Why to Calculate, When to Use, and How to Understand Curvature Measures of Nonlinearity

Marian Karolczak, Ph.D.
Adam Mickiewicz University
Faculty of Chemistry
Grunwaldzka 6
60-780 Poznań
Poland

Quantitative measures of nonlinearity are presented. A concise discussion shows how to understand them and when to use them. Some advanced tools of nonlinear least square analysis not usually addressed by commercial computer software packages are described. Linearity versus nonlinearity in nonlinear models is described. A "current separation" has been provided, separating chronoamperometric data into its faradaic and charging components, which is examined in the context of the statistical behavior of nonlinear curve fitting parameter estimates. The methodology outlined in this article may help the user make a decision about the acceptance or rejection of a particular physical model. General considerations on the problem of choosing an equation in nonlinear least square fitting supplement the article.

Recently, many computer software developers have been providing users the comfort of not having to worry about the math behind what is going on. In this article, we discuss some problems not usually considered in commercial software, and attempt to show how some advanced features of nonlinear least square analysis can be useful. In the previous article of this series (1), we noted that it is not always possible, and may sometimes be unwise, to transform an equation into a linear form, and that even evidently linear equations occasionally introduce nonlinear convolutions. One reason for the use of least square techniques in the elaboration of the experimental data, was "... support for identification of the physical phenomenon that led to the theoretical model from which the fitted line was derived."

In this article, we would like to return to this point, but we will consider it in the context of Nonlinear Least SQuare (NLSQ) curve fitting. In particular, we will focus on the information that can be obtained from studying the *behavior of least square estimates* (2). We will summarize and compile the major achievements of Ratkowsky (3) and co-workers. The study of the behavior of least square estimates in nonlinear models is a relatively new field. We hope that the use of the nomenclature, ideas, and methodology presented here will help the user to make a generalization and assist in the decision about the efficiency of a particular model.

Since this article invokes issues not accounted for by commercial software, it raises questions in many instances that have not been considered at all. This article will:

- Provide guidance and clarity to nomenclature used in linear and nonlinear problems.
- Introduce curvature measures of nonlinearity, discuss when to use them, and indicate their significance in the case of models that cannot be linearized.
- Discuss how to choose between different models and how well a model fits experimental points.

We assume that the reader is acquainted with the basics of curve fitting procedures (LLSQ and NLSQ methods) and has some fundamental knowledge of matrix algebra. For this reason we will not discuss any specific algorithms or techniques; interested readers may consult references 4 and 5 or the original literature on the subject (6,7).

Linearity vs. nonlinearity in curve fitting

The term "linear" in linear least square (LLSQ) fitting is often used to express two things: the first, and more common, refers to a straight-line relationship between variables (i.e., $y = A + BX$); the second refers directly to a physical model in which the quantities to be estimated (the parameters) appear linearly. According to the last distinction, models such as $y = A + BX + CX^2 + ...$ and $y = A + BX_1 + CX_2$ fall into a linear category, although in fact the first is a parabola (or higher polynomial) and the second represents a surface in a three-dimensional space.

It is obvious that regardless of the above meaning, in an intrinsic linear model all parameters enter the model (fitted) equation in a *linear way*. On the other hand, if any of the parameters were to enter the model equation in a *nonlinear way*, the model would be considered "nonlinear." The terms "linear way" and "nonlinear way" remain obscure, however, and the distinction between parameters that appear in a "linear way" requires further clarification. Two propositions have been advanced to address this (8). They are based on the differences in the behavior of the first derivative of the fitted function. Namely:

a)the first derivative of the fitted function with respect to the parameter under consideration is independent of *that* parameter,

b)that derivative is independent of *any other* parameter in the model.

According to this distinction, the parameters A and B in the apparently straight-line equation $y = A + BAX$ could be considered "linear" only by the first definition. Thus, the whole model falls into a nonlinear category. Another good example would be the function $y = A + BC^X$.

The grounds for the assessment of the unknown parameters in either a linear or a nonlinear model establishes the criterion of least squares (4). Provided that the errors of the variables under consideration are random (i.e., they are independent and normally distributed with mean equal zero and a finite variance), the significance of the LSQ criterion is due to the fact that the least square estimates of the parameters have some optimum statistical properties. *For the exclusive case of linear models*, this means that the estimates are:

(a)unbiased,

(b)the maximum likelihood estimates,

(c)the estimates having a minimum variance.

In this sense, the method of least squares provides "the best available parameters."

It should be realized that only estimates with the above properties allow one to calculate a reliable range for their variance, demonstrate the true correlation among them, provide predicted values of the dependent variable with predetermined probability, and assure statistically meaningful tests.

Nonlinear least square estimates, on the other hand, do not embody the above optimal properties. They are usually biased and nonrandomly distributed, and their variance most likely exceeds the minimum possible variance. The extent of the bias, the non-normality, and the excess of variance differ significantly from model to model, even within a model. It depends on a number of factors, including the amount of data and their distribution in the experimental space. It is actually more appropriate to speak about model/data combination when discussing nonlinearity and estimates for nonlinear models.

It is fortunate, however, that the least square estimates of the parameters in nonlinear models gain all the optimal statistical properties of true linear models asymptotically. Thus (provided that the depending variables (fitted values) have errors that are independent and normally distributed) the bigger the sample is, the closer all the desired properties of linear model estimates (and "linearity" in this sense) will be. Unfortunately, it is not possible to tell a priori how large a sample must be to approach these asymptotic properties sufficiently. One finds some nonlinear models for which the asymptotic properties are well approximated for small sample sizes (6–9), whereas for other models, sample size must be very large (50-100) to approximate such properties satisfactorily for statistical measures. So far, no rule exists — any improvement can thus be made a posteriori, but to do this, one needs some quantitative measures of the "deviation from linearity" (i.e., nonlinearity). The term "close to linear" (7) has been introduced and used for this purpose.

Several attempts have been made to measure the extent to which nonlinear models differ from linear models, and to provide a way to evaluate "close to linear" behavior. The first quantitative measures of nonlinear behavior were proposed in 1960 by Beale (9). Guttmann and Meeter (10) later discussed some limitations of Beale's method. In 1971, Box (11) derived formulas for estimating bias in least square estimates. His formulas were subsequently examined by Gillis and Ratkowski (12) in extensive simulation studies. Contemporary measures of nonlinearity were introduced in 1980 by Bates and Watts in their fundamental paper (13). They developed new measures of nonlinearity based on the intuitive concept of geometric curvature, which they applied for a multidimensional space. They also provided relationships between their measures and those previously proposed by Beale. Additionally, they showed how Box bias measures of nonlinearity were related to their proposition.

Bates and Watts established that the nonlinearity of a model can conveniently be separated into two components:

- "intrinsic" nonlinearity, and
- "parameter-effect" nonlinearity.

Intrinsic nonlinearity (IN) represents curvature of the multidimensional surface called the solution locus of the fitted model. IN is inherent to the particular model and data set, the number of data points, and their space distribution. It cannot be changed by mathematical transformation of parameters (reparametrization). It usually decreases as the amount of data increases for a particular model, although there is no rule on the relationship between the sample size and the magnitude of IN. It is equal to zero for a linear model. It is zero for nonlinear models, in the limit when the number of points goes to infinity. It should be emphasized in this context that a negligible value of intrinsic nonlinearity implies a negligible bias in predicted values of the dependent variable and makes the determined confidence limits for the predicted values meaningful.

Parameter-Effect Nonlinearity (PE) is a collective measure of nonlinear behavior associated with the mathematical form and the particular choice of the parameters. It can be changed by reparametrization. By suitable reparametrization, it can be continuously minimized to any desired value, ideally to zero. Geometrically, PE is related to the projection of the parameters' lines on a plane tangent to the solution locus. Provided the intrinsic nonlinearity is also negligible, the smaller the parameter-effect nonlinearity, the closer a given nonlinear model approaches a linear model in behavior. Consequently, fewer iterations are necessary to achieve convergence in parameter estimation (there is no iteration for a true linear model), and statistical formulas and other tests ordinarily used for linear models (e.g., t-test, parameter correlation) will be more valid. In short, the estimated parameters will be closer to their minimum variance parameter estimators.

Box bias curvature measure and the associated *percentage bias* were proposed by Box in 1971. They are measures of nonlinearity indirectly related to parameter-effect nonlinearity. Both represent individual nonlinearity measures. They are joined to the individual parameter and the way it enters the model. They reveal the parameters that behave most "nonlinearly" and suggest possible reparametrization. They change after reparametrization. They can be used for prognosing a reparametrization and judging its effect a priori. The *percentage bias* expresses the parameter bias as part of its least square estimate. It is a useful quantity which identifies the extent of a parameter's nonlinear behavior. Excluding the case of the parameter representing a constant term in model functions, the closer a parameter's percentage bias is to zero, the closer this parameter appears to linear behavior. An arbitrary rule has been established which says that absolute value of a percentage bias exceeding 1% is a good indication of nonlinear behavior. It is important to realize that a "linear-appearing" parameter in a nonlinear model does not necessarily imply linear behavior in estimation, zero Box bias, or other statistical properties expected for linear models.

A term closely related to curvature measures of nonlinearity and used in their calculation is Standard Radius (SR). It is the square root of the sum of squared residuals, divided by the degrees of freedom and multiplied by the number of the parameters. It scales real "curvature" of the solution locus. It is used during the calculation supplying the curvature measures of nonlinearity and parameter estimates. It also stands as the reference factor for the assessment of conservative confidence regions. This region is constructed so that the expected probability of the true parameter lying inside the region is at least $1-\alpha$.

In general, the characteristics discussed above provide information about how close the investigated model is approximated by a linear model. If it is close enough, one can use standard statistical tests, formulas, and methods. A "close enough" can be judged on statistical grounds by comparing IN and PE with $1/(2\sqrt{F})$, where $F = F(p,n-p;\alpha)$ is the F-distribution table value. The value $1/(2\sqrt{F})$ may be regarded as the radius of curvature of the $100(1-\alpha)$ percentage confidence region (9). The new measures of nonlinearity help us to estimate the efficacy of particular models. They also facilitate making a choice between competing models.

How curvature measures of nonlinearity are calculated

We have come to the point that requires a more quantitative description of these curvature measures of nonlinearity.

First, however, we must make some introductory remarks on the methods that contemporary NLSQ techniques use to estimate the unknown parameters in a model. There are a number of algorithms and hints offered in the literature on this subject (4,5). The most popular seems to be the method known as "Marquadt-Levenberg strategy" (5,6) (popularized by the majority of commercial curve fitting software), although its cousin, the "rank-two DFP" algorithm (5,7) has far superior convergence speed and numerical properties. Both methods have their roots in the "Gauss-Jordan method," which is the third most popular. Having the same basis (i.e., an approximation of a nonlinear function by the first-order expansion of its Taylor series), they differ in the algorithms in which the corrections to the estimates are introduced during subsequent iterations. The result is dif-

ference in speed of convergence, reliability, and sometimes convenience.

If one has a set of n observations (x_1, x_2, ..., x_n) and decides to fit a model represented by a nonlinear function $Y = f(\underline{X},\underline{\theta}]$, where $\underline{X} = (x_1, x_2, ..., x_n)^T$ and $\underline{\theta}$ represents (p×1) vector $\underline{\theta} = (\theta_1, \theta_2, ..., \theta_p\}^T$ of p parameters to be estimated, any NLSQ procedure requires calculation (either analytically or numerically) of the partial first derivatives of Y with respect to each parameter

$$a_{ij} = \left(\frac{\partial Y}{\partial \theta_i} \right) \qquad (1)$$

at some $\underline{X}_0$ and $\underline{\theta}_0$, and formation of an n×p matrix, say $A^{\bullet}$, which contains a_{ij} as elements. The index i goes from 1 to n and denotes the position of the data pair in the collection of experimental data. The index j goes from 1 to p and denotes the parameter under consideration. Evaluation of the above curvature measures of nonlinearity also requires the calculation of the second partial derivatives

$$a_{ijk} = \left(\frac{\partial^2 Y}{\partial \theta_i \partial \theta_k} \right) \qquad (2)$$

and the construction of an n×p×p matrix, say $A^{\bullet\bullet}$, from these derivatives.

Numerical practice has shown that it is advantageous in such calculations to deal with response-invariant standardized matrices. Thus, at the initial stage of matrix formation, all derivatives are divided by a scaling factor. In this calculation, the most convenient appears to be the standard radius, $\rho = \sigma\sqrt{p}$, where σ is the square root of the residual sum of squares divided by (n-p) (i.e., RSS/(n-p)).

Calculations of the curvature measures of nonlinearity can be greatly facilitated by first performing suitable transformations of the initial matrices. Such transformations are easy and quick on contemporary computers and consist of coordinate transformation that rotates the sample space so that the first coordinate vectors are parallel to the tangent plane, and the last n-p are orthogonal to it. In matrix language, this operation is equivalent to premultiplication of all derivative vectors in the sample space by an orthogonal Q matrix that is a part of QR decomposition (13) of the initial matrix $A^{\bullet}$, such that:

$$QA^{\bullet} = \left(\frac{\tilde{R}}{0} \right) \qquad (3)$$

If one defines (for convenience)

$$L = \tilde{R}^{-1} \qquad (4)$$

then the matrix

$$C = Q(L^T A^{\bullet\bullet} L) \qquad (5)$$

consists in the first p faces (the parameter-effect acceleration array) and the remaining n-p faces (the intrinsic acceleration array). In their original paper (11), Bates and Watts showed how to reduce each such array to a single measure and described an iterative method for finding the *maximum* intrinsic curvature measure, IN, and the *maximum* parameter-effect curvature measure, PE, from these arrays. This method can be incorporated into a computer program. This has been done in BASIC (15) and other programming languages.

Having constructed the matrices of first and second derivatives as described above, very little additional programming is necessary to provide simultaneously the bias in the maximum likelihood estimates of the parameters as defined by Box (9), since

$$\text{Bias}(\underline{\hat{\theta}}) = \frac{-\hat{\sigma}^2}{2} \left[\sum_{i=1}^{n} A_i^{\bullet} (A_i^{\bullet})^T \right]^{-1} \times \sum_{j=1}^{n} A_j^{\bullet} \, \text{tr} \left[\left(\sum_{i=1}^{n} A_i^{\bullet} (A_i^{\bullet})^T \right)^{-1} A_j^{\bullet\bullet} \right] \qquad (6)$$

where $A_i^{\bullet}, A_j^{\bullet}$ represents the p×1 vector of the first derivatives of $\hat{Y}$, and $A_j^{\bullet\bullet}$ is the p×p matrix of second derivatives concerning each element of the vector $\underline{\theta}$, evaluated at x_i, x_j, respectively. The bias given by the left hand side of this equation is the p×1 vector, representing the discrepancy between parameter estimates and true values.

Similarly, bias in $\hat{Y}$ can be calculated from the relationship

$$\text{Bias}(\hat{Y}) = (A^{\bullet})^T \text{Bias}(\underline{\hat{\theta}}) + \frac{1}{2} \text{tr} \left[A^{\bullet\bullet} \text{Cov}(\underline{\hat{\theta}}) \right] \qquad (7)$$

Box (9) also provides formulas that approximate (albeit less accurately) the variances of the predicted value of $\hat{Y}$ as well as the variance of the predicted value of the parameter $\hat{\theta}_j$.

Relationship of Student's t-value to the curvature measures of nonlinearity

The characteristic that expresses the ratio of a parameter's least square estimate to its standard error is known as Student's t-value (16). t-values are often used for examining the behavior of a model. Computer programs routinely deliver these values. t-values are ordinarily tested with reference to the Student's t-distribution having n-p degrees of freedom and 100(1-α) percent confidence region. It is a common belief that a high t-value tends to indicate that the parameter is well defined. It is less recognized however, that a high t-value does not necessarily mean that the parameter estimate will have other desirable statistical properties, such as independent and identical distribution of errors, "close to linear" behavior, etc. Also, especially in the case of multiparameter models, the t-value may be low because of high correlation between parameters within

the model. Although in many instances the magnitude of the t-value can be useful as an indicator of nonlinear behavior, its validity depends on several of the model's statistical properties, including satisfaction of the "close to linear" condition. "Close to linear" behavior is estimated, however, through IN and PE calculations. It can be especially dangerous to rely only on t-values, knowing nothing about PE and IN.

Correlation among least square estimates in nonlinear models

Some modern software packages provide correlation between the estimated parameters within a model, usually represented by the parameters' correlation coefficients. They should be either 1 (for the same parameter) or as close to zero as possible (for different parameters). Again, there is no reason to believe that a high correlation between parameters announces nonlinear behavior (bias, variance excess, nonnormal distribution) of the estimates in a nonlinear model. It can be demonstrated (7) that even for a strictly linear model, the correlation coefficients between the parameters can be made as large as desired simply by changing the independent variable's location (i.e., a constant shift in all x-values). Thus, for nonlinear models, high correlation between the least square estimates cannot indicate serious nonlinear behavior of these estimates. The explanation of a common belief contrary to this statement lies perhaps in the credit given to the fact that in badly behaved nonlinear models, the correlation among the parameters is often high. On the other hand, many examples can be found which prove the inaccuracy of such a belief. The conclusion that follows is that the magnitude of correlation coefficients of parameter estimates cannot serve as a diagnostic tool in the estimation of nonlinear behavior of least square estimates.

The choice of a model

One of the main purposes of fitting an equation to experimental data is to resume the multiplicity of data, in order to obtain either interpolation formulas or calibration curves which will be helpful in later calculations, control, or graphical representation. Other (often coinciding) purposes are to confirm (or refute) a theoretical relationship, to compare several sets of data in terms of constants in the representing equations, and to aid in the choice of a theoretical model. In this last context, the sufficient argument in favor of a choice is some evidence or physical phenomena that indicates that the model is appropriate. However, one often meets a situation in which several competing models appear to fit the experimental points equally well; here the use of a model is not particularly well resolved on physical grounds (although attractive or justified otherwise). The choice between such models must then consider other factors. We believe that IN and PE curvature measures may be of great value in this task. It is intuitively clear that having other things equal, or comparable, the nonlinear models that closest approach linear models should be given priority. Moreover, this intuitive affection has rigorous foundation on statistical grounds as indicated above.

There are other procedures and criteria that might aid in choosing the right model equations for the data set. They have been extensively offered in the literature; we shall critically review some of them. First, they consist of examination and testing of the so-called correlation coefficient (actually the square of it), R^2. This coefficient represents the portion of variation in the set Y_i explained by model function $\hat{Y}$. A justification for this approach relies on the inherent relationship between R^2 and Fisher-Snedecor, an F-statistic that in turn is related to a χ^2-statistic, both used in formal early tests of the goodness of fit. The examination of the correlation coefficient is a simple and attractive procedure and no one should complain about obtaining a high R^2. However, it is not commonly recognized that serious objections can be raised to relying solely, or very heavily, on this criterion. Some comments seem imperative:

1. It is always possible to increase R^2 by adding parameters. Thus one should be cautious of functions that must use many parameters to attain a high R^2. Adjustments to the correlation coefficient have been proposed to correct this defect. In particular, one conventional proposition is to use the sample statistic

$$D^2 = R^2 - \frac{p}{n-p-1}\left(1-R^2\right) \quad (8)$$

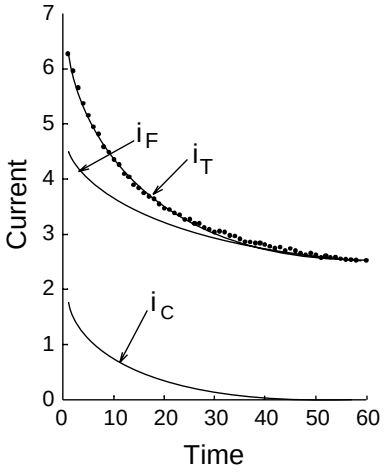
   instead of R^2. Equation 8 penalizes functions with many parameters (especially when n is small).
2. When one is bound to use a transformation of Y (e.g., log Y, etc.) (1), R^2 measures and explains the variation of the transformed function instead of the original one. It is hard to compare such different objects straightforwardly. In this way R^2 loses its potential value.
3. If the main intention is not merely a graphical representation, calibration curve, or interpolation formula, one should be wary of choosing a function with a high R^2 but many coefficients that either have no physical interpretation or have magnitudes or signs in conflict with a theory.

Residual patterns

A study of the residual patterns (graphical representation of the differences between observed and fitted data) sometimes assists in the choice of a model equation. Residual patterns can be more useful than R^2 in suggesting improvements of

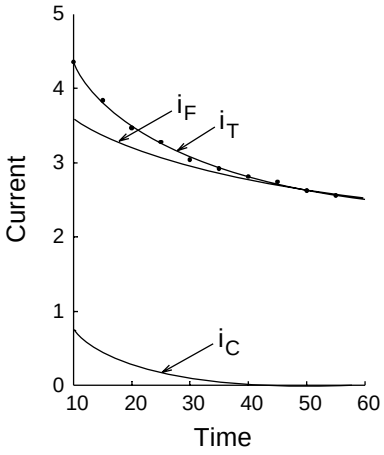
F1

Current separation. The total transient current (dots) together with its faradaic and charging contributions separated for the set of 61 data points. The least square characteristics are in T1.



F2

Current separation. The total transient current (dots) together with its faradaic and charging contributions separated for the set of 11 data points. The least square characteristics are in T2.



the functional form. If they show a determined pattern (e.g., linear, quadratic, etc.), then the addition of the corresponding function can be considered. Residual patterns sometimes reveal a nonhomogeneity in the sample that is not recognized otherwise. Finally, a supplementary study of autocorrelation in residual patterns answers the fundamental questions of LLSQ and NLSQ analysis, mentioned at the beginning of this article, and related to the form of distribution of errors, their randomness, etc. We note that the residual patterns are of little value and can be misleading if the sample size is small.

The current separation example

As an application of IN and PE curvature measures in electrochemistry, let us consider the case of "current separation" in a typical chronoamperometric experiment. Consider the case of an irreversible electrode process:

$$Ox + ne \underset{k_{bh}}{\overset{k_{fh}}{\rightleftharpoons}} Red$$

where k_{fh} and k_{bh} are heterogeneous rate constants of the electrode process: k_{fh} characterizes the reduction process; k_{bh} characterizes the oxidation process.

In the response to a potential step, the total current, i_T, comprises at least three components (17): the faradaic current i_F and the charging current i_C, both decaying with time t, and noise; that is,

$$i_T = i_F + i_C + \text{noise} \qquad (9)$$

where:

$$i_F = i_0 \exp\left(\lambda^2 t\right) \operatorname{erfc}\left(\lambda\sqrt{t}\right) \qquad (10)$$

$$i_C = i_1 \exp\left(-t/T\right) \qquad (11)$$

and $i_0 = nFAk_{fh}c_{ox}^0$, $\lambda = k_{fh}/D_{ox}$, and T = RC.

The other symbols in these equations have the following physical meanings:

A electrode area

c_{ox}^0 initial concentration of the oxidized form

C capacitance of the electrolytic cell, mainly the capacitance of the electrode

D_{ox} diffusion coefficient of the oxidized form

R resistance of electrolytic cell, mainly the uncompensated resistance of the electrolyte

i_1 initial charging current

$$\operatorname{erfc}(x) = \left(2/\sqrt{\pi}\right)\int_x \exp(-u^2)du$$

Equation 10 was derived by Delahay and Strassner (18) and independently by Evans and Hush (19). Total current given by equation 9 is represented by the data points in ***F1*** and ***F2***. These data were used in the NLSQ procedure which provided the least square estimates of i_0, λ, i_1, and T. These estimates allowed us to make a "current separation" (i.e., to calculate the theoretical, noise-free value of i_T, and its faradaic and charging components) via equations 10 and 11. We used two sets of data here: the full set of data pairs that contained 61 experimental points, and a subset of it (we removed the first ten points (to t=10), and we considered every tenth point after that). Thus, the subset consisted of 11 points.

Our program also provides standard deviations for the estimated parameter (SD), Student's t-values, Box's percentage bias, correlation coefficients, and of course, the maximum intrinsic curvature measure, IN, along with the maximum parameter-effect curvature measure, PE. All these values are presented in ***T1*** and ***T2***, for the original data set (61 points) and its subset (11 points), respectively. For comparison, the values of the curvature radius, $1/(2\sqrt{F})$ for F(4,7; 0.05), F(4,7;0.01) in the case of reduced data, and F(4,57;0.05), F(4,57;0.01) for the full data set are provided below each table.

The data in ***T1*** and ***T2*** show that IN is acceptable but PE is not, at both chosen probability levels, for the 11-point set. For the 61-point data set, however, both curvature measures of nonlinearity are acceptable; that is, they are less than $1/(2\sqrt{F})$, at the same probability levels. Thus, we may conclude that for this set of data, the investigated model is "close to linear" in statistical behavior. Besides, this result clearly tells us that although the estimated values of all four parameters for both data sets do not differ significantly, the SD values, t-values, and parameter cor-

T1

The least square estimates of the parameters in equations 9–11 for the full 61-point data set.

Param.	LSQ Estimate	SD	t	% bias	Correlation coefficients			
i_0	5.0403	0.0684	73.7	-0.009	1			
λ	0.1014	0.0027	38.2	-0.018	0.9925	1		
i_1	1.9655	0.0555	35.4	0.023	-0.9760	-0.9740	1	
T	9.9813	0.1290	77.3	0.031	-0.2778	-0.1944	0.1436	1

IN = 0.0095
PE = 0.2074
SR = 0.0412
$1/(2\sqrt{F}) = 0.2609$ (for $\alpha = 0.01$)
$1/(2\sqrt{F}) = 0.3147$ (for $\alpha = 0.05$)

T2

The least square estimates of the parameters in equations 9–11 for the reduced 11-point data set.

Param.	LSQ Estimate	SD	t	% bias	Correlation coefficients			
i_0	4.9106	0.6143	7.99	-1.06	1			
λ	0.0957	0.0226	4.24	-2.13	0.9994	1		
i_1	2.0951	0.2365	8.88	4.34	-0.7330	-0.7456	1	
T	9.8131	1.9401	5.03	1.97	-0.9424	-0.9333	0.4771	1

IN = 0.2253
PE = 8.0832
SR = 0.0420
$1/(2\sqrt{F}) = 0.1784$ (for $\alpha = 0.01$)
$1/(2\sqrt{F}) = 0.2463$ (for $\alpha = 0.05$)

relation coefficients of the 11-point set are meaningless from the statistical point of view (20). Consequently, all predictions and tests grounded on these estimates are statistically meaningless as well.

The results presented in ***T2*** also inform us of the danger of relying exclusively on calculations of t-values. An apparently high, and by some standards acceptable, t-value tells us nothing about the real behavior of the nonlinear estimate and the true value of SD. As this case shows, one can be easily misled by considering the t-values only.

For completeness, we note that we have also experimented (results not shown here) with a 51-point data set (i.e., we eliminated just the first 10 points from the original 61-point data set). We obtained IN = 0.1257, acceptable at both $\alpha = 0.05$ and $\alpha = 0.01$ probability levels, and PE = 4.8819, unacceptable at both probability levels. Since one can reduce PE to any desired value by suitable reparametrization, this finding could stimulate one to seek a reparametrization. We did not, however, exercise this point further. Rather, we have learned from this example that sometimes even 51 data points may not be sufficient to assure desired statistical properties for parameter estimates, and that the region containing the first few points is crucial for the type of problem we consider in this paper. Whether or not that region is accessible in practice and to what extent it is reliable is another important and interesting question, but it is out of the scope of this article.

A clear bonus of application of the NLSQ method and IN as well as PE calculations, for the type of problem we considered in this paper, is a convenient, statistically meaningful current separation procedure, as shown in ***F1*** and ***F2***. This procedure can be generalized and applied to similar problems. The physicochemical parameters i_0, λ, k_{fh}, D_{ox}, C, and R are amenable to further calculations, classifications, compilations, etc.

References

1. M. Karolczak, Current Separations 13(4):98–104.
2. The study of the behavior of least square estimates in nonlinear models is a relatively new field. We hope that the use of nomenclature, ideas, and methodology presented will assist the user in deciding about the efficiency of a particular model.
3. D.A. Ratkowsky, "Nonlinear Regression Modeling," Marcel Dekker, Inc., New York, 1983.
4. N.R. Draper and H. Smith, "Applied Regression Analysis," 2nd ed., Wiley, New York, 1981.
5. D.R. Sadler, "Numerical Methods for Nonlinear Regression," University of Queensland Press, St. Lucia, Quebec, 1975.
6. D.W. Marquardt, J. Soc. Ind. Appl. Math. 2:431 (1963).
7. R. Fletcher and M.D.J. Powell, Comput. J. 6:163 (1963).
8. W.H. Lawton and E.E. Sylvestre, Technometrics 13:461 (1971).
9. E.M.L. Beale, J. R. Statis. Soc., Ser. B. 22:41 (1960).
10. I. Guttman and D.A. Meeter, Technometrics 7:623 (1960).
11. M.J. Box, J. R. Statist. Soc., Ser. B. 33:171 (1971).
12. P.R. Gillis and D.A. Ratkowski, Biometrix 34:191 (1978).
13. D.M. Bates and D.G. Watts, J. R. Statist. Soc., Ser. B. 42:1 (1980).
14. P. Buisinger and G.H. Golub, Numerische Math. 7:269 (1965).
15. M. Karolczak, unpublished results.
16. Student was the pseudonym of W. Gosset, an English statistician of the 19th century.
17. We limit ourselves here to the consideration of semi-infinite and planar diffusion. We also neglect other plausible complications, such as transient currents during the first ~10^{-9}–10^{-5} s resulting from formation of the double layer, and all so-called coupling effects of faradaic and charging currents.
18. P. Delahay and J.E. Strassner, J. Am. Chem. Soc. 73:5218 (1951).
19. M.G. Evans and N.S. Hush, J. Chim. Phys. 49:159 (1952).
20. This is so for the kind of problem we deal with in this article. For other models, the 11-point data set/model combination may be quite satisfactory regarding IN and PE values, and thus linearity.