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)
(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),

\[ y = \alpha \beta^X, \]  

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

\[ y = \alpha \exp(\gamma X) \]  

and

\[ y = \exp(\delta + \gamma X). \]

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 \( \alpha \) appears in both Eqns (1) and (2); this is deliberate, as it is the same parameter in each model. That is, changing \( \beta \) in Eqn (1) to \( \gamma \) in Eqn (2) by use of the relationship \( \beta = \exp(\gamma) \) does not change the least-squares estimate of \( \alpha \), 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,

\[ y = y_1(y_2-y)/(x_2-x_1), y_2(y_2-y)/(x_2-x_1) \]

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.

(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

\[ y = \alpha \exp[-\exp(\beta - \gamma X)], \]

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-