Nonlinear regression in parameter estimation from polarographic signals[☆]

A.A.C.C. Pais *, J.L.G.C. Pereira, J.S. Redinha

Departamento de Química, Universidade de Coimbra P-3049 Coimbra Codex, Portugal

Abstract

In this work we describe a detailed treatment of polarographic data curves, including error analysis, by means of nonlinear least-squares in its standard form (or resorting to the errors in variables model). Error estimates for the related parameters are additionally verified by Monte-Carlo simulation and resampling techniques. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Least-squares; Monte-Carlo; Bootstrap; Jackknife; Parameter estimation; Polarography

1. Introduction

The study of complex formation reactions is crucial to the understanding of a variety of processes found in areas such as environmental analytical chemistry (Broekaert et al., 1989), medicine and pharmacology, food production and in several other industrial applications.

The most usual methods to study complex formation reactions are based on separation techniques and spectroscopic, electrochemical or thermodynamical approaches (Bond, 1971; Anderegg, 1982; Nancollas and Tomson, 1982; Kiss et al., 1991). However, the range of applicability of each technique is constrained by the actual characteristics of the chemical system, e.g., separation techniques usually fail with labile complexes, spectrophotometry can only be employed for chromophores, direct use of potentiometric methods is only possible for a few ions and polarographic methods are also restricted to a few electroactive cations.

Polarographic techniques, including differential pulse polarography (DPP), are simultaneously fast, reliable, and allow for the direct determination of the metal ion concentration which makes them suitable candidates for such studies (Bond, 1971; Heath and Hefter, 1977; Cukrowsky, 1996).

These techniques must, however, be complemented by careful data analysis in order to explore their full capabilities. Although some attempts have been made in this direction, for example in what concerns deconvolution of overlapping polarographic signals (Huang et al., 1995), parameter estimation from reliable physical models with the associated determination of errors and confidence intervals has not, to our knowledge, been carried out. We note that these models give rise to an intrinsically nonlinear regression analysis, which increases the complexity of the problem, especially in what concerns the estimation of confidence intervals for the estimated parameters.

The structure of this article is as follows: Section 2 presents and discusses the mathematical models available for polarographic curves, Section 3 describes the equipment and preliminary experimental error determinations in the relevant variables, Section 4 deals with the regression estimates of parameters and associated

---

* Corresponding author.

standard errors, comparing several approaches, and Section 6 gathers some conclusions.

2. Mathematical models

Polarographic applications to complexometric studies have been made for almost 60 years. A brief review can be found in Bond, (1971), Cukrowsky (1996). Cukrowsky, (1996) described an easy and convenient polarographic method for labile complexes at fixed ligand to metal ratio (CL/CM) and variable pH. Assuming constant temperature and constant ionic strength, this author presented an expression relating the peak current and peak potential with the fraction of free metal in solution for the reversible reduction of metal M^{n+} by DPP at a dropping mercury electrode,

$$\left(\frac{i_p^M}{i_p^{ML}}\right)\exp\left[\frac{nF}{RT}(E_p^M - E_p^{ML})\right] = \frac{C_M}{[M]} = 1 + \sum_{i=1}^{m}\beta_i[L]^i \tag{1}$$

were i_p^M and i_p^{ML} are DPP peak current and E_p^M and E_p^{ML} peak potential for metal reduction without and in the presence of the ligand, respectively, n is the number of electrons involved in the electrode process, R is the gas constant, T the absolute temperature, C_M the analytical concentration of metal in solution, $[M]$ and $[L]$ are free metal and ligand concentrations in solution and, finally, m is the stoichiometric coefficient for the ligand with β_m formation constant. Peak current and peak potential can be accurately estimated from polarographic curves through (Bond, 1980; Churáček et al., 1993)

$$i = i_p\left[\frac{P_A(\sigma + 1)^2}{\sigma P_A^2 + (\sigma^2 + 1)P_A + \sigma}\right] \tag{2}$$

where P_A and σ are defined by

$$P_A = \exp\left[\frac{nF}{RT}\left(E + \frac{\Delta E}{2} - E_{1/2}\right)\right] \tag{3}$$

and

$$\sigma = \exp\left[\frac{nF}{RT}\left(\frac{\Delta E}{2}\right)\right] \tag{4}$$

ΔE is the pulse amplitude, E is the potential applied to the working electrode and $E_{1/2}$ is the polarographic half-wave potential of the electroactive compound.

In the case of small pulse amplitudes ($n\Delta E < 2RT/F$) Eq. (2) can be simplified to

$$i = 4i_p\left[\frac{P}{(1 + P)^2}\right] \tag{5}$$

where

$$P = \exp\left[\frac{nF}{RT}(E - E_{1/2})\right] \tag{6}$$

Additionally, polarographic background current can be expressed, for a relatively narrow potential range, as a polynomial function of the applied potential,

$$i_o = A_0 + A_1E + A_2E^2 + \dots \tag{7}$$

3. Experimental

3.1. Instrumentation

Polarographic experiments were conducted using a 693 VA Processor/694 VA Stand (Metrohm) system equipped with a multimode working electrode, a double-junction reference electrode Ag/AgCl, 3.0 M NaCl and a Platinum wire (Metrohm) as counter electrode. This equipment was connected to a personal computer for data acquisition and processing.

DPP experiments were conducted on previously deareated electrolyte solutions of 50 μM Pb^{2+} with a dropping mercury electrode (1 s drop time, $0.370 \pm 0.002\ mm^2$). We used a pulse of small amplitude (20 mV) with a duration of 40 ms (current sampling in the last 10 ms of the pulse) and step increments of 2 mV in the cathodic direction ($-2\ mV\ s^{-1}$) from -260 to -510 mV.

All experiments were conducted at constant temperature conditions, $25.0 \pm 0.1°C$.

3.2. Experimental error characterization

Potential values in DPP experiments were checked by direct reading, with a digital micromultimeter (resolution $\pm 10\ \mu V$), of ten replicates of each potential value in the working range. Related errors were found to be additive, with zero mean, non-correlated, homochedastic and normally distributed according to auto-correlation, skewness and kurtosis tests (Churáček et al., 1993) conducted on our experimental data.

Current values in DPP experiments were checked recording 19 polarogram of Pb^{2+} in acidic media. Related errors are, in this case, additive, non-correlated, heterochedastic and normally distributed.

The relative error associated to the dependent variable is approximately three orders of magnitude larger than that of the independent ($\sigma^2 \approx 1.40 \times 10^{-10}\ V^2$, see Fig. 1). The absolute error associated to the former variable can adequately be represented by (see Rocke and Lorenzato, 1995 for a related structure)

$$\sigma^2(E) = A + B\left(\frac{di}{dE}\right)^2 \tag{8}$$

This form is very convenient because it not only provides a continuous and smooth representation of the error but can also be used in similar experimental conditions to estimate the error behaviour in the work-

ing potential range. It is based on the assumption of the existence of two contributions for the error in the dependent variable, one corresponding to the intrinsic instrumental fluctuation and the other to error propagation from the independent variable. We note that, in the zone where the background current dominates, extreme weighting of some points occurs by direct use of the experimental error vector. Also, the fact that it makes use of the first derivative of the intensity of the cathodic current in order to the potential also allows its use in situations where the peak potential is cathodically driven due to the presence of increasing concentrations of ligand in solution. In Fig. 2 we depict the experimental error and its representation by means of Eq. (8). For the above conditions the values of A and B are, respectively, 1.597×10^{-6} μA^2 and 2.299×10^{-7} V^2. These were obtained via unweighted linear least squares.

The experimental error distribution is not symmetric, unlike Eq. (8), probably due to stabilization of the
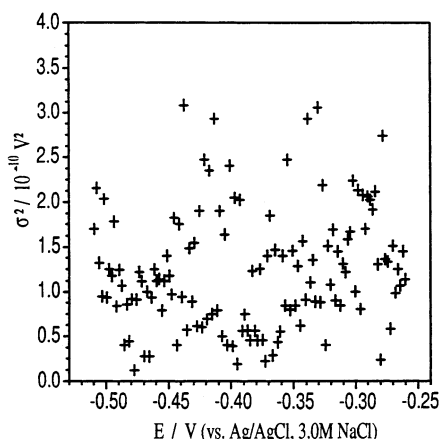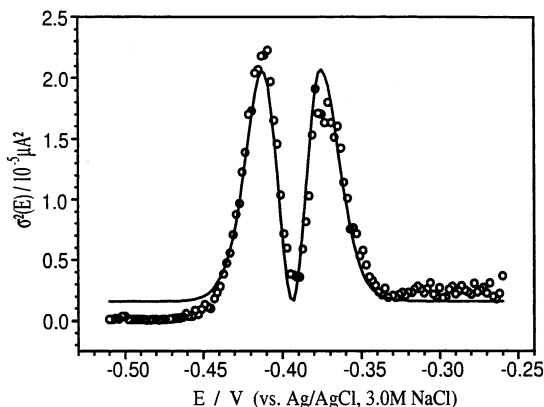

Fig. 1. Experimental error in potential values.

Fig. 2. Experimental error ($\bigcirc$) and its representation by Eq. (8) (solid line).

system after the initial perturbation. Although inclusion of a non-squared derivative term in the equation could approximately account for this effect, we have chosen to use the simpler form.

4. Regression procedure and estimation of confidence intervals

Nonlinear regression was carried out using both the Levenberg–Marquardt algorithm (Marquardt, 1963; Press et al., 1992) and the 'error-in-variables model' (Reilly et al., 1993). The approximate analytical expression for the experimental errors was used for weighting the squared deviations, except where indicated,

$$\chi^2 = \sum_{i=1}^{n} w_i(i_i - \eta_i)^2 = \sum_{i=1}^{n} \left(\frac{y_i - \eta_i}{\sigma(E_i)}\right)^2 \qquad (9)$$

considering n non-correlated errors. In the case of the 'error-in-variables model' the maximum likelihood expression simply accommodates errors in both variables.

We note that, in both cases, a design matrix ($A(npar \times npar) = [a_{jk}]$) is constructed by cross products of weighted $npar$ sensitivity coefficients,

$$[a_{j,k}] = \sum_{i=1}^{n} w_i \frac{\partial \eta_i}{\partial \theta_j} \frac{\partial \eta_i}{\partial \theta_k}, \qquad \forall j,k \in \{1,...,npar\} \qquad (10)$$

Matrix inversion of Eq. (10) leads to the covariance matrix error estimates of $npar$ parameters ($C(npar \times npar)$)

$$\hat{\sigma}_j = \sqrt{C_{jj}}, \quad \forall j \in \{1,...,npar\} \qquad (11)$$

The use of the approximate expression for the error is thus convenient in order to circumvent problems related to the inversion of ill-conditioned design matrices.

A rejection method (Press et al., 1992) was used to generate the normal distribution function, giving a slightly more rapid convergence rate then the Box-Muller algorithm (Box and Muller, 1958; Press et al., 1992). All uniform distributions were generated via a congruence formula (Press et al., 1992).

4.1. Parameter and error estimation methods

Once parametric error estimates are approximate for nonlinear regression (Beck and Arnold, 1977) we have also used non-parametric estimation methods including jackknife (Efron and Gong, 1983), bootstrap (Efron and Gong, 1983; Efron and Tibshirani, 1986; DiCiccio and Tibshirani, 1987) and Monte-Carlo simulation (Beck and Arnold, 1977; Press et al., 1992)

Jackknife error estimation methods are based on the effect of variability introduced in the estimated parameters by removing each data point, one at a time, from the original data set. If the parameter θ is a function of n data points, Jackknife estimate for the central position ($\hat{\theta}_J$) is,

$$\hat{\theta}_J = \frac{1}{n} \sum_{\substack{j \neq i}}^{n} \theta_{(i)} \tag{12}$$

while for the corresponding parameter dispersion ($\hat{\sigma}_J$),

$$\hat{\sigma}_J = \sqrt{\left(\frac{n-1}{n}\right) \sum_{i=1}^{n} (\theta_{(i)} - \hat{\theta}_J)^2} \tag{13}$$

is obtained, where $\theta_{(i)}$ is the parameter estimated with $(n-1)$ data points, when the ith data point is temporarily removed from original data set.

Bootstrap estimation can be carried out resorting to resampling of data points (Efron and Gong, 1983; Efron and Tibshirani, 1986) or residuals (Bonate, 1993; Jones et al., 1996). Estimation based on resampling of data points is accomplished by generating n_B pseudo-samples of the same size of the original set. These are obtained by drawing each point independently with replacement and equal probability from the sample. Bootstrap estimation with resampling of residuals is conducted in a similar way but the sample now refers to the residual pool. In this case, new residuals for each data point are computed by

$$e_i^* = \phi \, e_j \sqrt{w_j / w_i} \tag{14}$$

where r_j is the residual in the original jth point ($e_j = y_j - \eta_j$) and ϕ is a adjustment factor (Jones et al., 1996) ($\phi = \sqrt{n/(n-p)}$) for n data points and a p parameters model. These recalculated residuals are added to the respective model prediction (η_i),

$$y_i^* = \eta_i + e_i^* \tag{15}$$

The pseudo-samples yield n_B parameters which allow the calculation of bootstrap estimates of position ($\hat{\theta}_B$),

$$\hat{\theta}_B = \frac{1}{n_B} \sum_{i=1}^{n_B} \theta_i^* \tag{16}$$

and parameter dispersion ($\hat{\sigma}_B$),

$$\hat{\sigma}_B = \sqrt{\frac{1}{n_B - 1} \sum_{i=1}^{n_B} (\theta_i^* - \hat{\theta}_B)^2} \tag{17}$$

If the number of replicates is high (> 200), bootstrap confidence intervals can be more efficiently estimated from central percentile methods.

Monte-Carlo, estimates require a previous characterization of experimental errors followed by a simulation based on these determined errors. The data sets are constructed with random distributions centered at each experimental value with the respective distribution. Mean values of parameters can be obtained with expressions similar to those of the bootstrap method. Also, percentile confidence intervals are the most usual method for dispersion estimation.

4.2. DPP model

Polarographic DPP signals modeling consists in obtaining estimates for three polarographic parameters

plus those required to describe the background current. For the small pulse amplitude case, the fitting function is

$$i = 4\theta_1 \frac{\exp\left(\dfrac{E - \theta_2}{\theta_3}\right)}{\left[1 + \exp\left(\dfrac{E - \theta_2}{\theta_3}\right)\right]^2} + \sum_{j=4}^{npar} \theta_j E^{(j-4)} \tag{18}$$

where θ_1 and θ_2 are, respectively, i_p and E_p The θ_3 parameter is essentially useful to check redox reversibility. The remaining ($npar - 3$) parameters are required for the description of the background current.

5. Results and discussion

Least squares procedures were carried out over a polarographic set of (E,i) points corresponding to the average of the 19 acquisitions for each current value and ten determinations of the associated potential value. In all cases results were obtained for the complete model described by Eq. (2) and using the small amplitude case of Eq. (5) (also rewritten as Eq. (18)). The simplified approach has proved to be accurate enough for our purposes, yielding similar χ^2 values as the complete model. Also, similar results were obtained when errors in both the independent and dependent variables were considered and when only error in the dependent variable was taken into consideration (see Table 1). The maximum likelihood weights of Eq. (9) were described by Eq. (8). The background current was appropriately described by a straight line ($npar = 5$). Other tested values ($npar = 4$, 6 and 7) have proved to be either insufficient or inconsistent with determinations of the baseline from blank runs. These findings were corroborated by several criteria including parsimony (Beck and Arnold, 1977), Akaike, mean quadratic error of prediction, ANOVA lack-of-fit and others (Meloun et al., 1992) of which we omit the details.

Monte-Carlo simulation was conducted in two different ways: in one case, weighting and error distributions in each point were generated on the basis of Eq. (8), while in the other the actual measured error was used for both purposes. Obviously, we have avoided situations in which experimental and approximated error expressions were used in conjunction.

The results obtained for the most relevant polarographic parameters (θ_1 and θ_2) are summarized in Table 2. In this table, bootstrap and Monte-Carlo estimates are based on percentile determinations. In Fig. 3 we depict the experimentally determined points superimposed on the fitted function. Table 3 summarizes the convergence properties for methods involving an arbitrary number of simulations (bootstrap and Monte-Carlo). The simplified approach (Eq. (18)) was used in all cases.

Table 1
Estimation results for parameters θ_1 and θ_2, using different approaches[a]

npar	σ_{Fit}	θ_1	$\hat{\sigma}_1$	θ_2	$\hat{\sigma}_2$
Levenberg–Marquardt					
4	1.3546	0.32633	0.00068	−0.393938	0.000088
5	0.3688	0.32597	0.00068	−0.393647	0.000090
6	0.2754	0.32718	0.00082	−0.393629	0.000090
7	0.2689	0.32702	0.00084	−0.393602	0.000098
Error-in-variables model					
4	1.2247	0.32552	0.00073	−0.393884	0.000077
5	0.2730	0.32605	0.00073	−0.393660	0.000095
6	0.2532	0.32685	0.00075	−0.393605	0.000104
7	0.2142	0.32726	0.00077	−0.393545	0.000106

[a] Current intensities are in µA and potential values in V. The second column corresponds to $\sigma_{\text{Fit}} = \sqrt{\chi^2/(n-npar)}$, and $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are given by Eq. (11).

Table 2
Least squares results for parameters θ_1 and θ_2, using different approaches based on the Levenberg–Marquardt method[a]

Method	Values	Confidence intervals		RSD(%)	RE(%)
		68.3%	95.5%		
	Parameter θ_1				
Least squares (param.)	0.32597	±0.00068	±0.00136	0.21	–
Jackknife	0.32597	0.00032	0.00064	0.10	0.000
Bootstrap (data points, 37%)	0.32601	0.00028	0.00057	0.08	0.014
Bootstrap (data points, 100%)	0.32622	0.00041	0.00090	0.13	0.078
Bootstrap (residuals, 37%)	0.32598	0.00034	0.00065	0.10	0.000
Bootstrap (residuals, 100%)	0.32597	0.00035	0.00068	0.17	0.000
Monte-Carlo (Eq. (8))	0.32598	0.00062	0.00120	0.19	0.003
Monte-Carlo (experim.)	0.32602	0.00070	0.00134	0.22	0.017
	Parameter θ_2				
Least squares (param.)	−0.393647	±0.000090	±0.000181	0.023	–
Jackknife	−0.393647	0.000039	0.000078	0.010	0.0001
Bootstrap (data points, 37%)	−0.393647	0.000028	0.000057	0.007	0.0002
Bootstrap (data points, 100%)	−0.393638	0.000036	0.000072	0.009	0.0024
Bootstrap (residuals, 37%)	−0.393643	0.000079	0.000153	0.020	0.0010
Bootstrap (residuals, 100%)	−0.393647	0.000077	0.000148	0.020	0.0001
Monte-Carlo (Eq. (8))	−0.393645	0.000073	0.000145	0.018	0.0005
Monte-Carlo (experim.)	−0.393647	0.000077	0.000152	0.020	0.0001

[a] The relative standard deviations are given by RSD $= 100 \times \sigma_i/\theta_i$ and relative error (RE) is defined in terms of the corresponding parametric least squares values. Units as in Table 1. The resampling factor is indicated for the bootstrap methods. Monte-Carlo and bootstrap results correspond to 2000 simulations. The parameter values for the resampling and Monte-Carlo techniques are average values. Confidence intervals are based on Eq. (11) for the parametric estimation, Eq. (13) for the jackknife method and direct percentile determination for the bootstrap and Monte-Carlo approaches.

It is apparent from Table 2 that almost coinciding values[1] are obtained for each parameter using the contemplated methods (RE $\leq$ 0.078% for θ_1 and RE $\leq$

---

[1] The results presented for bootstrap and Monte-Carlo methods correspond to 2000 simulations. This number is probably more than enough to ensure almost complete convergence. See discussion later.

0.002% for θ_2). Some differences in the estimated confidence intervals can, however, be found amongst the different methods. First, the two Monte-Carlo approaches produce similar results and, in the case of θ_1, do not also essentially differ from the parametric least squares estimate. Jackknife and data points bootstrap seem to underestimate the confidence intervals for both parameters in comparison to all other approaches.

Residuals bootstrap presents, in relation to the Monte-Carlo, approaches, too small error intervals for θ_1. In what concerns θ_2, the error intervals are similar to those presented by Monte-Carlo.

In summary, for the system and experimental conditions that were considered here, we think that the parametric estimate of errors is essentially accurate, although somewhat excessive for θ_2. It is also computationally less expensive than the other approaches. The fact that we are dealing with a nonlinear model does not alter this conclusion. Jackknife and bootstrap over data points clearly point to underestimated error bars. Residual bootstrap[2] is almost insensitive to the level of resampling[3], 37 or 100%, and the confidence intervals are apparently underestimated for θ_1 and in accordance to Monte-Carlo findings for θ_2. Finally, and in what concerns the convergence properties of the bootstrap and Monte-Carlo methods we note that position values stabilize after ten simulations (RE $\leq 10\%$), irrespective of the method. However, for the dispersion estimates, convergence to $< 5\%$ is only attained for 100 simulations or more (see Table 3).

6. Conclusions

In this work we have applied several methods to obtain estimates for the parameters and associated errors describing a polagraphic curve. These parameters can be established within an uncertainty that is much
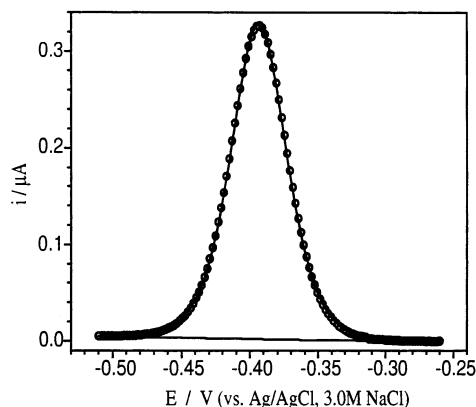

Fig. 3. Average polarographic data points ($\bigcirc$) and maximum likelihood least squares predicted values from Eq. (5) (solid line). The fit to the background current is also represented.

---

[2] Some improvement in confidence intervals can be obtained using skewness and kurtosis corrected percentiles (Efron, 1987; Efron and Tibshirani, 1993).

[3] See Dathe and Otto (1996) for an example of resampling affecting, for each bootstrap sample, only a definite part of the original data.

Table 3
Convergence of simulation results corresponding to the dispersion of θ_1 and θ_2 for bootstrap (residuals, resampling factor of 37%) and Monte-Carlo (using Eq. (8)) approaches[a]

#	RE/%		RE/%	
	θ_1		θ_2	
	Bootstrap	Monte-Carlo	Bootstrap	Monte-Carlo
10	−21.57	−25.12	−36.86	−43.9
20	−8.02	5.72	−24.45	−38.2
30	−22.94	7.21	−23.91	−18.61
50	−16.33	3.19	5.74	−11.67
100	−1.51	−3.97	1.66	−4.67
200	−3.47	−1.90	−3.94	−3.76
300	2.81	−1.30	−3.09	−2.26
500	−1.52	1.32	−0.51	−3.41
1000	3.69	−2.75	−1.44	−1.96

[a] The relative error (RE) is defined relatively to the corresponding 2000 simulations case. Units as in Table 1.

lower than that arising from direct use of the discrete data, with the advantage that modeling takes into account the background current behaviour. The use of DPP in complexometric studies thus allows a drastic improvement in precision and accuracy of determined constants.

We have also presented an approximate representation of the instrumental polarographic error that, apart from minimizing numerical difficulties can be used as a good estimate in similar experimental circumstances.

Although this type of modeling is intrinsically nonlinear, maximum likelihood parametric estimation of confidence intervals has shown to be almost as reliable as Monte-Carlo, simulation results, and the most suitable candidate for treating the systems here dealt with in a computationally inexpensive way. Jackknife and bootstrap of data points produce, comparatively, underestimated errors when applied to DPP models. Bootstrap of residuals, on the other hand, presents results that differ from those of Monte-Carlo approaches, especially for θ_1.

References

Anderegg, G., 1982. Pure Appl. Chem. 54, 2693.
Beck, J.V., Arnold, K.J., 1977. Parameter Estimation in Engineering and Science. Wiley, New York.

Bonate, P.L., 1993. Anal. Chem. 65, 1367.

Bond, A.M., 1971. Coord. Chem. Rev. 6, 377.

Bond, A.M., 1980. Modern Polarographic Methods in Electrochemistry. Dekker, New York.

Box, G.E.P., Muller, M.E., 1958. Ann. Math. Stat. 29, 610.

Broekaert, J.A.C., Güçer, Ş., Adam, F. (Eds.), 1989. Metal Speciation in the Environment, NATO-ASI Series. Springer-Verlag, Berlin.

Churáček, J., Cox, P.J., Masson, M.R., 1993. Advanced Instrumental Methods of Chemical Analysis. Ellis, New York.

Cukrowsky, L., 1996. Anal. Chim. Acta 336, 23.

Dathe, M., Otto, M., 1996. Fresenius J. Anal. Chem. 356, 17.

DiCiccio, T., Tibshirani, R., 1987. J. Am. Stat. Assoc. 82, 163.

Efron, B., 1987. J. Am. Stat. Assoc. 82, 171.

Efron, B., Gong, G., 1983. Ann. Math. Stat. 37, 36.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman, London.

Efron, B., Tibshirani, R., 1986. Stat. Sci. 1, 54.

Heath, G.A., Hefter, G., 1977. J. Electroanal. Chem. 84, 295.

Huang, W., Henderson, T.L.E., Bond, A.M., Oldham, K.B., 1995. Anal. Chim. Acta 304, 1.

Jones, G., Wortberg, M., Kreissig, S.B., Hammock, B.D., Rocke, D.M., 1996. Anal. Chem. 68, 763.

Kiss, T., Sóvágó, L., Gergely, A., 1991. Pure Appl. Chem. 63, 597.

Marquardt, D.W., 1963. J. Soc. Ind. Appl. Math. 11, 431.

Meloun, M., Militky, J., Forina, M., 1992. Chemometrics for Analytical Chemistry, vol. l. Ellis, New York.

Nancollas, G.H., Tomson, M.B., 1982. Pure Appl. Chem. 54, 2675.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in Fortran, 2nd ed. Cambridge University Press, Cambridge.

Reilly, P.M., Reilly, H.U., Keeler, S.E., 1993. Appl. Stat. 42, 693.

Rocke, D.M., Lorenzato, S., 1995. Technometrics 37, 176.