The relationship between mathematical expression and biological insight: Some issues in community ecology modeling, exemplified by the logistic equation

Peter J. Taylor
peter.taylor@umb.edu

Follow this and additional works at: https://scholarworks.umb.edu/cct_sicw

Part of the Philosophy of Science Commons, and the Population Biology Commons

Recommended Citation
https://scholarworks.umb.edu/cct_sicw/13
The relationship between mathematical expression and biological insight

Some issues in community ecology modeling, exemplified by the logistic equation

PETER J. TAYLOR
The relationship between mathematical expression and biological insight: Some issues in community ecology modeling, exemplified by the logistic equation

Peter J. Taylor
Department of Organismic and Evolutionary Biology
Harvard University, Cambridge, MA 02138, USA
August 1985

Abstract
The relationship between mathematical expression and biological insight is a central problem for modelers in community ecology. In this paper I distinguish three interpretations of this relationship: models as redescriptions as observations; models as capturing fundamental or generative biological relations; and models as an aid in exploring how to formulate the inquiries we put to nature. I aim to show that we can reduce ambiguities and confusion and ecological theory if we hold the different interpretations clearly in mind. I conclude by advancing an overall strategy of model building—de/composition—in which each of the interpretations plays a role. The logistic equation, a very simple mathematical model, is used to illustrate my arguments about models and modeling.

THE RELATIONSHIP BETWEEN MATHEMATICAL EXPRESSION AND
BIOLOGICAL INSIGHT: SOME ISSUES IN COMMUNITY ECOLOGY MODELING,
EXEMPLIFIED BY THE LOGISTIC EQUATION

The relationship between mathematical expression and biological insight is a central problem for modelers in community ecology. In this paper I will distinguish three interpretations of this relationship: models as redescriptions of observations, models as capturing fundamental (generative) biological relations, and models as an aid in exploring how to formulate the inquiries we put to nature. I do not propose that we keep model building within the boundaries of any one of these interpretations but, rather, I aim to show that we can reduce ambiguities and confusion in ecological theory if we hold the three different interpretations clearly in mind. I will conclude by advancing an overall strategy of model building—"de/composition"—in which each of the interpretations plays a role.

I have chosen to use the logistic equation, a very simple mathematical model, to illustrate my arguments about models and modeling. Given the existence of the definitive critical account of the logistic by Williams (1972) and the heights of sophistication of contemporary mathematical ecology (e.g. Schaffer 1985) I must provide some justification for this choice. One justification follows from the fact that I am addressing the problem of the relationship between models and biology and not, primarily, analyzing the logistic equation. The logistic is sufficient for me to tease out this problem. If I had used more complex models, I would have risked obscuring the conceptual
problems with mathematical detail. As a corollary, by keeping the mathematics to a minimum I hope to make the point that students can be introduced simultaneously to models and to alternative interpretations and conceptual problems.

Another source of justification for using the logistic is that it has occupied a central place in the historical development of population ecology (Kingsland 1982) and, despite strident criticisms (Smith 1952) the logistic remains part of contemporary population ecology. Students are conventionally introduced to mathematical models in ecology by way of the exponential and logistic (Wilson and Bossert 1971, Roughgarden 1979). The concepts of intrinsic rate of increase (r) and carrying capacity (K) have become incorporated in the larger discourse of population biology (Parry 1981). Many single species and multispecies models are simply elaborations of the logistic model. Typically the modeler presents them without addressing the problem that the early critics of the logistic demanded Peal and Reed face—can we interpret a mathematical expression as a law of growth for populations? (Lotka 1925a, Hogben 1931). A clear answer to this problem will provide a basis for a critical examination of more complicated models in community ecology. It is to this and related issues that I now turn.

I. REDESCRIPTION

The simplest interpretation of a mathematical expression is that it is a redescriptions of observations of a phenomenon. In the case of the logistic, if we are interested in a particular population elevated from
some ecosystem, and make observations of its size, \( N \), and rate of change, \( \dot{N} \), we can estimate the parameters \( a, b \) that allow these observations to best fit the logistic equation:

\[
\dot{N} = N(b + aN) \tag{1A}
\]

\[
\dot{N}/N = b + aN \tag{1B}
\]

Fitting equation (1B), that is, per capita growth rates to an expression which is linear in \( N \), is a univariate linear regression problem, a simple case of least squares fitting:

Let

\[
Y = \begin{bmatrix}
\dot{N}(1) \\
\dot{N}(2) \\
\vdots \\
\dot{N}(k)
\end{bmatrix} \quad T
\]

\[
X = \begin{bmatrix}
1 & 1 & \cdots & 1 \\
N(1) & N(2) & \cdots & N(k)
\end{bmatrix}
\]

\[
z = \begin{bmatrix}
b \\
a
\end{bmatrix} \quad ^T
\]

where \( k \) is the number of observations \((k \geq 2)\) and the superscript \((k)\) denotes the \( k \text{th} \) observation.

The least squares estimate of \( z \) for the equation \( Y = Xz \) is

\[
\hat{z} = (X^TX)^{-1}X^TY \tag{2}
\]

The estimates of variance of the elements of \( z \) are the diagonal elements of the variance/covariance matrix \((X^TX)^{-1}s^2\) where \( s^2 = [(Y - \hat{Y})(Y - \hat{Y})]/(k-2) \). The goodness of fit is indicated by the value of \( R^2 \), the coefficient of determination, where

\[
R^2 = 1 - [(Y - \hat{Y})^T(Y - \hat{Y})]/(\hat{Y}^T - \bar{Y}^2)
\]

(This least squares formulation is used because it generalizes to fitting multispecies observations to Generalized Lotka-Volterra equations, and, in principle, to more complicated mathematical expressions.)
The fitted parameters $a, b$ have implicitly incorporated the effects of any other, "hidden," variables that exist in the ecosystem. Let us consider the conditions necessary for this procedure to yield a good fit; in statistical terms, a fit with low variance estimates or high $R^2$. I will construct three simple, hypothetical systems. These systems will be governed by known equations so that I can be omniscient about the effects of using the logistic model for one population elevated from these systems.

Although these systems are simple they should serve to make a point which holds in more complex systems. A good fit depends, in general, both on the relationship of the population of interest with the rest of the ecosystem and on the initial conditions in that ecosystem. I will give in parentheses real analogs to these systems but I am not suggesting that my system equations are the best models for these real analogs.

System 1: Population with non-renewed resource (batch culture)

Consider a system composed of an initial resource of $R_0$ and an initial population inoculum of size $N_0$. This population grows by consuming its resource without the resource being renewed or the organisms dying, governed by Eqs. 3 and 4:

\[ \frac{dR}{dt} = \lambda RN/Y \quad (3) \]
\[ \frac{dN}{dt} = \lambda RN \quad (4) \]

where $\lambda =$ per capita growth rate of organism per unit resource available
\[ y = \text{ratio of population increase to resource consumed, that is,} \]
\[ \text{yield} \]
\[ \dot{R} = -\frac{N}{y} \text{so at any point in time a conservation relation (Williams, 1972) holds:} \]
\[ R = R_0 - \frac{(N - N_0)}{y} \]  \hspace{1cm} (5)

Substituting into Eq. 4 gives
\[ \dot{N} = \lambda N \left[ R_0 - \frac{(N - N_0)}{y} \right] = N \left[ \lambda \left( \frac{R_0 + N_0}{y} \right) - \left( \lambda N/y \right) \right] \]  \hspace{1cm} (6)

Eq. 6 is the logistic exactly, without any variance of the estimates. The effect of the resource variable has been included indirectly but faithfully. If the organisms begin to die or the yield, \( y \), varies the fit will begin to break down.

System 2: Population with one renewed resource (continuous culture, one-stage chemostat)

A simple example of a population with a renewed resource is a system governed by the following equation for level of resource (glucose), \( R \) and population size of cultured organism (bacteria), \( N \):
\[ \dot{R} = i - dR - \frac{mRN}{(k+R)/y} \]  \hspace{1cm} (7)
\[ \dot{N} = -dN + \frac{mRN}{(k+R)} \]  \hspace{1cm} (8)

where \( i = \text{input rate of resource} \)
\( d = \text{dilution rate} \)
\( m = \text{maxima per capita growth rate of organism on the resource} \)
\( k = \text{level of resource at which growth rate equals } m/2 \)
\( y = \text{yield} \)
There are three analytic methods for deriving the fitted logistic, applicable under different conditions, in addition to the least squares method. I will describe these methods and compare them.

Waldon (1975) has shown analytically that $N$ can be fitted with a logistic equation when $N_0 \ll iy/D$, $R_0 \approx i/d$ and $R$ has declined to being $\ll k$, in which

\begin{align*}
a &= -m/yk \\
b &= m/k - d
\end{align*}

(9) \hspace{1cm} (10)

In general Waldon's conditions are restrictive. An alternative analytically derived fit preserves the dominant stability character of the system local to its equilibrium, $N^*$, (see appendix). When $4d^2 \ll i(m-D)^2/km$ the logistic equation can be fitted with

\begin{align*}
a &= -d/N^* \approx -iy \\
b &= d
\end{align*}

(11) \hspace{1cm} (12)

If the observations are not close to equilibrium the fit is much poorer than a direct fit to the observations (see Fig. 1).

A final analytic derivation is from Schaffer (1981). A logistic equation is constructed with two features: 1) the same value for $N^*$ as the full system; and 2) the same value of $\dot{N}/N$ at time = 0 as the full system when $R_0$ is set at the equilibrium value of $R$ in the absence of $N$. For system 2 we obtain

\begin{align*}
a &= b/N^* \\
b &= -d + im/(i+dk)
\end{align*}

(13) \hspace{1cm} (14)

Consider the system governed by Eqs. 7 and 8 with the following values for the parameters:

\begin{align*}
i &= 0.1gh^{-1} \\
d &= 0.1 h^{-1}
\end{align*}
\[ m = 0.4 \text{ h}^{-1} \]
\[ k = 2 \times 10^{-3} \text{ g} \]
\[ y = 0.5 \]

where \( g \) = unit of mass (gram)
\( h \) = unit of time (hour)

The analytic values for \( a \) and \( b \) are, following Waldon, \(-400\) and \(200\), following the method given in the appendix, \(-2\) and \(1\), and following Schaffer, \(-6\) and \(3\). Values that give least squares fit depend on the initial resource and population sizes. Five cases of system 2 are given in Table 1. Fig. 1 illustrates the actual population growth, the fitted logistic approximation for case 2.1, and the analytic logistic approximations. Fig. 2 illustrates the resource level corresponding to the actual population growth. Clearly, Waldon's method is unsatisfactory (his conditions do not obtain) and the method given in the appendix does not fit well when the population is far from equilibrium.

Table 1.
\[ i = 10g \text{h}^{-1} \], \[ d = 1h^{-1} \], \[ m = 4h^{-1} \], \[ k = 2 \times 10^{-3} \text{ g} \],
\[ y = 0.5 \]. Time interval of simulation 0-25 h.

<table>
<thead>
<tr>
<th>Case</th>
<th>Initial values</th>
<th>Parameter estimates</th>
<th>Sample SD of estimates</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R_0 )</td>
<td>( N_0 )</td>
<td>( \hat{b} )</td>
<td>( \hat{a} )</td>
</tr>
<tr>
<td>2.1</td>
<td>.1</td>
<td>.005</td>
<td>.3449</td>
<td>-.6957</td>
</tr>
<tr>
<td>2.2</td>
<td>.1</td>
<td>.05</td>
<td>.3825</td>
<td>-.8633</td>
</tr>
<tr>
<td>2.3</td>
<td>1</td>
<td>.01</td>
<td>.3589</td>
<td>-.6588</td>
</tr>
<tr>
<td>2.4</td>
<td>1</td>
<td>.05</td>
<td>.4174</td>
<td>-.7851</td>
</tr>
<tr>
<td>2.5</td>
<td>1</td>
<td>.1</td>
<td>.4779</td>
<td>-.8886</td>
</tr>
</tbody>
</table>
System 3: Population with renewed resource and linear uptake rate

Consider another system, similar to system 2, but having a linear rate of uptake of resource:

\[
\dot{R} = i - dR - \lambda R N / y
\]

\[
\dot{N} = - dN + \lambda R N
\]

(15) (16)

where \( R, N, i, d, \lambda, y \) are as before and \( \lambda = 150 \text{ g g}^{-1} \text{h}^{-1} \)

Least squares fitted values for a and b in Eq. 1B for \( R_0 = 0.1 \), \( N_0 = 0.005 \) are 5.924 and -15.55, with sample standard deviations of 1.687 and 4.773, respectively and \( R^2 = 0.3067 \). Fig. 3 illustrates the actual population growth and the fitted logistic approximation. It is clear that the logistic is not a close redevelopment in this case.

From the preceding three example systems I draw the following implications:

a. The logistic equation can be fitted to population growth arising from very different underlying circumstances. The first case was a closed system (batch culture) and the second an open system (continuous culture). (Other resource utilization models and growth processes can generate a population growth curve which is redescribed well by the logistic, for example, Schaffer 1981.) Conversely, similar situations, for example, the second and third systems, can generate different goodness of fit.

b. Although the least squares fit for each of the cases for system 2 is good, the fitted parameters vary significantly. Furthermore, the actual population growth in case 2.5 is qualitatively different from the
other cases. There is a marked overshoot of the equilibrium, shown in Fig. 4. Since I have omniscience I know that this is a consequence of the different initial values of R and N. To someone observing only the trajectory of N that cause would not be obvious. They might attribute the effect to the elevated initial value of N. This explanation can be easily refuted. (Case 2.1 passes through N = .1 when R = .47. If we take these as initial values the trajectory of N will not overshoot.) Alternatively they might attribute the effect to a time lag in response of \( \dot{N} \) to N, an explanation I consider in Sect.III.

c. We cannot, therefore, use goodness-of-fit as evidence to conclude that we have captured the biological relations that generated the observations. Such a conclusion requires further elements of confirmation (see Sect.IV.3).

II. EXTRAPOLATION, USING A GOOD FIT

If we have achieved a good fit between the actual observations and predictions derived from the mathematical dynamics we may hope to use the dynamics in other situations—to predict ahead in time or extrapolate to other circumstances.

For the three systems in the previous section we can predict ahead in time successfully because both the full system and the logistic redescription approach equilibria. In the second and third systems, the discrepancies between the actual and logistic equilibria would become more obvious, that is, easier to distinguish from random error, as time progressed (see Table 2).
Table 2. Actual equilibria and equilibria predicted from logistic redemptions.

<table>
<thead>
<tr>
<th>System</th>
<th>Equilibrium values (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual</td>
</tr>
<tr>
<td>2</td>
<td>.4996</td>
</tr>
<tr>
<td>3</td>
<td>.4996</td>
</tr>
</tbody>
</table>

In the first system, the prediction ahead in time would break down if there were any death term in Eq. 4 because N would decay to zero.

In contrast to predictions ahead in time, we cannot automatically extrapolate to other circumstances. Different initial values require us to use different parameter estimates (see Table 1), and, even then, this is not sufficient, sometimes, to capture the behavior (e.g., case 2.5). Extrapolation will work if we reproduce, perhaps without knowing, conditions similar to the original. It may be, for example, that we never, in practice, choose initial values as extreme as those of case 2.5 and the anomalous overshoot of case 2.5 does not arise.

Extrapolation further, to circumstances in which the original system is embedded in a larger system, is more problematic because the circumstances are not "similar to the original." Nevertheless, accurate extrapolation is not impossible. Vandermeer (1969) fitted the logistic to each of four protozoan populations growing on identical bacterial resources. Pairwise cultures of these populations were fitted well by adding interaction terms without altering the original logistic parameter estimates. Such extrapolations are gratifying, but, again, they are not evidence that the model represents the generative biological relations (this dissertation, Chap. 3).
III. IMPROVING ON AN INADEQUATE REDESCRIPTION

We might interpret an inadequate redescription as a first approximation, requiring us to add extra terms to the model in order to improve the fit.

The logistic does not fit the overshoot or hump in case 2.5, but the hump can be fitted if we add a time lag to Eq. 1B:

\[ \frac{\dot{N}}{N(t)} = b + aN(t-\tau) \]  

where \( t = \) time

and \( \tau = \) time lag.

Testing a variety of time lags I derived a much improved fit with a lag of \( \tau = 1.4h \), parameter estimates \( b = .4500 \), \( a = -.8626 \), sample standard deviations .0083 and .0173, and \( R^2 = .9139 \). The actual population trajectory, the logistic description and the lagged logistic redescription are shown in Fig. 4.

The Θ-logistic (Pomerantz, Thomas and Gilpin 1980) is another elaboration of the logistic:

\[ \frac{\dot{N}}{N} = b + aN^\Theta \]  

The fit is improved, sometimes markedly, with the addition of the parameter. Similarly, the addition of a quadratic term improves the fit:

\[ \frac{\dot{N}}{N} = b + aN + cN^2 \]  

Eqs. 18 and 19 cannot capture the hump in case 2.5 because the hump requires that \( \frac{\dot{N}}{N} \) be able to take on two different values for the same value of \( N \).

Another elaboration of the logistic is to make the parameters time-dependent, that is, for \( b \) substitute \( b(t) \) and so on. With this alteration the hump, and, in theory, any growth curve could be fitted.
In general, the most appropriate way to add terms is not obvious. The addition of a time lag can be given a biological interpretation more readily than the $\theta$ parameter or the quadratic term. With omniscience, however, I know that there is no time lag in the actual dynamics of system 2. The ease of making a biological interpretation is, therefore, no guarantee that the actual biology corresponds to that interpretation. In the systems of Sect.I the terms we need to add are those specifying the dynamics of the resource. This elaboration may, in other circumstances, be misleading. In some cases there may be time lags in the response of $\dot{N}/N$ to $N$ or variation of biological parameters over time. Furthermore, when there is a multiplicity of influences on the actual growth, it may be more revealing to explore those influences than to postulate a unitary "resource" and suppose that the consumption of that resource generates logistic population growth. The use of the logistic for human population growth, for example, provoked a dispute along these lines (Kingsland 1982).

In summary, the addition of terms will improve the fit of model to observations, sometimes significantly. It may improve the quality of extrapolation and predictions based on the redescription. Improving fit does not, however, guarantee convergence to a better explanation of the observations. To be sure of representing the generative biological relations requires a stronger relationship between the model and the observations than any redescription of the observations can provide.
IV. REPRESENTATION OF GENERATIVE RELATIONS

Most modelers want their model to be more than a well-fitting redescription. They want to explain why the observations occurred. This requires that the model represents the biological relations that generated past observations and will, in turn, generate future observations. I will examine four ways we might support the claim that the logistic equation is more than a redescription and actually does represent the generative relations.

1. Truncation of a Taylor's series expansion

Suppose the actual dynamics of the population growth in some real ecosystem were governed by the system of equations

\[
\dot{N} = f(N, X) \quad (20)
\]
\[
\dot{X}_i = g_i(N, X) \quad i = 1, \ldots, m \quad (21)
\]

where \( X \) is the vector of variables \( X_i \) which affect \( N \) and \( f, g_i \) are general but unknown functions.

Modelers commonly claim that they can substitute for these unknown functions a truncation of the Taylor's series around some given point, usually the equilibrium, and that the logistic is such a truncation for \( \dot{N} \) (e.g., Hutchinson 1978:2). I will show that unless the claim is substantially qualified, it is incorrect on two counts, and that even when qualified the truncation method is not very useful.

The truncations to quadratic terms of the Taylor's series expansions of Eqs. 20 and 21 around an equilibrium \( N^*, X^* \) are complex expressions in \( N, X_i \)'s and products of these variables. For example, Eq. 20 can be reduced to
\[ N \approx A + BN + CN^2 + \sum_{i} D_{i} X_{i} + \sum_{i} E_{i} X_{i} + \sum_{i} F_{i} X_{i}^2 \]
\[ + \sum_{i \neq j} G_{ij} X_{i} X_{j} \]  

(22)

where \( A, B, C, D_i, E_i, F_i, G_{ij} \) are constants determined by the equilibrium values and the values of the partial derivatives of \( f \) and \( g_i \)'s at the equilibrium.

The terms in \( X_i \)'s cannot be eliminated unless the partial derivatives of \( f \) with respect to \( X_i \)'s are zero (a trivial case) or the \( X_i \)'s operate on time scales sufficiently different from \( N \) so that the \( X_i \)'s can enter \( f \) as parameters, that is, the system of equations can be decomposed (Gößer and Seelig 1975). I will use the term "quasi-independent" to describe this relationship of \( N \) to the \( X_i \)'s.

Nevertheless, suppose that such a decomposition is possible and that \( N \) is governed by dynamics which can be expressed as a function of \( N \) alone:

\[ N = f(N) \]

There are still problems with claim that the logistic is a truncated Taylor's series. If we expand the Taylor's series around the trivial equilibrium, \( N^* = 0 \), or around the feasible equilibrium, and then truncate these expressions we derive a simple expression approximating the case when \( N \) is close to the corresponding equilibrium. The relevant truncation around \( N^* \) is

\[ N \approx A + BN + CN^2 \]

(23)

where

\[ A = -N^* f_N(N^*) + 1/2 N^* \frac{d^2 f_N(N^*)}{dN^2} \]

\[ B = f_N(N^*) - N^* \frac{d f_N(N^*)}{dN} \]

\[ C = 1/2 \frac{d^2 f_N(N^*)}{dN^2} \]

and the subscripts denote the partial derivatives.
The problems are quickly evident. At the feasible equilibrium, A will not, in general, be zero and Eq. 23 will not simplify to the logistic. (Lotka (1925b:65) simply drops this term; other authors appear to follow his lead.) At the trivial equilibrium A is zero but B and C need not be positive and negative, respectively. For example the sigmoidal growth curve governed by

\[ \dot{N} = N(b + aN^2) \]  

(24)

where \( b > 0, a < 0 \) gives \( B > 0 \) but \( C = 0 \). These objections can be overcome if \( f(N) \) can be factored out as \( N h(N) \) and \( h(N) \) is truncated to linear terms, \( A + BN \), around the feasible equilibrium.

Suppose now both that a decomposition is possible and that \( f(N) \) can be factored out as \( N h(N) \). Notwithstanding the conditions the truncation method is of limited value. The truncation of a Taylor's series expansion is an approximation applicable near the equilibrium and may break down away from equilibrium. For example, the analysis used to derive Eqs. 11 and 12 uses, in effect, a decomposition of system 2 and a truncation around the feasible equilibrium. Fig. 1 illustrates the poorness of the approximation away from equilibrium. We have to know the actual dynamics, therefore, to know how far away from equilibrium that approximation holds. If we knew the actual dynamics, however, we would not need to reduce the representation to a simple expression in one variable alone.

We can, of course, invoke the Taylor's series truncation to perform the standard analysis of local stability of known systems near equilibrium. According to my arguments we should not, however, consider a truncated Taylor's series to provide a first approximation to the unknown generative relations of a system.
2. **Apparent interactions**

If there is no competing model of the generative relations and if the fit is good, we might treat as unimportant the distinction between redescription and representation of generative relations. We could define self-damping to be that relation which yields negative values for \( a \) in the logistic. It would not matter whether the self-damping results from crowding, exploitation of a common resource or something else. Similarly a positive \( b \) would denote, by definition, the "intrinsic growth potential" of the population in the absence of self-damping. Departures from fit must then be "noise," that is, variation of no biological significance.

We have, however, established the fit only for a particular circumstance(s); these "apparent" causes (and conclusions, Chap. 3) are certain only in this circumstance(s). The five cases of system 2 illustrate that the range of circumstances in which any particular fit holds is narrow, unless we are prepared to attribute most of the deviation from fit to noise. In order to treat redescription as equivalent to generative relations, we need to be explicit about the range of conditions in which we intend the model to represent the generative relations and then accumulate evidence that it fits in a variety of circumstances drawn from this range. Typically, however, for an explicit, qualified use of the model, modelers substitute an unqualified exposition. The implicit conditions become forgotten and the apparent interactions take on a generative power independent of their context.
3. **Confirmation**

In order to confirm that a model represents generative relations we must show that two criteria are fulfilled (Lloyd 1984). The first is that the model fits the observations, or, more precisely, the model with its peculiar features fits better than the corresponding model without those features. The second criterion is that the assumptions of the model—the accessory conditions—have been established independently of the goodness-of-fit.

In the case of the logistic the accessory conditions include, for example, the condition that "[a]ll organisms with respect to their impact on the environment or on each other are identical through time" (Williams 1972). When the logistic is used to fit heterogeneous populations, for example, fisheries or humans, the assumption is patently false. We may argue that the falsity of an accessory condition is (provisionally) unimportant but we should not base this argument on the fact that the model fits. In the absence of an alternative argument the model represents no more than apparent interactions, strictly applicable only in the range of circumstances in which the model has been shown to fit.

If we find circumstances in which the model no longer fits it may be that the peculiar feature of the model needs modification. On the other hand it may be that the accessory conditions no longer hold (or their falsity has become important). Lack of confirmation may motivate us to search for alternative models, the subject of Sect.VII on de/composition.
4. **Controlled, isolated experiments**

Biologists sometimes restrict their attention to controlled experiments on isolated systems of a few components only. In such cases the translation from redescription to representation of generative relations is easier. We can explain this fact in the terms of each of the preceding three sections: 1. Most of the $X_i$ terms have been held constant so a Taylor's series truncation is plausible. 2. Most of the natural variability has been removed and conditions are carefully replicated so that the experimental data are more likely to remain in the range of conditions in which the distinction between redescription (apparent interactions) and generative representation can be overlooked. 3. Most of the assumptions which need to be established independently of fit have either been achieved by experimental control (e.g. elimination of hidden variables) or their falsity may plausibly be considered unimportant. Confirmation of the model then requires simply that the peculiar feature improves fit.

In the case of the logistic, however, when per-capita growth rates have been compared with population density for controlled, experimental systems, typically there were systematic departures from linearity (Smith 1952, Williams 1972). Williams (1972) took this result as grounds for breaking down the accessory conditions that were patently false into ones with more detail. By successive replacement of one model by another he derived a very plausible generative model for microbial population growth. This model incorporated both explicit reference to the nutrient supply and utilization and partitioning of biomass into separate cells.
We should not readily attach ecological significance to a model that fits an isolated, controlled experimental system well. It is true that if the experiments expose previously unknown factors then we have increased our understanding of the larger system from which the few components were originally isolated. We cannot, however, use the model to represent the relations among the components when they are back in that larger system. Such a re-insertion generates the accessory condition that the non-modeled components do not confound the relationships in the simple model. This is, in general, true only in the special circumstances of quasi-independence.

In summary, we can consider the logistic equation to represent generative relations only in a provisional and qualified sense. Depending on false or unconfirmable assumptions, the logistic is subject to replacement by a model with assumptions that can be independently established. If we allow redescription and representation of generative relations to be considered equivalent, the logistic captures "apparent" interactions only in the variety of circumstances in which a particular fit has been established. In isolated, controlled experimental systems the unconfirmed assumptions may be fewer and unimportant. Even in these circumstances, however, the lack of fit of the logistic is sufficient to motivate us to search for a better model of causal relations.
V. AN AID TO EXPLORATION

Modelers often use models in an approach different both from fitting the observations well or representing the actual generative relations. The models are used for exploration—to know the mathematical consequences of the model system under consideration. What is the model's range of behaviors, for example, its stability, oscillations and parameter sensitivity? Using a model in this way we can see how to ask questions, what terms and concepts to employ, how we might build more elaborate or alternative models. This interpretation of models sets up an uneasy tension between the facility of manipulation of the model and the requirement that it be subject eventually to confirmation. We have to make strategic choices about when to break the cycle of theoretical interrogation and to attempt to confirm the model so that we can claim biological insight. Often theoretical investigation is continued without confirmation; the modeler hopes that nature will oblige by resembling the model. Nevertheless exploration can be an invaluable initial strategy towards understanding biological complexity (this dissertation, Chap. 5).

In the case of the logistic, however, there is little value in exploration. Students can be introduced to elementary differential equations. Beyond this, the exploratory use of the logistic quickly takes us past the appropriate point for shifting to confirmation and then, finding lack of confirmation, to search for a new model. I will use two examples to support this conclusion:
1. **r,K and time lags**

Suppose we want to explore the consequences of a model which includes a term for per capita increase at low values of N, "r," and a reduction of this increase in proportion to the ratio of N to a second term, "K," that is, the logistic in the form:

\[
\dot{N}/N = r(1-N/K)
\]  

(25)

One consequence is that if \( N_0 < K \) then \( N \) can never exceed \( K \). The overshoot of case 2.5 is impossible. We could add time lags or discrete time steps to make an overshoot possible. With omniscience, I know, however, that the actual situation in system 2 is continuous, and there are no time lags. Instead of pursuing an exploration of the r,K formulation by adding details to it, we should consider a range of different models including, for example, a two-component model. If we attempted empirical confirmation we could eliminate many from the range to be explored, and, in particular, would probably eliminate all one-component models.

2. **Levins's paradox**

Levins (quoted in Hutchinson 1978) presented a paradox implicit in the r,K formulation of the logistic equation. Suppose we have a population with a positive "carrying capacity," \( K \), and a negative intrinsic rate of increase, \( r \), that is, in the absence of other resources the death rate exceeds the birth rate. While we would expect a population with negative \( r \) to decline, in fact, if the initial value, \( N_0 \), is greater than \( K \) the model predicts that the population will grow super-exponentially. (This circumstance might eventuate, for example, if the physical conditions changed, lowering the carrying capacity \( K \) to
below the current value of \( N \) and increasing the death rate.) Hutchinson resolved this paradox by adding the postulate that (intrinsic) birth rate must exceed death rate (Hutchinson 1978:4), thus ensuring that the population declines when \( N_0 > K \).

The paradox, however, does not exist. A logistic model with those values of \( r, K \) and \( N_0 \) will give explosive growth. While a real population could have death rate exceeding birth rate, it would never have those values over any appreciable length of time. The two-component system 2 illustrates the general case for real populations:

\[
\begin{align*}
birth &= mR/k + R \\
death &= -d
\end{align*}
\]  \hspace{1cm} (26)

where\( birth = \) per capita birth rate of \( N \)
\( death = \) per capita death rate of \( N \)

Now \( r = birth - death < 0 \)
\( \Rightarrow mR/(k + R) < d \)
\( \Rightarrow R(m - d) < kd \)  \hspace{1cm} (28)

There are two circumstances in which \( r < 0 \):

a) \( m \leq d \)

b) \( m > d \) and \( R < kd/(m - d) = R^* \)

In the first case \( N \) declines to zero [corresponding to a "washout" in a chemostat (Williams 1972:415)] since \( \dot{N} < 0 \) for all positive values of \( R \) and \( N \) (from Eq. 8). In the second case \( N \) declines and \( R \) increases. Either \( N \) and \( R \) converge directly to the feasible equilibrium or \( N \) declines to below \( N^* \) and \( R \) grows above \( R^* \), in which case \( r \) is no longer negative.

In the earlier sections I have presented many reasons why the logistic is limited in application. Levins (pers. comm.) also intended
to indicate a limitation of the logistic. Nevertheless, both Levins's paradox and Hutchinson's addition of a biological postulate to resolve that paradox rest on not distinguishing between exploration of a model system and using the model as representation of biological relations. In the following section I discuss that conflation further.

VI. THE MACARTHUR-LEVINS-HUTCHINSON STRATEGY OF MODEL BUILDING

Hutchinson's (1978) strategy of model building is much the same as that advocated by MacArthur (1972) and Levins (1966). Briefly stated--I will elaborate as this section proceeds--simple models are proposed; when major discrepancies are shown to occur between their predictions and observations additional detail is brought into the model. Much of theoretical community ecology is based on this strategy (Kingsland 1985, chap. 8). That Hutchinson perceives and attempts to resolve a non-existent paradox demands, therefore, our attention. I will use my three interpretations to characterize the confusion embodied in Hutchinson's treatment of the logistic and suggest that the same confusion is implicit in much of the use of simple models in community ecology.

1. Addition of "biological" postulates to simple models

Hutchinson called the additional postulate required to resolve Levins' paradox a "biological" postulate (Hutchinson 1978:5). Following my interpretations, Hutchinson is conflating the model as an aid to exploration with the model as capturing generative biological relations.
In contrast to Hutchinson, I interpret the additional postulate as a mathematical postulate, subject to (dis)confirmation if we switch from exploration to attempted biological explanation. In Sect.IV.3, however, I argued that the logistic itself cannot be tightly confirmed. That being so, the logistic reduces to a redescription, or, if we specify the range of conditions in which a redescription holds, we might take the logistic to be a representation of "apparent" interactions. In either case, since the logistic cannot be confirmed, it will not add any biological insight to confirm the logistic with the additional postulate that \( r(=b) > 0 \). Nevertheless, suppose that we do attempt to do so. If we examine what might happen we will derive an important insight about strategy in model building.

Imagine that we survey a variety of circumstances in which a particular population grows from a low level to an equilibrium. The initial per capita rate of change \( \hat{N}_0/N_0 \) can serve as an approximation to \( r \), and be measured independently of fitting the entire set of observations to the logistic. It is possible that in all circumstances \( r \) will be positive and Hutchinson's postulate will appear to be confirmed. This confirmation is, however, problematic in two ways. Using system 2 it is possible to generate circumstances in which \( \hat{N}_0/N_0 \) is negative yet the population eventually grows to an equilibrium (case b in Sect.V.2). The necessary initial conditions are unusual for actual chemostat experiments. Usually, for example, \( R_0 \gg R^* \). Nevertheless, they are possible to generate. In addition, a decline to zero can be generated for system 2. If we restrict our attention to circumstances in which a positive equilibrium is attained we will consider such potentially disconfirming circumstances to be out of
bounds. A lesson can be drawn from these two possibilities: tight confirmation requires that we do not overlook unusual circumstances. We might conclude that discrepancies of the model and the observations in such circumstances mean that some additional biological feature is present in those circumstances while absent or negligible in the usual ones. There are alternative conclusions, however. We might, instead, conclude that the logistic is a redescription and not a representation of generative biological relations. In that case these unusual circumstances could stimulate us to search for a confirmable model covering both the usual and the unusual circumstances.

2. MacArthurian simple models

The distinctions I have drawn in discussing Hutchinson's use of the logistic can be extended to MacArthur's strategy of model building in general (MacArthur 1972, Kingsland 1985: Chap. 8). I will characterize this strategy and present my objections and reinterpretation.

For MacArthur, complexity in ecology need not be treated as "a tedious set of case histories" (MacArthur 1972:169). Complexity can be brought to a level where we can see pattern and unity among disparate processes. We achieve this by relating mathematical models to ecological observations. The models must be simple, that is, incorporate only a few components and be expressed in terms of general processes and concepts. These simple models are analogous to the perfect crystals or ideal gases of physics and chemistry. Significant departures of nature from the model are cause for us to invoke additional biological postulates. (Significance depends on the relative value we place on eliminating discrepancies versus simplicity or manageability of the
model.) In this way we can generate interesting questions in ecology. Even if the "perfect crystal" is subsequently abandoned as being biologically inapplicable or irrelevant, the model will have been useful if it has provoked ecological investigation and clarification of concepts.

Hutchinson's (1978, chap. 1) presentation of the logistic equation exemplifies this strategy. The logistic has been fitted to many observed population growth curves (e.g. Allee et al. 1949) and it appears to unify many disparate biological processes. The model is simple and is expressed in terms of the general processes of growth and a limit to growth. And so on.

My objection to MacArthur's strategy is that simple models are not actually simple nor general when applied to biological processes. As a mathematical expression a model may be simple and this simplicity facilitates our exploration of the model as a mathematical expression (Sect. V) or our use of the model to redescribe observations (Sects. I-III). Nevertheless, when we interpret the mathematical terms as representing biological processes then, either the simplicity is deceptive (Sect. IV.1 and 2), we have to restrict our attention to systems showing quasi-independence (Sect. IV.4), or the simplicity dissolves into complex accessory conditions (Sect. IV.3).

Accessory conditions are not always apparent. They include, for example, the rarely acknowledged conditions required so that we can capture the circumstances in which the components excluded from the simple model--the "hidden variables"--do not enter into relationships which confound those postulated and incorporated into that simple model. In order to characterize such circumstances of quasi-independence we
must, in general, explicitly refer to the larger system. We may have to restrict the domain of applicability of the model so far that we must abandon claims for unity and generality, or at least heavily qualify them. For example, in system 2 the range of circumstances in which any given logistic (that is, the logistic form instantiated with particular parameter values) holds is narrow. That range is best characterized by including reference to the initial value of the hidden variable R. In natural ecosystems, more complex than my systems, these limitations will be even more stringent.

There is a pragmatic justification for using MacArthurian simple models. They have been invaluable in stimulating ecologists to generate and investigate hypotheses about ecology. At first sight, this fact might seem to defuse my objection that simple models are both difficult to confirm and, when confirmation is attempted, have a limited domain of applicability, that is, they are not general. Employing the distinctions I have made in this paper, however, my response is that the stimulation a simple model provides is a result of the model being used as an aid to exploration or, even more simply, as a "schema" (this dissertation, Chap. 1)—holding in juxtaposition for our attention a particular set of biological processes. Nevertheless, modelers do not confine their use of simple models to exploration or schematizing. When we express the model in mathematical terms and use it to generate predictions to be compared with biological observations, we are granting the model a generative power. For the comparison of mathematical outcome and biological observations to have any biological meaning the generative status of the model must be confirmed.
VII. DE/COMPOSITION

In the preceding section I maintained that a model must be confirmed for it to have biological meaning. Nevertheless unconfirmed models have roles in the development of ecological theory. In this concluding section I will discuss these roles. I have examined three interpretations of the logistic equation and, by extension, mathematical models in ecology. A model can be considered to be a redescription of the observed phenomenon (Sects.I-III), a representation of the generative biological relations (Sect.IV), or an aid to exploration of the consequences of a particular kind of mathematical formulation (Sect.V). Each interpretation stimulates its own questions. As a redescription, how far can we use the model to predict ahead in time or to extrapolate to other conditions (Sect.II)? How valuable is it to elaborate the model to derive a closer fit (Sect.III)? As a representation of generative relations, how well confirmed is the model? In what range of conditions do we intend the model to represent these generative relations? Have we established a good fit in a representative sample of circumstances drawn from this range? Do we have evidence for the accessory conditions; evidence derived independently from the fit? Is the range of conditions broad enough to generalize the generative interpretation into a principle of ecology? Is the model as a representation of generative relations context dependent or local? How provisional is its status as a generative representation? These were the questions of Sect.IV. Finally, as a tool for exploration, when should we break the cycle of investigating the
model as model to attempt to confirm the model and thereby claim biological insights from the model (Sects. V and VI)?

The interpretations are different and stimulate different questions. Nevertheless, there is a basis for unity among them. Each has a role in an overall strategy of theory development, which I call "de/composition." We start from observed phenomena in ecological contexts. De/composition emphasizes the dual movements of breaking down the initial categories with which we describe a complex phenomenon and reconstituting new ones with which we move closer to explaining the phenomenon in its rich web of interrelations and conditions. At any stage in this process a model can be given a particular status and a prescribed domain of applicability. We may be satisfied with this status and seek no further de/composition. In this case it is important that the interpretation and domain of applicability are made explicit, both to maintain clarity in ecological theory and to enable other modelers to assess the need for de/composition, that is, whether the model is satisfactory for their purposes.

In the case of the logistic the parameters implicitly incorporate the resources. An obvious de/composition is to unpack these parameters and refer explicitly to the resources and their dynamics. For each of the hypothetical systems of Sect.I this de/composition would allow us to account for all the behavior of the system. For real systems having only a consumer and a food resource we may have to de/compose further to refer to the sub-components of the food resource, to accommodate spatial distribution, and so on. But let me discuss how we might go about de/composition in terms more general than the specific case of the logistic.
Consider redescriptions and generative representations as they apply to de/composition. Redescriptions hold in the circumstances in which the model fits the observations and generative representations hold in the range of conditions for which confirmation has been established. Indications of how to de-compose may take the form of discrepancies or anomalies exposed when we attempt to stretch the domain of applicability. For example, a redescription used to make predictions into the future or extrapolations into different circumstances and may not fit the new observations well. Similarly, a generative representation may not be well confirmed in new circumstances. The close fit of a model to data from a controlled experiment may break down when the controls become variables. Finally, when we attempt to change the status of the model from redescription to generative representation the falsity of any accessory conditions may indicate how we should redescribe the phenomena so that the new model may eventually be confirmed.

Indications for de/composition may come, in addition, from comparisons of alternative models. The refinement of fit achieved by adding parameters to a model may suggest new considerations. These may not always lead to a confirmed model. (The case of the time lag added to the logistic for case 2.5 shows this; Sect.III.) Similarly when accessory conditions cannot be well specified or tightly established we should construct a range of models and compare their degree of confirmation.

Consider now how models in their exploratory role can contribute to de/composition. The models have no domain of biological applicability until confirmed or used as redescriptions. They can, however, be a
source of predictions about circumstances in which anomalies are most likely to appear and a source of ideas about alternative models. Occasionally they even serve to overturn our framework for conceptualizing the phenomena (this dissertation, Chap. 5).

In summary, de/composition focuses on the provisional character of any explanatory account, whether this account is expressed through a model, a metaphor or a plausible story. We dispute the correspondence of explanatory categories and processes to natural ones and either replace them or analyze them into greater detail. The use of mathematical models is one approach to de/composition; their role varies according to the interpretation being given to the model.

As ecological modelers, we should be receptive to a sobering possibility. Ecological phenomena have too many locally and historically contingent conditions; even with imaginative de/composition we may not be able to develop generative representations, let alone general laws. For this reason it is particularly important that we hold in mind the alternative interpretations of mathematical models and do not slip from exploration or redescription to premature claims of generative representation or biological insight. If we can only describe, make short-term predictions or derive locally specific explanations then this may be the constraints of our subject matter. We should resist the temptation to compare ourselves with our mathematical modelers in experimental and reductionist sciences. Rather we should patiently explore the challenges and limitations of a field for which diversity and complexity constitute its inherent appeal.
To most field ecologists it has always been obvious that there are far more interesting stories to tell than those captured in simple mathematical models. What was not so obvious was that one of the simplest of growth models—the logistic—could be an exemplar of problems in the use of mathematical models and, more generally, of theory development in ecology. The interpretations I have advanced, the questions stimulated by these interpretations, and the conclusions about de/composition should stimulate us as mathematical modelers to articulate more clearly our strategies for the use of mathematical models in unpacking ecological relationships.

Acknowledgments: In the course of writing this paper I have benefited from discussions with David Glaser, Lisa Lloyd, Sharon Kingsland, Steven Strogatz, Charles Puccia and Richard Lewins.
Literature Cited


Appendix

The feasible equilibrium for Eqs. 7 and 8 is

\[ R^* = \frac{k \mu}{m - d} \]  \hspace{1cm} (A.1)

\[ N^* = \left( \frac{1}{R^*} - d \right) \frac{y}{m} (k+R^*) \]  \hspace{1cm} (A.2)

Linearizing around this equilibrium gives the \( \alpha \)-matrix

\[
\begin{bmatrix}
-d - \frac{mkN^*}{y(k+R^*)^2} & -\frac{mR^*}{y(k+R^*)} \\
\frac{mkN^*}{(k+R^*)^2} & 0
\end{bmatrix}
\]

The eigenvalues of \( \alpha \) are

\[ \lambda = \frac{1}{2} \left[ \alpha_{RR} \pm \left( \alpha_{RR}^2 + 4\alpha_{RN} \alpha_{NR} \right)^{1/2} \right] \]  \hspace{1cm} (A.3)

If \( \alpha_{RR}^2 >> -4\alpha_{RN} \alpha_{NR} \)

then \( \lambda_1 = \frac{-\alpha_{RN} \alpha_{NR}}{\alpha_{RR}} \) or \( \lambda_2 = \alpha_{RR} - \frac{\alpha_{RN} \alpha_{NR}}{\alpha_{RR}} \)  \hspace{1cm} (A.5)

Now if \( 4d^2 << \frac{i(m-d)^2}{km} \)  \hspace{1cm} (A.6)

then also \( d^2 < 4d^2 << \frac{i(m-d)}{k} \)  \hspace{1cm} (A.7)

Eq. A.7 is equivalent to \( \frac{i}{R^*} - d \approx \frac{i}{R^*} \)  \hspace{1cm} (A.8)

Substituting this simplified expression into A.2 we obtain

\[ N^* \approx \frac{iy}{d} \]  \hspace{1cm} (A.9)

and, in turn,

\[ \alpha_{RR} \approx -\frac{i(m-d)^2}{dkm} \]  \hspace{1cm} (A.10)
and

\[ \alpha_{RN} \alpha_{NR} \approx \frac{-i(m-d)^2}{\text{km}} \]  

(A.11)

The condition A.4 becomes

\[ \frac{i(m-d)^2}{d^2 \text{km}} \gg 4 \]

which is satisfied if A.6 is true.

Since \( \alpha_{RR} < 0 \), \( \lambda_1 \) is the dominant eigenvalue and simplifies to \( \lambda_1 = -d \).

A logistic equation with equilibrium \( N^* \) and eigenvalue \(-d\) is

\[ \dot{N} = N(d - \frac{dN}{N^*}) \]

\[ = N(d - iyN) \]

Note also that condition A.6 is satisfied by the parameters used for "ecosystem" 2.
Figure 1. Population growth in system 2, case 1. □ Actual trajectory; *
least squares fitted logistic approximation. Analytic logistic approximates: , Waldon; + Appendix; − Schaffer.
Figure 2. Resource utilization in system 2, case 1.
Figure 3. Population growth in system 3. □ Actual trajectory; * least squares fitted logistic approximation.
Figure 4. Population growth in system 2, case 5. □ Actual trajectory; * least squares fitted logistic approximation; + lagged logistic approximation. The initial value for the lagged variable is .057 which ensures that it passes through 0.1 at time t = 1.4h.