

---

---

## CHAPTER 24

# Nonlinear Least-Squares Fitting Methods

**Michael L. Johnson**

Department of Pharmacology and Internal Medicine  
University of Virginia Health System  
Charlottesville, Virginia 22908

---

Abstract

- I. Introduction
  - II. Formulate a Hypothesis-Based Mathematical Model
  - III. Determining the Optimal Parameters of the Model
  - IV. Distinguishing Between Multiple Mathematical Models
  - V. Estimate the Precision of the Model Parameters
  - VI. Cross-Correlation of the Estimated Parameters
  - VII. Uniqueness of the Parameters
  - VIII. Conclusions
- References

---

---

### Abstract

This chapter provides an overview of the techniques involved in “fitting equations to experimental data” with a particular emphasis on the what can be learned with these techniques, what are the requirements of the experimental data for these techniques, and what are the underlying assumptions of these techniques. The layout of this chapter is to start with a set of experimental data, and then walk the reader through the analysis of this set of data. The rigorous mathematical methods are referenced but not presented in detail.

---

---

### I. Introduction

This chapter addresses the often asked question, “*What is the best equation that describes this data?*” As written, this question is actually ill-posed because the meaning of “the best equation” is not clear. As will be seen below, it implies the best function form (i.e., the type of mathematical equation) and it also

implies the best parameters (i.e., coefficients) of that equation. While these are interrelated, they have very distinct meanings.

As an example to help clarify this point, consider a simulated ligand-binding experiment where the amount bound is a function of the ligand concentration, as shown in Table I. This could represent the binding of a hormone to a receptor, or the binding of oxygen to hemoglobin, or any other binding experiment of interest. The mathematical and statistical concepts presented in this chapter do not depend on the actual biochemical system being studied or even that a ligand binding system is being studied, with the exception of the specific binding equations. Furthermore, the methods are independent of the concentration units and, as a consequence, the units of the test data set are arbitrary.

The data shown in Table I is plotted as the open squares in Fig. 1 along with calculated curves based on two possible “best equations” that might describe this data. The dashed line appears to describe the data and is the fifth degree polynomial:

$$Y(X) = -0.002633 + 0.199246X - 0.097234X^2 - 0.051963X^3 + 0.056020X^4 - 0.012065X^5 \quad (1)$$

where  $Y(X)$  corresponds to the predicted amount bound at a concentration of  $X$ . This appears, at least visually, to provide a reasonable description of this data within the range of the data. Note that the coefficient of the fifth order term is

**Table I**  
**Simulated Data Used for the Examples**

Bound, $Y_i \pm \sigma_i$	Free ligand concentration, $X_i$
0.0008 ± 0.0020	0.0
0.0145 ± 0.0020	0.1
0.0269 ± 0.0020	0.2
0.0476 ± 0.0020	0.3
0.0612 ± 0.0020	0.4
0.0735 ± 0.0020	0.5
0.0793 ± 0.0025	0.6
0.0834 ± 0.0025	0.7
0.0839 ± 0.0025	0.8
0.0877 ± 0.0025	0.9
0.0901 ± 0.0025	1.0
0.0886 ± 0.0025	1.1
0.0931 ± 0.0030	1.2
0.1006 ± 0.0030	1.3
0.0942 ± 0.0030	1.4
0.0925 ± 0.0030	1.5
0.0977 ± 0.0030	1.6
0.0926 ± 0.0030	1.7
0.0977 ± 0.0035	1.8

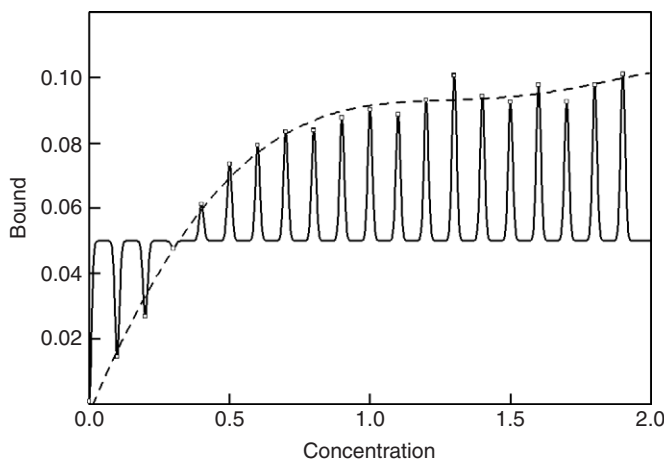
negative, and as a consequence this fifth degree polynomial will be negative at much higher ligand concentrations. This is clearly a physically impossible result!

But, is it the “best equation that describes this data”? To answer this we need a measure to compare different fits of the same data to determine which provides the best description of the data. The most commonly used, and abused, measure of how well an equation describes a set of data is the weighted sum-of-squared-residuals (WSSR) (which is proportional to the sample variance-of-fit,  $s^2$ ):

$$\text{WSSR} = \sum_i \left( \frac{Y(X_i) - Y_i}{\sigma_i} \right)^2 = \sum_i R_i^2 \quad (2)$$

$$s^2 = \frac{\text{WSSR}}{\text{NDF}} = \frac{\text{WSSR}}{n - n_f} \quad (3)$$

where  $R_i$  is the residual for the  $i$ th data point;  $Y(X_i)$  corresponds to the calculated value of the amount bound [as in Eq. (1)];  $X_i$ ,  $Y_i$ , and  $\sigma_i$  are the  $i$ th observed data value (as is given in Table I); and NDF is the number of degrees of freedom which is normally evaluated as the number of data points,  $n$ , minus the number of parameters being estimated,  $n_f$ . The consequences of a large and/or variable experimental measurement errors are corrected for by including the estimated uncertainties,  $\sigma_i$ , in Eq. (2). Equation (3) does not have  $a - 1$  in the denominator because the mean of the residuals has not been calculated; it is assumed to be zero. In mathematical jargon Eq. (2) is the L2 norm. These values are always positive and a lower value generally corresponds to a better description of the data. The WSSR = 32.68084 for the dashed line in Fig. 1.



**Fig. 1** The simulated data example found in Table I and two possible “best fit” equations to describe the data points. The dashed line is given by Eq. (1) and the solid line is given by Eq. (4).

The mathematical form for the alternative description of the data shown as the solid line in Fig. 1 is:

$$Y(X) = 0.05 + \sum_i (Y_i - 0.05) e^{-1/2((X-X_i)/0.008)^2} \quad (4)$$

Visually this contrived (i.e., arbitrary) functional form appears to be a terrible description of this data, but the corresponding WSSR is exactly zero. It passes precisely through every data points! Thus, if “best equation” is defined as the lowest WSSR, then the solid line in Fig. 1 [i.e., Eq. (4)] is the best description of the data that can be obtained. In addition, if a 19th degree polynomial ( $n - 1$  degree where  $n$  is the number of data points) is used, then the corresponding WSSR is also exactly zero. It also will pass precisely through every data point. There are actually an infinite number of different mathematical forms that yield a WSSR of zero for this, or any other, set of data! These forms cannot be the best description of this binding data because we have some knowledge of the mathematical form of the binding equations and neither of these are consistent with those forms [e.g., Eq. (5) for a single binding site and Eq. (6) for two binding sites per receptor] (Johnson and Straume, 2000):

$$\bar{Y}(X) = \frac{K_a X}{1 + K_a X} \quad (5)$$

$$\bar{Y}(X) = \frac{1}{2} \frac{K_1 X + 2K_2 X^2}{1 + K_1 X + K_2 X^2} = \frac{1}{2} \frac{10^{\log K_1} X + 2 \times 10^{\log K_2} X^2}{1 + 10^{\log K_1} X + 10^{\log K_2} X^2} \quad (6)$$

where  $\bar{Y}(X)$  is the fraction of binding sites occupied;  $K_a$  is the single site model association (i.e., binding) constant;  $K_1$  and  $K_2$  are the binding constants to the sites of the two-site model; and  $X$  is the free, or unbound, concentration of the ligand. Equations (5) and (6) are written in terms of the stoichiometric, or macroscopic, association constants and not the site, or microscopic, constants.

Also, the exact choice of model parameters is of great importance. For example, Eq. (6) is written in two algebraically equivalent forms, one in terms of the binding constants and the other in terms of the base 10 logarithms of the binding constants. The preferable form of Eq. (6) is the second form (on the right) logarithmic terms because it constrains the equilibrium constants to have only positive values. Ten raised to the power of the logarithm of the equilibrium constant,  $\log K$ , will always be positive even though  $\log K$  can have any real value positive or negative. Consequently, the left form of Eq. (6) allows physically meaningless negative values of the equilibrium constants which the right form (involving the logarithms) does not allow.

Neither Eq. (1) nor Eq. (4) provides any information about the nature of the ligand binding process. The problem with these equations is that they are not based

on a hypothesis about the underlying biochemical processes. *A general conclusion is that arbitrary models (i.e., models that are not hypothesis driven) will not provide the desired information.*

The answer to our initial question, “What is the best equation that describes this data?” is actually answered by two other questions:

*What do you want to learn from the data?*

*What are the properties of the data?*

The answers to these two additional questions will uniquely determine the mathematical form of the equations required to fit the data [e.g., Eqs. (1), (4)–(6)]. If we have a well-defined hypothesis about the biochemical mechanism being described by the data then the hypothesis can be transformed into the desired mathematical form (i.e., the appropriate fitting equation). The hypothesis can then be tested by assessing how well the equation can describe the experimental data. The forms of the equations that were used in the above examples were arbitrarily chosen to describe the data instead of to describe the hypotheses about the data. It is the answer to the second question that determines how to evaluate the best model parameters for a specific model and set of data.

---

---

---

## II. Formulate a Hypothesis-Based Mathematical Model

The first step is to derive a mathematical model which is based on the hypothesis about the nature of the biochemical processes that are being measured by the experimental observations. For the present ligand binding example, we assume that the data is represented by either a single binding site model, as in Eq. (5), or a two binding site model, as in Eq. (6). The nature of the data in Table I, however, is not directly compatible with Eqs. (5) and (6). The amount bound (i.e., the dependent variable) is in terms of fractional saturation in the equations and is in concentration units in the data. Consequently, the model equation must be transformed to match the data, as in the following equation:

$$Y(X) = A\bar{Y}(X) \quad (7)$$

where the fitting function is  $Y(X)$  is related to the fractional saturation by a multiplicative constant,  $A$ . This multiplicative constant is related to the concentration of the binding protein (e.g., receptor or hemoglobin), and possibly the design of the instrumentation that is utilized for the measurements. For the present example,  $A$  will be estimated in the fit because it is an unknown constant, or at least it is a constant that is not known to an infinite precision.

There is always the alternative of scaling the data to match the form of the equation but this is usually not a good idea. For this present example, this scaling would involve dividing each of the bound concentrations by the saturating amount

bound at an infinite ligand concentration (i.e., the independent variable). This limiting value is not known *a priori*, however. It must be estimated from the data and as a consequence will have an associated uncertainty. For the present example, the experimental uncertainties that are always superimposed on the experimental data can be described by a Normal (i.e., Gaussian or bell shaped) distribution. Specifically, if the particular data points were independently measured a near infinite number of the times, the distribution of observed values could be represented as a Gaussian distribution. If the data that include uncertainties are divided by a limiting value that also includes uncertainties, then the resulting ratios (i.e.,  $\bar{Y}$ ) will have an uncertainty that is not a Gaussian. The sum or difference of two Gaussian distributions is a Gaussian distribution. However, the ratio of two Gaussian distributions is a Cauchy distribution, not a Gaussian distribution! As will be discussed below, a Gaussian form of the distribution of these uncertainties is critically important in determining how the parameters of the model will be determined. In addition, any systematic errors in the evaluation of the limiting value will be propagated into the resulting ratios and it is always a poor idea to introduce systematic errors. Normally the only justifiable reason to perform a transformation of the data is to convert a non-Gaussian distribution of uncertainties into a Gaussian distribution of uncertainties (Abbott and Gutgesell, 1994; Acton, 1959).

For the present example, the parameters of both the one-site model, the combination of Eqs. (5) and (7), and the two-site model, the combination of Eqs. (6) and (7), will be estimated by fitting the model to the data. *These procedures do not fit the data to the model, they fit the model to the data!*

Note that the independent variable,  $X_i$ , of the present example is the free, or unbound, ligand concentration. This is consistent with experimental protocols where the ligand concentration is directly measured, such as oxygen binding to hemoglobin where the free oxygen concentration can be measured with an oxygen electrode. However, when the binding of a hormone to its receptor is measured with a competition experiment, the resulting independent variable is the total, not the free, hormone concentration. In a competition experiment a small amount of radioactively labeled ligand is initially bound to the receptor and then is titrated with an excess of unlabeled ligand, thus providing a competition between the labeled and unlabeled ligand for the binding sites, and an independent variable which is the sum of the total concentration of labeled and unlabeled ligands. Equations (5) and (6) do not apply to the competitive hormone binding case because they are formulated in terms of the free concentration of the ligand and not the total concentration. The solution is to substitute the free concentration with the total concentration minus the calculated amount bound. Equation (8) presents the single-site binding equation in terms of the total concentration,  $X_t$ :

$$Y(X_t) = A \frac{K_a(X_t - Y(X_t))}{1 + K_a(X_t - Y(X_t))} \quad (8)$$

Equation (8) is a transcendental equation [i.e.,  $Y(X_i)$  occurs on both sides of the equal sign and as a consequence is somewhat more difficult, but not impossible, to calculate (Johnson and Frasier, 1984, 1985)].

The alternative of simply calculating the observed free concentration as the observed total minus observed bound concentrations is problematic. The observed bound concentration will have a significant experimental uncertainty and as a consequence the calculated free concentration will also have a significant uncertainty. As will be discussed below, the common parameter estimation procedures do not allow an independent variable (i.e., the free concentration) that contains significant errors. Thus, this alternative approach will yield a simpler function form for the fit [i.e., Eq. (5) instead of Eq. (8)], but it will also preclude the use of the most common methods of actually performing the fit.

The two take-home lessons from this section are that (1) *the mathematical models must be hypothesis based* and (2) *the data should only be “transformed” if it will make the measurement uncertainties more consistent with a Gaussian distribution* (Abbott and Gutgesell, 1994).

### III. Determining the Optimal Parameters of the Model

Once the mathematical form for the mathematical model (i.e., the fitting function) has been determined, the next step is to find the optimal parameters (i.e., coefficients) of the model. Conceptually, this simply involves trying different combinations of the parameters until the optimal values have been determined. But, what is meant by optimal? The most common definition of optimal is the least-square parameter values which correspond to the lowest WSSR, shown in Eq. (2). However, other definitions are also used which yield different results. For example, another definition of optimal is L1 norm where the sum of the absolute values of the residuals, the  $R_i$  in Eq. (2), is minimized:

$$L1 = \sum_i \left| \frac{Y(X_i) - Y_i}{\sigma_i} \right| = \sum_i |R_i| \quad (9)$$

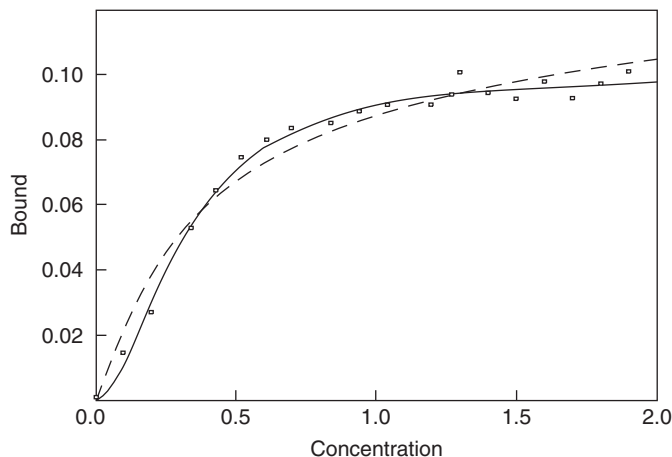
Yet another definition of optimal is the min–max norm. For the min–max norm, the parameters are adjusted until the largest value of the residuals,  $R_i$ , is minimized. *It is the property of the data, and more specifically the experimental uncertainties contained within the data, that determines the proper choice of least squares, versus L1, versus min–max optimization, versus whatever.*

Table II presents several permutations of results for the analysis of the data in Table I. The various combinations include the choice of fitting equations, weighted versus unweighted analysis (discussed in detail below), and either least squares or L1 optimization. The calculated curves for the least-squares optimizations of  $A$  and  $K_a$  in the one-site model [Eqs. (5) and (7)] and  $A$ ,  $K_1$ , and  $K_2$  in the two-site model [Eqs. (6) and (7)] are presented in Fig. 2. *But, how do we determine which of*

**Table II**  
**Parameter Estimates Based Upon Different Fitting Equations, Optimization Norms, and Weighting Schemes**

Optimization	Equations	Norm	$A$	$K_1$ or $K_a$	$K_2$
LS	Fifth degree poly	32.6808			
Exact	4	0.0			
Exact	19th Degree poly	0.0			
LS	5 and 7	101.6051	0.1301	2.0720	
LS	6 and 7	19.9977	0.1012	0.4611	9.9712
LS, unweighted	6 and 7	21.5875	0.1016	0.5032	9.9010
LS	8	87.5728	0.1241	2.6930	
L1	5 and 7	33.1078	0.1222	2.5089	
L1	6 and 7	15.4942	0.1020	0.6832	9.7946
L1, unweighted	6 and 7	15.8463	0.1008	0.3327	9.8742
L1	8	30.3504	0.1189	3.1282	

*Note:* That the unweighted parameter estimations were performed with a constant uncertainty of 0.0025. The LS parameter estimations were performed by minimizing the sum-of-squared-residuals [i.e., the norm given by Eq. (2)] and the L1 parameter estimations were performed by minimizing the sum-of-absolute-values [i.e., the norm given by Eq. (9)]. The calculations in terms of Eq. (8) assumed that the ligand concentration given in Table I was actually the total ligand concentration, instead of the free ligand concentration.



**Fig. 2** The analysis of the data presented in Table I with either a one-component binding isotherm [dashed line and Eqs. (5) and (7)], or a two-component binding isotherm [solid line and Eqs. (6) and (7)]. Both were fit by a least-squares optimization procedure and used the weighting factors shown in Table I.

*these parameter estimations provide the best description of the data?* It is not simply the one which corresponds to the lowest WSSR because that criterion predicts either the 19th degree polynomial or the contrived Eq. (4), both of which correspond to an exact zero for the WSSR, both of which are arbitrary and contrived equations, and both of which provide essentially no information of the binding

process being investigated. Hence, the answer to this question is provided by the properties of the experimental data and the properties of the measurement errors (i.e., experimental uncertainties) that are always present within every set of experimental observations. For example, the data presented in Table I is in terms of the free ligand concentration, not the total ligand concentration, and thus Eq. (8) is not appropriate for the analysis. Similarly, since the uncertainties of the observations are not constant (i.e., in statistical jargon they are heteroscedastic), it is not appropriate to accept the unweighted parameter estimations.

The objective is to find the parameter values with the highest probability of being correct, that is, it is the maximum-likelihood results that are desired. To accomplish this, we will write an equation for the probability as a function of the parameter values and then maximize this function to obtain the maximum-likelihood parameter values. Specifically, the probability of the parameter values based on the  $i$ th data point is given by the following equation:

$$P_i(Y(X_i), \text{parameters}, X_i, Y_i, \sigma_i) = Qe^{-1/2((Y_i - Y(X_i))/\sigma_i)^2} \quad (10)$$

where  $Q$  is a proportionality constant. The probability in Eq. (10) is a function of the form of the fitting equation, the parameter values, and the specific data point. The validity of this equation is based on several assumptions, specifically:

1. All of the measurement errors are contained in the dependent,  $Y_i$ , values.
2. These experimental uncertainties can be approximated by a Gaussian distribution.
3. The null hypothesis is that the fitting function,  $Y(X_i)$ , is correct.
4. The measured uncertainties are not correlated.

Most experimental protocols can be manipulated such that most, if not all, of the measurement uncertainties are present with the dependent variables (i.e., along the  $Y$ -axis). In addition, the central limit theorem of calculus implies that the distribution of experimental uncertainties is likely to be either Normal (i.e., Gaussian or bell shaped) or Log Normal. Note that Log Normal distributions can be converted to Normal distribution with a logarithmic transformation. Thus, the first three assumptions are very reasonable.

The fourth assumption is required to combine the probabilities based on the individual data points, Eq. (10), into the overall probability,  $P(\text{parameters}, \text{data})$ , for the entire set of  $n$  data points:

$$\begin{aligned} P(\text{parameters}, \text{data}) &= \prod_{i=1}^n P_i(Y(X_i), \text{parameters}, X_i, Y_i, \sigma_i) \\ &= Q^n \prod_{i=1}^n e^{-1/2((Y_i - Y(X_i))/\sigma_i)^2} \end{aligned} \quad (11)$$

Since the product of the exponentials is equal to the exponential of the sum, this overall probability can also be written as in the following equation:

$$P(\text{parameters, data}) = Q^n e^{-1/2 \sum_{i=1}^n ((Y_i - Y(X_i))/\sigma_i)^2} \quad (12)$$

Finally, Eq. (12) implies that if the above assumptions are valid then the overall probability is a maximum when the summation in the negative exponential is at a minimum. This summation is the WSSR, shown in Eq. (2). Consequently, if the assumptions are valid and appropriate then the least-squares approach will provide the maximum-likelihood set of parameter values (Johnson and Frasier, 1985). Conversely, if the data are consistent with this set of reasonable assumptions then any method of analysis that results in answers which are different from those obtained by least squares will produce results that have less than the maximum likelihood of being correct. *Thus, for most cases the least-squares method is the method of choice.* This is the reason that Fig. 2 contains only the calculated curve corresponding to the weighted least-squares fit of both the combination of Eqs. (5) and (7) (i.e., the model with one binding site), and the combination of Eqs. (6) and (7) (i.e., the two-site model) to the data values.

By now it should be clear that a set of data is more than a set of observations (i.e.,  $Y$  or dependent variables) as a function of some experimental condition (i.e.,  $X$ , or independent variables). The data always includes experimental measurement uncertainties. *An integral part of the data is a determination of how accurately the values were measured.* Thus, the measured second data point in Table I is  $Y = 0.0145 \pm 0.0020$  at  $X = 0.1$ , not just  $Y = 0.0145$  at  $X = 0.1$ . It is this determination of the measurement uncertainties that provides the relative weighting factors for the data points [i.e., the  $\sigma_i$  in Table I and Eqs. (2), (3), and (12)].

Least-squares calculations are commonly grouped into two categories, either linear or nonlinear least squares. In reality, linear least squares are simply a special case of the more general nonlinear least-squares approach. This chapter concentrates on the nonlinear category and makes note of some of the differences when they occur. The rigorous mathematical distinction between a linear and a nonlinear least-squares fit is related to the form of a fitting function. Specifically, if the second and higher order derivatives of the fitting function with respect to the parameters being estimated are all equal to zero, then the fit is a linear fit. If any of these derivatives are not equal to zero then it is a nonlinear fit.

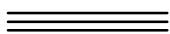
A linear least-squares fit does not imply that a straight line is being fit to the experimental data. For example, fitting polynomials containing a single independent variable [e.g., Eq. (1)] of any order is a linear process. A Fourier transform with a constant base period or frequency is equivalent to a linear fit. However, if the period or frequency is simultaneously being estimated by the fitting procedure, then the Fourier analysis will be a nonlinear fit. Fitting to the sum of exponentials

with known constant half-lives and/or rate constants is a linear process. Similarly, if the half-lives and/or rate constants are also being estimated by the fitting procedure then the exponential analysis is a nonlinear process.

There are many divergent numerical algorithms which will adjust the parameter values to obtain a minimum of the WSSR, or in other words perform a weighted nonlinear least-squares (WNLLS) fit of an equation to a set of data (Johnson and Frasier, 1985). These include the damped Gauss–Newton, the Marquardt–Levenberg, and the Nelder–Mead algorithms. When correctly implemented all of these algorithms will yield equivalent results.

No simple solution of the least-squares fitting procedure exists for nonlinear fitting equations. Equations (7) and (8) cannot be fit to a set of data by plugging the values of  $X$ ,  $Y$ , and  $\sigma$  into equations to find the binding constants. WNLLS fits of equations to data are performed by successive approximation methods. Given an initial set of answers, the WNLLS procedures will provide a better set of answers. The algorithm is repeated starting with this better set of answers iteratively until the answers do not change to within a small convergence limit (e.g., one part in a million). Consequently, WNLLS software usually requires that the user enter initial starting values for the answers. Furthermore, realistic values for these initial estimates should be utilized. The damped Gauss–Newton, the Marquardt–Levenberg, and the Nelder–Mead algorithms have different rates of convergence and different tolerances to unrealistic initial values. But they all perform better when they are started near the desired answers.

When fitting to linear models (those with second and higher order derivatives equal to zero), the iterative procedure requires only a single iteration and will be extremely tolerant of poor initial values. For example, the equations for the fitting polynomial found in most statistics textbooks can be derived by applying the Gauss–Newton WNLLS algorithm for one iteration and with initial parameter estimates of zero. Fourier transforms equations can also be derived in the same manner.



#### IV. Distinguishing Between Multiple Mathematical Models

Table II presents 11 distinct analyses of this data, but several of these are not applicable. The first three do not apply because the mathematical models are arbitrary, that is, not hypothesis based. The sixth does not apply because the data has a variable measurement uncertainty and this analysis method assumes a constant uncertainty, that is, an unweighted analysis. The seventh does not apply because the independent variable for this data set is the free concentration, not the total concentration. The last four do not apply because they correspond to a minimization of the L1 Norm instead of the preferred (see above) Least-Squares Norm, L2. Figure 2 presents the optimal fits for the remaining one-site and two-site ligand binding models to the data shown in Table I. The question at hand is does either, or both, of the two remaining analyses that are shown in Fig. 2 provide a

statistically invalid description of this data. This would allow the invalid ones to be eliminated from further consideration. The question at hand is not whether either, or both, of the two remaining analyses that are shown in Fig. 2 provide a statistically valid description of this data? You can never prove a hypothesis, you can only disprove a hypothesis!

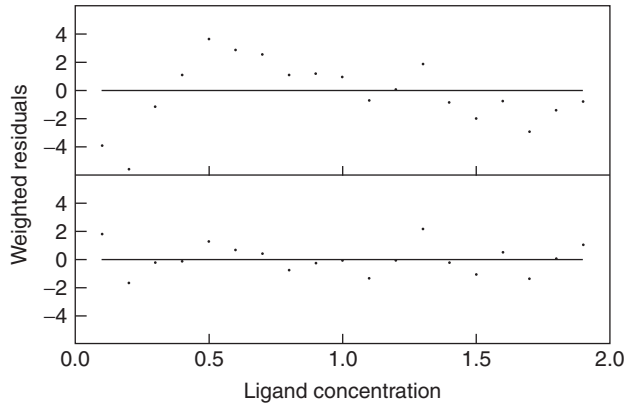
The traditional approach is to compare the WSSRs, and the analysis with the lowest WSSR is presumed to provide the best description. However, it should be recalled that there are an infinite number of function forms, such as in Fig. 1, that will provide a WSSR of zero. While zero is as low as possible and thus “as good as you can get,” the fit based on Eq. (4) clearly does not do a good job of describing the data. Of course, neither Eq. (4) nor the 19th degree polynomial are hypothesis-based models.

Some would assume that surely the WSSR of 101.6 from Eqs. (5) and (7) (Table II, line 4) is certainly much worse than the WSSR of 20.0 from Eqs. (6) and (7) (Table II, line 5). This is true but a probability level cannot be provided (i.e., a  $P$  value of 0.05) to indicate that line 5 is better than line 4. A comparison of the two variance-of-fits simply cannot be evaluated with an  $F$  statistic because the residuals (i.e., the weighted differences between the fitted function and the experimental data points) are not independent (i.e., orthogonal) of each other. In mathematical terms, Cochran's Theorem does not apply for these nonlinear functional forms. Thus, the apparent variance-of-fit cannot be separated into two independent components where one is due to the addition of the second binding site and the other being the intrinsic experimental measurement errors found within the data. Other approaches are needed to assist in differentiating the two analyses presented in Fig. 2.

Recall that the null hypothesis for these analyses was that the fitting equations were correct. If this and the other assumptions about the nature of the experimental measurement errors contained within the data are correct then the residuals (the weighted differences between the fitted curves and the data points) should be random with a Gaussian, or bell-shaped, distribution with a mean of zero. Furthermore, if the magnitudes of the weighting factors ( $1/\sigma_i$ ) are correct then this distribution should have a variance of one. No consideration of the shape of the distribution of the residuals is taken into account when using only the WSSR or  $s^2$  as a measure of goodness-of-fit.

The simplest approach to distinguish the two analyses is simply to plot the weighted residuals as a function of the independent variable and/or the dependent variable. Such a plot is shown in Fig. 3. The upper panel is a plot of the weighted residuals for the one-site model as a function of the ligand concentration and the lower panel is the corresponding graph for the two-site model. The null hypothesis (i.e., the particular fitting equation being correct) can be ruled out if the residuals are obviously not random. This is presumably the case for the upper panel of Fig. 3, but no probability level can be assigned simply by visual inspection.

There are numerous statistical tests to assign a probability that the distribution of residuals is not normally distributed with a mean of zero (Straume and Johnson, 1992). Each of these tests evaluate different aspects of the distribution of residuals



**Fig. 3** The residuals from the analyses presented in Fig. 2 as a function of the ligand concentration. The upper panel corresponds to the one-site model analysis and the lower panel is for the two-site analysis.

and provide a probability level that the distribution is not normally distributed. For the present example we will examine Autocorrelation and the Runs Test.

The existence of trends in residuals with respect to either the independent (i.e., experimental) or dependent (i.e., the experimental observable) variables suggests that some systematic behavior is present in the data that is not accounted for by the analytical model. Trends in residuals will often manifest themselves by causing too few runs (sequences of consecutive residual values of the same sign) or, in cases where negative serial correlation occurs, by causing too many runs. A convenient way to assess quantitatively this quality of a distribution of residuals is to perform a Runs Test. The method involves calculating the expected number of runs given the total number of residuals as well as an estimate of variance in this expected number of runs. The expected number of runs,  $R$ , may be calculated from the total number of positive and negative valued residuals,  $n_p$  and  $n_n$ , as:

$$R = \frac{2n_p n_n}{n_p + n_n} + 1 \quad (13)$$

The corresponding variance of the expected number of runs is given in the following equation:

$$\sigma_R^2 = \frac{2n_p n_n (2n_p n_n - n_p - n_n)}{(n_p + n_n)^2 (n_p + n_n - 1)} \quad (14)$$

A quantitative comparison is then made between the expected number of runs,  $R$ , and the observed number of runs,  $n_R$ , by calculating an estimate for the standard normal deviate as:

$$Z = \frac{n_R - R \pm 0.5}{\sigma_R} \quad (15)$$

If  $n_p$  and  $n_n$  are both greater than  $\sim 10$  then  $Z$  will be distributed approximately as a standard normal deviate. In other words, the calculated value of  $Z$  is the number of standard deviations that the observed number of runs is from the expected number of runs for a randomly distributed set of residuals of the total number being considered. The value of 0.5 is a continuity correction to account for biases introduced by approximating a discrete distribution with a continuous one. This correction is  $+0.5$  when testing for too few runs and is  $-0.5$  when testing for too many runs. The test is therefore estimating the probability that the number of runs observed is different from that expected from randomly distributed residuals. The greater the value of  $Z$ , the greater the likelihood that there exists some form of correlation in the residuals relative to the particular variable being considered. The Runs Test is not dependent on the magnitude of the residuals and thus it can be used with either weighted or unweighted residuals.

The application of the Runs Test to the current analysis example is presented in [Table III](#). For both the one-site and the two-site analyses the observed number of runs is less than the expected number of runs, thus the appropriate  $Z$  score and probability level is for too few runs. For the one-site model the  $Z = 2.068$  which corresponds to a one-sided probability level of about 0.02 that there are enough runs. By comparison, for the two-site model  $Z = 0.6501$  and a probability level of 0.26 that there are enough runs. Based on the Runs Test, the data and/or the assumptions about the data are not consistent with the one-site model and are consistent with the two-site model.

It is important to note the phrasing here. The Runs Test does not validate the two-site model, it simply says that the data and/or assumptions about the data are not consistent with the one-site model. Remember that the justification for using least-squares minimization and testing for Gaussian distributed residuals is that all of the measurement errors contained within the data are in the amount bound and that they are normally distributed with a mean of zero. If these assumptions are not valid then neither are the conclusions of any test of the normality of the residuals (e.g., the Runs Test and the autocorrelation analysis given below).

**Table III**  
**Runs Test**

	One site	Two site
Observed	6	9
Expected	$11.0 \pm 2.2$	$10.9 \pm 2.2$
$Z$	2.068	0.6501
Probability	0.9807	0.7422

Experimental data will sometimes exhibit serial correlations. These serial correlations arise when the random uncertainties superimposed on the experimental data tend to have values related to the uncertainties of other data points that are close in the independent variable, in this case the ligand concentration. For example, if the weight of a test animal is being measured once a month and the data are expressed as a weight gain per month, negative serial correlation may be expected because of the subtraction. This negative serial correlation is expected because a positive experimental error in an estimated weight gain for 1 month (e.g., an overestimate) would cause the weight gain for the next month to be systematically incorrect (e.g., an underestimated).

A basic assumption of parameter-estimation procedures is that the experimental data points are independent observations. Therefore, if the weighted differences between experimental data points and the fitted function (the residuals) exhibit such a serial correlation, then either the observations are not independent or the mathematical model did not correctly describe the experimental data. Thus, the serial correlation of the residuals for adjacent and nearby points provides a measure of the quality of the fit.

The autocorrelation function provides a simple method to quantify this serial correlation for a series of different lags,  $k$ . The lag refers to the number of data points between the observations for a particular autocorrelation. For a series of  $N$  observations,  $Y_i$ , with a mean value of  $\mu$ , the autocorrelation function is defined as the ratio of two autocovariance functions:

$$\beta_k = \frac{\hat{\sigma}_k}{\hat{\sigma}_0} \quad (16)$$

The autocovariance function is:

$$\hat{\sigma}_k = \frac{1}{n} \sum_{i=1}^{n-k} (Y_i - \mu)(Y_{i+k} - \mu) \quad (17)$$

for  $k = 0, 1, 2, \dots, K$ . In these equations,  $K$  is a maximal lag less than  $n$ . Typically  $K$  is less than  $n/2$ . The autocorrelation function has a range between  $-1$  and  $+1$ . The null hypothesis here is that the autocorrelation is equal to zero for a normally distributed random process. Note that the autocorrelation function for a zero lag is equal to 1 by definition. The expected variance of the autocorrelation coefficient of a random process with independent, identically distributed random (i.e., normal) errors is:

$$\text{var}(\beta_k) = \frac{n-k}{n(n+2)} \quad (18)$$

where the mean,  $\mu$ , is assumed to be zero. This variance of the autocorrelation coefficient is used to test the null hypothesis that the autocorrelation is equal to zero. If the autocorrelation is not equal to zero then serial correlation exists in the residuals.

The autocorrelations are actually a series of statistical tests, one for each value of  $k$ . Autocorrelations are commonly presented graphically or in tabular form as a function of  $k$ . This allows an investigator to easily compare the autocorrelation at a large series of lags  $k$  with the corresponding associated standard errors (square root of the variance) to decide if any significant autocorrelations exist.

Table IV presents an autocorrelation analysis for the residuals from the current analyses based on the one-site and two-site model examples. The autocorrelation is dependent on the magnitude of the residuals. Thus, it is calculated from the weighted residuals. For the one-site model the residual at  $k = 1$  is very significantly nonzero with a  $P > 0.002$ . Note that the maximum lag in this example is 5 so this is a case with five repeated tests or measures. Consequently, the  $P > 0.002$  is not quite as significant as it might appear due to the existence of the repeated measures. However, as long as the data is consistent with the above assumptions, then the one-site model can clearly be eliminated based on the probability level for the observed autocorrelation at  $k = 1$ .

The maximum lag,  $K$ , for the autocorrelation analysis was set to 5 for this specific analysis for two reasons. One is that trends in the residuals will usually appear within the very low numbered lags. The other reason is that this specific example only has 20 data points. In general, the goodness-of-fit test is based on the distribution properties of the residuals of the fit and requires a large number of residuals, and thus data points. This specific example, with only 20 data points, is at the low limit of applicability for these, and most other, statistical tests.

The Runs Test and autocorrelation examples presented here are illustrations of the general class of goodness-of-fit tests that are based on the expected Gaussian distribution of the residuals of the fit (Straume and Johnson, 1992).

## V. Estimate the Precision of the Model Parameters

Up to this point, we have reached three conclusions about the analysis of the current data set (Table I). The first that least-squares procedures which provide the maximum-likelihood estimate of the parameter values, are based on the properties of the measurement uncertainties that are contained within the data. The second is that

**Table IV**  
**Autocorrelation Analysis**

One-site model			Two-site model		
$K$	$\beta_k$	$P$	$k$	$\beta_k$	$P$
1	$0.5974 \pm 0.2078$	0.0020	1	$-0.1191 \pm 0.2078$	0.2833
2	$0.1538 \pm 0.2023$	0.2235	2	$-0.2961 \pm 0.2023$	0.0716
3	$-0.0788 \pm 0.1966$	0.3442	3	$-0.0558 \pm 0.1966$	0.3883
4	$-0.2376 \pm 0.1907$	0.1064	4	$-0.1568 \pm 0.1907$	0.2055
5	$-0.2236 \pm 0.1846$	0.1130	5	$-0.0951 \pm 0.1846$	0.3032

the hypothesis based one-site binding model, the combination of Eqs. (5) and (7), is not consistent with the data. The third is that the hypothesis based two-site binding model, the combination of Eqs. (6) and (7), is consistent with the data. There are surely many other hypothesis-driven binding models that are consistent with the data so the two-site model cannot be proven. All that can be concluded is that the combination of Eqs. (6) and (7) is consistent with the data.

Now that the two-site mathematical model and its concomitant hypothesis have been found to be consistent with the data, estimates of the precision of this model's parameters can be evaluated. Nonlinear models require an iterative solution to evaluate the parameter values and as a consequence an exact method to evaluate the precision of the model parameters does not exist. However, for linear models a simple set of noniterative equations to evaluate the parameter values can be derived and an exact solution for the precision of these parameter values also exists, that is, the Asymptotic Standard Errors (ASE). This linear solution is exact in the limit of a large number of data points and when no parameter correlation exists (see below), hence the title ASE.

Most available nonlinear fitting software packages report the ASE as if they were the actual precision of the determined nonlinear parameters. This can be very deceiving and problematic because for nonlinear models and/or in the presence of parameter correlation (see below) the ASE will commonly underestimate the magnitude of the parameter uncertainties, and as a consequence lead the investigator to overestimate the significance of results and thus reach incorrect conclusions. *ASE should not be used for nonlinear models!* However, they are commonly used and their basis will be described here.

The reason that the ASE are reported is that they are very easy to calculate. The most common nonlinear least-squares minimization procedures perform the calculations required to evaluate the ASE in order to accomplish the parameter estimation. As will be seen below, the more accurate procedures can require orders of magnitude more computer time than was required for the original least-squares parameter estimate. But, thankfully computers are getting very fast.

The ASE are based on the "information matrix" that is evaluated during the parameter minimization process. The minimization procedures that are based on the Gauss-Newton procedure (Johnson and Faunt, 1992; Johnson and Frasier, 1985; Straume and Johnson, 1992), all evaluate a matrix,  $H$ , of partial derivatives where the  $jk$  elements of this matrix are the sum, over each of the  $i$  data points, of the products of weighted partial derivatives of the fitting function with respect to the parameters being estimated. Specifically,

$$H_{jk} = \sum_{i=1}^n \left[ \frac{1}{\sigma_i} \frac{\partial F(\text{parameters}, X_i)}{\partial \text{parameter}_j} \right] \left[ \frac{1}{\sigma_i} \frac{\partial F(\text{parameters}, X_i)}{\partial \text{parameter}_k} \right] \quad (19)$$

where  $F(\text{parameters}, X_i)$  is the fitting function evaluated at the  $i$ th data point and *parameters* with a subscript refers to a specific parameter being estimated. The "information matrix" is the inverse of this  $H$  matrix. The Gauss-Newton

minimization algorithm employs this matrix in order to find the direction and magnitude to change the parameters that will decrease the overall WSSR. Once the Gauss–Newton procedure has converged, the ASE estimates of the precision for the  $j$ th estimated parameter is evaluated from this same matrix as the square root of  $j$ th diagonal element of the inverse of  $H$  times the variance-of-fit,  $s^2$  in Eq. (3):

$$\text{ASE}_j = \sqrt{s^2(H^{-1})_{jj}} \quad (20)$$

As noted above, the ASE commonly underestimate the actual uncertainty of the estimated parameters. This is in part because they ignore the off-diagonal elements of the information matrix. One of the previous chapters in this volume speculates that the ASE are acceptable as long as they are less than 10% of the parameter values (Tellinghuisen, 2006). This should be considered only as a rule of thumb because there are many exceptions where this is not the case.

There are several better methods to evaluate the precision of estimated parameters, such as joint confidence intervals, support plane methods, Monte-Carlo methods, and bootstrap methods (Johnson and Faunt, 1992; Johnson and Frasier, 1985; Straume and Johnson, 1992). In the present chapter, the ASE will be contrasted with the bootstrap techniques.

Probably the best method is to use the Bootstrap (Efron and Tibshirani, 1993) approach as it has the fewest assumptions. Unfortunately, this method is seldom used because of the substantial amounts of computer time required. The first step of the bootstrap procedure is to fit the hypothesis-based equation to the data. This fit produces a *noise-free fitted curve* calculated from the optimal values and a *table of residuals*, the weighted differences between the data points and the fitted curve. These values are used to create a series of surrogate data sets consisting of the noise-free fitted curve calculated from the optimal values with added randomized experimental noise. The noise is generated by repeatedly randomly selecting residuals from original table of residuals and adding them to the surrogate data sets in a random order. It is important to note that once a residual has been selected from the table it is not actually removed from the table. This is called sampling with replacement. As a consequence, for any specific surrogate data set some specific residuals will be used more than once while others will not be used at all. The fitting equation is then fit to each of the generated surrogate data sets. The fits to these surrogate data sets provide a series of estimated values for each of the estimated parameters that are used to create a probability distribution for each of the parameters. The confidence regions corresponding to any desired probability are then determined from these probability distributions.

This implementation of the Bootstrap is very similar to a Monte-Carlo analysis. The difference is the method utilized to create the simulated experimental uncertainties for the surrogate data sets. The Bootstrap uses sampling with replacement from the observed table of residuals while Monte-Carlo approach uses Gaussian distributed pseudo random numbers that are generated to match the sample variance-of-fit.

**Table V**  
**Comparison of Parameter Uncertainties Estimation Methods Based Upon The Data Shown in Table I**

	ASE	Bootstrap
$A$	0.1012 ( $\pm 3.1\%$ )	0.1012 ( $-1.5\%$ , $+2.4\%$ )
$\log K_1$	$-0.3362$ ( $\pm 95.7\%$ )	$-0.3362$ ( $-210\%$ , $+85\%$ )
$K_1$	0.4611 ( $-52\%$ , $+110\%$ )	0.4611 ( $-80\%$ , $+93\%$ )
$\log K_2$	0.9987 ( $\pm 3.8\%$ )	0.9987 ( $-3.2\%$ , $+2.8\%$ )
$K_2$	9.9712 ( $-8.4\%$ , $+9.2\%$ )	9.9712 ( $-7.0\%$ , $+6.6\%$ )

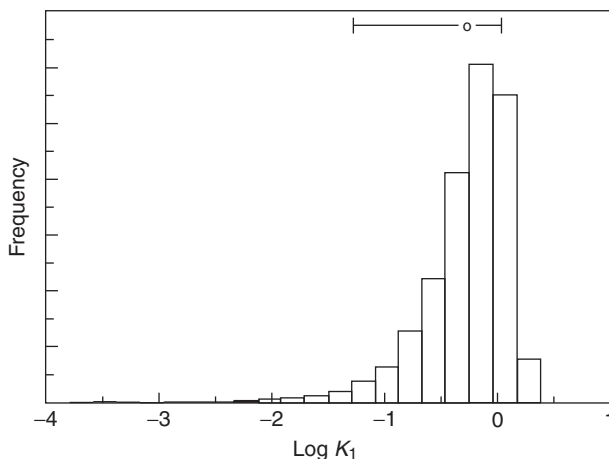
Table V presents this comparison. In this case the parameters that are being estimated are  $A$ ,  $\log K_1$ , and  $\log K_2$ . The binding constants are thus constrained to only positive values by estimating the logarithms of the binding constants. In this case 10,000 surrogate data sets were generated. Thus, the bootstrap analysis required approximately 10,000 times as much computer time as the original least-squares parameter estimation.

One of the most noteworthy aspects of Table V is that the confidence regions for the fitted parameters are symmetrical when evaluated by the ASE method and they are asymmetrical when evaluated by the Bootstrap method. In Table V the confidence regions correspond to  $\pm 1$  SEM.

For linear models it can be analytically demonstrated that the confidence regions will be symmetrical, and since the ASE is a linear approximation, it predicts that they are symmetrical. However, this proof is not valid for nonlinear models where it is common for the confidence intervals to be asymmetrical. This is even more obvious in Table VI where the analysis was performed on only the first 10 data points in Table I. Figure 4 presents the complete probability distribution for  $\log K_1$  based on the bootstrap analysis of the first 10 data points in Table I. Clearly, this probability distribution is skewed asymmetrically as is expected for nonlinear fitting models. This is contradictory to the prediction of the linear ASE method. The horizontal error bar at the top of this figure corresponds to the optimal value of  $\log K_1$  and to the  $\pm 1$  SD asymmetrical confidence region.

**Table VI**  
**Comparison of Parameter Uncertainties Estimation Methods Based Upon the First Ten Data Points Shown in Table I**

	ASE	Bootstrap
$A$	0.1028 ( $\pm 8.5\%$ )	0.1028 ( $-3.6\%$ , $+6.2\%$ )
$\log K_1$	$-0.2662$ ( $\pm 293.2\%$ )	$-0.2662$ ( $-377.3\%$ , $+114.9\%$ )
$K_1$	0.5420 ( $-83.4\%$ , $+502.0\%$ )	0.5420 ( $-90.1\%$ , $+102.3\%$ )
$\log K_2$	0.9827 ( $\pm 10.4\%$ )	0.9827 ( $-7.0\%$ , $+5.2\%$ )
$K_2$	9.6090 ( $-20.4\%$ , $+25.5\%$ )	9.6090 ( $-14.7\%$ , $+12.5\%$ )



**Fig. 4** Probability distribution from the bootstrap analysis presented in Table VI. The horizontal error bar corresponds to the optimal value and the  $\pm 1$  SD confidence region for the parameter.

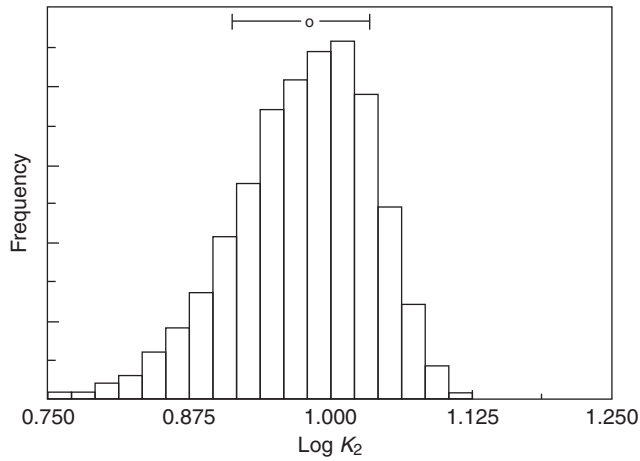
The same expected asymmetrical confidence regions are also observed for the other two estimated parameters, as are shown in Figs. 5 and 6. In this case the asymmetrical confidence regions are observed even though the ASE for  $\log K_2$  (Table VI and Fig. 5) and  $A$  (Table VI and Fig. 6) fall within the “10% rule” that has been proposed (Tellinghuisen, 2006).

A recapitulation of this section is that (1) *ASE will commonly underestimate the actual parameter uncertainties* and (2) *that for nonlinear equations, the actual confidence regions of the estimated parameters are expected to be asymmetrical.*

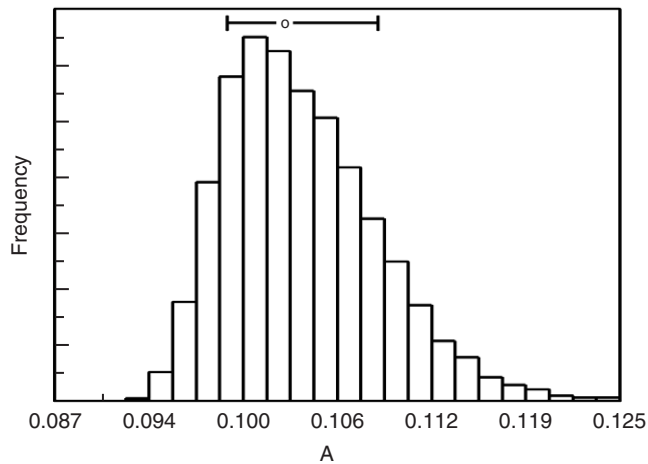
## VI. Cross-Correlation of the Estimated Parameters

Apparent parameter correlation is almost always present when multiple parameters are simultaneously estimated by any parameter estimation procedure. The correlation is a consequence of fitting a complex equation to a small number of data points that span a limited range of the independent variables. Please note that parameter correlation is also typically present when fitting simple linear equations such as a straight line. *These parameter correlations are not necessarily telling us anything about the underlying chemistry or physiology.* It is important to be aware of the magnitude of these parameter correlations because they are closely related to the difficulties that are encountered by any data fitting procedure. A larger correlation will result in an analysis that is computationally more complex or impossible.

The cross-correlation coefficient,  $CC_{jk}$ , between the  $j$ th and  $k$ th estimated parameters are evaluated from the elements of the inverse of the  $H$  matrix [Eq. (19)] that was



**Fig. 5** Probability distribution from the bootstrap analysis presented in Table VI. The horizontal error bar corresponds to the optimal value and the  $\pm 1$  SD confidence region for the parameter.



**Fig. 6** Probability distribution from the bootstrap analysis presented in Table VI. The horizontal error bar corresponds to the optimal value and the  $\pm 1$  SD confidence region for the parameter.

already evaluated by the Gauss–Newton least-squares parameter estimation procedure. Remember, this is the linear approximation.

$$CC_{jk} = \frac{(H^{-1})_{jk}}{\sqrt{(H^{-1})_{jj}(H^{-1})_{kk}}} \quad j \neq k \quad (21)$$

These cross-correlation have a range of  $\pm 1$  with the optimal being zero. As the cross-correlation approaches  $+1$  or  $-1$  the fitting procedure will become increasingly more difficult and the results questionable because the  $H$  matrix is becoming nearly singular and cannot easily be inverted. For practical purposes, if the magnitudes of the cross-correlation coefficients are less than approximately  $\pm 0.97$  the least-squares procedure can usually function adequately. However, the  $\pm 0.97$  should not be considered as an absolute threshold with everything acceptable below  $\pm 0.97$  and everything unacceptable outside of this range. All fitting procedures get progressively worse as the magnitude of the cross-correlations increase toward 1. If equal to  $\pm 1$ , the matrix will be singular and the parameter estimation procedures will fail.

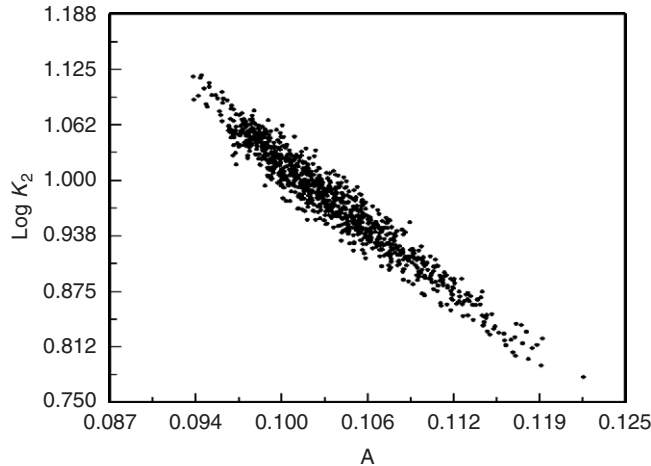
Table VII presents the cross-correlation coefficients when Eqs. (6) and (7) were fit to all of the data (i.e., the parameter estimation shown in Table V) and to only the first 10 data points (i.e., the parameter estimation shown in Table VI). Clearly, as the number of data points and the range of the independent variable (i.e., the ligand concentration) decreases, the cross-correlation coefficients rapidly approach the limiting value of  $\pm 1$ . If only the first 8 data points are used then the cross-correlation between  $A$  and  $\log K_2$  is  $-0.990$ , and so on.

The simulated data set analyses that were performed for the evaluation of the parameter precision by the bootstrap method can also be used to visualize the cross-correlation between the estimated parameters. The bootstrap procedure creates thousands of sets of simulated surrogate data series and then analyzes each to obtain thousands of sets of parameter values that are estimates of the actual parameter values. These are tabulated to obtain the confidence regions and probability distributions of the estimated parameters (e.g., Figs. 4–6). Figure 7 is a plot of 1000 pairs of  $A$  and  $\log K_2$  obtained from the bootstrap evaluation of the confidence intervals when fitting to only the first 10 data points. In Fig. 7 the pairs of values appear to fall along a line that is not coincident with the parameter axes, indicating that these two parameters are highly correlated. If the parameters were not correlated then the points would appear to be aligned with the axes.

In some cases, the cross-correlation between the parameters can be minimized, or eliminated, by a careful choice for the form of the fitting equation. As an example, consider a straight line,  $Y(X) = a + bX$ , fit to the first 10 data points in Table I. The resulting least-squares estimated values are:  $a = 0.0095$  and  $b = 0.1044$  with a  $CC_{ab} = -0.8221$ . In this case, the cross-correlation between the parameters can be changed by altering the form of the fitting equation. For example, by altering the form of the straight line from  $Y(X) = a + bX$  to  $Y(X) = a + b(X - \beta)$ ,

**Table VII**  
**Alteration of the Cross-correlation Coefficients Caused By Limiting the Data Set**

$CC$	All data	First 10 points
$A$ and $\log K_1$	0.844	0.956
$A$ and $\log K_2$	$-0.762$	$-0.974$
$\log K_1$ and $\log K_2$	$-0.774$	$-0.923$



**Fig. 7** Bootstrap analysis of the correlation between parameter values.

where  $\beta = 0.3995$  the weighted average value of  $X$ , the least-squares estimated parameters for this later case are,  $a = 0.5122$  and  $b = 0.1044$  and the cross-correlation coefficient is exactly zero. This is actually the identical line as in the previous fit, however, the intercept is now at  $X = \beta$  instead of  $X = 0$ .

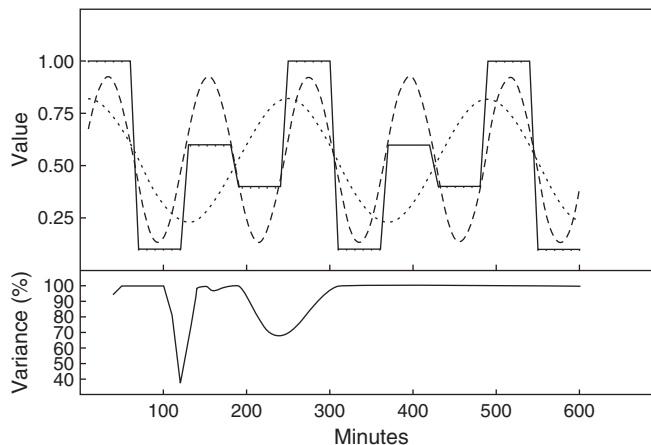
## VII. Uniqueness of the Parameters

For linear fitting equations it can be algebraically demonstrated that only a single set of model parameters exist that correspond to a minimum least squares. However, this cannot be demonstrated for nonlinear fitting equations and thus it is possible that multiple sets of parameters which corresponding to a minimum in the WSSR might exist. Unfortunately, the existence of multiple minima for the analysis of a data set cannot be demonstrated with the data set presented in [Table I](#).

One method to test for multiple minima is to start the iterative nonlinear fitting procedure at many different initial starting sets of values. If the iterative procedure always converges to the same set of answers then you can have some assurance that multiple minima do not occur. However, this approach does not guarantee that multiple minima do not exist.

[Figure 8](#) presents a simulated time series that demonstrates multiple minima when analyzed as a sign wave with a single harmonic. When a cosine wave with an unknown period, [Eq. \(22\)](#), is fit to this data the resulting values will depend on the initial starting values where the least-squares algorithm was initiated:

$$Y(X) = A + B \cos\left(\frac{2\pi(\text{time} - \phi)}{L}\right) \quad (22)$$



**Fig. 8** An example of least-squares parameter estimation exhibiting multiple minima. Upper panel: The solid curve represents the data points, the long dashed lines is a cosine wave with a period of  $\sim 120$  min, and the short dashed line corresponds to a cosine wave with a period of  $\sim 240$  min. Lower panel: The percent of the variance remaining after a cosine wave of the specific period is fit to the data.

Four parameters are being simultaneously estimated,  $A$ ,  $B$ ,  $L$ , and  $\phi$ . This is a nonlinear analysis because the period,  $L$ , is being simultaneously estimated with the other three parameters. If the initialization value of  $L$ , the period, is near 120 the least-squares algorithm will converge to the 120-min period (the long dashed line in Fig. 8). However, if it is initially near 239 it will converge to 239 period (the short dashed line in Fig. 8). The least-squares algorithms are not guaranteed to find the lowest variance-of-fit, only to find a minimum variance-of-fit.

In the example shown in Fig. 8 there are actually four periods that correspond to a minimum in the variance-of-fit, specifically  $L = 82, 120, 162,$  or  $239$  min. The ones at 82 and 162 min are very shallow minima, but they are minima and some algorithms will find them. Remember that the variance-of-fit is not a good measure of goodness-of-fit so it is not a good idea to specify that the minimum at 120 min is better than the one at 239 min because its variance-of-fit is lower.

The potential for multiple minima always exists when fitting to nonlinear equations but, unfortunately, no method exists that will guarantee locating all of these minima. It is, however, common for some of the multiple minima to have parameter values that are physically unrealistic. For example, a negative molecular weight has no physical meaning. If multiple physically meaningful minima are found then they must all be described when you report your results.

---

## VIII. Conclusions

This chapter discusses WNLLS parameter estimations methods for the analysis of experimental data. A Bootstrap method is presented as the optimal approach for the evaluation the precision of the least-squares estimated parameter values.

Goodness-of-fit, parameter cross-correlation, and uniqueness of parameter values were also discussed.

### Acknowledgments

The author thanks the National Institutes of Health for partial support from the following grants: RR00847, RR08119, HD011474, RR019991, DK063609, and DK064122.

### References

- Abbott, R. D., and Gutgesell, H. P. (1994). Effects of heteroscedasticity and skewness on prediction in regression: Modeling growth of the human heart. *Meth. Enzymol.* **240**, 37–51.
- Acton, F. S. (1959). “Analysis of Straight Line Data.” John Wiley & Sons, New York, p. 219.
- Efron, B., and Tibshirani, R. J. (1993). “An Introduction to the Bootstrap.” Chapman and Hall, New York.
- Johnson, M. L., and Faunt, L. M. (1992). Parameter estimation by least-squares methods. *In* “Methods in Enzymology 210” (L. Brand, and M. L. Johnson, eds.), pp. 1–37. Academic Press, New York.
- Johnson, M. L., and Frasier, S. G. (1984). Analysis of hormone binding data. *In* “Methods in Diabetes Research” (J. Larner, and S. Pohl, eds.), Vol. I, pp. 45–61. Wiley and Sons, New York.
- Johnson, M. L., and Frasier, S. G. (1985). Nonlinear least-squares analysis. *In* “Methods in Enzymology 117” (S. N. Timasheff, and C. H. W. Hirs, eds.), pp. 301–342. Academic Press, New York.
- Johnson, M. L., and Straume, M. (2000). Deriving complex ligand binding formulas. *Meth. Enzymol.* **323**, 155–167.
- Straume, M., and Johnson, M. L. (1992). Analysis of residuals: Criteria for determining goodness-of-fit. *In* “Methods in Enzymology 210” (L. Brand, and M. L. Johnson, eds.), pp. 87–105. Academic Press, New York.
- Tellinghuisen, J. (2008). Stupid Statistics! *In* “Methods in Cell Biology” (John J. Correia, and H. William Detrich, III, eds.), pp. 739–780. Amsterdam, Academic Press.