1431

Data Analysis in the Determination of Stoichiometries and Stability Constants of Complexes

Juan M. BOSQUE-SENDRA,[†] Eva ALMANSA-LÓPEZ, Ana Mª GARCÍA-CAMPAÑA, and Luis CUADROS-RODRÍGUEZ

School of Qualimetrics, Department of Analytical Chemistry, University of Granada, Fuentenueva, s/n, E-18071 Granada, Spain

The determination of stoichiometries and stability constants of complexes by means of UV-visible spectrophotometry applying traditional methods does not evaluate the quality of the values obtained, since the classic application of these methods does not provide coverage intervals. However, the use of chemometric techniques in different steps of the application of these methods makes it possible to obtain not only a real value of the characteristics of the complexes but a validation of such a value. In this paper a methodology is proposed that combines some traditional methods, three different regression models (LMSR, LSR and LSPR) and a small number of mathematical algorithms. This methodology is able to estimate, in a simple and rigorous way, the stoichiometry and stability constant of a complex and its corresponding uncertainties.

The determination of the stoichiometries and the stability constants of complexes is an important subject in analytical chemistry and other branches of chemistry. Many computer programs, such as HYPERQUAD,[1] SUPERQUAD,[2] SIRKO,[3] and POLET[4] have been proposed to obtain their values. However, the stoichiometry and the global stability constant of a complex are frequently estimated from the measurement and treatment of analytical data using a wide selection of traditional methods. These include the following: Job,[5] Yoe and Jones,[6] Harvey and Manning,[7] Holme and Langmyhr,[8] Rose and Drago,[9] Bench and French,[10] and Diehl and Lidstrom[11] methods as well as stoichiometric[12,13] and non-stoichiometric[14] dilution methods. Lately, more than 80 papers have been published where some of these methods have been used to calculated the stoichiometry and/or the stability constant of metallic complexes or similar compounds.

The traditional methods are based on the graphical representation of curves derived from more or less complicated functions, obtained by means of the experimental measurement of analytical signals from a chemical system in equilibrium. From the study of these curves, conclusions are reached and the numerical results achieved are dependent on a large subjective observer-determined component. Therefore, these results do not guarantee the quality required from a rigorous study of statistical validation.

Of the above, Job's and Yoe and Jones' methods are the most widely used due to the simplicity of their theoretical foundation and their straightforward experimental application. They are based on the study of a graphical representation of analytical signals *versus* ligand molar fraction (Job's method) or molar ratio ligand/metal (Yoe and Jones' method). In the representation the analytical signal increases with the ligand molar fraction (or the molar ratio) until a maximum is reached; at this point, the analytical signal diminishes (Job's method) or it is maintained constant (Yoe and Jones' method). The

maximum of the curve corresponds to the maximum formation of complex and it indicates the molar ligand fraction, which is the stoichiometry of the complex.

The curvature of the line obtained depends on the stability of the complex; the more stable it is, the closer the lateral segments of the experimental curve approach a straight line because formation of the complex is practically total. From the adjustment of the two parts of the curve, we obtain, therefore, two straight lines: one with a positive slope due to the increase of the proportion of complex in the solution, and another with a negative slope (Job's method) or with a different positive slope (Yoe and Jones' method) due to the lesser formation of complex when the stoichiometric point is exceeded.

When the complex is very stable, two straight lines can be traced directly on the experimental points obtained; these intersect at a point that corresponds to the stoichiometry of the complex. On the other hand, if the complex shows less stability, it is necessary to plot the straight lines using only the experimental points most distant from the maximum; the stoichiometry is obtained from the intersection of the extensions of these straight lines. In the case of a weak complex, the experimental curve does not present any straight segments and the application of the method is impractical (Fig. 1).

In a similar way, the stability constant of the complex can be estimated from the vertical distance that separates the intersection point of the straight lines plotted and the maximum of the experimental curve.

The choice of the straight segments of the experimental curve is the crucial stage in the application of these methods, since the results obtained depend entirely on this process. Two methodologies are traditionally used to achieve these straight lines: a) Using a plot. In this case, the numerical values obtained do not have any statistical corroboration, since the straight lines used are not appropriately validated. In addition, the use of a graphical method prevents us from obtaining a coverage interval for the value of the stoichiometry and for the constant of stability. b) Fitting experimental points using least

---

[†] To whom correspondence should be addressed.

square regression. However, as the complex formation reaction is not quantitative, a curvature is obtained in the proximity of the stoichiometric point and this can impede an appropriate choice of linear segments. This fact produces a high degree of subjectivity in the conclusions of each study. Indeed, the presence of anomalous points or outliers in experimental data can modify the plotting of these straight lines and therefore the values of stoichiometry and stability constant obtained.

Job's and Yoe–Jones' methods are the most widely reported in the literature and the methodological problems derived from their application can be considered representative of those related to the estimation of the stoichiometry and stability constants of a chemical compound by means of graphical procedures. Therefore, in this paper a rigorous protocol for the application of these methods is proposed in order to offer an alternative to computer programs when the stoichiometry and stability constants of a complex must be estimated and suitable computer programs are not available. It is a simple and easy procedure, during which calculations may be performed and graphs produced either in Excel or in any other spreadsheet. The proposed procedure uses three different regression models to fit the experimental points, avoiding the previously described ambiguous situations and determining validated values of stoichiometries and stability constants of complexes in solution.

Theoretical Aspects

We assume the reaction of formation of the complex: $m\text{M} + g\text{L} \longrightarrow \text{M}_m\text{L}_g$, where M represents the metallic ion, L the ligand and m and g the corresponding stoichiometric coefficients, respectively. On this assumption, solutions of metal (C_M) and ligand (C_L) are prepared and measured in accordance with the literature.[5,6]

Job's method

The analytical signals measured (S) are plotted *versus* the corresponding reagent molar fractions (x), where $x = C_\text{L}/C$ and $C = C_\text{M} + C_\text{L}$ is a constant. A curve with a signal maximum for a certain value of molar fraction is obtained, from which the stoichiometric molar fraction (SMF) can be deduced:

$$x_{\max} = \text{SMF} = g/(g + m) \tag{1}$$

which indicates the stoichiometry of the complex (g:m).

The curvature of the experimental lines is due to the fact that the complex formation reaction is not quantitative, so the complex formation constant can be estimated from the deviations of the theoretical straight lines. The stability constant, K_est, is given by the expression:

$$K_\text{est} = \frac{[\text{M}_m\text{L}_g]}{[\text{M}]^m[\text{L}]^g} \tag{2}$$

and each one of these concentrations can be expressed in terms of the initial concentrations of metal (C_M) and/or ligand (C_L):

$$[\text{M}_m\text{L}_g] = (1/m)C_\text{M}(1 - \alpha) = (1/g)C_\text{L}(1 - \alpha)$$

$$[\text{M}] = C_\text{M} - m[\text{M}_m\text{L}_g]; \quad [\text{L}] = C_\text{L} - g[\text{M}_m\text{L}_g]$$

where α is the degree of dissociation of the complex.

The concentrations C_L and C_M can also be expressed as:

$$C_\text{L} = Cx; \; C_\text{M} = C - C_\text{L} = C(1 - x)$$

By replacing these expressions in Eq. (2) and regrouping terms, the following equation[12] is obtained:

$$K_\text{est} = \frac{(1 - \alpha)C}{g(\alpha C)^{m+g}} \; \frac{x^{1-g}}{(1 - x)^m} \tag{3}$$

Substituting the value of SMF (Eq. (1)) in Eq. (3), it is possible to obtain an equation of the stability constant in terms of the total concentration (C), the degree of dissociation (α), and the stoichiometric coefficients, m and g:

$$K_\text{est} = \frac{(1 - \alpha)C}{(\alpha C)^{m+g}} \; \frac{(m + g)^{m+g-1}}{g^g m^m} \tag{4}$$

The value of the stability constant must be accompanied by its corresponding uncertainty, $K_\text{est} \pm k u_\text{Kest}$, where u_Kest is the standard uncertainty associated with the estimated value of the stability constant. It is calculated by applying the variance propagation rule to the stability constant equation (Eq. (4)), thus obtaining the expression:

$$u_\text{Kest}^2 = u^2 \left\{ \frac{(m + g)^{m+g-1}}{g^g m^m C^{m+g-1}} \cdot \frac{1 - \alpha}{\alpha^{m+g}} \right\} = \left(\frac{(m + g)^{m+g-1}}{g^g m^m C^{m+g-1}} \right)^2 \left(\frac{\partial \left(\frac{1 - \alpha}{\alpha^{m+g}} \right)}{\partial \alpha} \right) \cdot u_\alpha^2$$

By operating and regrouping the terms we obtain:

$$u_\text{Kest}^2 = \left[\frac{(m + g)^{m+g-1}}{g^g m^m C^{m+g-1}} \cdot \frac{(m + g - 1)\alpha^{(m+g)} - (m + g)\alpha^{(m+g-1)}}{\alpha^{2(m+g)}} \right]^2 u_\alpha^2 \tag{5}$$

where α is the degree of dissociation and u_α is the uncertainty associated with this estimation.

To estimate the stability constant it is necessary to know the degree of dissociation of the complex, α, which is obtained by the following equation:

$$\alpha = \frac{S_\text{IP} - S_{\max}}{S_\text{IP}} = 1 - \frac{S_{\max}}{S_\text{IP}} \tag{6}$$

where $S_{\max}$ is the analytical signal of the maximum at the experimental curve that represents the maximum quantity of the complex that is formed with a degree of dissociation (α) and S_IP is the analytical signal corresponding to the intersection point of the straight lines, *i.e.* the analytical signal due to the complex when the maximum quantity of complex is formed with a value of $\alpha = 0$.

For a less strong complex (Fig. 1 I.A), tangents to the curve (Fig. 1 I.B) intersect at a point which corresponds to $S_{\max}$, and are located below the intersection point of the straight lines, corresponding to S_IP, which represents the behaviour of the complex when the value of tends to zero (Fig. 1 I.C).

This difference between the intersection points of the experimental and theoretical straight lines (Fig. 1 I) means that the value of the degree of dissociation obtained with the intersection point B is considerably smaller than what would be obtained with the intersection point A. This decrease in the degree of dissociation would produce an overestimation of the stability constant; therefore, the use of $S_{\max}$, corresponding to the intersection point of experimental straight lines, is only valid when the difference between the analytical signals of the two points (A and B) is very small, *i.e.*, when the difference between the degrees of dissociation is negligible. Since the degree of dissociation depends on the value of the stability constant and
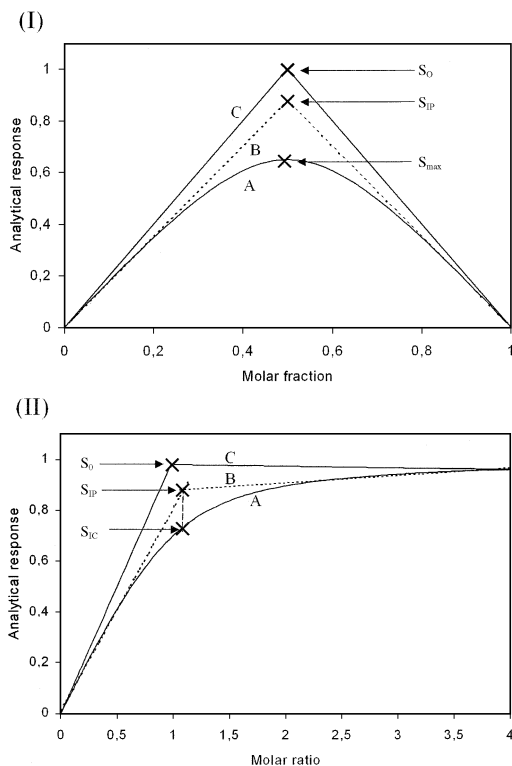
(I)

(II)

Fig. 1   (A) Simulated data for a less strong complex, (B) tangents to the curve and (C) the corresponding straight lines for a strong complex. (I) Job's method, (II) Yoe and Jones' method.

on the concentration of C ($\alpha = f(K_{est}, C)$), a limitation[12] on the use of S_{max} can be established. Thus, only when $K_{est}C^{m+g-1} \geq 100$ would be possible to use this value, because only in this case is there 99% confidence that the metal will be complexed.

In practice, and to avoid these situations, the degree of dissociation can be calculated from the value of the analytical signal of the maximum of the experimental curve, S_{max} and from the theoretical value of the analytical signal,[15] S_0, under conditions of maximum complexation ($\alpha = 0$), that is when the whole metal has been transformed into complex:

$$\alpha = \frac{S_0 - S_{max}}{S_0} = 1 - \frac{S_{max}}{S_0} \tag{7}$$

The relation between the analytical signal and the concentration complex is linear: $S = \varphi_0 + \varphi[M_mL_g]$, where φ_0 is the independent term of the function and φ is the linear coefficient that represents the proportional factor between the analytical signal and the concentration of complex. When one considers that the maximum concentration of complex ($\alpha = 0$) has been formed, then $[M_mL_g] = C_M/m$ and the expression of the analytical signal can be written in the following way:

$$S = \varphi_0 + \frac{\varphi}{m}C_M \tag{8}$$

This equation is a particular expression of the calibration function when a linear model is completed. Therefore, when an excess of reagent is used, a straight line would be obtained in the analytical signal *versus* metal concentration (C_M) plot, where the intercept is φ_0 and the slope is φ/m.

The theoretical maximum analytical signal, S_0, is obtained by

replacing the metal concentration value at the maximum of the Job's plot in the function defined by Eq. (8):

$$S_0 = \varphi_0 + (\varphi/m)C_{M,max} \tag{9}$$

where $C_{M,max}$ is the metal concentration at the stoichiometric point: $C_{M,max} = C(1 - x_{max})$.

The uncertainty of the degree of dissociation, u_α, is calculated by applying the uncertainties propagation rule to Eq. (7). This uncertainty depends on the analytical signal corresponding to the maximum of the curve, S_{max}, and the value of the maximum theoretical analytical signal, S_0, and their corresponding variances, according to the expression:

$$u_\alpha^2 = s^2\left(\frac{S_{max}}{S_0}\right) = \left(\frac{S_{max}}{S_0}\right)^2\left[\frac{u_{S,max}^2}{S_{max}^2} + \frac{u_{So}^2}{S_0^2}\right] \tag{10}$$

$u_{S,max}$ and u_{So} are the uncertainties associated with the above cited magnitudes, which are the corresponding standard deviations ($s_{S,max}$ and s_{So}) calculated by regression.

The value of $s_{S,max}$ is calculated by direct application of Eq. (24). On the other hand, s_{So} is obtained by applying Eq. (11) to Eq. (9):

$$s_{So} = \left(\frac{1}{n}s_{resid}^2 + (C_{M,max} - \overline{C}_M)^2s_{\varphi/m}^2\right)^{1/2} \tag{11}$$

where $C_{M,max}$ is the metal concentration estimated from the molar fraction in the maximum of the experimental curve and $\overline{C}_M$ is the mean of the concentrations used in the calibration curve.

The number of effective degrees of freedom (v) associated with the estimation of the stability constant (which coincide with those of the estimated degree of dissociation) necessary to set the value of the coverage constant (k)[16] in the uncertainty estimation ($K_{est} \pm k \cdot u_{Kest}$) can be calculated by applying the Welch[17]-Satterthwaite[18] equation to the components u_{Smax} and u_{So}:

$$v = \frac{\sum_i (s_i^2/n_i)^2}{\sum_i\left(\frac{(s_i^2/n_i)^2}{n_i - 1}\right)} \tag{12}$$

Yoe and Jones' method

The analytical signals measured (S) are plotted *versus* the corresponding molar ratios (r), where $r = C_L/C_M$, and a curve is obtained. The stoichiometry of the complex is estimated from the point where this curve changes its slope:

$$SMR = g/m \tag{13}$$

The stability constant, K_{est}, is given by the expression (2) and each one of these concentrations can be expressed in terms of the initial concentrations of metal (C_M) and/or ligand (C_L):

$$[M_mL_g] = (1/m)C_M(1 - \alpha)$$

$$[M] = C_M - m[M_mL_g]; \quad [L] = C_L - g[M_mL_g] = rC_M - g[M_mL_g]$$

where α is the degree of dissociation of the complex.

By substituting these expressions in Eq. (2) and regrouping terms, one obtains the following equation:

$$K_{est} = \frac{(1-\alpha)}{mC_M^{m+g-1}\alpha^m\left[r - \frac{g}{m}(1-\alpha)\right]} \quad (14)$$

The standard uncertainty associated (u_{Kest}) with the estimated value of the stability constant is obtained from the expression:

$$u_{Kest}^2 = \left[\frac{(C_M)^{1-m-g}}{g^g m^{1-g}} \cdot \frac{(m+g-1)\alpha^{(m+g)} - (m+g)\alpha^{(m+g-1)}}{\alpha^{2(m+g)}}\right]^2 u_\alpha^2 \quad (15)$$

Similarly to Job's method, it is necessary to know the degree of dissociation of the complex, α, which is obtained by the following equation:

$$\alpha = \frac{S_{IP} - S_{IC}}{S_{IP}} = 1 - \frac{S_{IC}}{S_{IP}} \quad (16)$$

where S_{IC} is the analytical signal of the corresponding molar ratio at the intersection point of the experimental curve and S_{IP} is the analytical signal corresponding to the intersection point of the straight lines (Fig. 1 II).

In practice, the degree of dissociation is calculated from the value of the analytical signal of the corresponding stoichiometric relation, S_{IC}, and from the theoretical value, S_0:

$$\alpha = \frac{S_0 - S_{IC}}{S_0} = 1 - \frac{S_{IC}}{S_0} \quad (17)$$

S_0 and the uncertainty of the dissociation degree, u_α, are calculated in a similar way to that described in the Job's method using similar equations to 9 and 10.

Applied regression models

To choose the experimental points, which can be adjusted to straight lines, one can use a least median squares regression (LMSR).[19] Standardized residuals calculated from this regression allow us to discern among the points that are fitted to a straight line and those that are considered anomalous or belonging to the curved area.

When the experimental points corresponding to linear segments are selected, the two equations that are best fitted to these data are calculated by means of linear least squares regression (LSR).[20] The extensions of these two straight lines, one with a positive slope and the other with a negative slope, intersect at a point from which the stoichiometry complex can be deduced. Due to the presence of errors in the calculation of the value of the intersection point, this might not coincide with any of the theoretical values of stoichiometry. But if the coverage interval of the intersection point is calculated and it is confirmed that this interval contains only a theoretical value of the stoichiometry, it would be possible to assign this theoretical value as the stoichiometry complex.

The experimental points corresponding to the central segment, which were eliminated from the previous calculations, can be adjusted by means of least squares polynomial regression (LSPR).[20] The value of the maximum obtained from polynomial regression is used to check that the molar fraction is included in the confidence interval of the intersection point obtained with the purpose of calculating the stoichiometry of the complex.

Finally, we carry out a strategy to model the linear relationship between the analytical signal and the concentration complex. Thus, the LSR model is applied to the experimental points obtained. The theoretical maximum signal is calculated

using the linear function estimated. The degree of dissociation can be obtained from the vertical distance between the value of the theoretical maximum analytical signal and the corresponding experimental analytical signal. Then, the stability constant of the complex can be estimated.

The calculation algorithms used are described below:

A) Least Squares Linear Regression Model

The linear regression model is based on the following equation: $S = a + bx + \Delta e$. In this case, the predictor variable (x) is the ligand molar fraction, and the variable response (S) represents the analytical signal measured, while a and b are the regression coefficients estimated and Δe denotes the random error term that affects the experimental response.

This model allows estimating more exact coefficients (a and b), and will present a smaller variance if the random errors are independent of the predictor variable, follow a normal distribution and present homocedasticity.[20] The a and b values should be such that: $\min[\Sigma\{S_i - (a + bx_i)\}^2, i = 1, 2, \cdots, n]$. However, if for some reason one or more data do not follow the linear tendency, for example, if there are anomalous values or the experimental line presents curvature, its contribution to the sum of squares is very important, as it causes an non-accurate estimation of the values of the independent term and slope.

B) Least Median Squares Regression Model

LMSR, which was proposed by Massart and Rousseeuw,[21,22] is one of the robust regression techniques that presents greatest breakdown (50%); thus, if at least 50% of the experimental points are fitted to a linear model, the application of LMSR enables the estimation of the model, without taking into account the presence or otherwise of anomalous points.

LMSR estimates the coefficients, a and b, of the linear model, such that the median of the sum of squares of the residuals reaches a minimum value: $\min$ median $[\Sigma\{y_i - (a + bx_i)\}^2, i = 1, 2, \cdots, n]$. Since the median is insensitive to the presence of large residuals, the existence of anomalous points does not modify the values of the coefficients that permit one to obtain the minimum in the previous equation. Therefore, anomalous points can be identified and then the rest of the data can be fitted to a linear model using LSR.

The 2-step process used to identify the points considered anomalous or points that present deviation from the linear model is the following: 1) Calculation of an initial estimator:[19] $s° = 1.4826\cdot[1 + 5/(n-2)]\cdot(\text{med}(R_i^2))^{1/2}$, where $1.4826 \times [1 + 5/(n-2)]$ is a correction factor, n is the number of experimental points and R_i are the residuals obtained. Those points that have a value of $|R_i/s°|$ smaller than or equal to 2.5 are considered acceptable. 2) Calculation of a final estimator: $s^* = [\Sigma R_i^2/(n^* - 2)]^{1/2}$, where n^* is the number of points selected in the previous step. The proportion $|R_i/s^*|$ is termed the *standardized residual* (SR) and so if SR > 2.5 the point is considered anomalous and should be eliminated from the linear model estimation.

C) Intersection point and coverage interval

By fitting the left and right linear segments of Job's or Yoe–Jones' plot to straight lines, the intersection point of their extensions can be obtained, and then the stoichiometry (SMF or SMR) of the complex is calculated by the abscissa value of the intersection point (x_{IP}): $S = a_1 + b_1x$ and $S = a_2 + b_2x$:

$$x_{IP} = \frac{a_2 - a_1}{b_1 - b_2} = \frac{\Delta a}{\Delta b} \quad (18)$$

As each line is associated with two coverage bands (a hyperbolic arch at each side of the line), the intersection point (x_{IP}) is a value with which a certain coverage interval can also be associated.

Of the different procedures proposed to determine the interval of coverage of the intersection point of two straight lines, the method proposed by Lark et al.[23,24] has been chosen, because it is relatively simple and allows a good estimate of the coverage interval, presenting, at the same time, solid statistical support.[25]

To obtain the coverage interval, it is necessary to solve an equation that depends on the parameters of the two regression straight lines, i.e. the slopes and their associated variances, the independent terms and their associated variances and a covariance term, because there is a correlation between the slope and the independent term in each one of the straight lines:

$$x^2[(\Delta b)^2 - k^2 s_{\Delta b}^2] - 2x[\Delta a \cdot \Delta b - k^2 s_{\Delta a \Delta b}] + [(\Delta a)^2 - k^2 s_{\Delta a}^2] = 0 \quad (19)$$

where $s_{\Delta a}^2 = s_{a1}^2 + s_{a2}^2$, is the variance of the difference of the independent terms, $s_{\Delta b}^2 = s_{b1}^2 + s_{b2}^2$, is the variance of the difference of the slopes and $s_{\Delta a, \Delta b}$ is the covariance term following the expression:

$$s_{\Delta a, \Delta b} = \bar{s}_{resid}^2 \left(\frac{\bar{x}_1}{(Q_{x^2})_1} \right) \left(\frac{\bar{x}_2}{(Q_{x^2})_2} \right) \quad (20)$$

where:

$$\bar{s}_{resid}^2 = \frac{v_1(s_{resid}^2)_1 + v_2(s_{resid}^2)_2}{v_1 + v_2} \quad (21)$$

Here, s_{resid}^2 is the residual variance in the regression, n is the number of experimental points on each one of the straight lines and $v = n - 2$ are the number of degrees of freedom on each regression line. The term Q_{x^2} is the well-known reduced sum of squares of x, given by the expression: $Q_{x^2} = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

The two values corresponding to the coverage interval associated with the intersection point in the abscissa value of the straight lines, $(x_{IP} + ks_{x,IP})$ and $(x_{IP} - ks_{x,IP})$, are obtained by solving Eq. (19). To do this, the corresponding Student's t value for a particular significance level and $v = n_1 + n_2 - 4$ degrees of freedom should be chosen as the value of the coverage constant, k. An estimate of the standard deviation associated with the intersection point ($s_{x,IP}$) can be obtained from this interval.

If the coverage interval calculated for SMF or SMR only includes one of the theoretical stoichiometry values, this value can be assigned to the complex. Otherwise, it would not be possible to assign any of the stoichiometries and it would be advisable to repeat the process after increasing the number of experimental points.

D) Least Squares Polynomial Regression Model

LSPR[26] is used to fit the experimental points located in the curved segment into a polynomial equation, the applied model is: $S = a + bx + cx^2 + \cdots + zx^p + \Delta e$, where p represents the degree of the polynomial.

By selecting the degree of the polynomial and obtaining the equation that best fits to the experimental data, it is possible to calculate the maximum of the curve, MAX = (x_{max}, S_{max}), where x_{max} would correspond to SMF or SMR and S_{max} would be the value of the predicted analytical signal for the value of SMF or SMR. To calculate the equation maximum, the first-derivate function with respect to the independent variable is made equal to zero and the roots of the resulting equation are obtained.

When there is a second order polynomial ($S = a + bx + cx^2$), the molar fraction and the analytical signal in the maximum are calculated according to the following equations:

$$x_{max} = \frac{-b}{2c}; \quad S_{max} = a + bx_{max} + cx_{max}^2 \quad (22)$$

However, in the case of a third order polynomial ($S = a + bx + cx^2 + dx^3$) when the derived equation is solved, two solutions are obtained from which we select the value with a negative second-derivate function with regard to x, according to the following equations:

$$x_{max} = \frac{-c}{3d} \pm \sqrt{\left(\frac{c}{3d} \right) - \frac{b}{3d}} \; ; S_{max} = a + bx_{max} + cx_{max}^2 + dx_{max}^3 \quad (23)$$

The standard deviation[19] associated with the maximum signal value, $s_{s,max}$, in the estimated polynomial curve is calculated by:

$$s_{S,max} = s_{resid} \sqrt{X_o^T (X^T X)^{-1} X_o} \quad (24)$$

where s_{resid} is the corresponding residual standard deviation, X_o is the matrix formed by the independent variables ($x, x^2, x^3, \cdots$) when substituting the value of the maximum, X_o^T is the transpose matrix of X_o:

$$X_o = \begin{pmatrix} 1 \\ x_{max} \\ x_{max}^2 \\ \vdots \\ x_{max}^p \end{pmatrix} \quad (25)$$

and the matrix $(X^T X)^{-1}$ is obtained from the variance-covariance matrix (V) of the regression coefficients:

$$(X^T X)^{-1} = \frac{V}{s_{resid}^2} \rightarrow V = \begin{pmatrix} s_a^2 & cov(a,b) & cov(a,c) & \cdots & cov(a,z) \\ cov(b,a) & s_b^2 & cov(b,c) & \cdots & cov(b,z) \\ cov(c,a) & cov(c,b) & s_c^2 & \cdots & cov(c,z) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ cov(z,a) & cov(z,b) & cov(z,c) & \cdots & s_z^2 \end{pmatrix} \quad (26)$$

where the main diagonal ($s_a^2, s_b^2, s_c^2, \cdots, s_z^2$) represents the variances associated with the regression coefficients ($a, b, c, \cdots, z$, respectively) and the rest of the matrix is composed of the covariances among these coefficients.

Proposed Protocol

To reach the final objective of calculating the stoichiometry and stability constant of complexes with their associated intervals of coverage, a general protocol is proposed (Fig. 2).

To check the applicability of the proposed methodology, first, the proposed protocol is applied to simulated data considering colored complexes that have an absorption maximum in the UV-visible region and second, it is applied to experimental data obtained from the application of Job's and Yoe and Jones' methods to the colored complex formed between a metallic ion and an azo-compound.
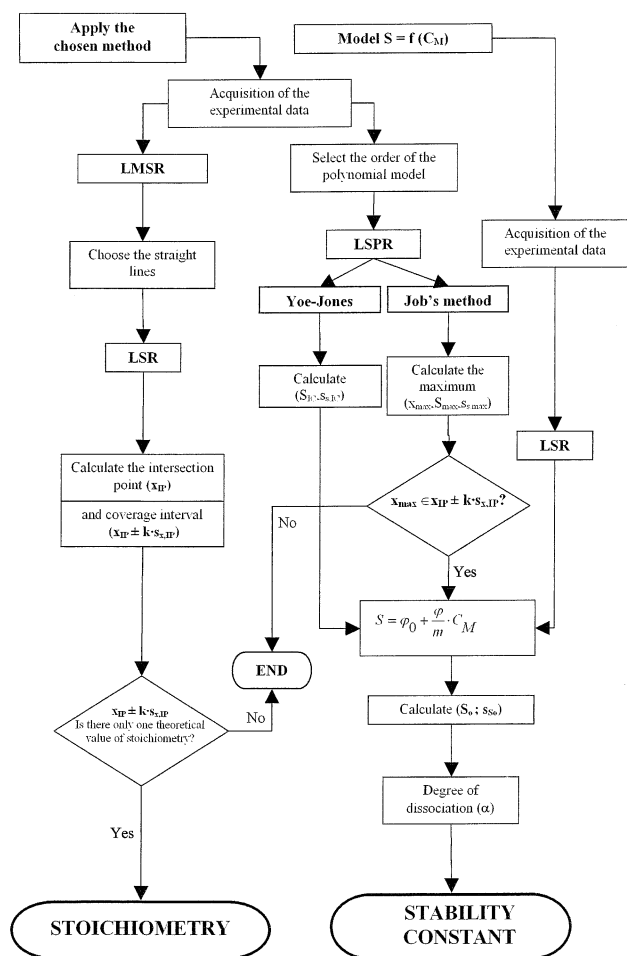
Fig. 2   Experimental flow diagram of the proposed protocol to calculate the stoichiometry and stability constant of complexes.
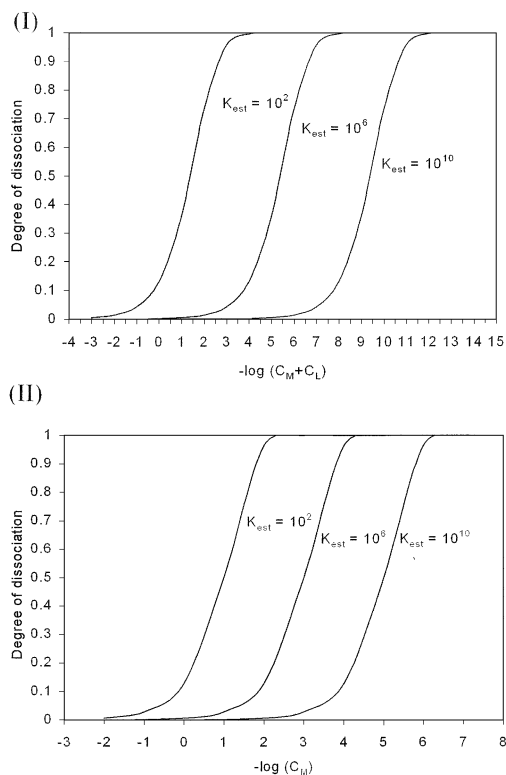

Fig. 3   Graphical representation of degree of dissociation *versus* $-\log(C_M + C_L)$ for 1:1 stoichiometry complexes and different K_{est} (10^{10}, 10^6 and 10^2 M^{-1}) using simulated data. (I) Job's method, (II) Yoe and Jones' method.

Validation of the Proposed Methodology with Simulated Data

Applying the theoretical equation of the stability constant, one can obtain simulated data for different stability constants, K_{est}, by fixing the stoichiometry m and g), the maximum absorbance $A_0 = 1$, the concentration $C = C_M + C_L$ (Job's method) or the concentration C_M (Yoe and Jones' method), which are constants in the studied cases. As the value of the dissociation degree is a function of the product of the stability constant and the concentration, $\alpha = f(K_{est}C^{m+g-1})$, if we represent α *versus* $-\log C$ (Job's method) or $-\log C_M$ (Yoe and Jones' method) for different stability constants, the value of the absorption maximum, A_{max} (Fig. 3), can be determined for any value of C and C_M in the range of stability constants studied.

The corresponding theoretical absorbance for each value of molar fraction is obtained by varying the molar fraction between 0.01 and 0.99 (99 data). At the curvature area, replicas of the points were obtained in order to get a better fit of the experimental data to the polynomial model. A variability component (SD = 0.01 or 0.02) is added to these theoretical absorbances. The variability component is obtained from a normal distribution of mean zero, N(0, SD), where SD is the theoretical standard deviation of repeatability from the absorbance values between 1% – 15% of the maximum absorbance.

LMSR was applied to the data that could constitute linear segments, including points of the curved area that might be considered to correspond to the segments under study. Points with standardized residuals greater than 2.5 were removed and LSR was applied to the points not rejected, to obtain the parameters of linear models (Tables 1 and 3) associated with each line (straight lines 1 and 2).

The intersection point was obtained by using the equations of these straight lines (Tables 1 and 3), it provides the ligand/metal molar fraction and therefore the stoichiometry of the complex.

The points close to the maximum of the curve and their corresponding replicates were fitted to a 2nd order equation by means of LSPR (Tables 1 and 3). The maximum point of this function and the standard deviation of the maximum absorbance in the curve were obtained (Tables 2 and 4).

The proportionality factor, the molar absorptivity (ε) in this case, was estimated using simulated data pairs (absorbance, metal concentration). The theoretical values (for $A_0 = 1$) and experimental values of ε are shown in Tables 2 and 4. By means of LSR, a straight line can be obtained whose slope is the value of the molar absorptivity (Tables 1 and 3, straight line 3).

The theoretical maximum absorbances and its standard deviations are estimated from straight line 3 (Tables 1 and 3) for the concentration at the stoichiometric point.

From A_{max} and A_o, the degree of dissociation and its corresponding standard deviation were calculated (Tables 2 and 4). Thus, using Eqs. (4) and (5) (Job's method) and (14) and (15) (Yoe and Jones' method), one can estimate the stability constant and its standard deviation (Tables 2 and 4).

For Job's method, in two of the cases, 1:1 and 2:1 stoichiometries, it can be seen (Table 2) that the proposed

Table 1   Job's method (parameters of the linear and polynomial models obtained from simulated data)

1:1 Stoichiometry. (K=10^6 M^{-1}; C=1.8×10^{-4} M; SD=0.01)

	Straight line 1	*Straight line 2*	**Polynomial**	**Straight line 3**
N	45	48	22	16
a	0.0026	1.9010	−3.9171	−0.0007
s_a	0.0027	0.0102	0.5936	0.0042
b	1.9545	−1.8840	19.3174	11121.68
s_b	0.0103	0.0132	2.3810	44.58
c			−19.3606	
s_c			2.3801	
R^2	0.9988	0.9977	0.7780	0.9998
s_{resid}	0.0090	0.0127	0.0098	0.0092

2:1 Stoichiometry. (K=10^{10} M^{-2}; C=1.5×10^{-4} M; SD=0.02)

	Straight line 1	*Straight line 2*	**Polynomial**	**Straight line 3**
N	64	32	62	16
A	−0.0050	2.6820	1.9169	−0.0002
s_a	0.0052	0.0480	0.8087	0.0082
B	1.3211	−2.6512	−10.4847	20088.66
s_b	0.0014	0.0572	4.1350	157.54
C			24.0784	
s_c			6.9641	
D			−16.3015	
s_d			3.8653	
R^2	0.9931	0.9862	0.9379	0.9991
s_{resid}	0.0207	0.0299	0.0170	0.0180

3:1 Stoichiometry. (K_{est} = 10^{15} M^{-3}; C = 1.06×10^{-4} M; SD = 0.01)

	Straight line 1	*Straight line 2*
N	68	23
a	−0.0157	3.6755
s_a	0.0028	0.0811
b	1.1745	−3.6543
s_b	0.0070	0.0919
R^2	0.9976	0.9868
s_{resid}	0.0115	0.0292

4:1 Stoichiometry. (K_{est} = 10^{10} M^{-4}; C = 1.6×10^{-2} M; SD = 0.01)

	Straight line 1	*Straight line 2*
N	68	15
A	−0.0575	4.6161
s_a	0.0028	0.1092
b	1.0204	4.6044
s_b	0.0064	0.1186
R^2	0.9974	0.9914
s_{resid}	0.0105	0.0198

Table 3   Yoe and Jones's method (parameters of the linear and polynomial models obtained from simulated data)

1:1 Stoichiometry. (K=10^{10} M^{-1}; C=9×10^{-9} M; SD=0.01)

	Straight line 1	*Straight line 2*	**Polynomial**	**Straight line 3**
N	38	51	50	16
A	0.0090	0.8925	−1.5923	−0.0092
s_a	0.0032	0.0082	0.6032	0.0054
B	0.9546	0.0511	5.5479	1.12×10^8
s_b	0.0062	0.0053	1.7392	5.79×10^5
C			−4.0709	
s_c			1.6551	
D			1.0144	
s_d			0.5201	
R^2	0.9985	0.9586	0.9791	0.9996
s_{resid}	0.0086	0.0112	0.0086	0.0119

2:1 Stoichiometry. (K=10^{10} M^{-2}; C=1.5×10^{-4} M; SD=0.01)

	Straight line 1	*Straight line 2*	**Polynomial**	**Straight line 3**
N	68	37	32	16
A	0.0125	0.6874	3.0108	0.0044
s_a	0.0033	0.0240	8.1130	0.0029
B	0.4508	0.1134	−4.0517	6622.77
s_b	0.0027	0.0097	12.0150	18.36
C			2.3121	
s_c			5.9171	
D			−0.4073	
s_d			0.9690	
R^2	0.9977	0.9950	0.9347	0.9999
s_{resid}	0.0129	0.0191	0.0111	0.0063

3:1 Stoichiometry. (K_{est} = 10^{15} M^{-3}; C = 2.65×10^{-5} M; SD = 0.02)

	Straight line 1	*Straight line 2*	**Polynomial**	**Straight line 3**
N	57	42	26	16
A	−0.0359	0.5729	41.5658	−0.0005
s_a	0.0040	0.0278	20.2117	0.0083
B	0.2807	0.0912	−40.7100	3.77×10^4
s_b	0.0022	0.0069	20.2896	0.03×10^4
C			13.4424	
s_c			6.7757	
D			−1.4676	
s_d			0.7527	
R^2	0.9965	0.9965	0.8609	0.9991
s_{resid}	0.0148	0.0271	0.0190	0.0182

4:1 Stoichiometry. (K_{est} = 10^{20} M^{-4}; C = 4.3×10^{-5} M; SD = 0.01)

	Straight line 1	*Straight line 2*	**Polynomial**	**Straight line 3**
N	69	31	36	16
A	−0.0240	0.8431	0.3495	−0.0051
s_a	0.0026	0.0230	4.0701	0.0049
B	0.2334	0.0289	−0.0864	3.32×10^4
s_b	0.0011	0.0045	3.0478	0.01×10^4
C			0.1041	
s_c			0.7584	
D			−0.0120	
s_d			0.0627	
R^2	0.9986	0.9860	0.9752	0.9997
s_{resid}	0.0107	0.0135	0.0088	0.0107

Table 2   Job's method (parameters calculated from simulated data)

	1:1 stoichiometry	*2:1 stoichiometry*	*3:1 stoichiometry*	*4:1 stoichiometry*
x_{IP}	0.4946	0.6765	0.7644	0.8309
s_{xIP}	0.0034	0.0155	0.0221	0.0266
Coverage interval	0.488 – 0.502	0.646 – 0.708	0.7201 – 0.8087	0.7787 – 0.8831
x_{max}	0.4989	0.6597		
A_{max}	0.9015	0.7990		
$s_{A,max}$	0.0032	0.0038		
Theoretical ε	1.111×10^4	2.00×10^4		
Experimental ε	1.112×10^4	2.01×10^4		
A_0	1.0017	1.0043		
s_{A0}	0.0023	0.0046		
α	0.100	0.204		
$s_α$	0.004	0.005		
K_{est}	9.99×10^5	9.32×10^9		
s_K	1.63×10^5	1.64×10^9		
ν*	36	21		

k = 2 for a significance level of 95% and ν* degrees of freedom

Table 4   Yoe and Jone's method (parameters calculated from simulated data)

	1:1 stoichiometry	*2:1 stoichiometry*	*3:1 stoichiometry*	*4:1 stoichiometry*
r_{IP}	0.978	2.001	3.213	4.240
s_{rIP}	0.013	0.542	0.381	0.146
Coverage interval	0.952 – 1.004	1.823 – 2.192	2.850 – 3.622	3.955 – 4.542
A_{IC}	0.8991	0.897	0.792	0.902
s_A	0.0020	0.003	0.006	0.002
Theoretical ε	1.111×10^8	6.67×10^3	3.78×10^4	2.31×10^4
Experimental ε	1.12×10^8	6.62×10^3	3.77×10^4	2.32×10^4
A_0	0.999	0.998	0.997	0.998
s_{A0}	0.003	0.002	0.005	0.003
α	0.100	0.100	0.206	0.096
$s_α$	0.003	0.003	0.007	0.003
K_{est}	1.00×10^{10}	9.77×10^9	8.82×10^{14}	1.23×10^{20}
s_K	0.52×10^{10}	1.04×10^9	2.50×10^{14}	0.45×10^{20}
ν*	19	46	36	24

k = 2 for a significance level of 95% and ν* degrees of freedom

methodology is valid and that the conditions imposed are fulfilled, *i.e.* the coverage interval of the molar fraction includes only one theoretical stoichiometry value (0.50 and 0.67, respectively), the maximum of the curve is also included in this interval and the coverage interval of the stability constant includes the previously fixed value.   Figure 4A shows the intersection points for a 1:1 stoichiometry and Fig. 4B shows the corresponding coverage intervals.

The results shown in Table 4 allow one conclude that, for the Yoe and Jones' method and in all cases, the coverage interval estimated for the molar ratio includes only the corresponding theoretical stoichiometry value and the previously fixed stability constant is also included in the coverage interval of the stability constant obtained.   Therefore, the proposed methodology is applicable in the four cases studied: 1:1, 2:1, 3:1 and 4:1 stoichiometries.

When the variability level is represented graphically (SD) *versus* the dissociation degree (Fig. 5) boundary curves are obtained.   These curves represent the applicability limit of the proposed procedure for stoichiometry tested.   The area bounded by the curves gives us the area where the method can be applied, *i.e.* the maximum standard deviation that the method can present for each degree of dissociation.   For example, in Fig. 5-I, if there is a standard deviation of 5%, the method is applicable to a complex with a degree of dissociation between 0.07 and 0.85 for a 1:1 stoichiometry and only to a complex with a degree of dissociation between 0.10 and 0.36 for a 2:1 stoichiometry.
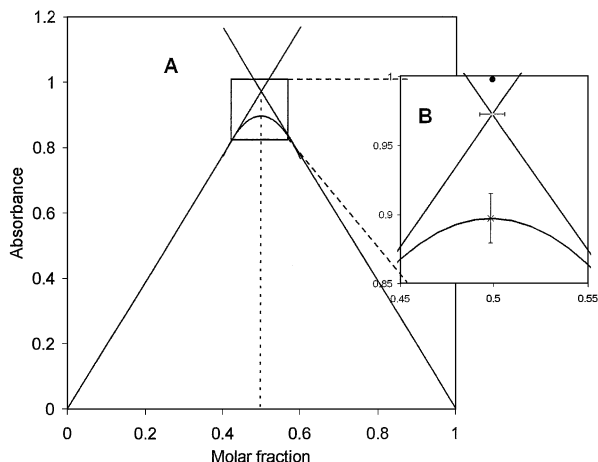
Fig. 4   Job's method.   Intersection point for a 1:1 stoichiometry complex and the corresponding coverage intervals.

Application of the Proposed Methodology to Real Data

To illustrate the proposed methodology, we have carried out a complete application for the determination of the stoichiometry and stability constant of the complexes formed by Ni(II) and Co(II) and *p*-nitrophenylazo-2-naphthol-3,6-disulfonic acid (disodium salt) (pRNa) (pH 9.30 and 9.60, respectively) using the Job's and Yoe and Jones' method, respectively.

The application of the proposed methodology to the experimental data obtained gave the following results: LMSR criteria allow us to establish the experimental points belonging to linear segments. In this way, it ensures objectivity in the selection of the points that will be fitted, using LSR, to a straight line in each experimental area, with the previous elimination of the anomalous points. The parameters of the adjusted straight line 1 and straight line 2 are shown in Tables 5 and 6.

The values of the molar fraction of the Ni(II) complex and of the molar ratio of the Co(II) complex are: 0.695 ± 0.031 ($k = 2$) and 2.149 ± 0.297 ($k = 2$), respectively. They have been obtained from the calculation of the abscissa coordinate of the intersection point and the corresponding horizontal interval. These values clearly allow us to attribute a 2:1 stoichiometry to the complexes, since they do not differ significantly from the theoretical value, 0.666 and 2, respectively, and the molar fractions and the molar ratios that would indicate other stoichiometries are not included in the above coverage intervals ($0.665 - 0.727$ and $1.872 - 2.465$, respectively).

Experimental points on the curved segments were then adjusted to a polynomial function by means of LSPR. The coefficients of a third-order equation are shown in Tables 5 and 6.

MAX = (x_{max}, A_{max}) = (0.669; 0.603) represents the maximum of the curve for the Ni(II)-pRNa complex. The x_{max} value is included inside the interval of horizontal coverage associated with the intersection point, $x_{IP} \pm kS_{xIP}$ ($0.669 \in [0.665 - 0.727]$) and consequently it is possible to calculate the degree of dissociation. The standard deviation of the absorbance of this maximum is $s_{S,max} = 0.003$.

The value of absorbance associated with the molar ratio of the Co(II)-pRNa complex is: $A_{IC} = 0.554 \pm 0.004$.

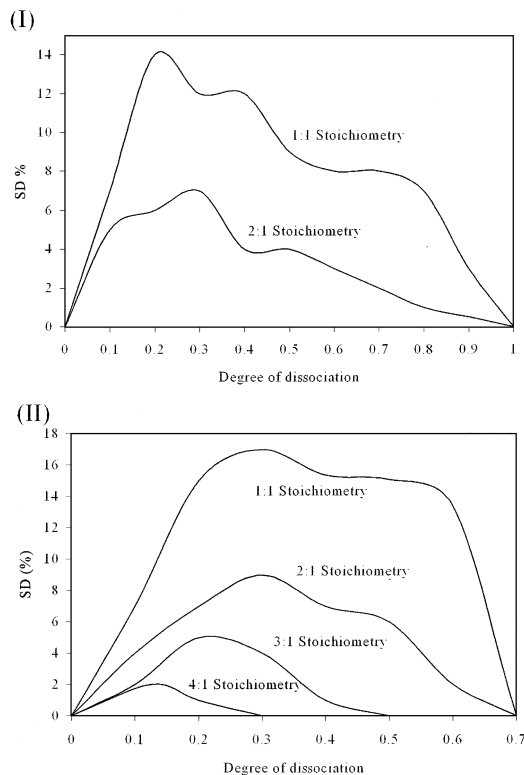The values of molar absorptivities for the Ni(II) and Co(II)

Fig. 5   Applicability limits of the proposed procedure for: (I) Job's method. 1:1 and 2:1 stoichiometry complexes. (II) Yoe and Jones' method. 1:1, 2:1, 3:1 and 4:1 stoichiometry complexes. SD = theoretical standard deviation of repeatability.

complexes, $\varepsilon = 11650$ and 7427.5 dm^3 mol^{-1} cm^{-1}, respectively, were estimated from the slopes of the appropriate calibration graphs. The parameters of the linear patterns are shown in Tables 5 and 6 (straight line 3).

The theoretical maximum values of absorbance (A_0) and its standard deviation (s_{Ao}) for the metal concentration ($C_{Ni} = 6.67 \times 10^{-5}$ M and $C_{Co} = 8.0 \times 10^{-5}$ M) in the stoichiometric points are: 0.681, 0.004, and 0.597, 0.001, respectively. Then, the degrees of dissociation and its standard deviations were calculated: 0.115, 0.007 and 0.071, 0.005. Finally, the estimated values of the stability constants and its coverage intervals are $K_{est} \pm ks_K = 3.30 \times 10^{10} \pm 1.21 \times 10^{10}$ M^{-2}, with $k = 2$, a significance level of 95% and 10 degrees of freedom (Ni(II) complex) and $K_{est} \pm ks_K = 8.95 \times 10^{11} \pm 3.47 \times 10^{11}$ M^{-2}, with $k = 2$, a significance level of 95% and 22 degrees of freedom (Co(II) complex).

The uncertainty associated with the stability constant is so large (a relative uncertainty of 37% and 39%, respectively) because too few experimental points were used to calculate the stoichiometry. Therefore, to reduce this uncertainty it would be necessary to carry out additional experiments.

Acknowledgements

Table 5   Ni(II)-pRNa$_2$ complex (parameters of the linear and polynomial models obtained from real data)

	Straight line 1	Straight line 2	Polynomial	Straight line 3
N	11	5	16	10
a	-0.072	2.142	1.0288	-0.0958
s_a	0.004	0.035	0.2192	0.0070
b	1.083	-2.100	-6.1441	1165.00
s_b	0.011	0.042	1.3155	131.95
c			15.5167	
s_c			2.5630	
d			-10.8889	
s_d			1.6244	
R^2	0.9990	0.9988	0.9966	0.9990
s_{resid}	0.006	0.007	0.0070	0.0094

Table 6   Co(II)-pRNa$_2$ complex (parameters of the linear and polynomial models obtained from real data)

	Straight line 1	Straight line 2	Polynomial	Straight line 3
N	15	7	16	8
a	-0.0468	0.3101	-1.3331	0.0027
s_a	0.0052	0.0174	0.9952	0.0039
b	0.3064	0.1403	1.8376	7427.5
s_b	0.0042	0.0066	1.3111	53.6
c			-0.5724	
s_c			0.5711	
d			0.0627	
s_d			0.0822	
R^2	0.9950	0.9781	0.9875	0.9996
s_{resid}	0.0087	0.0044	0.0055	0.0034

Glossary of Symbols and Abbreviations

C_M: concentration of metal M

C_L: concentration of ligand L

x: molar ligand fraction

r: molar ratio

SMF: stoichiometric molar fraction

SMR: stoichiometric molar ratio

m and g: stoichiometric coefficients of the complex M_mL_g

K_{est}: stability constant of the complex M_mL_g

u_{Kest}: uncertainty associated with stability constant of the complex M_mL_g

α: degree of dissociation

u_α: uncertainty associated with the degree of dissociation

LMSR: least median squares regression

LSR: least squares regression

LSPR: least squares polynomial regression

$a, b, \cdots, z$: parameters of regression

$s_a, s_b, \cdots, s_z$: standard deviation of parameters of regression

n: number of experimental data

Q_{x^2}: reduced sum of squares of x

s_S or s_A and s_x: standard deviations of analytical response and molar fraction

s_{resid}: residual standard deviation in the regression

v: degrees of freedom

SR: standardized residual

x_{IP}: molar fraction in the intersection point between regression curves

S_{IP} or A_{IP}: analytical response in the intersection point between regression curves

S_{IC} or A_{IC}: analytical response of the corresponding molar ratio at the intersection point of the experimental curves

$s_{x,IP}$: standard deviation of the molar fraction in the intersection point between regression curves

k: coverage factor

x_{max}: molar fraction in the maximum of the polynomial regression

S_{max} or A_{max}: analytical response in the maximum of the polynomial regression

$s_{S,max}$ or $s_{A,max}$: standard deviation of the molar fraction in the maximum of the polynomial regression

S_0 or A_0: theoretical maximum analytical response

s_{So} or s_{Ao}: standard deviation of theoretical maximum analytical response

SD: theoretical standard deviation of repeatability

References

1. P. Gans, A. Sabatini, and A. Vacca, *Talanta*, **1996**, *43*, 1739.

2. P. Gans, A. Sabatini, and A. Vacca, *J. Chem. Soc. Dalton Trans.*, **1985**, 1195.

3. V. I. Vetrogon, N. G. Lukynenko, M.-J. Schwing-Weill, and F. Arnaud-Neu, *Talanta*, **1994**, *41*, 2105.

4. J. Havel and M. Meloun, *Talanta*, **1986**, *33*, 525.

5. P. Job, *Ann. Chim.*, **1928**, *113*, 203.

6. J. H. Yoe and A. L. Jones, *Ind. Eng. Chem. Anal. Ed.*, **1944**, *16*, 111.

7. A. E. Harvey and D. L. Manning, *J. Am. Chem. Soc.*, **1950**, *72*, 4488.

8. A. Holme and F. J. Langmyhr, *Anal. Chim. Acta*, **1966**, *36*, 383.

9. N. J. Rose and R. S. Drago, *J. Am. Chem. Soc.*, **1959**, *81*, 6138.

10. H. E. Bent and C. L. French, *J. Am. Chem. Soc.*, **1941**, *63*, 568.

11. H. Diehl and F. Lindstrom, *Anal. Chem.*, **1959**, *31*, 414.

12. D. V. González García, A. Arrebola Ramírez, and M. Román Ceba, *Talanta*, **1979**, *26*, 215.

13. T. Román Galan, A. Arrebola Ramírez, and M. Román Ceba, *Talanta*, **1980**, *27*, 545.

14. J. J. Berzas Nevado, A. Arrebola Ramírez, and M. Román Ceba, *Anal. Chim. Acta*, **1981**, *124*, 201.

15. W. Likussar and D. F. Boltz, *Anal. Chem.*, **1971**, *43*, 1265.

16. International Organisation for Standadisation, "*Guide to the Expression of Uncertainty in Measurement*", **1993**.

17. B. L. Welch, *J. R. Stat. Soc. Suppl.*, **1936**, *3*, 29; *Biometrika*, **1938**, *29*, 350; *ibid.*, **1947**, 34, 28.

18. F. E. Satterthwaite, *Psychometrika*, **1941**, *6*, 309; *Biometrics Bull.*, **1946**, *2*, 110.

19. D. L. Massart, B. G. M. Vandeginste, L. M. C. Buydens, S. De Jong, P. J. Lewi, and J. Smeyers-Verbeke, "*Handbook of Chemometrics and Qualimetrics: Part A*", **1997**, Elsevier, Amsterdam, London, New York, Tokyo, Chap. 10, 263 and Chap. 12, 339.

20. R. De Levie, *Crit. Rev. Anal. Chem.*, **2000**, *30*, 59.

21. D. L. Massart, L. Kaufman, P. J. Rousseeuw, and A. Leroy, *Anal. Chim. Acta*, **1986**, *187*, 171.

22. P. J. Rousseeuw and A. Leroy, "*Robust Regression and Outlier Deterction*", **1987**, John Wiley and Sons, New York, Chichester, Brisbane, Toronto.

23. P. D. Lark, B. R. Graven, and R. C. L. Bosworth, "*The Handling of Chemical Data*", **1968**, Pergamon, London, 177.

24. C. Liteanu and I. Rica, "*Statistical Theory and Methodology of Trace Analysis*", **1990**, Ellis Horwood, Chichester, 141.

25. K. N. Carter, Jr., D. M. Scott, J. K. Salmon, and G. S. Zarcone, *Anal. Chem.*, **1991**, *63*, 1270.

26. W. Funk, V. Dammann, and G. Donnevert, "*Quality Assurance in Analytical Chemistry*", **1995**, VCH, Weinheim, 185.