Journal of Industrial Microbiology, 12 (1993) 195–199

Published by The Macmillan Press Ltd

Principles of nonlinear regression modeling

David A. Ratkowsky

Department of Primary Industry, Tasmania, Mt Pleasant Laboratories, PO Box 46, Kings Meadows, Tas. 7249, Australia

(Received 27 November 1992)

Key words: Nonlinear models; Parsimony; Parameterization; Range of applicability; Stochastic assumption; Interpretability

SUMMARY

Five principles govern the selection of nonlinear regression models for bacterial growth. Examples are given of the various ways in which researchers have approached the problems of nonlinear regression modeling together with some discussion of linear modeling.

As a consequence of the fact that physical and biological models often arise as solutions of differential equations, regression models that describe natural processes are often *nonlinear* in the model parameters (coefficients). An important consequence of the fact that a regression model is nonlinear is that the least-squares estimators of its parameters do not possess the desirable properties of their counterparts in linear regression models, that is, they are *not* unbiased, minimum variance, normally distributed estimators. Instead, the parameter estimators in nonlinear models may be greatly biased, have considerable excess variance above the minimum variance bound, and have a markedly skewed distribution. Nonlinear regression models differ greatly among themselves in the extent to which the estimators exhibit these manifestations of nonlinear behavior. Ratkowsky [18] labeled models with a small amount of nonlinearity 'close to linear', since the parameter estimators of such models will be almost unbiased, have variances only slightly in excess of the minimum attainable variance, and be almost normally distributed. Models whose parameters give rise to estimators that don't possess this property were labeled 'far from linear', for obvious reasons.

Reparameterization, that is, changing the form in which the parameters appear in the model, may improve a model's estimation properties and make the model close to linear. Sometimes it is only a single parameter that is the cause of a model exhibiting far-from-linear behavior. Numerous examples of the way in which reparameterization can improve a model's properties are given in Ratkowsky [18,19]. In [19], a general procedure for obtaining a class of close-to-linear parameters known as 'expected-value' parameters is described.

In predictive microbiology, a wide range of nonlinear regression models are in use. Isothermal growth of bacteria has often been modeled by use of nonlinear regression models such as the Gompertz [5,10] or the logistic [10]. A multitude of nonlinear regression models has been proposed and tested for modeling the temperature dependence of the bacterial growth rate constant. These include the square-root models [20], the Arrhenius-based Johnson–Lewin model [12], the Sharpe–DeMichele [25] and Schoolfield et al. models [24], and the damage/repair model [11]. Extensions of some of the above models to include water activity effects [6,7,15] or pH effects [2] also result in nonlinear regression models. Therefore, since predictive microbiological models are almost invariably nonlinear regression models, it is important to be able to ascertain the extent to which the parameter estimators are biased and non-normally distributed. This is especially true as the parameters of these models are often said to have physical or biological meaning. Clearly, a parameter representing some physical quantity would be of little practical use if its estimator were grossly biased.

In this paper, we describe five principles, considerations or desiderata that modelers should be aware of, when indulging in a nonlinear regression modeling exercise. For more elaboration on the ideas to be presented below, see [3, Ch. 3] and [19, Chs. 2 and 10]. The five points are listed briefly below, then discussed more fully in subsequent paragraphs.

(i) Parsimony (models should contain as few parameters as possible)
(ii) Parameterization (find the one which has the best estimation properties)
(iii) Range of applicability (the data should cover the full range of X and Y)
(iv) Stochastic specification (the error term needs to be modeled, too)

Correspondence to: D.A. Ratkowsky, c/o Prof. T.A. McMeekin, Dept. of Agricultural Science, University of Tasmania, G.P.O. Box 252C, Hobart 7001, Tasmania, Australia.
Reference to a brand or firm name does not constitute endorsement by the US Department of Agriculture over others of a similar nature not mentioned.

(v) Interpretability (parameters should be meaningful, as far as possible).

(i) Parsimony

The principle of parsimony embodied in Ockham's Razor, a philosophic principle enunciated by medieval clergyman William of Ockham (or Occam) which translates loosely as 'Entities are not to be multiplied beyond necessity', forms a basis for regression modeling just as it does for other scientific endeavors. A simple model is thus seen by that principle to be more desirable than a complex one. (Surprisingly, there are a large number of scientists, perhaps even a majority, who think that a complex model is 'better', usually in some undefined or vague sense, than a simple one.) Belief in parsimony will automatically direct a modeler towards developing as simple a model as possible which explains the phenomenon under study.

There is a very practical reason for seeking as simple a model as possible that will describe a phenomenon, as, in general terms, the greater the number of parameters, the greater the extent of nonlinear behavior. For example, most one-parameter models are either close to linear or exhibit only a small amount of nonlinearity. Two-parameter models will usually require that their original form be reparameterized to obtain a close-to-linear model. Examples of three- and four-parameter models that are close to linear are rare, and the author does not know of any close-to-linear five-parameter models. Keeping the model as simple as possible with few parameters is the way of being likely to obtain a model with a small amount of nonlinearity.

(ii) Parameterization

It is a rare nonlinear regression model of two or more parameters that is close to linear in its original parameterization. Consider the following two-parameter (α and β) model for describing a convex curve (see Fig. 1 for the shapes described by this curve),

$$y = \alpha\beta^X, \tag{1}$$

where X is the explanatory (regressor) variable and y is the expected value of the response (dependent) variable at the specific X value. Amongst the alternative parameterizations of this model are

$$y = \alpha \exp(\gamma X) \tag{2}$$

and

$$y = \exp(\delta + \gamma X). \tag{3}$$

These models differ from each other only in the form in which the parameters appear in the model; in all other respects, the three models are identical. That is, given a set of data to which these models are fitted, the fitted values at each value of X will be the same for each of the three models. Note that parameter α appears in both Eqns (1) and (2); this is deliberate, as it is the same parameter in each model. That is, changing β in Eqn (1) to γ in Eqn (2) by use of the relationship $\beta = \exp(\gamma)$ does not change the least-squares estimate of α, its standard error, and other estimation properties. Therefore, if one parameter in a model is far from linear, with the other parameters being close to linear, one may reparameterize the offending parameter without disturbing the properties of the other parameters. Generally speaking, Eqn (3) has better properties than either Eqn (1) or Eqn (2), determined by testing the three models on various generated data sets (Ratkowsky, unpublished results). A particularly useful form of reparameterization is via the use of 'expected-value' parameters, also called 'stable' parameters [22,23]. Many examples of how to use expected-value parameters may be found in [19]. Suffice it to say that Eqns (1)–(3) can be reparameterized to give the following model,

$$y = y_1^{(x-x_2)/(x_1-x_2)}\, y_2^{(x_1-x)/(x_1-x_2)}$$

in which the 'new' parameters are y_1 and y_2, representing the expected values corresponding to $X = X_1$ and $X = X_2$, respectively. Provided X_1 and X_2 are chosen to be well within the range of the observed data, the parameters y_1 and y_2 will have excellent estimation properties, being close to linear in behavior.

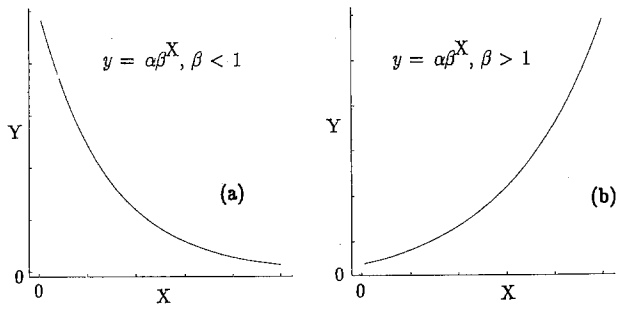

Fig. 1. Example of a two-parameter model for describing a convex curve: (a) $\beta < 1$; (b) $\beta > 1$.

(iii) Range of applicability

It is important that the data set to which the model is fitted covers the full range for which the model applies. One major source of poor estimation behavior lies with the fact that sometimes a fit is attempted to a model when only fragmentary data are available. For example, consider Fig. 2, which shows a Gompertz model, one parameterization of which is given by

$$y = \alpha \exp[-\exp(\beta - \gamma X)], \tag{4}$$

trying to represent data that includes almost no readings above the inflection point of the model. Attempts to fit such a model to the data depicted often fails to achieve convergence to the least-squares estimates. Even if conver-

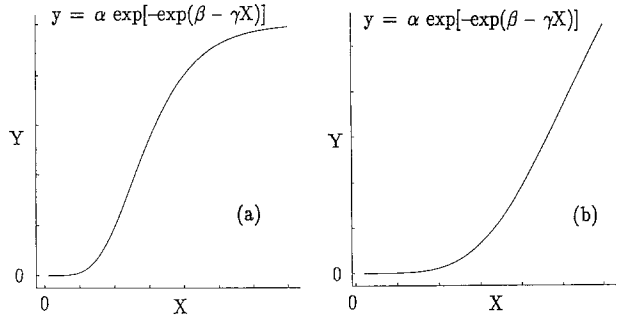

Fig. 2. Examples of Gompertz model. (a) Data describing full range of applicability of model; (b) Data in lower region only.

gence does occur, the estimation properties are likely to be very poor, with the three parameters α, β and γ all being far from linear.

(iv) Stochastic specification

A regression model, whether it be a linear or a nonlinear model, consists of two components, a deterministic part and a stochastic part. The deterministic part represents the true relationship between the response variable and the explanatory variable. Hence, a model such as Eqn (4), if indeed a Gompertz model is appropriate, could better be written as

$$Y = \alpha \exp[-\exp(\beta - \gamma X)] + \epsilon \tag{5}$$

where ϵ represents the 'error' or stochastic term, and Y denotes the random variable representing the response. Customarily, for a fixed value of X, one uses an averaging process via the expectation operator E, to obtain

$$E(Y) = \alpha \exp[-\exp(\beta - \gamma X)],$$

thereby averaging over the values of the random error term ϵ. Often, E(Y) is written y, as in Eqns (1)–(4).

The usual method of obtaining the estimators of the parameters is by use of ordinary least squares. This gives equal weight to each of the data points and implies that ϵ is *homogeneous*, that is, having the same mean and variance for each possible value of X. In addition, it is usually assumed that these errors are normally distributed, an assumption that is needed for carrying out tests of significance. If ϵ is not homogeneous, then one may use weighted least squares or seek a transformation of the response variable.

As an example of a transformation of the response, consider a Gompertz curve with a multiplicative, rather than an additive, stochastic term:

$$Y = \alpha \exp[-\exp(\beta - \gamma X)]\epsilon'.$$

Taking the logarithm of both sides results in

$$\ln Y = \ln \{\alpha \exp[-\exp(\beta - \gamma X)]\} + \ln \epsilon'$$

where $\ln \epsilon'$ may now represent independent, normally distributed random error, and thereby allow the above regression model to be fitted by ordinary least squares using ln Y as the response variable.

In predictive microbiology, one is usually interested in modeling the growth rate constant as a function of temperature and other factors. One set of such models, the so-called square-root models, have the following forms,

$$\sqrt{k} = b(T - T_{\min}) \tag{6}$$

and

$$\sqrt{k} = b(T - T_{\min})\{1 - \exp[c(T - T_{\max})]\}, \tag{7}$$

with the former to be used in the low temperature region (below the optimum temperature for growth) and the latter to be used throughout the entire biokinetic temperature range. In these models, T is the absolute temperature, k is the growth rate constant (which might be obtained as the reciprocal of the lag phase duration or of the time required to reach a specified level of spoilage), and b, c, $T_{\min}$ and $T_{\max}$ are parameters, the latter two being notional minimum and maximum temperatures for growth, respectively. The stochastic assumption that is being implicitly made in Eqns (6) and (7) is that the error term (which is not explicitly written down in those equations) is homogeneous in $\sqrt{k}$ and not in k itself or in some other transformation of k. In the early development of these equations [20,21], the data sets lacked replication at each temperature, and the question of what was the correct form of the stochastic term was decided by examining the residuals after fitting various transformations of the response variable. More recently, however, Zwietering et al. [27], in modeling the temperature dependence of growth of *Lactobacillus plantarum*, concluded that the correct stochastic assumption was that ϵ was homogeneous in k, that is, the appropriate model form should be written and fitted as

$$k = [b(T - T_{\min})]^2\{1 - \exp[c(T - T_{\max})]\}^2. \tag{8}$$

A further modification due to those authors is to drop the exponent 2 from the right-hand-most term, that is, to model growth throughout the entire bacterial growth range using

$$k = [b(T - T_{\min})]^2\{1 - \exp[c(T - T_{\max})]\}. \tag{9}$$

Heitzer et al. [11] also concluded that the variance of k was constant, rather than the variance of $\sqrt{k}$, and used Eqn (8) to model the growth rate constants of *K. pneumoniae* NCIB 418, *E. coli* NC3, the coccobacillus strain NA17, and *Bacillus* sp. strain NCIB 12522.

Other authors, in using other growth models, have made other stochastic assumptions. Consider first the Sharpe–DeMichele model

$$k = \frac{T \exp\left[\Phi - \frac{\Delta H^{\ddagger}}{RT}\right]}{1 + \exp\left[(\Delta S_L - \Delta H_L/T)/R\right] + \exp\left[(\Delta S_H - \Delta H_H/T)/R\right]} \quad (10)$$

where Φ, $\Delta H^{\ddagger}$, ΔS_L, ΔH_L, ΔS_H and ΔH_H are parameters which were said to reflect individual thermodynamic characteristics of the organism's control enzyme system. In this model, the response variable appears on the left-hand side of the equation as k. Although Sharpe and DeMichele [25] fitted Eqn (10) to eight data sets (see their Figure 2) using 'a non-linear regression technique', they did not actually state whether the k was transformed or not. Schoolfield et al. [24] modified the original formulation with the aim of improving the model's 'regression suitability'. Their reparameterization is written as

$$k = \frac{\rho\,(25\,^{\circ}C)\,\frac{T}{298}\exp\left[\frac{\Delta H_A^{\ddagger}}{R}\left(\frac{1}{298} - \frac{1}{T}\right)\right]}{1 + \exp\left[\frac{\Delta H_L}{R}\left(\frac{1}{T_{1/2_L}} - \frac{1}{T}\right)\right] + \exp\left[\frac{\Delta H_H}{R}\left(\frac{1}{T_{1/2_H}} - \frac{1}{T}\right)\right]} \quad (11)$$

in which the entropy parameters have been replaced by temperatures said to correspond to conditions at which the enzyme is half inactivated. Weighted nonlinear regression was used in which k was 'weighted according to the reciprocal of the rate values' on the grounds that 'low rates tend to be measured with greater accuracy than high rates'.

Later, Adair et al. [1], in comparing the Arrhenius model to the square-root models, reverted back to a parameterization which is more reminiscent of the original Sharpe–DeMichele form, after taking logarithms of both sides, than of the Schoolfield et al. [24] reparameterization. The form they used for the full temperature range was equivalent to the following (since ln (1/time) = ln k),

$$\ln k = -A - B/T + \ln(T) - \ln\{1 + \exp[F + (D/T) + \exp[G + (H/T)]\}, \quad (12)$$

where A, B, D, F, G and H are functions of the thermodynamic parameters of Eqn (10). Adair et al. [1] fitted this model using standard unweighted non-linear regression with ln (lag or generation time) = $-\ln k$ as the dependent variable.

Thus, the stochastic assumptions made by the various workers in modeling the growth rate constant are very diverse.

(v) Interpretability

Often there is a conflict between interpretability of the parameters and the goodness of their estimation properties. That is, a parameter may have good biological interpretability but be far from linear in estimation behavior. This was the reason why Schoolfield et al. [24] sought to reparameterize the Sharpe–DeMichele model, Eqn (10). However, Lowry and Ratkowsky [14] showed that the reparameterization represented by Eqn (11) did not achieve its objective, with the new parameters also exhibiting a high degree of non-linear behavior.

The square-root model given by Eqns (7) or (9) contains four parameters, two of which (T_{min} and T_{max}) are readily interpretable as minimum and maximum notional temperatures for growth. These are the temperatures at which the model gives a zero value for k, the name 'notional' being used so as not to confuse them with the actual temperatures at which growth may cease. For example, T_{min} for *Escherichia coli* is about 3.5 °C, but the actual minimum temperature for growth of this organism is 7.8 °C [16,26]. The other two parameters of Eqns (7) or (9), b and c, are less readily interpretable, except to note that b controls the rate at which k rises from T_{min} to its maximum at a temperature denoted as T_{opt}, and c denotes the rate at which k declines between T_{opt} and T_{max}.

The parameters of Eqns (10) and (11) purport to represent various thermodynamic constants, such as enthalpy and entropy changes. However, Brandts [4] showed that these thermodynamic quantities are strong functions of temperature, and cannot be expected to be constant over the entire biokinetic temperature range. This helps explain the often anomalous results obtained by Adair et al. [1]. Although Adair et al. [1] did not present values of the estimated thermodynamic constants these values were calculated from the raw data of Adair et al. [1] by McMeekin et al. [17]. The estimated thermodynamic parameters for replicate data sets on the same commodity are inconsistent, thermodynamically impossible or unlikely (wrong sign), or of a magnitude that would exclude growth in the low temperature region.

Another model, not discussed in previous sections, as it is a linear regression model, is the 'linear Arrhenius' model of Davey et al. [9], which adds a quadratic term to the usual Arrhenius form to produce

$$\ln k = C_0 + \frac{C_1}{T} + \frac{C_2}{T^2}. \quad (13)$$

As this model is only intended for the low temperature region, it is not as parsimonious as the simple square-root model given by Eqn (6). The coefficients C_0, C_1 and C_2 are also not readily interpretable, but one advantage of the model is that it can be fitted by linear regression computer programs. Also, it was readily extended to model water activity as well as temperature effects [8] with the following relationship,

$$\ln k = C_0 + \frac{C_1}{T} + \frac{C_2}{T^2} + C_3 a_w + C_4 a_w^2, \quad (14)$$

also a linear regression model. The parameters are not readily interpretable.

The simple square-root model has also been modified to model water activity as well as temperature [6, 7, 15] using the following model form,

$$\sqrt{k} = c(T - T_{min}) \sqrt{(a_w - a_{w\,min})}. \quad (15)$$

This model contains only three parameters, two of which, T_{min} and $a_{w\,min}$, are readily and obviously interpretable. Adams et al. [2] have shown that a model of identical form can be applied to changes in pH rather than a_w,

$$\sqrt{k} = c(T - T_{min}) \sqrt{(pH - pH_{min})}. \quad (16)$$

Ratkowsky [19] emphasized the importance of examining whether the parameters in a nonlinear regression model exhibited close-to-linear behavior. One way to examine this behavior is to carry out a simulation study treating the parameter estimates as though they were the true values of the parameters. Kohler et al. [13] carried out such a simulation study for the square-root model using the form given by Eqn (8), that is, with untransformed k as the dependent variable. The distributions of the estimates obtained from 500 trials were presented in their Figure 3, and showed that, for each of the four parameters, the estimates were close to being unbiased and normally distributed. Hence, one may have some confidence, when modeling with the square-root models, that the estimates of the parameters will be close to being unbiased, normally distributed, minimum variance estimators.

REFERENCES

1 Adair, C., D.C. Kilsby and P.T. Whitall. 1989. Comparison of the Schoolfield (non-linear Arrhenius) model and the Square Root model for predicting bacterial growth in foods. Food Microbiol. 6: 7–18.

2 Adams, M.R., C.L. Little and M.C. Easter. 1991. Modelling the effect of pH acidulant and temperature on the growth rate of *Yersinia enterocolitica*. J. Appl. Bacteriol. 71: 65–71.

3 Bates, D.M. and D.G. Watts. 1988. Nonlinear Regression Analysis and its Applications. John Wiley and Sons, New York.

4 Brandts, J.F. 1967. Heat effects on proteins and enzymes. In: Thermobiology (Rose, A.H. ed.), pp. 25–72, Academic Press, London.

5 Buchanan, R.L. and J.G. Phillips. 1990. Response surface model for predicting the effects of temperature, pH, sodium chloride content, sodium nitrite concentration and atmosphere on the growth of *Listeria monocytogenes*. J. Food. Protect. 53: 370–376, 381.

6 Chandler, R.E. and T.A. McMeekin. 1989a. Modelling the growth response of *Staphylococcus xylosus* to changes in temperature and glycerol concentration/water activity. J. Appl. Bacteriol. 66: 543–548.

7 Chandler, R.E. and T.A. McMeekin. 1989b. Combined effect of temperature and salt concentration/water activity on the growth rate of *Halobacterium* spp. J. Appl. Bacteriol. 67: 71–76.

8 Davey, K.R. 1989. A predictive model for combined temperature and water activity on microbial growth during the growth phase. J. Appl. Bacteriol. 67: 483–488.

9 Davey, K.R., S.H. Lin and D.G. Wood. 1978. The effect of pH on continuous high-temperature/short-time sterilization of liquid. Am. Inst. Chem. Eng. J. 24: 537–540.

10 Gibson, A.M., N. Bratchell and T.A. Roberts. 1987. The effect of sodium chloride and temperature on the rate and extent of growth of *Clostridium botulinum* type A in pasteurized pork slurry. J. Appl. Bacteriol. 62: 479–490.

11 Heitzer, A., H.-P.E. Kohler, P. Reichert and G. Hamer. 1991. Utility of phenomenological models for describing temperature dependence of bacterial growth. Appl. Environ. Microbiol. 57: 2656–2665.

12 Johnson, F.H. and I. Lewin. 1946. The growth rate of *E. coli* in relation to temperature, quinine and coenzyme. J. Cell. Comp. Physiol. 28: 47–75.

13 Kohler, H.-P.E., A. Heitzer and G. Hamer. 1991. Improved unstructured model describing temperature dependence of bacterial maximum specific growth rates. In: International Symposium on Environmental Biotechnology (Verachtert, H. and W. Verstraete, eds), pp. 511–514, Royal Flemish Society of Engineers, Ostend, Belgium.

14 Lowry, R.K. and D.A. Ratkowsky. 1983. A note on models for poikilotherm development. J. Theor. Biol. 105: 453–459.

15 McMeekin, T.A., R.E. Chandler, P.E. Doe, C.D. Garland, J. Olley, S. Putro and D.A. Ratkowsky. 1987. Model for the combined effect of temperature and salt concentration/water activity on the growth rate of *Staphylococcus xylosus*. J. Bacteriol. 62: 543–550.

16 McMeekin, T.A., J. Olley and D.A. Ratkowsky. 1988. Temperature effects on bacterial growth rates. In: Physiological Models in Microbiology (Bazin, M.J. and J.L. Prosser, eds), pp. 75–89, CRC Press, Boca Raton, Florida.

17 McMeekin, T.A., J. Olley, D.A. Ratkowsky and T. Ross. 1989. Comparison of the Schoolfield (non-linear Arrhenius) model and the square root model for predicting bacterial growth in foods – a reply to C. Adair et al. Food Microbiol. 6: 304–308.

18 Ratkowsky, D.A. 1983. Nonlinear Regression Modeling: a Unified Practical Approach. Marcel Dekker, New York.

19 Ratkowsky, D.A. 1990. Handbook of Nonlinear Regression Models. Marcel Dekker, New York.

20 Ratkowsky, D.A., R.K. Lowry, T.A. McMeekin, A.N. Stokes and R.E. Chandler. 1983. Model for bacterial culture growth rate throughout the entire biokinetic temperature range. J. Bacteriol. 154: 1222–1226.

21 Ratkowsky, D.A., J. Olley, T.A. McMeekin and A. Ball. 1982. Relationship between temperature and growth rate of bacterial cultures. J. Bacteriol. 149: 1–5.

22 Ross, G.J.S. 1975. Simple non-linear modelling for the general user. Proc. 40th Session Int. Statist. Inst., Warsaw, Contributed Papers, Vol. 2: 585–593.

23 Ross, G.J.S. 1990. Nonlinear Estimation. Springer-Verlag, New York.

24 Schoolfield, R.M., P.J.H. Sharpe and C.E. Magnuson. 1981. Non-linear regression of biological temperature-dependent rate models based on absolute reaction-rate theory. J. Theor. Biol. 88: 719–731.

25 Sharpe, P.J.H. and D.W. DeMichele. 1977. Reaction kinetics and poikilotherm development. J. Theor. Biol. 64: 649–670.

26 Shaw, M.K., A.G. Marr and J.L. Ingraham. 1971. Determination of the minimal temperature for growth of *Escherichia coli*. J. Bacteriol. 105: 683–684.

27 Zwietering, M.H., J.T. de Koos, B.E. Hasenack, J.C. de Wit and K. van't Riet. 1991. Modeling of bacterial growth as a function of temperature. Appl. Environ. Microbiol. 57: 1094–1101.