USE OF STATISTICAL METHODS IN KINETICS

$$\frac{df_A}{dt} = \frac{k\left(\frac{n_A^0}{V^0}\right)(1 - f_A)(M - f_A)}{(1 + \varepsilon_A f_A)} \qquad (1.5.16)$$

IN WHAT FOLLOWS $f_A \equiv X_A$ FOR CONVERSION

EXAMPLE 1.5.4

Equal volumes of 0.2 M trimethylamine and 0.2 M n-propylbromine (both in benzene) were mixed, sealed in glass tubes, and placed into a constant temperature bath at 412 K. After various times, the tubes were removed and quickly cooled to room temperature to stop the reaction:

$$N(CH_3)_3 + C_3H_7Br \Rightarrow C_3H_7\overset{+}{N}(CH_3)_3Br^-$$

The quaternization of a tertiary amine gives a quaternary ammonium salt that is not soluble in nonpolar solvents such as benzene. Thus, the salt can easily be filtered from the remaining reactants and the benzene. From the amount of salt collected, the conversion can be calculated and the data are:

Time at 412 K (min)	Conversion (%)
5	4.9
13	11.2
25	20.4
34	25.6
45	31.6
59	36.7
80	45.3
100	50.7
120	55.2

Are these data consistent with a first- or second-order reaction rate?

■ **Answer**

The reaction occurs in the liquid phase and the concentrations are dilute. Thus, a good assumption is that the volume of the system is constant. Since $C_A^0 = C_B^0$:

$$\text{(first-order)} \qquad \ln\left[\frac{1}{1 - f_A}\right] = kt$$

$$\text{(second-order)} \qquad \frac{f_A}{1 - f_A} = kC_A^0 t$$

In order to test the first-order model, the $\ln[1/(1 - f_A)]$ is plotted versus t while for the second-order model, $f_A/(1 - f_A)$ is plotted versus t (see Figures 1.5.1 and 1.5.2). Notice that both models conform to the equation $y = \overline{\alpha}_1 t + \overline{\alpha}_2$. Thus, the data can be fitted via linear

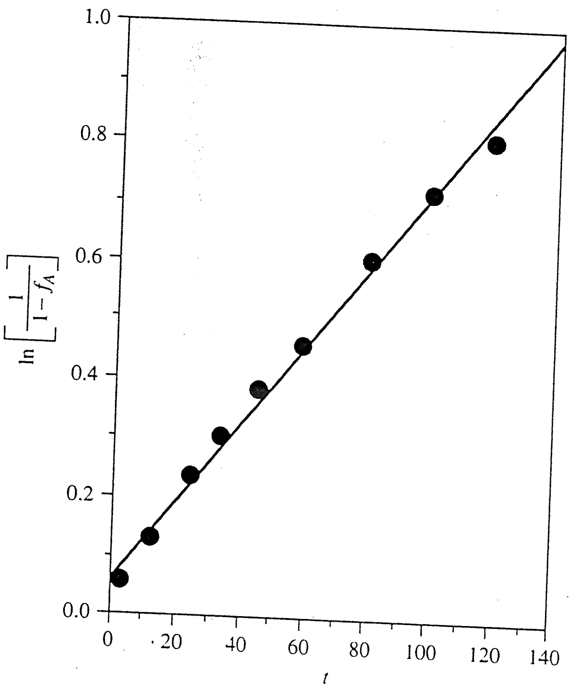

Figure 1.5.1 |
Reaction rate data for first-order kinetic model.

regression to both models (see Appendix B). From visual inspection of Figures 1.5.1 and 1.5.2, the second-order model appears to give a better fit. However, the results from the linear regression are (SE is the standard error):

first-order $\bar{\alpha}_1 = 6.54 \times 10^{-3}$ $SE(\bar{\alpha}_1) = 2.51 \times 10^{-4}$
$\bar{\alpha}_2 = 5.55 \times 10^{-2}$ $SE(\bar{\alpha}_2) = 1.63 \times 10^{-2}$
$\bar{R}_{cc} = 0.995$

second-order $\bar{\alpha}_1 = 1.03 \times 10^{-2}$ $SE(\bar{\alpha}_1) = 8.81 \times 10^{-5}$
$\bar{\alpha}_2 = -5.18 \times 10^{-3}$ $SE(\bar{\alpha}_2) = 5.74 \times 10^{-3}$
$\bar{R}_{cc} = 0.999$

Both models give high correlation coefficients ($\bar{R}_{cc}$), and this problem shows how the correlation coefficient may not be useful in determining "goodness of fit." An appropriate way to determine "goodness of fit" is to see if the models give $\bar{\alpha}_2$ that is not statistically different from zero. This is the reason for manipulating the rate expressions into forms that have zero intercepts (i.e., a known point from which to check statistical significance). If a student $\bar{t}^*$-test is used to test significance (see Appendix B), then:

$$\bar{t}^* = \frac{|\bar{\alpha}_2 - 0|}{SE(\bar{\alpha}_2)}$$

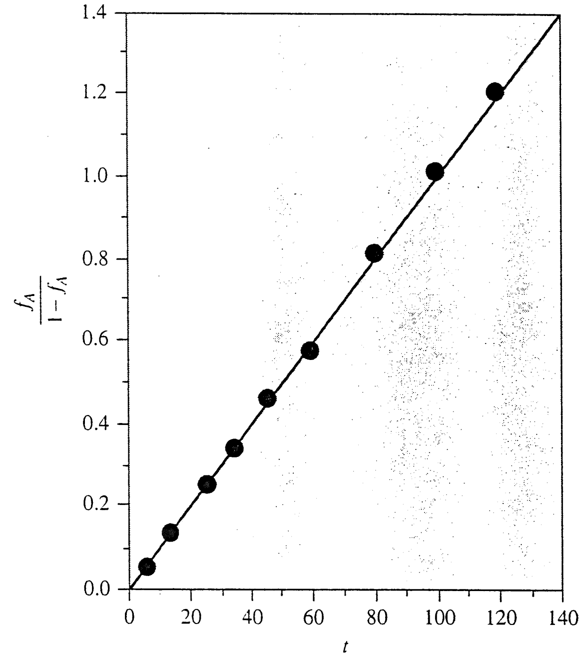

Figure 1.5.2 |
Reaction rate data for second-order kinetic model.

The values of $\bar{t}^*$ for the first- and second-order models are:

$$\bar{t}_1^* = \frac{|5.55 \times 10^{-2} - 0|}{1.63 \times 10^{-2}} = 3.39$$

$$\bar{t}_2^* = \frac{|-5.18 \times 10^{-3} - 0|}{5.74 \times 10^{-3}} = 0.96$$

For 95 percent confidence with 9 data points or 7 degrees of freedom (from table of student $\bar{t}^*$ values):

$$\bar{t}_{exp}^* = \frac{\text{expected deviation}}{\text{standard error}} = 1.895$$

Since $\bar{t}_1^* > \bar{t}_{exp}^*$ and $\bar{t}_2^* < \bar{t}_{exp}^*$, the first-order model is rejected while the second-order model is accepted. Thus,

$$kC_A^0 = 1.030 \times 10^{-2}$$

and

$$k = \frac{1.030 \times 10^{-2}}{0.1\,\text{M}} = 0.1030 \frac{1}{\text{M} \cdot \text{min}}$$

THERE FORE

$$r = [(0.1030 \pm 0.0009)(M^{-1}min^{-1})] C_A C_B$$

APPENDIX B

Regression Analysis

B.1 | Method of Least Squares

Below is illustrated the method of least squares to fit a straight line to a set of data points (y_i, x_i). Extensions to nonlinear least squares fits are discussed in Section B.4.

Consider the problem of fitting a set of data (y_i, x_i) where y and x are the dependent and independent variables, respectively, to an equation of the form:

$$y = \overline{\alpha}_1 + \overline{\alpha}_2 x \tag{B.1.1}$$

by determining the coefficients $\overline{\alpha}_1$ and $\overline{\alpha}_2$ so that the differences between y_i and $y_i = \overline{\alpha}_1 + \overline{\alpha}_2 x_i$ are minimized. Given $\overline{\alpha}_1$ and $\overline{\alpha}_2$, the deviations Δy_i can be calculated as:

$$\Delta y_i = y_i - \overline{\alpha}_1 - \overline{\alpha}_2 x_i \tag{B.1.2}$$

For any value of $x = x_i$, the probability PP_i for making the observed measurement y_i with a Gaussian distribution and a standard deviation σ_i for the observations about the actual value $y(x_i)$ is (P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969, p. 101):

$$PP_i = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right] \tag{B.1.3}$$

The probability of making the observed data set of measurements of the N values of y_i is the product of the individual PP_i or:

$$PP(\overline{\alpha}_1, \overline{\alpha}_2) = \prod_i^N PP_i = \prod_i^N \left(\frac{1}{\sigma_i\sqrt{2\pi}}\right) \exp\left[-\frac{1}{2}\sum_i^N \left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2\right] \tag{B.1.4}$$

The best estimates for $\overline{\alpha}_1$ and $\overline{\alpha}_2$ are the values that maximize $PP(\overline{\alpha}_1, \overline{\alpha}_2)$ (method of maximum likelihood). Define:

$$\overline{X}^2 = \sum_i^N \left[\frac{y_i - y(x_i)}{\sigma_i}\right]^2 = \sum_i^N \left[\frac{1}{\sigma_i^2}(y_i - \overline{\alpha}_1 - \overline{\alpha}_2 x_i)^2\right] \qquad \text{(B.1.5)}$$

Note that in order to maximize $PP(\overline{\alpha}_1, \overline{\alpha}_2)$, $\overline{X}^2$ is minimized. Thus, the method to find the optimum fit to the data is to minimize the sum of the squares of the deviations (i.e., least-squares fit).

As an example of how to calculate $\overline{\alpha}_1$ and $\overline{\alpha}_2$, consider here the case where $\sigma_i = \sigma =$ constant. To minimize $\overline{X}^2$, the partial derivatives of $\overline{X}^2$ with respect to $\overline{\alpha}_1$ and $\overline{\alpha}_2$ must be set equal to zero:

$$\frac{\partial \overline{X}^2}{\partial \overline{\alpha}_1} = 0 = \frac{\partial}{\partial \overline{\alpha}_1}\left[\frac{1}{\sigma^2}\sum (y_i - \overline{\alpha}_1 - \overline{\alpha}_2 x_i)^2\right] = \frac{-2}{\sigma^2}\sum (y_i - \overline{\alpha}_1 - \overline{\alpha}_2 x_i)$$

$$\frac{\partial \overline{X}^2}{\partial \overline{\alpha}_2} = 0 = \frac{\partial}{\partial \overline{\alpha}_2}\left[\frac{1}{\sigma^2}\sum (y_i - \overline{\alpha}_1 - \overline{\alpha}_2 x_i)^2\right] = \frac{-2}{\sigma^2}\sum x_i(y_i - \overline{\alpha}_1 - \overline{\alpha}_2 x_i)$$

These equations can be rearranged to give:

$$\sum y_i = \sum \overline{\alpha}_1 + \sum \overline{\alpha}_2 x_i = \overline{\alpha}_1 N + \overline{\alpha}_2 \sum x_i$$

$$\sum x_i y_i = \sum \overline{\alpha}_1 x_i + \sum \overline{\alpha}_2 x_i^2 = \overline{\alpha}_1 \sum x_i + \overline{\alpha}_2 \sum x_i^2$$

The solutions to these equations yield $\overline{\alpha}_1$ and $\overline{\alpha}_2$ in the following manner:

$$\left.\begin{aligned} \overline{\alpha}_1 &= \frac{\begin{vmatrix} \Sigma y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma x_i^2 \end{vmatrix}}{\begin{vmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix}} = \frac{\Sigma y_i \, \Sigma x_i^2 - \Sigma x_i \, \Sigma x_i y_i}{N \Sigma x_i^2 - (\Sigma x_i)^2} \\ \overline{\alpha}_2 &= \frac{\begin{vmatrix} N & \Sigma y_i \\ \Sigma x_i & \Sigma x_i y_i \end{vmatrix}}{\begin{vmatrix} N & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{vmatrix}} = \frac{N \Sigma x_i y_i - \Sigma x_i \, \Sigma y_i}{N \Sigma x_i^2 - (\Sigma x_i)^2} \end{aligned}\right\} \qquad \text{(B.1.6)}$$

The calculation is straightforward. First compute Σx_i, Σy_i, Σx_i^2, and $\Sigma x_i y_i$. Second use the summed values in Equation (B.1.6) to obtain $\overline{\alpha}_1$ and $\overline{\alpha}_2$.

B.2 | Linear Correlation Coefficient

Referring to Equations (B.1.1) and (B.1.6), if there is no correlation between x and y, then there are no trends for y to either increase or decrease with increasing x. Therefore, the least-squares fit must yield $\overline{\alpha}_2 = 0$. Now, consider the question of whether the data correspond to a straight line of the form:

$$x = \overline{\alpha}_1' + \overline{\alpha}_2' y \qquad \text{(B.2.1)}$$

The solution for $\overline{\alpha}_2'$ is:

$$\overline{\alpha}_2' = \frac{N \Sigma x_i y_i - \Sigma x_i \Sigma y_i}{N \Sigma y_i^2 - (\Sigma y_i)^2} \qquad \text{(B.2.2)}$$

Again, if there is no correlation between x and y, then $\overline{\alpha}_2' = 0$. At the other extreme, if there is complete correlation between x and y, then there is a relationship between $\overline{\alpha}_1$, $\overline{\alpha}_2$, $\overline{\alpha}_1'$, and $\overline{\alpha}_2'$ that is:

$$y = -\frac{\overline{\alpha}_1'}{\overline{\alpha}_2'} + \frac{1}{\overline{\alpha}_2'}x = \overline{\alpha}_1 + \overline{\alpha}_2 x$$

$$\overline{\alpha}_1 = -\frac{\overline{\alpha}_1'}{\overline{\alpha}_2'}$$

$$\overline{\alpha}_2 = \frac{1}{\overline{\alpha}_2'}$$

Thus, a perfect correlation gives $\overline{\alpha}_2\overline{\alpha}_2' = 1$ and no correlation yields $\overline{\alpha}_2\overline{\alpha}_2' = 0$ since both $\overline{\alpha}_2$ and $\overline{\alpha}_2'$ are zero for this condition. The linear correlation coefficient is therefore defined as:

$$\overline{R}_{cc} = \sqrt{\overline{\alpha}_2\overline{\alpha}_2'} = \frac{N\sum x_i y_i - \sum x_i \sum y_i}{\left[N\sum x_i^2 - (\sum x_i)^2\right]^{1/2}\left[N\sum y_i^2 - (\sum y_i)^2\right]^{1/2}} \tag{B.2.3}$$

The values of $\overline{R}_{cc}$ are $-1 \le \overline{R}_{cc} \le 1$ with $\overline{R}_{cc} = 1$ and $\overline{R}_{cc} = 0$ defining perfect and no correlations, respectively. Although the linear correlation coefficient is commonly quoted as a measure of "goodness of fit," it is really not appropriate as a direct measure of the degree of correlation. If the data can be represented in a manner such that the fit should result in a y-intercept equal to zero, then a simple method can be used to determine the "goodness of fit."

B.3 | Correlation Probability with a Zero Y-Intercept

Numerous kinetic expressions can be placed into a form that would yield a zero y-intercept when using the linear least-squares method. A survey of a few of these models is provided in Table B.3.1. Given that the y-intercept is a known value (i.e., zero), if a perfect correlation could be achieved, the hypothesis that the true value of the parameter, $\overline{\alpha}_1$, is equal to the specified value, $\overline{\alpha}_1^*$, could be tested by referring the quantity:

$$\bar{t}^* = \frac{\overline{\alpha}_1 - \overline{\alpha}_1^*}{SE(\overline{\alpha}_1)} \tag{B.3.1}$$

to the table of Student's $\bar{t}^*$ values with N-2 degrees of freedom. The standard error, SE, can be calculated as follows. The standard deviation σ_z of the determination of a parameter z is via the chain rule:

$$\sigma_z^2 = \sum\left[\sigma_i^2\left(\frac{\partial z}{\partial y_i}\right)^2\right] \tag{B.3.2}$$

where σ_i is the standard deviation of each datum point i. If $\sigma_i = \sigma =$ constant, then σ_z^2 is approximately equal to the sample variance, which is (P. R. Bevington, *Data*

Table B.3.1 | Examples of kinetic relationships yielding zero intercepts.

Kinetics	Reactions	Relationship
1. 0th-order, irreversible (one-way), (constant volume)	$A \xrightarrow{k} \overline{P}$	$C_A^0 - C_A = kt$
2. First-order, irreversible (one-way)	$A \xrightarrow{k} \overline{P}$	$\ln\left[\frac{1}{1-f_A}\right] = kt$
3. First-order, reversible (two-way)	$A \underset{k_2}{\overset{k_1}{\rightleftarrows}} \overline{P}$	
• no product present at $t = 0$		$\ln\left[\frac{1}{1-(f_A/f_A^{eq})}\right] = \left(\frac{k_1}{f_A^{eq}}\right)t$
• product present at $t = 0$, C_P^0		$\ln\left[\frac{1}{1-(f_A/f_A^{eq})}\right] = \left[\frac{k_1\left(\frac{C_P^0}{C_A^0}+1\right)}{\left[\frac{C_P^0}{C_A^0}+f_A^{eq}\right]}\right]t$
4. Second-order, irreversible (one-way), (constant volume)	$A + B \xrightarrow{k} \overline{P}$	
• $C_A^0 = C_B^0$		$\frac{1}{C_A} - \frac{1}{C_A^0} = kt$
• $C_A^0 \neq C_B^0$		$\ln\left[\frac{C_B C_A^0}{C_A C_B^0}\right] = (C_B^0 - C_A^0)kt$
5. Second-order, reversible (two-way) $C_A^0 = C_B^0, C_C^0 = C_D^0 = 0$	$A + B \underset{k_2}{\overset{k_1}{\rightleftarrows}} C + D$	$\ln\left[\frac{f_A^{eq} - (2f_A^{eq} - 1)f_A}{(f_A^{eq} - f_A)}\right] = 2k_1\left[\frac{1}{f_A^{eq}} - 1\right]C_A^0 t$
6. Third-order, irreversible (one-way), (constant volume)	$A + 2B \xrightarrow{k} \overline{P}$	$\ln\left[\frac{C_B C_A^0}{C_A C_B^0}\right] - \frac{2\left[\frac{C_B^0}{C_A^0} - 2\right](C_A^0 - C_A)}{(C_B^0/C_A^0)C_B} = (C_A^0)^2\left[\frac{C_B^0}{C_A^0} - 2\right]kt$
	$A + B + C \xrightarrow{k} \overline{P}$	$\ln\left(\frac{C_A^0}{C_A}\right) + \left[\frac{C_C^0/C_A^0 - 1}{\left(\frac{C_B^0}{C_A^0} - \frac{C_C^0}{C_A^0}\right)}\right]\ln\left(\frac{C_B^0}{C_B}\right) + \left[\frac{C_B^0/C_A^0 - 1}{\left(\frac{C_B^0}{C_A^0} - \frac{C_C^0}{C_A^0}\right)}\right]\ln\left(\frac{C_C}{C_C^0}\right) = (C_A^0)^2\left[\frac{C_B^0}{C_A^0} - 1\right]\left[\frac{C_C^0}{C_A^0} - 1\right]kt$
7. nth-order, irreversible (one-way), (constant volume)	$A \xrightarrow{k} \overline{P}$	$(C_A)^{(1-n)} - (C_A^0)^{(1-n)} = (n-1)kt$

Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, New York, 1969, p. 114):

$$\sigma^2 \cong \frac{1}{N-2}\sum\left(y_i - \overline{\alpha}_1 - \overline{\alpha}_2 x_i\right)^2 \tag{B.3.3}$$

for the linear equation (B.1.1). (The sample variance is the sum of squares of the residuals divided by the number of data points minus the number of parameters fitted.) Now Equation (B.3.2) written for the linear equation (B.1.1) gives:

$$\left.\begin{aligned} \sigma_{\overline{\alpha}_1}{}^2 &= \sigma^2 \sum\left(\frac{\partial \overline{\alpha}_1}{\partial y_i}\right)^2 \\ \sigma_{\overline{\alpha}_2}{}^2 &= \sigma^2 \sum\left(\frac{\partial \overline{\alpha}_2}{\partial y_i}\right)^2 \end{aligned}\right\} \tag{B.3.4}$$

Using Equation (B.1.6) to calculate the partial derivatives in Equation (B.3.4) yields:

$$\sigma_{\overline{\alpha}_1}{}^2 = \frac{\sigma^2}{\Delta}\left(\sum x_i^2\right) \tag{B.3.5}$$

$$\sigma_{\overline{\alpha}_2}{}^2 = \frac{N\sigma^2}{\Delta} \tag{B.3.6}$$

where

$$\Delta = N \sum x_i^2 - \left(\sum x_i\right)^2$$

Thus, $SE(\overline{\alpha}_1) = \sigma_{\overline{\alpha}_1}$ and $SE(\overline{\alpha}_2) = \sigma_{\overline{\alpha}_2}$. Now returning to Equation (B.3.1), $\overline{t}^*$ is the experimental deviation over the standard error and if this value is larger than the value in a Student's $\overline{t}^*$-distribution table (see any text on statistics for this table) for a given degree of confidence, for example, 95 percent ($\overline{t}^*_{\text{exp}}$ = expected deviation/standard error), then the hypothesis is rejected, that is, the y-intercept is significantly different than zero. If $\overline{t}^* < \overline{t}^*_{\text{exp}}$ then the hypothesis is accepted and $\overline{\alpha}_2$ can be reported as:

$$\overline{\alpha}_2 \pm \sigma_{\overline{\alpha}_2} \tag{B.3.7}$$

B.4 | Nonlinear Regression

There are numerous methods for performing nonlinear regression. Here, a simple analysis is presented in order to provide the reader the general concepts used in performing a nonlinear regression analysis.

To begin a nonlinear regression analysis, the model function must be known. Let:

$$y = f(x, \boldsymbol{a}) \tag{B.4.1}$$

where the function f is nonlinear in the dependent variable x and unknown parameters designated by the set $\boldsymbol{a} = [a_1, a_2, \ldots, a_n]$. A least-squares fit of the observed measurements y_i to the function shown in Equation (B.4.1) can be performed as follows. First, define $\overline{X}^2$ [for linear regression see Equation (B.1.5)] as:

$$\overline{X}^2 = \sum_{i=1}^{N}\left[\frac{y_i - y(x_i)}{\sigma_i}\right] = \sum_{i=1}^{N}\left\{\frac{1}{\sigma^2}\left[y_i - f(x_i, \boldsymbol{a})\right]^2\right\} \qquad \text{(B.4.2)}$$

As with linear least squares analysis, $\overline{X}^2$ is minimized as follows. The partial derivatives of $\overline{X}^2$ with respect to the parameters of $\boldsymbol{a}$ are set equal to zero, for example, with a_1:

$$0 = \frac{\partial \overline{X}^2}{\partial a_1} = \frac{\partial}{\partial a_1}\left\{\frac{1}{\sigma^2}\sum_{i=1}^{N}[y_i - f(x_i, \boldsymbol{a})]^2\right\} = \frac{-2}{\sigma^2}\sum_{i=1}^{N}[y_i - f(x_i, \boldsymbol{a})]\frac{\partial f(x_i, \boldsymbol{a})}{\partial a_1} \qquad \text{(B.4.3)}$$

Thus, there will be n equations containing the n parameters of $\boldsymbol{a}$. These equations involve the function $f(x_i, \boldsymbol{a})$ and the partial derivatives of the function, that is,

$$\frac{\partial f(x_i, \boldsymbol{a})}{\partial a_j}, \quad \text{for } j = 1, \ldots, n$$

The set of n equations of the type shown in Equation (B.4.3) needs to be solved. This set of equations is nonlinear if $f(x_i, \boldsymbol{a})$ is nonlinear. Thus, the solution of this set of equations requires a nonlinear algebraic equation solver. These are readily available. For information on the type of solution, consult any text on numerical analysis. Since the solution involves a set nonlinear algebraic equation, it is performed by an iterative process. That is, initial guesses for the parameters a are required. Often, the solution will terminate at local minimum rather than the global minimum. Thus, numerous initial guesses should be used to assure that the final solution is independent of the initial guess.

The issue of "goodness-of-fit" with nonlinear regression is not straightforward. Numerous methods can be used to explore the "goodness-of-fit" of the model to the data (e.g., residual analysis, variance analysis, and Chi-squared analysis). It is always a good idea to inspect the plot of the predicted [$y(x_i)$] versus observed y_i values to watch for systematic deviations. Additionally, some analytical measure for "goodness-of-fit" should also be employed.