an appropriate regression model for the calibration dataset is very important to achieve the primary goal of any analytical procedure, which is the reliable measurements of real unknown samples [1, 2, 7]. For each of the calibrators used to build a calibration graph, it is possible, by using the regression parameters, to obtain the back-calculated concentration results [46, 47]. In fact, this method of curve fitting and back-calculations looks foolproof according to the fitness for purpose criterion of method validation [48]. Poor-quality calibration, especially at the high and low ends of the working range, will yield unreliable results. In this manner, *x*-residuals or deviations of model fit as regards to standard calibrators have been considered as

a way of estimating GoF, but without applicability in the research field [49]. Two important subjects relating to the estimated back-calculation of concentration values, confidence interval (C.I) and percentage of relative error (*%RE*), are explained as follows.

3.3.2.1. Confidence interval from *x*-concentration (or *x*-amount)

The main purpose of a calibration equation is to enable accurate calculation of analyte concentrations from instrument responses throughout the regression model. However, slope and intercept are estimated coefficients based on a finite number of measurements, so there is some uncertainty in their values. This uncertainty can be quantified statistically by displaying the confidence interval together with the interpolated results obtained from any regression model selected for calibration [14]. A confidence interval is calculated for a "problem" sample using the standard error of each interpolated value and the corresponding critical values of the *t*-distribution (according to the degrees of freedom and the confidence level):

$$(5)\ \text{Confidence Interval (C.I)} = x_p \pm t \cdot s_{xp}$$

It is important to highlight that on many occasions the differences between calibration models are insignificant when interpolated *x*-results provided by the alternative models are compared. In such cases, slight modifications in the regression parameters can have a large effect on the confidence intervals obtained by the different models [22].

3.3.2.2. Utilization of *x*-residuals as basis to calculate accuracy results

Different proposals were reported in the literature relating to the use of back-calculated concentrations from calibration models [46, 50]. A simple and straight way to use back-calculated concentrations is the determination of *%RE* [31, 47, 48]. The equation is:

$$(6)\ \%RE = 100 \cdot (x_{BC} - x_{NC}) / x_{NC}$$

Where $x_{BC}$ is the back-calculated concentration value (or amount) and $x_{NC}$ is the nominal or theoretical one.

The *%RE* can be used in different ways for model evaluation such as individual values (*%RE*), summation of absolute individual values ($\Sigma\%RE_{ABS}$) *and average values* ($\%RE_{AVER}$) [27-29, 31]. Alternatively, the *%RE p*arameter has been suggested as an appreciated indicator of linearity when graphically represented versus concentration values using both normal or log units [47].

Some method validation guidelines have proposed the utilization of deviation of each calibration standard, in the form of *%RE* from back-calculated concentrations, to evaluate the GoF of calibration models [31, 49]. Specifically, the criterion proposed has been that *%RE* values should not exceed a cut-off limit of 15% of the nominal or theoretical concentrations, and 20% near the LOQ.

## 4. Guidelines for the selection of calibration model

This assessment includes three steps which are easy to accomplish using statistical software. The different steps are shown in figure 1 and described as follows.

4.1. Step 1. Plotting of calibration dataset

Graphs *x-y* are often used to evaluate the general trend in the form of a linear or non-linear relationship and also to highlight data points that may be outliers.

4.2. Step 2. Preliminary diagnosis by applying OLSR model:

The most common way to apply analytical calibration by using the LSR method is to select the first order (linear) and unweighted fitting technique. This is due to the fact that the OLSR model can be considered as the simplest way to fit an experimental calibration dataset.

4.2.1. Obtaining the main regression coefficients

The regression coefficients describe the mathematical relationship between both variables. The slope indicates the gradient of a line and the intercept indicates the location where it intersects in *y*-axis. In addition, $R^2$ can be useful as an overall numerical value used as a qualitative index.

4.2.2. Residuals plot

Examine the residuals plot obtained by linear regression. If the data presents in:

i) a normal random pattern, an OLSR model may be appropriate;
ii) a curved pattern, a QLSR model may be appropriate;
iii) a funnel-like pattern, a WLSR model may be appropriate.

4.3. Step 3A. Selection of regression model order: OLSR versus QLSR

4.3A.1. Complementary statistical tests for linearity

At this step the important question to be answered is whether to select QLSR instead of the OLSR model. Using the statistical criterion, the question can be formulated as: can the hypothesis of linearity be rejected in favor of a quadratic model? Hypothesis testing such as Mandel´s test or SQT test can be applied to determine this.

4.3A.2. Obtaining alternative QLSR model coefficients

When a calibration dataset shows a curvature, a quadratic equation should probably be selected. This is a second-degree polynomial equation in which there are three coefficients present, i.e. quadratic ($b_2$), linear ($b_1$) and constant ($b_0$). Additionally, $R^2$ should be determined and used for comparative purposes.

4.3A.3. Calculation of $x_{BC}$ & *%RE* (individual & mean values)

The back-calculated concentrations for linear and quadratic models can be obtained by interpolation, applying the equations:

$$(6)\ (x_{BC})_{OLSR} = (y_i - b_0) / b_1 \qquad [40]$$

$$(7)\ (x_{BC})_{QLSR} = (-b_1/2\cdot b_2) \pm \{ (b_1/2\cdot b_2)^2 - [(b_0 - y_i)/b_2] \}^{1/2} \qquad [41]$$

When $x_{BC}$ values have been obtained, it is possible to calculate the corresponding *%RE* for each experimental calibration point.

4.3. Step 3B. Selection of regression model fitting technique: OLSR versus WLSR

At this step the key question to be answered is if a weighted calibration is necessary [51]. It is noticeable that both models (unweighted and weighted) provide quite similar coefficients estimates [9, 14]. Theoretically, the potential advantage of the WLSR model is that the predictive power of the resulting regression line may be superior to the OLSR one [11]. This fact could become important in the calculation of low range concentration values [9, 52].

4.3B.1. Complementary statistical tests of variance: Scedasticity testing

It has been described that one of the basic assumptions to allow application of the OLSR model is that the instrument response variances should be homogeneous and the errors equally distributed i.e. the data are homoscedastic. To determine if application the OLSR model is appropriate the data can be evaluated for the presence of heteroscedasticity. The authors used Statgraphic® software to evaluate datasets for scedasticity using different tests but that is outside of the scope of this document.

4.3B.2. Selection of weighting factor

Determining the correct weighting to use for calibration can be a challenging task because the selected weights will provide the extent of the difference between WLSR and OLSR models. It is important to highlight that $1/x^2$ has been suggested as the best weighting factor for quantitative analysis techniques using analytical calibration based on the WLSR model. Therefore, this is the only weighting factor used throughout this manuscript.

4.3B.3. Obtaining alternative WLSR model coefficients

The presence of heteroscedasticity does not cause large differences in the estimated values of regression coefficients. On the one hand, slope values are very similar from OLSR and WLSR models. On the other hand, intercept values can be less accurate for unweighted model depending on the extension of the calibration range selected. In any case, small differences between intercept values could lead to a large bias between experimental and theoretical calibration datasets.

4.3B.4. Calculation of $x_{BC}$ & *%RE* (individual & mean values)

The back-calculated concentrations for linear weighted models can be obtained by an equation equivalent to an unweighted one:

$$(8)\ (x_{BC})_{WLSR} = (y_i - b_{0w}) / b_{1w}$$

After $x_{BC}$ values are obtained, the *%RE* can be calculated.

## 5. Selection of analytical calibration models: Case studies

The OLSR model has been routinely used in quantitative assays of all types of analytes in different matrices. In many of these cases, it may have been more appropriate to choose a different calibration model [8, 11]. To demonstrate this, some calibration datasets were selected as examples and the approach described in this paper applied to ensure an appropriate calibration model was chosen. These examples demonstrate how greatly the choice of the model impacts the outcome of an experimental calibration from a practical point of view [4].

5.1. Selection of regression model order: first-order (OLSR) versus second-order (QLSR)

The decision of using a quadratic equation instead of employing a linear equation includes some degree of complexity that is characteristic in non-linear fitting [20].

### 5.1.1. Case study 1. Non-significant QLSR model

- Calibration dataset:

100/1.064; 100//1.070; 120/1.177; 120/1.200; 160/1.414; 160/1.425; 200/1.642; 200/1.660; 240/1.852; 240/1.870; 260/1.936; 260/1.960; 280/2.046; 280/2.080

- Step 1. Plot of calibration dataset

Figure 2.1a shows the calibration dataset of measurements that appear to fall on a straight-line. However, it is typical for calibration curves to be linear at low concentration but non-linear at high concentration as the detector of the instrument becomes saturated. In this example, virtually no difference was observed in the graphical trends of the regression by using linear or curved profiles. Following visual inspection of the scatter plot graph, the simplest linear model seems to be suitable.

- Step 2. Preliminary diagnosis by applying OLSR model

2.1. OLSR model coefficients

The main coefficients of calibrations are summarized in table 1. The error in slope and intercept values are around 1.3 % and 2.7 %CV, respectively.

2.2. Residuals plot

The errors plot (figure 2.1b) shows a typical inverted-U shaped form indicating that the data is not linear. Linearity is necessary for application of the OLSR model. Thus, review of the residuals plot indicates that a quadratic model is appropriate for this calibration dataset.

- Step 3. Selection of regression model order

3.1. Complementary statistical significance tests for linearity

Considering the discrepancy between both earlier graphs, it may be questioned whether the use of the QLSR model is justified with respect to OLSR when performing the test for selecting model order. To choose the best model some complementary statistical tests are available to check if a dataset, that looks linear, can therefore be identified as quadratic:

- Mandel´s test: $F_{exp}$ and $F_{cri}$ values are 6.45 and 4.84, respectively. The null hypothesis that applying a superior order model makes no difference, can be rejected, indicating that applying the quadratic model to this calibration dataset is appropriate.
- SQT test: $t_{exp}$ value is -3.03, which is compared to 2.16 as $t_{cri}$ value. The corresponding $p$-value of $t_{exp}$ is 0.011. The confidence interval does not include the zero value. This agrees with the outcome of the Mandel´s test i.e. that applying the quadratic model to this calibration dataset is appropriate.

3.2. Obtaining alternative QLSR model coefficients

The coefficients for quadratic regression are summarized in table 1. Quadratic polynomial can be considered as a superior calibration model considering the slightly lower value obtained by $s_{y/x}$.

3.3. Calculation of $x_{BC}$ and *%RE*

The $x_{BC}$ values and their corresponding *%RE* were determined by applying both models. The results were summarized in table 2. It is possible to see a small difference between the results; indicating that a slightly better fit is seen with the QLSR model than with the OLSR model. In any case, the criterion of *%RE* has been met for both models, meaning the OLSR model is accurate enough with regards to this relationship.

- Summary: Final decision

Relating to the model selection, it is very important to consider what the model is being used for, and the likely data that will be produced. Even in a case such as this, where the true relationship is obviously curved, the linear model probably is valid over the working range selected. Therefore, the OLSR model is appropriate for this example according to *%RE* values because all values are lower than the cut-off limit and the QLSR model does not fit substantially better.

**5.1.2. Case study 2. Significant QLSR model**

- Calibration dataset
  0.05/27; 0.05/29; 0.10/49; 0.10/50; 0.20/78; 0.20/80; 0.30/102; 0.30/105; 0.40/120; 0.40/122; 0.45/128; 0.45/129; 0.50/132; 0.50/134

- Step 1. Plot of calibration dataset

Figure 2.2a shows a calibration dataset where visual inspection indicates the prevalence of a curvature. It is apparent that the linear calibration model systematically over and under-predicts the experimental data at different points of the curve. From the above, it seems that the second-order trend line fits the curvature better than the straight-line.

- Step 2. Preliminary diagnosis by applying OLSR model

2A. OLSR model coefficients

The main coefficients of calibration are summarized in table 1. In this case, the slope and intercept error values are around 5% and 15% CV, respectively.

2B. Residuals plot

Figure 2.2b shows that residuals are not random, an assumption of the OLSR model. In fact, the residuals plot clearly exhibits a typical curvilinear pattern (inverted-U shaped form). Thus, rejection of a linear model in favour of a quadratic one is suggested by the plot of the residuals.

- Step 3. Selection of regression model order

3.1. Complementary statistical tests for linearity

There is agreement between both statistical tests:

- Mandel´s test: The value of $F_{exp}$ is 160.01 while for $F_{cri}$ is 4.84. The null hypothesis, that applying a superior order model makes no difference is rejected, indicating that a linear model is not appropriate for this calibration dataset.
- SQT test: The $t_{exp}$ value is -15.48, which is compared to 2.16 as $t_{cri}$ value. The *p*-value corresponding to $t_{exp}$ is much lower than 0.05. The confidence interval of the $b_2$ coefficient does not include the zero value. The SQT test, therefore, also indicates that a QLSR model should be used with this calibration dataset.

3.2. Obtaining alternative QLSR model coefficients

The coefficients for the quadratic model are summarized in table 1. In this case, the QLSR model provides a lower $s_{y/x}$ value, around 22% of the OLSR model.

3.3. Calculation of $x_{BC}$ and *%RE*

Values summarized in table 3 for this calibration dataset demonstrated the lack of reliability when using OLSR instead of a QLSR model. Overall larger errors, in the forms of $\Sigma\%RE_{ABS}$ and $\Sigma\%RE_{AVER,}$ are obtained when using a linear model in comparison to a quadratic fit. By using the OLSR model, large differences among the individual *%RE* values were observed. In this case, the largest errors were not experienced typically at higher concentrations. They were obtained at the low extreme of the calibration curve with values around -82%, well above 15-20%, which

is considered to be the acceptable limit. The QLSR model gives comparatively better individual values of *%RE* over the whole concentration range selected, with a good fit obtained for the lower part of the concentration range.

- Summary: Final decision

The dataset evaluated showed a marked degree of non-linearity. The different methodologies for model GoF evidenced that a non-linear model fit is better. Firstly, the residuals plot showed a curved pattern instead of randomness; secondly, both statistical tests for linearity rejected the significance of the OLSR model, and thirdly *%RE* quantitatively showed that deviations were much lower when the quadratic model was used. Therefore, the results clearly demonstrate that the QLSR model describes the response characteristics considerably better than the OLSR model.

5.2. Selection of regression model fitting technique: unweighted (OLSR) versus weighted (WLSR)

Calibration data may be homoscedastic or heteroscedastic, suggesting the use of OLSR and WLSR models, respectively [6]. From a practical perspective, contradictory information has been reported. In some cases, a significant difference between the OLSR and WLSR models is detected, while in other situations the difference is not significant and regression output showed a complete agreement [53]. In addition, it was stated that in practice weighted fitting of calibration is only necessary when various experimental conditions are fulfilled [51]. In the following examples, the effect of the weighted criterion in the statistical treatment of calibration data has been evaluated.

### 5.2.1. Case study 3. Non-significant WLSR model

- Calibration dataset

3/1433; 3/1440; 3/1443; 4/1908; 4/1918; 4/1928; 5/2385; 5/2400; 5/2410; 6/2862; 6/2880; 6/2900; 7/3339; 7/3360; 7/3380; 8/3816; 8/3840; 8/3880; 9/4293; 9/4320; 9/4355

- Step 1. Plot of calibration dataset

Figure 3.1a shows a linear trend with no data considered as outliers.

- Step 2. Preliminary diagnosis by applying OLSR model

As usual, the preliminary evaluation of the calibration dataset is typically considered from the linear unweighted viewpoint.

2.1. OLSR model coefficients

The values for OLSR model are summarized in table 1. The slope and intercept error values are around 0.4 % and 257 %CV, respectively.

2.2. Residuals plot

Figure 3.1b shows a typical trend in the form of funnel from lower to higher concentration, visually demonstrating the increasing variance with analyte concentration. This trend can be interpreted as a preliminary indicator that the data are heteroscedastic.

- Step 3. Selection of regression model fitting technique

3.1. Complementary statistical tests to check the equality of variance
Further statistical testing showed a strong indication that the variances are equal. In fact, the majority of tests showed the experimental data to be homoscedastic (see supplementary material). This allows rejection of the preliminary assessment of heteroscedasticity provided by the residuals plot as it was not confirmed statistically.

3.2. Selection of weighting factor
In this document $1/x^2$ has been suggested as the only weighting factor for analytical calibration based on the WLSR model.
3.3. Obtaining alternative WLSR model coefficients
The similarity between the OLSR and WLSR models is clearly observed. The coefficients of calibration for both equations are comparable by looking at the scatterplot of the data. It can be seen that the two regression lines are superimposed and the values of $R^2$ are also very similar. However, it was demonstrated that the WLSR model gives a better confidence interval for estimates.
3.4. Calculation of $x_{BC}$ and *%RE*
The $x_{BC}$ values and their corresponding *%RE* are summarized in table 4. In this example, there is no difference between the results calculated from the OLSR and WLSR models, which reported a very good fit for the whole short calibration range selected. A very small $\%RE_{MAX}$ value of 0.8 was obtained.

- Summary: Final decision

In this example, the OLSR model should be selected because the OLSR general assumptions were shown to hold true. In addition, before choosing to use WLSR in place of the OLSR model, it is important to evaluate the difference it makes in the values of *%RE*. It should be noted that from a quantitative perspective, the WLSR model does not necessarily provide more accurate results over a narrow calibration range.

**5.2.2. Case study 4. Significant WLSR calibration model**

- Calibration dataset

10/0.0350; 10/0.0360; 10/0.0370; 50/0.2359; 50/0.2365; 50/0.2368; 100/0.4831; 100/0.4856; 100/0.4888; 250/1.2665; 250/1.2884; 250/1.2901; 500/2.5275; 500/2.5834; 500/2.6004; 750/3.7000; 750/3.8000; 750/3.8386; 1000/4.9000; 1000/5.0526; 1000/5.1000

- Step 1. Plot of calibration dataset

The general trend of this dataset covering a wide concentration range is shown in figure 3.2a. It looks like a linear model with no outlier points.

- Step 2. Preliminary diagnosis by applying OLSR model

2.1. OLSR model coefficients
Coefficient values are summarized in table 1. The slope and intercept values of error are around 0.6 % and 848 %CV, respectively.
2.2. Residuals plot
Figure 3.2b shows a typical trend as funnel-shaped, suggesting that variances of data are not equal and randomly distributed. This is a preliminary indication that the data are heteroscedastic.

- Step 3. Selection of regression method fitting technique

3.1. Complementary statistical tests of variance

The different variance tests statistically confirmed the preliminary finding that the data are heteroscedastic.

3.2. Selection of weighting factor

In this document $1/x^2$ has been suggested as the only weighting factor for analytical calibration based on the WLSR model.

3.3. Obtaining alternative WLSR model coefficients

Similarities are seen between the two models because slope values reported are comparable. However, the intercept value of the WLSR model is one order of magnitude higher than that of the OLSR model. In addition, a major difference is seen in the estimated standard error of the intercept. For the OLSR model, the coefficient is higher than the intercept value and not precise at all (849 %CV), while for the WLSR model the precision is good (around 4 %CV). So, the influence of weighting will be very important when determining confidence intervals. Finally, the value of $s_{y/x}$ was much lower for WLSR when compared to OLSR model (0.0024 versus 0.0503).

3.4. Calculation of $x_{BC}$ and *%RE*

Results were summarized in table 5. It was demonstrated that, as usual, in the case of a wide calibration range and heteroscedastic data, the OLSR model failed to produce accurate estimates in the lower range of calibration, near the LOQ. Values were slightly higher than the suggested deviation limit of 15-20 *%RE*. In addition, $\Sigma\%RE_{ABS}$ and $\Sigma\%RE_{AVER}$ values were greater when using OLSR in comparison to the WLSR model.

- Summary: Final decision

In this example, all the methodologies for checking model GoF agreed. Residual plots showed funnel-like form; statistical tests about the equality of variance can be significantly rejected, and finally, *%RE* values revealed the quantitative difference between OLSR and WLSR models. In this situation, the best option to obtain accurate results is the WLSR model.

## 6. Conclusions

The next conclusions have been achieved from this overall evaluation about analytical calibration procedure:

- To select the best calibration model, different methodologies of GoF can be used, and unfortunately, every so often, there is not always agreement among them.
- $R^2$, used in isolation, should be excluded as an accurate parameter to evaluate GoF of calibration models because of its lack of reliability.
- Bias information, from *x*-residual values in the form of *%RE*, can be considered as the principal tool to make a well-founded decision regarding calibration model selection that might otherwise be overlooked.
- The OLSR model is not always the best option for analytical calibration. In fact, QLSR and WLSR models must be undoubtedly selected in some circumstances to achieve accurate results from unknown samples.
- Experimental conditions and the design of calibration procedures are very important when selecting the best calibration model.

## References


[1] P. Kościelniak, M. Wieczorek, Univariate analytical calibration methods and procedures. A review, Anal. Chim. Acta. 944 (2016) 14–28. https://doi.org/10.1016/j.aca.2016.09.024.

[2] L. Cuadros-Rodríguez, L. Gámiz-Gracia, E.M. Almansa-López, J.M. Bosque-Sendra, Calibration in chemical measurement processes. II. A methodological approach, TrAC - Trends Anal. Chem. 20 (2001) 620–636. https://doi.org/10.1016/S0165-9936(01)00111-X.

[3] S. H. Frisbie, Using polynomial regression to objectively test the fit of calibration curves in analytical chemistry, Int. J. Appl. Math. Theor. Phys. 1 (2015) 14. https://doi.org/10.11648/j.ijamtp.20150102.11.

[4] R.I. Rawski, P.T. Sanecki, K.M. Kijowska, P.M. Skitał, D.E. Saletnik, Regression analysis in analytical chemistry. Determination and validation of linear and quadratic regression dependencies, South African J. Chem. 69 (2016) 166–173. https://doi.org/10.17159/0379-4350/2016/v69a20.

[5] S.N. Ketkar, T.J. Bzik, Calibration of analytical instruments. Impact of nonconstant variance in calibration data, Anal. Chem. 72 (2000) 4762–4765. https://doi.org/10.1021/ac000018s.

[6] R.S. Nascimento, R.E.S. Froes, N.O.C. e Silva, R.L.P. Naveira, D.B.C. Mendes, W.B. Neto, J.B.B. Silva, Comparison between ordinary least squares regression and weighted least squares regression in the calibration of metals present in human milk determined by ICP-OES, Talanta. 80 (2010) 1102–1109. https://doi.org/10.1016/j.talanta.2009.08.043.

[7] N. V. Nagaraja, J.K. Paliwal, R.C. Gupta, Choosing the calibration model in assay validation, J. Pharm. Biomed. Anal. 20 (1999) 433–438. https://doi.org/10.1016/S0731-7085(98)00256-8.

[8] J. M. Sanchez, Ordinary Least Squares with laboratory calibrations: A practical way to show students that this fitting model may easily yield biased results when used indiscriminately, World J. Anal. Chem. 5 (2017) 1–8. https://doi.org/10.12691/wjac-5-1-1.

[9] I. Lavagnini, F. Magno, A statistical overview on univariate calibration, inverse regression, and detection limits: Application to gas chromatography/mass spectrometry technique, Mass Spectrom. Rev. 26 (2007) 1–18. https://doi.org/10.1002/mas.20100.

[10] J. Burrows, K. Watson, Linearity of chromatographic systems in drug analysis part III: Examples of nonlinear drug assays, Bioanalysis. 7 (2015) 1763–1774. https://doi.org/10.4155/bio.15.105.

[11] E. Szabo, G.K; Browne, H.K; Ajami, A.; E.G. Josephs, Alternatives to least squares linear regression analysis for computation of standard curves for quantitation by high performance liquid chromatography: Applications to clinical pharmacology, J. Clin. Pharmacol. 34 (1994) 242–249. https://doi.org/10.1002/j.1552-4604.1994.tb03993.x

[12] A. Maydeu-Olivares, C. Garcia-Forero, Goodness-of-Fit testing, Int. Encycl. Educ. 7 (2010) 190–196.

[13] B. Desharnais, F. Camirand-Lemyre, P. Mireault, C.D. Skinner, Procedure for the selection and validation of a calibration model II-theoretical basis, J. Anal. Toxicol. 41 (2017) 269–276. https://doi.org/10.1093/jat/bkx002

[14] S. Burke, Regression and calibration, LC-GC Eur. (2001) 13–18.

[15] J. Martin, D.D.R. de Adana, A.G. Asuero, Fitting models to data: Residual analysis, a primer, Uncertain. Quantif. and Model Calibration (2017).

[16] J. Burrows, K. Watson, Linearity of chromatographic systems in drug analysis part I: Theory of nonlinearity and quantification of curvature, Bioanalysis 7 (2015) 1731–1743. https://doi.org/10.4155/bio.15.103

[17] M. Mulholland, D.B. Hibbert, Linearity and the limitations of least squares calibration, J. Chromatogr. A. 762 (1997) 73–82. https://doi.org/10.1016/S0021-9673(96)00976-4

[18] I. Lavagnini, G. Favaro, F. Magno, Non-linear and non-constant variance calibration curves in analysis of volatile organic compounds for testing of water by the purge-and-trap method coupled with gas chromatography/mass spectrometry, Rapid Commun. Mass Spectrom. 18 (2004) 1383–1391. https://doi.org/10.1002/rcm.1498.

[19] J.N. Miller, Is it a straight line?, *Spectrosc. Int.* 3 (1991) 41–43.

[20] L. Kirkup, M. Mulholland, Comparison of linear and non-linear equations for univariate calibration, J. Chromatogr. A. 1029 (2004) 1–11. https://doi.org/10.1016/j.chroma.2003.12.013.

[21] J. Martin, A.R. Gracia, A.G. Asuero, Fitting nonlinear calibration curves: No models perfect, J. Anal. Sci. Methods Instrum. 07 (2017) 1–17. https://doi.org/10.4236/jasmi.2017.71001.

[22] J.M. Mermet, Calibration in atomic spectrometry: A tutorial review dealing with quality criteria, weighting procedures and possible curvatures, Spectrochim. Acta - Part B At. Spectrosc. 65 (2010) 509–523. https://doi.org/10.1016/j.sab.2010.05.007.

[23] A.G. Asuero, G. González, Fitting straight lines with replicated observations by linear regression. III. Weighting data, Crit. Rev. Anal. Chem. 37 (2007) 143–172. https://doi.org/10.1080/10408340701244615.

[24] T. Singtoroj, J. Tarning, a. Annerberg, M. Ashton, Y. Bergqvist, N.J. White, N. Lindegardh, N.P.J. Day, A new approach to evaluate regression models during validation of bioanalytical assays, J. Pharm. Biomed. Anal. 41 (2006) 219–227. https://doi.org/10.1016/j.jpba.2005.11.006.

[25] ISO 11095:1996. Linear calibration using reference materials

[26] J.M. Sanchez, Estimating detection limits in chromatography from calibration data: Ordinary least squares regression vs. weighted least squares, Separations. 5 (2018) 49–61. https://doi.org/10.3390/separations5040049.

[27] A.M. Almeida, M.M. Castel-Branco, A.C. Falcão, Linear regression for calibration lines revisited: Weighting schemes for bioanalytical methods, J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 774 (2002) 215–222. http://www.scopus.com/inward/record.url?eid=2-s2.0-0037099575&partnerID=40&md5=bfe66ab8503995b62c2952d6768c8f07.

[28] H. Gu, G. Liu, J. Wang, A.-F. Aubry, M.E. Arnold, Selecting the correct weighting factors for linear and quadratic calibration curves with least-squares regression algorithm in bioanalytical LC-MS/MS assays and impacts of using incorrect weighting factors on curve stability, data quality, and assay perfo, Anal. Chem. 86 (2014) 8959–8966. http://www.scopus.com/inward/record.url?eid=2-s2.0-84913580209&partnerID=40&md5=d5a011a5cb3b6d791c7bc61f2c26d37a.

[29] M. Kisser, J.W. Dolan, Selecting the best curve fit, LC-GC Europe 17 (2004) 138-143

[30] U.S. Food and Drug Administration/Foods and Veterinary Medicine (FDA/FVM). Guidelines for the Validation of Chemical Methods for the FDA Foods Program. 3rd Edition. (2019). https://www.fda.gov/media/81810/download

[31] F. Raposo, Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: A tutorial review, TrAC - Trends Anal. Chem. 77 (2016) 167–185. https://doi.org/10.1016/j.trac.2015.12.006

[32] B. Desharnais, F. Camirand-Lemyre, P. Mireault, C.D. Skinner, Procedure for the selection and validation of a calibration model I-description and application, J. Anal. Toxicol. 41 (2017) 261–268. https://doi.org/10.1093/jat/bkx001

[33] T. Belloto Jr, R.J; Sokoloski, Residual analysis in regression, Am. J. Pharm. Educ. 49 (1985) 295–303.

[34] J.L. Feménias, Goodness of fit: Analysis of residuals, J. Mol. Spectrosc. 217 (2003) 32–42. https://doi.org/10.1016/S0022-2852(02)00013-9

[35] J. Mandel. “The statistical analysis of experimental data”, Dover Publications, New York, 1964.

[36] J.M. Andrade, M.P. Gómez-Carracedo, Notes on the use of Mandel's test to check for nonlinearity in laboratory calibrations, Anal. Methods. 5 (2013) 1145–1149. https://doi.org/10.1039/c2ay26400e.

[37] D.L. Massart, B.G.M. Vandegiste, L.M.C. Buydens, S. de Jong, P.J.Lewi, J.Smeyers-Verbeke, "Handbook of chemometrics and qualimetrics", Elsevier, Amsterdam, 1997.

[38] S.L.R. Ellison, V.J. Barwick and T.J. Duguid, Practical Statisticcs for the analytical scientist, The Royal Society of Chemistry, Cambridge, 2009.

[39] A. Sayago, M. Boccio, A. G. Asuero, Fitting straight lines with replicated observations by linear regression: Part I. the last squares postulates, *Crit. Rev. Anal. Chem.* 34 (2004) 39–50. https://doi.org/10.1080/10408340490273744

[40] ISO 8466-1, Water quality – Calibration and evaluation of analytical methods and estimation of performance characteristics-Part 1: Statistical evaluation of the linear calibration function (1990)

[41] ISO 8466-2, Water quality – Calibration and evaluation of analytical methods and estimation of performance characteristics-Part 2: Calibration strategy for non-linear second-order calibration functions (2001).

[42] A. Sayago and A. G. Asuero, "Fitting straight lines with replicated observations by linear regression: Part II. Testing for homogeneity of variances", *Crit. Rev. Anal. Chem.* 34 (2004) 133–146. https://doi.org/10.1080/10408340490888599

[43] R. Sahai, H.; Singh, The use of R2 as a measure of goodness of fit, Virginia J. Sci. 40 (1989) 5–9.

[44] J. Tellinghuisen, C.H. Bolster, Using R2 to compare least-squares fit models: When it must fail, Chemom. Intell. Lab. Syst. 105 (2011) 220–222. https://doi.org/10.1016/j.chemolab.2011.01.004.

[45] Analytical Methods Committee, Is my calibration linear? *The Analyst* 119 (1994) 2363-2366. https://doi.org/10.1039/AN9941902363

[46] B.A. Logue, E. Manandhar, Percent residual accuracy for quantifying goodness-of-fit of linear calibration curves, Talanta. 189 (2018) 527–533. https://doi.org/10.1016/j.talanta.2018.07.046.

[47] J.W. Dolan, Calibration curves, part 3: A different view, LC GC Eur. 22 (2009) 304. http://www.scopus.com/inward/record.url?eid=2-s2.0-77952075969&partnerID=40&md5=ad1453868fcb625c540dd2ea14d42679.

[48] J.M. Jurado, A. Alcázar, R. Muñiz-Valencia, S.G. Ceballos-Magaña, F. Raposo, Some practical considerations for linearity assessment of calibration curves as function of concentration levels according to the fitness-for-purpose approach, Talanta. 172 (2017) 221–229. https://doi.org/10.1016/j.talanta.2017.05.049.

[49] F. Raposo, C. Ibelli-Bianco, Performance parameters for analytical method validation: Controversies and discrepancies among numerous guidelines, TrAC - Trends Anal. Chem. 129 (2020) 115913. https://doi.org/10.1016/j.trac.2020.115913.

[50] The National Environmental Laboratory Accreditation Institute (NELAC-TNI). Vol. 1, Module 4, Quality system for chemical testing (2016) https://nelac-institute.org/index.php

[51] V.R. Meyer, Weighted linear least-squares fit-A need? Monte Carlo simulation gives the answer, LC-GC Europe 28 (2015) 204-209

[52] J.M. Sanchez, Linear calibrations in chromatography: The incorrect use of ordinary least squares for determinations at low levels, and the need to redefine the limit of quantification with this regression model, J. Sep. Sci. 43 (2020) 2708–2717. https://doi.org/10.1002/jssc.202000094.

[53] P.G.A. Barbosa, F.I.C.C. Martins, L.K. Lima, M.A.L. Milhome, R.M. Cavalcante, R.F. do Nascimento, Statistical analysis for quality adjustment of the analytical curve for determination of pesticide multiresidue in pineapple samples, Food Anal. Methods. 11 (2018) 466–478. https://doi.org/10.1007/s12161-017-1017-9.

**Table 1. Coefficients of regression – OLSR versus QLSR & WLSR**

| | $b_2$ | $s_{b2}$ | $b_1$ | $s_{b1}$ | $b_0$ | $s_{b0}$ | $s_{y/x}$ | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| **Case study 1** | | | | | | | | |
| **OLSR** | - | - | 0.005508 | $7.09 \cdot 10^{-5}$ | 0.529518 | 0.014521 | 0.017184 | 0.9980 |
| **QLSR** | $-3.61 \cdot 10^{-6}$ | $1.19 \cdot 10^{-6}$ | 0.006876 | 0.000454 | 0.415168 | 0.039365 | 0.013251 | 0.9989 |
| **Case study 2** | | | | | | | | |
| **OLSR** | - | - | 230.72 | 11.52 | 25.87 | 3.78 | 6.9806 | 0.9709 |
| **QLSR** | -339.84 | 21.96 | 416.84 | 12.29 | 9.34 | 1.35 | 1.5278 | 0.9987 |
| **Case study 3** | | | | | | | | |
| **OLSR** | - | - | 480.8 | 1.86 | -4.6 | 11.8 | 17 | 0.9997 |
| **WLSR** | - | - | 480.6 | 1.49 | -3.7 | 7.4 | 12 | 0.9998 |
| **Case study 4** | | | | | | | | |
| **OLSR** | - | - | 0.005045 | $3.1 \cdot 10^{-5}$ | -0.001900 | 0.016122 | 0.0503 | 0.9993 |
| **WLSR** | - | - | 0.005083 | $2.4 \cdot 10^{-5}$ | -0.014965 | 0.000620 | 0.0024 | 0.9996 |

Table 2. Values of $x_{NC}$, $x_{BC}$ and *%RE* for case study 1

| | OLSR | | QLSR | |
|---|---|---|---|---|
| $x_{NC}$ | $x_{BC}$ | %RE | $x_{BC}$ | %RE |
| 100 | 97.0 | -3.0 | 99.6 | -0.4 |
| 100 | 98.1 | -1.9 | 100.5 | 0.5 |
| 120 | 117.5 | -2.0 | 118.1 | -1.6 |
| 120 | 121.7 | 1.4 | 121.9 | 1.6 |
| 160 | 160.6 | 0.4 | 158.4 | -1.0 |
| 160 | 162.6 | 1.6 | 160.4 | 0.2 |
| 200 | 202.0 | 1.0 | 199.3 | -0.4 |
| 200 | 205.2 | 2,6 | 202.6 | 1.3 |
| 240 | 240.1 | 0.0 | 238.9 | -0.4 |
| 240 | 243.4 | 1.4 | 242.4 | 1.0 |
| 260 | 255.3 | -1.8 | 255.4 | -1.8 |
| 260 | 259.7 | -0.1 | 260.2 | 0.1 |
| 280 | 275.3 | -1.7 | 277.6 | -0.8 |
| 280 | 281.5 | 0.5 | 284.6 | 1.7 |
| | $\Sigma\%RE_{ABS}$ | 19.4 | $\Sigma\%RE_{ABS}$ | 12.8 |
| | $\Sigma\%RE_{AVER}$ | 1.4 | $\Sigma\%RE_{AVER}$ | 0.9 |

Table 3. Values of $x_{NC}$, $x_{BC}$ and *%RE* for case study 2

| | **OLSR** | | **QLSR** | |
|---|---|---|---|---|
| $x_{NC}$ | $x_{BC}$ | **%RE** | $x_{BC}$ | **%RE** |
| 0.05 | 0.005 | -90.2 | 0.044 | -12.1 |
| 0.05 | 0.014 | -72.9 | 0.049 | -1.8 |
| 0.10 | 0.100 | 0.3 | 0.104 | 3.9 |
| 0.10 | 0.105 | 4.6 | 0.107 | 6.8 |
| 0.20 | 0.235 | 17.3 | 0.203 | 1.6 |
| 0.20 | 0.226 | 13.0 | 0.196 | -2.0 |
| 0.30 | 0.343 | 14.3 | 0.306 | 1.9 |
| 0.30 | 0.330 | 10.0 | 0.292 | -2.8 |
| 0.40 | 0.408 | 2.0 | 0.389 | -2.9 |
| 0.40 | 0.417 | 4.2 | 0.402 | 0.5 |
| 0.45 | 0.443 | -1.6 | 0.449 | -0.2 |
| 0.45 | 0.447 | -0.7 | 0.458 | 1.8 |
| 0.50 | 0.460 | -8.0 | 0.490 | -2.0 |
| 0.50 | 0.469 | -6.3 | 0.517 | 3.4 |
| | $\Sigma\%RE_{ABS}$ | 245.2 | $\Sigma\%RE_{ABS}$ | 43.1 |
| | $\Sigma\%RE_{AVER}$ | 17.5 | $\Sigma\%RE_{AVER}$ | 3.7 |

Table 4. Values of $x_{NC}$, $x_{BC}$ and *%RE* for case study 3

| | OLSR | | WLSR | |
|---|---|---|---|---|
| $x_{NC}$ | $x_{BC}$ | %RE | $x_{BC}$ | %RE |
| 3 | 2.990 | -0.3 | 2.989 | -0.4 |
| 3 | 3.005 | 0.2 | 3.004 | 0.1 |
| 3 | 3.011 | 0.4 | 3.010 | 0.3 |
| 4 | 3.978 | -0.5 | 3.978 | -0.6 |
| 4 | 3.999 | 0.0 | 3.998 | 0.0 |
| 4 | 4.020 | 0.5 | 4.019 | 0.5 |
| 5 | 4.970 | -0.6 | 4.970 | -0.6 |
| 5 | 5.002 | 0.0 | 5.001 | 0.0 |
| 5 | 5.022 | 0.4 | 5.022 | 0.4 |
| 6 | 5.963 | -0.6 | 5.963 | -0.6 |
| 6 | 6.000 | 0.0 | 6.000 | 0.0 |
| 6 | 6.042 | 0.7 | 6.042 | 0.7 |
| 7 | 6.955 | -0.6 | 6.955 | -0.6 |
| 7 | 6.998 | 0.0 | 6.999 | 0.0 |
| 7 | 7.040 | 0.6 | 7.040 | 0.6 |
| 8 | 7.947 | -0.7 | 7.948 | -0.7 |
| 8 | 7.997 | 0.0 | 7.998 | 0.0 |
| 8 | 8.059 | 0.7 | 8.060 | 0.8 |
| 9 | 8.939 | -0.7 | 8.940 | -0.7 |
| 9 | 8.995 | -0,1 | 8,996 | 0,0 |
| 9 | 9.068 | 0.8 | 9.069 | 0.8 |
| | $\Sigma\%RE_{ABS}$ | 8.48 | $\Sigma\%RE_{ABS}$ | 8.43 |
| | $\Sigma\%RE_{AVER}$ | 0.40 | $\Sigma\%RE_{AVER}$ | 0.40 |

Table 5. Values of $x_{NC}$, $x_{BC}$ and *%RE* for case study 4

| | **OLSR** | | **WLSR** | |
|---|---|---|---|---|
| **$x_{NC}$** | **$x_{BC}$** | **%RE** | **$x_{BC}$** | **%RE** |
| 10 | 7.314 | -26.9 | 9.830 | -1.7 |
| 10 | 7.710 | -22.9 | 10.223 | 2.2 |
| 10 | 7.512 | -24.9 | 10.027 | 0.3 |
| 50 | 47.252 | -5.5 | 49.473 | -1.1 |
| 50 | 47.307 | -5.4 | 49.528 | -0.9 |
| 50 | 47.136 | -5.7 | 49.358 | -1.3 |
| 100 | 96.123 | -3.9 | 97.982 | -2.0 |
| 100 | 96.615 | -3.4 | 98.471 | -1.5 |
| 100 | 97.250 | -2.7 | 99.102 | -0.9 |
| 250 | 255.749 | 2.3 | 256.428 | 2.6 |
| 250 | 256.069 | 2.4 | 256.746 | 2.7 |
| 250 | 251.396 | 0.6 | 252.107 | 0.8 |
| 500 | 515.777 | 3.2 | 514.533 | 2.9 |
| 500 | 501.331 | 0.3 | 500.194 | 0.0 |
| 500 | 512.419 | 2.5 | 511.200 | 2.2 |
| 750 | 733.724 | -2.2 | 730.869 | -2.6 |
| 750 | 761.193 | 1.5 | 758.135 | 1.1 |
| 750 | 753.545 | 0.5 | 750.542 | 0.1 |
| 1000 | 971.567 | -2.8 | 966.952 | -3.3 |
| 1000 | 1001.803 | 0.2 | 996.965 | -0.3 |
| 1000 | 1011.207 | 1.1 | 1006.300 | 0.6 |
| | $\Sigma\%RE_{ABS}$ | 120.74 | $\Sigma\%RE_{ABS}$ | 31.17 |
| | $\Sigma\%RE_{AVER}$ | 5.7 | $\Sigma\%RE_{AVER}$ | 1.5 |

Figure 1. Steps for the selection of calibration model

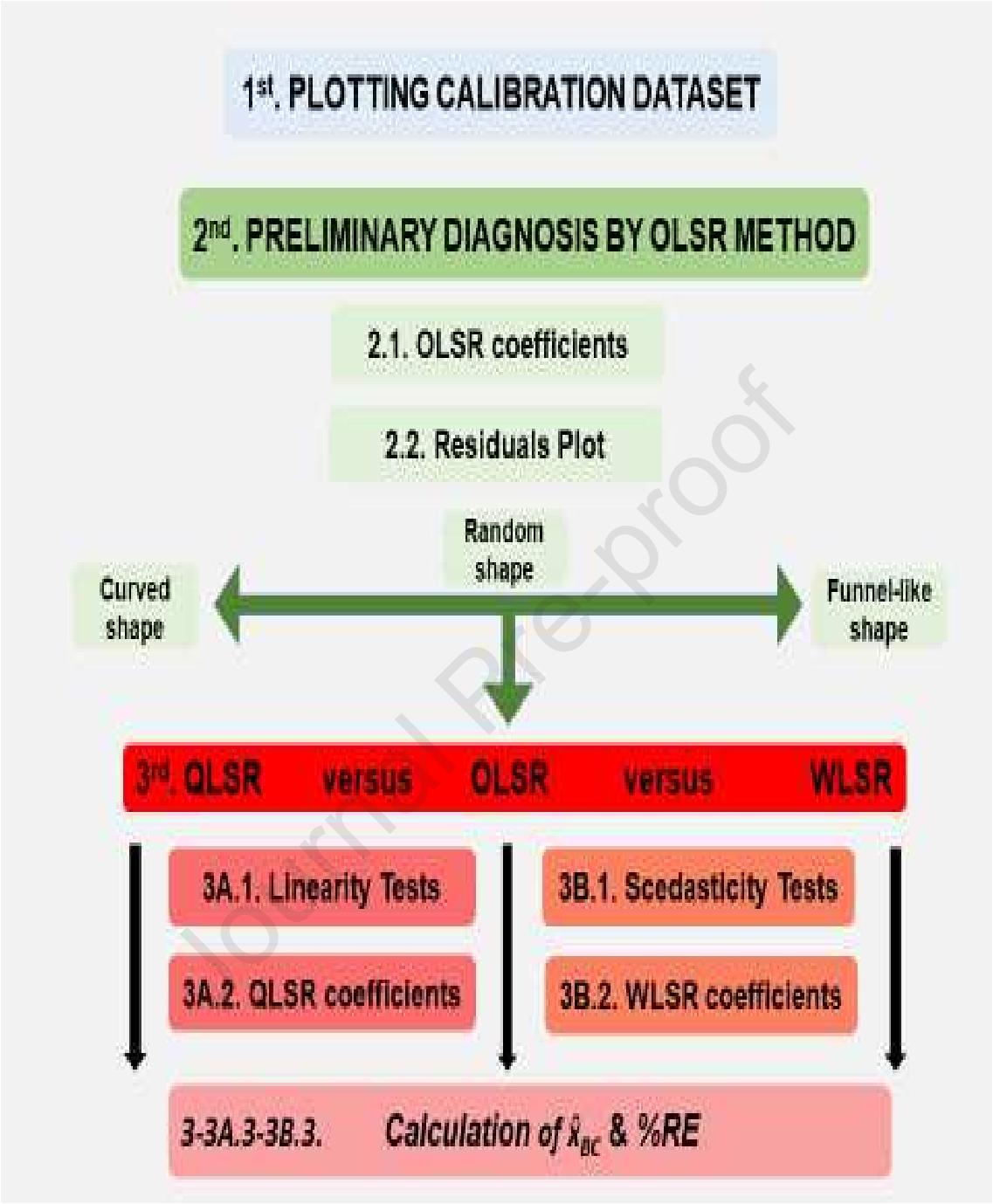
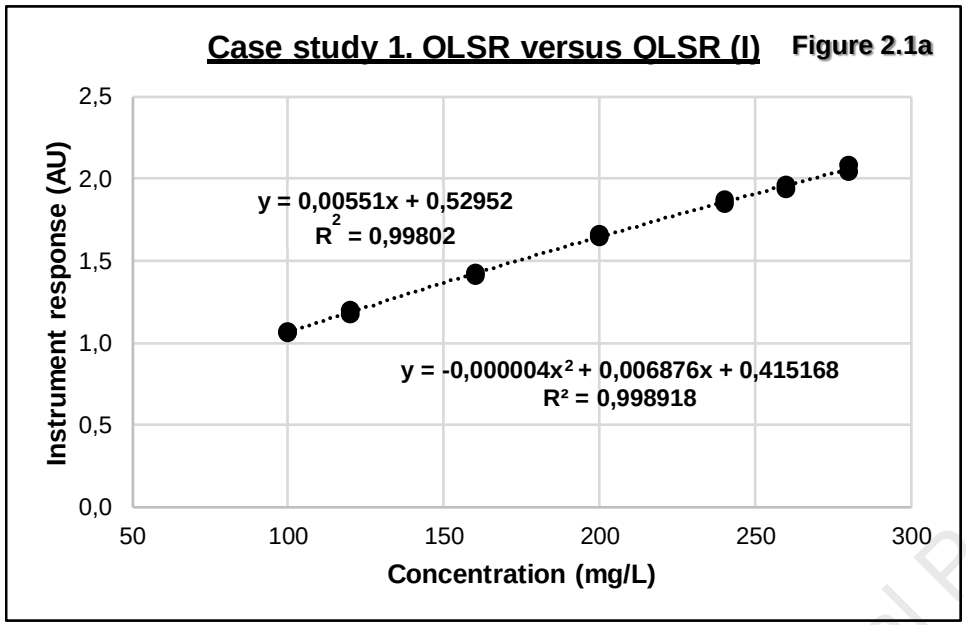
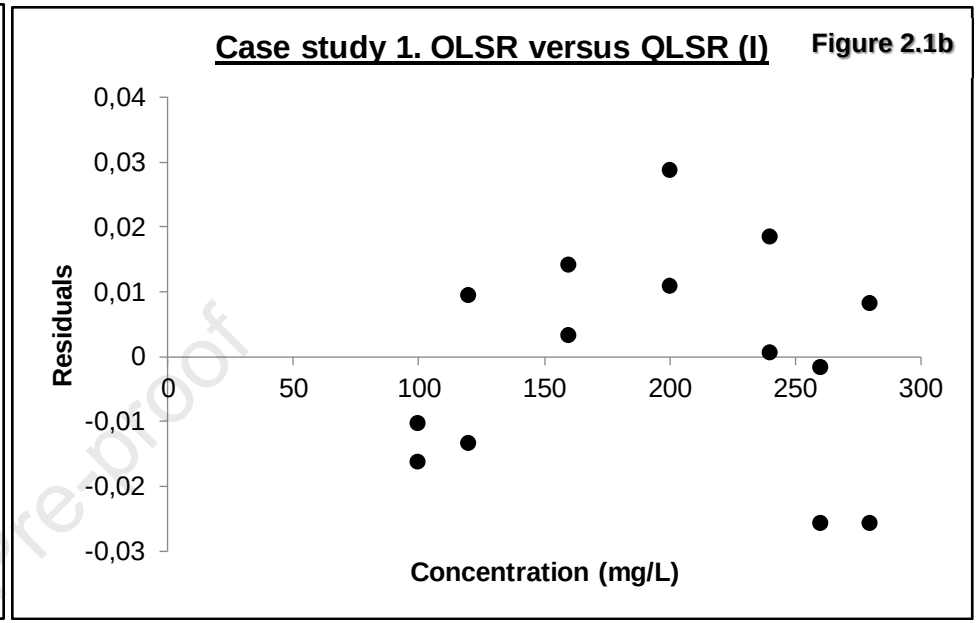
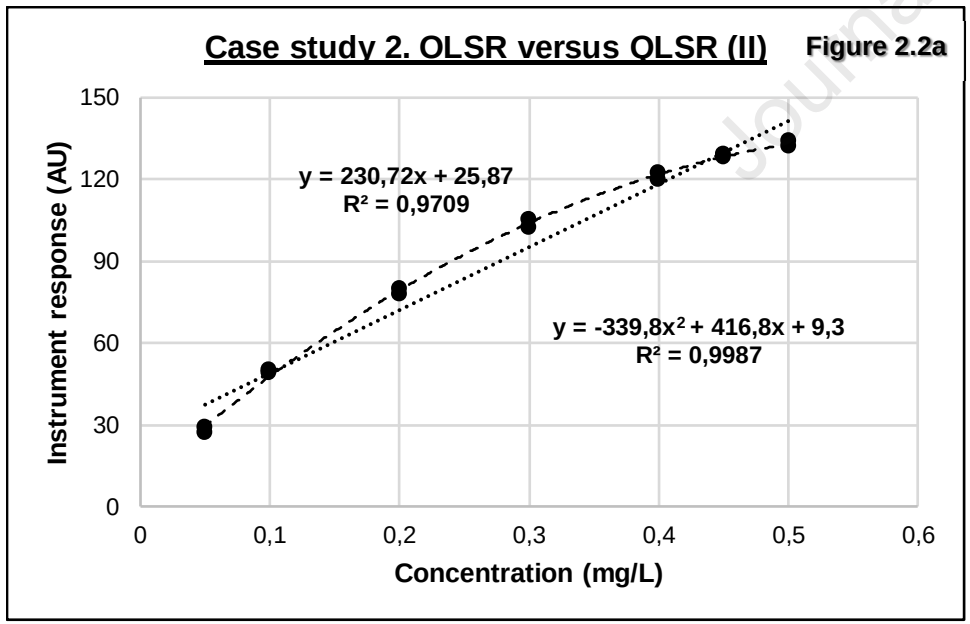


Figure 2. OLSR versus QLSR



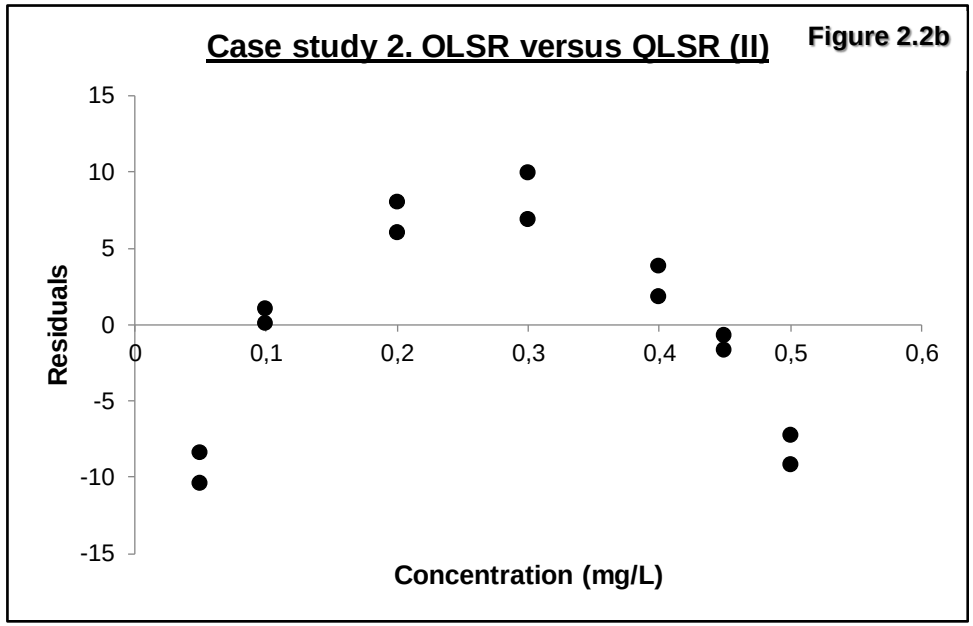
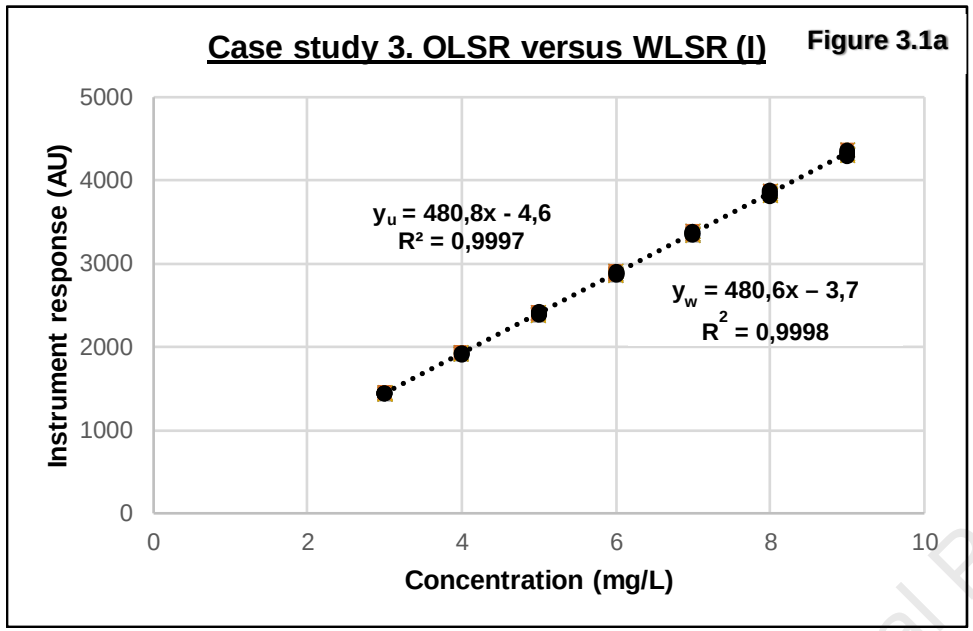
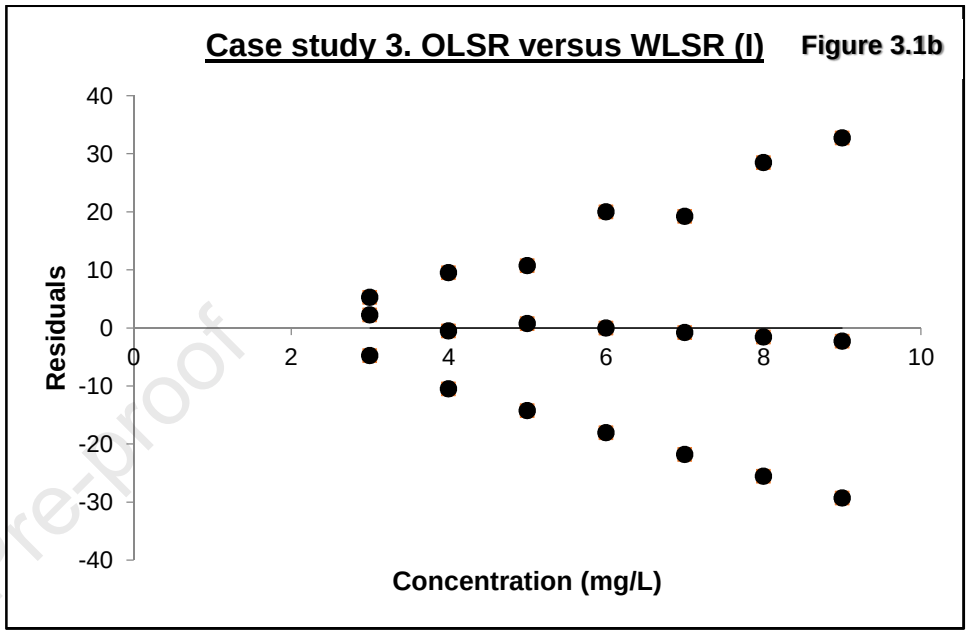
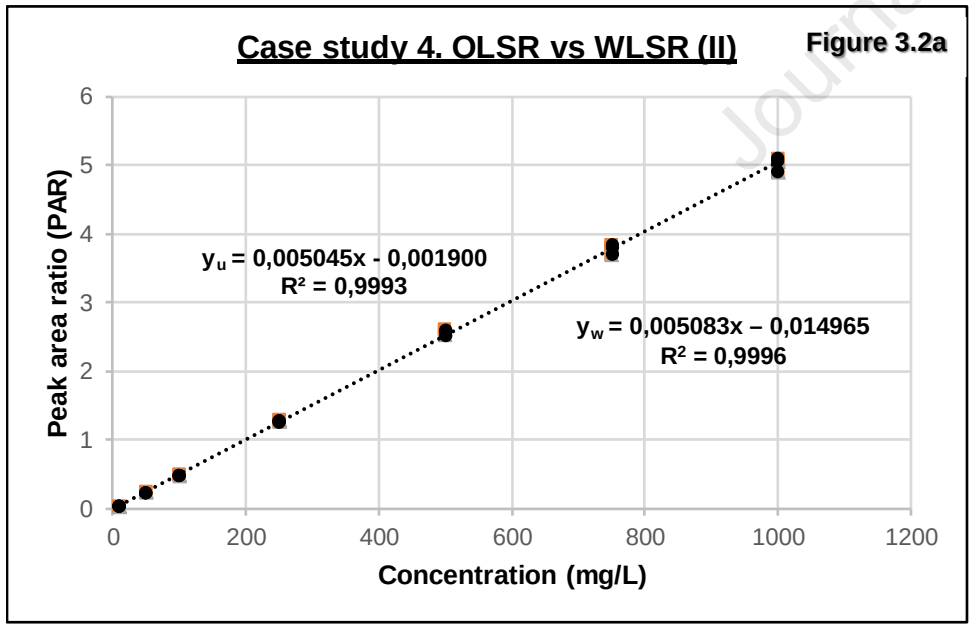
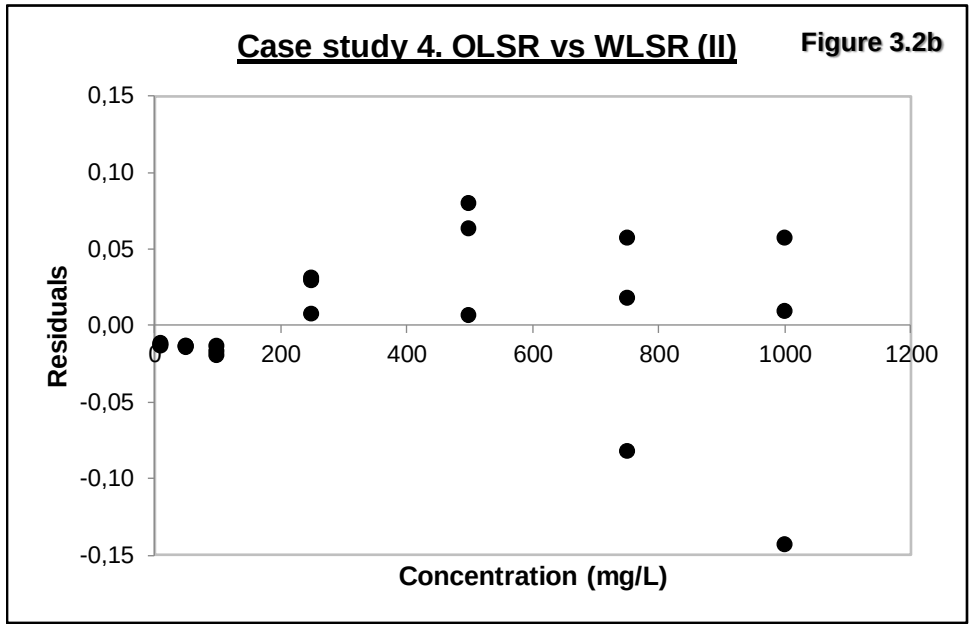



Figure 3. OLSR versus WLSR



**HIGHLIGHTS:**

- OLSR, QLSR and WLSR are the main analytical calibration models.
- Graphical plots, statistical tests and numerical parameters are used to evaluate GoF of calibration models.
- *%RE* values are useful to evaluate bias and therefore to appropriately select the calibration model.
- Model selection is dependent on experimental conditions and design of calibration procedures.

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: