

Stage-structured matrix models for organisms with non-geometric development times

ANDREW BIRT,¹ RICHARD M. FELDMAN,^{2,5} DAVID M. CAIRNS,³ ROBERT N. COULSON,¹ MARIA TCHAKERIAN,¹
WEIMIN XI,¹ AND JAMES M. GULDIN⁴

¹*Department of Entomology, Texas A&M University, College Station, Texas 77843 USA*

²*Department of Industrial and Systems Engineering, Texas A&M University, College Station, Texas 77843 USA*

³*Department of Geography, Texas A&M University, College Station, Texas 77843 USA*

⁴*Southern Research Station, USDA Forest Service, Hot Springs, Arkansas 71901 USA*

Abstract. Matrix models have been used to model population growth of organisms for many decades. They are popular because of both their conceptual simplicity and their computational efficiency. For some types of organisms they are relatively accurate in predicting population growth; however, for others the matrix approach does not adequately model growth rate. One of the reasons for the lack of accuracy is that most matrix-based models implicitly assume a specific degree of variability in development times for the organism. Because the variability is implicit, the implied variances are often not verified with experimental data. In this paper, we shall present extensions to the stage-classified matrix models so that organisms with arbitrary means and standard deviations of development times can be modeled.

Key words: development time; matrix models; population models.

INTRODUCTION

The time taken for an organism to reach maturity is a fundamental component of its life history and of interest to both practical and theoretical ecologists. Studies have shown development rates of organisms are affected by a range of environmental factors including food, moisture and nutrient availability (e.g., Fischer and Fiedler 2000), pollutants (Michaud and Grant 2003, Wang et al. 2004), inter- (Teng and Apperson 2000) and intra- (Haukioja et al. 1988, Schneider et al. 2000, Ruohomaki et al. 2003) species competition and for poikilotherms especially, temperature (Davidson 1944, Taylor 1981, Pipe and Walker 1987). In turn, ecologists have attempted to assess the practical and theoretical consequences of maturation times. For example, phenological population models have been a central tenet for insect pest management and sustainable resource management programs. Here the primary interest is how environmentally driven changes in maturation time affect the population dynamics of a species. Confronted with the same phenomenon, theoretical ecologists have studied plasticity in development traits (and the associated trade-offs between other life-history components) in an attempt to explain differences in the vital rates of organisms inhabiting different habitats (e.g., Reznick and Endler 1982, Reznick et al. 1990, Blanckenhorn 1998). In both cases, mathematical models play a critical role in evaluating the ecological significance of develop-

ment rates by placing them into more relevant population level effects.

Despite general interest in the development or growth rates of organisms, most work has involved measurement and interpretation of the “average” response of an individual within a population rather than variations in the responses of individuals that make up the population (intrapopulation variation [Peacor et al. 2007]). Intrapopulation variation in measured development rates (or any other life-history parameters) may arise from (1) sampling errors (which we shall ignore in this discussion), (2) phenotypic variation, and (3) genetic variation (Van Tienderen 1995). The latter two mechanisms have been demonstrated and quantified in a number of studies. For example, models of southern pine beetle (*Dendroctonus frontalis*) development times can be explained by a combination of temperature and intrinsic variability in the development rates of individuals (Wagner et al. 1985). Liu and Meng (1999) used a similar model to simulate development in an aphid (*Myzus persicae*) and concluded that the distribution of development times at constant temperatures could be used to predict the distribution of development times under fluctuating temperature regimes. Both studies suggest innate variations in the growth rates of cohorts of individuals experiencing identical environments. Other studies have demonstrated environmentally driven variation in individual development times. For example, Ryer and Olla (1995) measured a greater variation in the development rates of salmon cohorts subjected to competition for food, while Twombly (1993) detected both genetic and environmentally driven

Manuscript received 21 April 2008; accepted 30 April 2008.
Corresponding Editor: L. Stone.

⁵ Corresponding author. E-mail: richf@tamu.edu

intrapopulation variation in the development rates of a copepod. From a modeling point of view it can be concluded that evaluating the population level effects of development rate variability is driven by two factors. The first is an appreciation of the innate and environmentally driven variability in individual development times. This requires a strong definition of the focus population including an appreciation of implied versus explicit handling of environmental variation. Second, mathematical tools are required that allow accurate representation of development time and intrapopulation level variability and allow them to be translated into more meaningful ecological outputs.

Matrix models were first used to model populations by Lewis (1942) and Leslie (1945, 1948). Since the initial efforts at using matrices to model populations, there have been many extensions in an effort to improve the matrix method and insure their applicability for a variety of organisms. In this paper, we are particularly interested in the stage-classified matrix model (see Caswell [2001] for a comprehensive review). The appeal of this approach is that it allows the life cycle of an organism be divided into discrete stages. Such divisions are natural for a wide variety of organisms and essentially allow different life-history parameters to be attributed to each stage. For example, a model of insect populations might be delineated by morphological stages. For other organisms life history parameters might be correlated with size or weight.

A major difficulty with the standard matrix population modeling approach is the assumption that the organism spends a geometrically distributed (random) time within each stage; whereas, insects rarely have geometrically distributed development times. Caswell (2001) provides two extensions to the stage-classified matrix model to overcome the restriction on development times; namely, a variable stage duration model and a negative binomial stage duration model. The variable stage duration model provides characteristics for the limiting distribution if the population reaches an equilibrium age distribution and is not well suited to obtain characteristics of a population not at steady state. The negative binomial stage duration model is ideal if the ratio of the mean squared to the sum of the variance plus the mean is approximately equal to a constant (see Caswell 2001: Eq. 6.118). The goal of this paper is to generalize the negative binomial case so that organisms with general mean and variance of development times can be modeled. For those familiar with Markov processes, this paper is analogous to generalizing Erlang models to phase-type models. Note that the Erlang distribution is the sum of exponentials while the negative binomial is the sum of geometrics. The usual formation of phase-type models is the extension of Erlang models to include arbitrary means and variance while we propose a discrete version of the phase-type models as the extension of the negative binomial. Of course to

create a population model, we shall also include fecundity terms to the phase-type structure.

The remainder of this paper is organized as follows. The next section presents Caswell's matrix model and gives estimates for the matrix parameters correcting an error from Caswell's text regarding parameter estimation, then we give a brief overview of discrete phase-type distributions that will be used to overcome the implicit assumption of geometrically distributed development times. Next, we present the phase-type population model, give a numerical example, and finally discuss of the importance of including variances as well as means for the description of life stages. The Appendix presents the mathematics necessary to demonstrate an error contained in Caswell's (2001: Eqs. 6.97 and 6.98) formula for parameter estimation.

A STAGE-CLASSIFIED MATRIX MODEL

Many plant and animal populations have discrete life stages each associated with different vital rates (survival, fecundity, or development rates). The stage-classified matrix model allows the representation of animal or plant populations according to these distinct morphological or physiological stages. For example, modeling an insect population lends itself naturally to a stage-classified matrix model. In the case of insects, their life histories are defined by distinct instars that drive distinct behaviors and ultimately differences in vital rates. For other populations, stages might be more appropriately delineated by size, physiology, or developmental status.

In this paper, we use an insect population as an example. We will attempt to keep the notation simple by only using three stages: (1) eggs, (2) immatures, and (3) adults. It is a simple matter to generalize from this situation to any number of stages. We shall also only keep track of females under the assumption that survival to the adult stage implies being fertilized. The purpose of the model is to track the expected number of (female) insects over time, and our time scale will be a single day.

The number of females in each stage at time 0 is assumed known and is given by the vector $\mathbf{n}(0) = [n_e(0), n_i(0), n_a(0)]$ denoting the number of eggs, immatures, and adults, respectively, at time 0. The number of females in each stage at time t is denoted by the vector $\mathbf{n}(t) = [n_e(t), n_i(t), n_a(t)]$ and is calculated according to the recursion

$$\mathbf{n}(t+1) = \mathbf{A}\mathbf{n}(t) \quad (1)$$

where

$$\mathbf{A} = \begin{bmatrix} p_{e,e} & 0 & f \\ p_{i,e} & p_{i,i} & 0 \\ 0 & p_{a,i} & p_{a,a} \end{bmatrix}.$$

The p terms represent transition probabilities, in other words, $p_{e,e}$ is the probability that an egg one day will remain as an egg the next day, $p_{i,e}$ is the probability that an egg one day will become an immature the next day, $p_{i,i}$ is the probability that an immature one day will

remain as an immature the next day, $p_{a,i}$ is the probability that an immature one day will become an adult the next day, and $p_{a,a}$ is the probability that an adult one day will remain as an adult the next day. Finally, f denotes the fecundity of adults and is equal to the expected number of female eggs laid each day by an adult. Note that we must always have the following relationships: $p_{e,e} + p_{i,e} < 1$, $p_{i,i} + p_{a,i} < 1$, and $p_{a,a} < 1$. (Notice that these relationships are stated as strict inequalities implying that there will always be some mortality associated with any stage. We also point out that for those who are familiar with Markov models, that both the Leslie matrix and the Caswell matrix approaches use the transpose of the usual Markov matrix formulation. Of course, the fecundity term keeps the matrices from being Markov.)

In this section, our goal is to build a model of a population at a fixed temperature. With this assumption, suppose that experiments have been conducted at a fixed temperature in which development times for a fixed number of eggs are recorded. From these data, the mean time (standard deviation will be discussed in *Phase-type population model*) to develop from egg to the immature insect is obtained and the fraction of survival is determined. Note that survival is implicit within the development times so that daily survival is not directly known but only the fraction of eggs that ultimately survive to immatures is known. Let \bar{t}_e denote the mean time for an insect to develop from the egg stage to an immature given that the insect survives, and let s_e denote the fraction of eggs that eventually reach the immature stage. For example, if we start with a cohort of 1000 eggs and only 650 eggs survive to become immature insects, then \bar{t}_e is the average development time for those 650 eggs and $s_e = 0.65$. Likewise, \bar{t}_i and s_i are the mean development time for the immature to develop into an adult with its survival given by s_i ; and \bar{t}_a is the mean life length for an adult.

Clearly from our model, the time spent in any one stage (denoted by T_k for stage k) is random and is described by a geometric distribution. In particular, the sojourn time of stage k , for $k \in \{e, i, a\}$ is given as

$$P\{T_k = t\} = (1 - p_{k,k}) p_{k,k}^{t-1} \quad \text{for } t = 1, 2, \dots \quad (2)$$

The mean of the distribution is thus given as

$$E[T_k] = \frac{1}{1 - p_{k,k}}$$

which leads to the following parameter estimates:

$$p_{e,e} = 1 - (1/\bar{t}_e) \quad (3)$$

$$p_{i,i} = 1 - (1/\bar{t}_i) \quad (4)$$

$$p_{a,a} = 1 - (1/\bar{t}_a) \quad (5)$$

assuming that all mean development times are greater than 1. When the insect leaves one stage, it will either transition to the next stage or die. The ultimate probability of surviving to the next stage is the ratio of the probability of transition to the next stage divided by

the probability of leaving the stage; thus we have the following:

$$p_{i,e} = s_e/\bar{t}_e \quad (6)$$

$$p_{a,i} = s_i/\bar{t}_i. \quad (7)$$

It should be noted that Eqs. 3–7 are different estimators than given by Caswell (2001: Eqs. 6.97 and 6.98) due to an error in his formula in the middle of page 160. Caswell defines two terms (the survival probability given by σ_k and a development rate given by γ_k) and then defines the transition probability in terms of those terms (namely, he sets $p_{k,k} = \sigma_k(1 - \gamma_k)$). However, Caswell uses only development (i.e., γ_k) in the duration probability formula (equivalent to our Eq. 2) instead of using $\sigma_k(1 - \gamma_k)$; thus incorrectly ignoring the effect of survival on the stage sojourn time. See the Appendix for a more detailed discussion of the importance of not using a survival term within the transition probabilities.

We shall define fecundity, f , to be the expected number of female eggs laid per adult female per day. To give the estimator for f , let the number of female eggs per female be denoted by n_{eggs} . For example, assume that from a starting cohort of 100 (female) adults there are a total of 1800 eggs oviposited over the life of the cohort. Further assume that some insects only lived a day or so, some lived as many as 10 days, but the average life span of the adults was five days. In this example, we would have $n_{\text{eggs}} = 9$ (notice this assumes a 50:50 split in male and female eggs) yielding $f = 1.8$. In other words, the fecundity terms is given by

$$f = n_{\text{eggs}}/\bar{t}_a. \quad (8)$$

PHASE-TYPE DISTRIBUTIONS

Phase-type distributions are usually defined as continuous distributions that are generalizations of Erlang distributions, i.e., the sum of exponential random variables (for a review, see Neuts [1981]). Since the matrix approach for population modeling uses discrete time, we shall define a discrete version of phase-type distributions in order to model discrete development processes that are not geometric. Conceptually, we divide each developmental stage into phases. Note that a “stage” refers to an observable biological period within the life cycle of the insect; whereas, a “phase” refers to a fictitious construct used in modeling the time that an insect spends within a stage. As an example, assume we have collected data and determined that the average time in the immature stage is 10 days with a standard deviation of 6.32 days. If we model the immature stage using one phase, as is assumed in the above section, the probability of remaining in that phase would be set at 0.9 (i.e., $p_{i,i} = 0.9$) so that the mean would be 10. However, since the distribution is geometric, the standard deviation must be 9.49 days (i.e., $\sqrt{p_{i,i}/(1 - p_{i,i})}$) instead of 6.32 days so we know that a single phase yields an incorrect distribution. However, we could model immature development using two phases. Upon

leaving the egg stage, the immature insect enters the first phase and each day there is a probability of 0.8 of remaining in that phase. Upon leaving the first phase, the immature enters the second phase and again there is a probability of 0.8 of remaining in the second phase each day. Upon leaving the second phase, the insect is considered to be in the adult stage. Thus, the time spent in the immature stage is the sum of two geometric distributions yielding a total mean time of 10 days and a standard deviation of 6.32 days. (Note that each phase is geometric having a mean of 5 days and a variance of 20 so that the mean of the total time is the sum of the two individual means and the standard deviation of the total time is the square root of the sum of the two variances.) Thus, using two phases to model the immature stage yields a model consistent with the sample mean and sample variance for this insect. The idea behind the phases is to choose the number of phases and the probabilities associated with the phases such that the mean and standard deviation might match the mean and standard deviation of our sample data; however, since each phase retains the geometric assumption, the general approach of the matrix models can still be used.

To state the above description mathematically, we need an initial probability vector, α , and the matrix

$$\begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \tilde{\mathbf{q}} & 1 \end{bmatrix} \quad (9)$$

where α is a column vector of dimension v of nonnegative numbers whose sums equal 1, \mathbf{Q} is a $v \times v$ nonnegative matrix with each column sum being less than or equal to 1, and $\tilde{\mathbf{q}}$ is a row vector of dimension v where each element is such that the column sum of the full matrix equals 1. (In other words, $\tilde{\mathbf{q}}(j) = 1 - \sum_{i=1}^v \mathbf{Q}(i, j)$ for each $j \in \{1, \dots, v\}$.) The idea of the phase-type distribution is that the submatrix \mathbf{Q} represents a collection of v phases associated with a stage and the final column represents an absorbing state. We think of a process that probabilistically starts in one of the phases according to the probabilities given by α , spends some time moving among the v phases, and eventually gets absorbed in the final state. Thus, a phase-type distribution is defined to be the time that it takes to move into the final absorbing state. Note that we have implicitly assumed that the final state is the only absorbing state in the system so that the matrix \mathbf{Q} must be structured so that all its states are transient.

As an example, consider the two-phase model representing the development with a mean of 10 days and a standard deviation of 6.32 days. This would yield the following:

$$\alpha = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} 0.8 & 0 \\ 0.2 & 0.8 \end{bmatrix}$$

and

$$\tilde{\mathbf{q}} = (0 \quad 0.2).$$

A phase-type distribution is identified by reference to

the ordered pair (α, \mathbf{Q}) . (Note that vector $\tilde{\mathbf{q}}$ is implicitly defined once the matrix is given. We also point out for those familiar with Markov chain terminology, that the matrix \mathbf{Q} is the transpose of the usual (sub) Markov matrix. We use the transpose to be consistent with the historical use of matrix population models.)

It is not too difficult to develop expressions for the mean and standard deviation for a discrete phase-type distribution, and more importantly, these expressions are relatively easy to compute within a computer program or in a spreadsheet program. Before giving the expressions for the mean and variance, we observe that the distribution function for a random variable, T , that has a phase-type distribution defined by (α, \mathbf{Q}) is

$$P\{T = t\} = \tilde{\mathbf{q}} \mathbf{Q}^{t-1} \alpha \quad \text{for } t = 1, 2, \dots.$$

The main difficulty in our computations of the moments is the need to obtain the so-called potential matrix, denoted by \mathbf{R} , which is defined as

$$\mathbf{R} = (\mathbf{I} - \mathbf{Q})^{-1}. \quad (10)$$

Note that the inverse always exists as long as \mathbf{Q} is nonnegative and the column sums are less than or equal to 1 with at least one column having a strict inequality and all states represented by \mathbf{Q} are transient.

Let T be a discrete phase-type random variable defined by (α, \mathbf{Q}) . We now give the definitions of the first three factorial moments of T in terms of α and \mathbf{Q} , and for purposes of completeness, give the relationship of the centralized moments in terms of the factorial moments:

$$E[T] = \tilde{\mathbf{q}} \mathbf{R}^2 \alpha \quad (11)$$

$$E[T(T-1)] = 2\tilde{\mathbf{q}} \mathbf{Q} \mathbf{R}^3 \alpha \quad (12)$$

$$E[T(T-1)(T-2)] = 6\tilde{\mathbf{q}} \mathbf{Q}^2 \mathbf{R}^4 \alpha \quad (13)$$

$$\sigma^2 = E[T(T-1)] - \mu(\mu-1) \quad (14)$$

$$E[(T-\mu)^3] = E[T(T-1)(T-2)] - (\mu-1)(3\sigma^2 + \mu^2 - 2\mu) \quad (15)$$

where $\mu = E[T]$ and \mathbf{R} is the potential matrix associated with \mathbf{Q} from Eq. 10.

The advantage of phase-type distributions is the great flexibility that this class provides. It can also be a source of frustration since there are often multiple parameter sets that yield the same mean and standard deviation; thus, parameter estimation within this class is not straightforward. In general, as the number of phases (i.e., the dimension of the matrix \mathbf{Q}) increases, the coefficient of variation (i.e., the standard deviation divided by the mean) decreases. We provide here some general rules and leave the specific question of the best statistical estimation procedure up to a future paper.

Let us assume that we wish to model development times having a mean given by \bar{t} and a variance given by

σ^2 . The first step is to determine the number of phases necessary for the model. If the expression

$$k = \bar{t}^2 / (\sigma^2 + \bar{t}) \tag{16}$$

yields an integer, the resulting model is relatively easy; namely, it is the sum of k geometric distributions. In other words, for an integer result in Eq. 16 use a phase-type distribution with $v=k$, the initial probability vector $\alpha = (1, 0, \dots, 0)$, and the transition matrix given by

$$\mathbf{Q} = \begin{bmatrix} q & 0 & 0 & \cdots & 0 & 0 \\ 1-q & q & 0 & \cdots & 0 & 0 \\ 0 & 1-q & q & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & & \cdots & q & 0 \\ 0 & 0 & & \cdots & 1-q & q \end{bmatrix} \tag{17}$$

where

$$q = 1 - k/\bar{t} \tag{18}$$

and the dimension of \mathbf{Q} is $v \times v$. To understand Eqs. 17 and 18, remember that a phase with associated probability q yields a mean time of $1/(1 - q)$ and a variance of $q/(1 - q)^2$; thus the above yields a model that is the sum of k phases, each with mean \bar{t}/k and variance $\bar{t}(t - k)/k^2$. Since the phases are independent, the variances add as well as the means yielding the desired mean and variance.

Determining the number of phases and appropriate parameter values are more difficult when the variance of development times is such that Eq. 16 does not yield an integer. For example, assume we would like to model development times with a mean of 10 days and a standard deviation of 5.75 days. Using these values in Eq. 16 yields $k = 2.322$. Notice that the sum of two identical geometric distributions yielding a mean of 10 would produce a standard deviation of 6.32 and the sum of three identical geometric distributions yielding a mean of 10 would produce a standard deviation of 4.83. However, it is possible by taking a mixture of these two sums to produce any standard deviation between 4.83 and 6.32. Specifically for this example, the phase-type distribution defined by (α, \mathbf{Q}) , where

$$\alpha = (0.452, 0, 0.548, 0, 0)^\top$$

$$\mathbf{Q} = \begin{bmatrix} 0.7452 & 0 & 0 & 0 & 0 \\ 0.2548 & 0.7452 & 0 & 0 & 0 \\ 0 & 0 & 0.7452 & 0 & 0 \\ 0 & 0 & 0.2548 & 0.7452 & 0 \\ 0 & 0 & 0 & 0.2548 & 0.7452 \end{bmatrix}$$

will yield a mean of 10 and a standard deviation of 5.75. This phase-type distribution is equivalent to a mixture of two distributions where there is 45.2% chance of picking the first distribution of the mixture and a 54.8% chance of choosing the second. The first distribution in the mixture is the sum of two geometric distributions and the second distribution in the mixture is the sum of three geometric distributions.

The following steps generalize the above method to accommodate most means and variances that would be relevant for population modeling since it is more likely to have variances that are less than that produced by the geometric distribution than variances that are greater. In the steps below, it is assumed that we wish to establish a phase-type distribution to model development times with a mean denoted by \bar{t} and a variance denoted by σ^2 such that $\sigma^2 \leq \bar{t}^2 - \bar{t}$.

Step 1.—Determine k according to Eq. 16. If k is an integer, use the phase-type distribution defined by Eqs. 17 and 18; otherwise, proceed to step 2.

Step 2.—Let $\lfloor k \rfloor$ denote the value of k rounded down to the nearest integer, and let $\lceil k \rceil$ denote the value of k rounded up to the nearest integer. The phase-type distribution will be the mixture of the sum of $\lfloor k \rfloor$ geometric distributions and a sum of $\lceil k \rceil$ geometric distributions; thus the number of phases is given by

$$v = \lfloor k \rfloor + \lceil k \rceil = 2\lceil k \rceil - 1.$$

Step 3.—Define the probability vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_v)$, where α_1 is given by

$$\alpha_1 = \frac{\lceil k \rceil (\sigma^2 + \bar{t}) - \bar{t} \sqrt{\lfloor k \rfloor \bar{t}^2 - \lceil k \rceil (\sigma^2 + \bar{t})}}{\sigma^2 + \bar{t} + \bar{t}^2} \tag{19}$$

where $\alpha_{\lceil k \rceil} = 1 - \alpha_1$, and $\alpha_i = 0$ for all i except $i = 1$ and $i = \lceil k \rceil$.

Step 4.—Define the matrix $\mathbf{Q} = (q_{ij})$ for $i, j = 1, \dots, v$ by

$$q_{ij} = \begin{cases} 1 - \frac{\lceil k \rceil - \alpha_1}{\bar{t}} & \text{if } i = j, j = 1, \dots, v \\ \frac{\lceil k \rceil - \alpha_1}{\bar{t}} & \text{if } i = j + 1, j = 1, \dots, \\ & v - 1, j \neq \lceil k \rceil \\ 0 & \text{otherwise.} \end{cases} \tag{20}$$

Each phase within the matrix \mathbf{Q} defined above has the same probability of staying within the phase and the initial probabilities are structured so that the process either starts in phase 1 with probability given by α_1 and proceeds through the first $\lfloor k \rfloor$ phases and then is absorbed or starts in the $\lceil k \rceil$ th phase with probability $1 - \alpha_1$ and proceeds through the final $\lceil k \rceil$ phases and then is absorbed. Thus, the above phase type distribution is a mixture of two distributions, each of which is the sum of geometric distributions so that the resulting mean equals \bar{t} and the variance equals σ^2 .

PHASE-TYPE POPULATION MODEL

We return to our basic insect model involving three stages of insect development: eggs, immatures, and adults. Furthermore, we now assume that data have been collected so that we have estimates for the mean and standard deviation of the development times in each of three stages, plus we have survival rates and fecundity estimates. The development times for eggs are assumed

to follow a phase-type distribution defined by (α_e, \mathbf{Q}_e) of dimension v_e , the time that the insect spends as an immature follows a phase-type distribution defined by (α_i, \mathbf{Q}_i) of dimension v_i , and the time as an adult is a phase-type distribution defined by (α_a, \mathbf{Q}_a) of dimension v_a .

The population model still follows the same iteration scheme as defined by Eq. 1, except now the stage-phase matrix is defined by

$$\mathbf{A} = \begin{bmatrix} \mathbf{Q}_e & 0 & \alpha_e \mathbf{f} \\ s_e \alpha_i \tilde{\mathbf{q}}_e & \mathbf{Q}_i & 0 \\ 0 & s_i \alpha_a \tilde{\mathbf{q}}_i & \mathbf{Q}_a \end{bmatrix} \quad (21)$$

where \mathbf{f} is a row vector of dimension v_a with each component equaling $f = n_{\text{eggs}}/\bar{l}_a$.

As a reminder, the product of a column vector times a row vector is a rectangular matrix. In other words, the product of the column vector $\mathbf{c} = (c_1, \dots, c_m)$ times the row vector $\mathbf{r} = (r_1, \dots, r_n)$ yields an $m \times n$ matrix whose $i - j$ element is given by $c_i \times r_j$. An equivalent way to think of this product is to consider the column vector \mathbf{c} as an $m \times 1$ matrix and the row vector \mathbf{r} as a $1 \times n$ so that their product must yield an $m \times n$ matrix. Notice that all three products within the matrix \mathbf{A} involve a column vector times a row vector.

The population vector, \mathbf{n} also needs to be partitioned into phases; thus, the population on day t is represented as

$$\mathbf{n}(t) = [\mathbf{n}_e(t) \mid \mathbf{n}_i(t) \mid \mathbf{n}_a(t)]$$

where $\mathbf{n}_e(t)$, $\mathbf{n}_i(t)$, and $\mathbf{n}_a(t)$ are column vectors of dimension v_e , v_i , and v_a , respectively. In order to determine the population size of eggs, immatures, or adults, we simply sum the components of the individual appropriate vectors.

To illustrate the use of phase-type distributions in the population model, assume that following data: the mean and standard deviation for egg development are 8 days and 4.9 days, respectively, with a 40% survival rate from egg to immatures. The mean and standard deviation of development times for immatures are 2 days and 1.414 days, respectively, with a 60% survival rate from the immature to the adult stage. Finally, the mean and standard deviation of the lifetime of adults is 5 days and 3.5 days, respectively. Adults average 5.2 female eggs distributed uniformly over their adult life span yielding a fecundity of 1.04 female eggs per day per female.

The phase-type distribution for egg development times are straightforward because the application of Eq. 16 results in an integer; namely, $k = 2$ so that we set v_e equal to 2, yielding

$$\alpha_e = \begin{bmatrix} 1.0 \\ 0 \end{bmatrix} \quad \mathbf{Q}_e = \begin{bmatrix} 0.75 & 0 \\ 0.25 & 0.75 \end{bmatrix}.$$

The standard deviation for immature development is the same as a geometric distribution so multiple phases are not needed for modeling the immatures; thus, we have

$$\alpha_i = [1.0] \quad \mathbf{Q}_i = [0.5].$$

The phase-type distribution representing adult lifetimes requires three phases because the application of Eq. 16 resulted in a non-integer; namely, $k = 1.45$ so that we set v_a equal to 3. The application of Eqs. 19 and 20 yields

$$\alpha_a = \begin{bmatrix} 0.35 \\ 0.65 \\ 0 \end{bmatrix} \quad \mathbf{Q}_a = \begin{bmatrix} 0.67 & 0 & 0 \\ 0 & 0.67 & 0 \\ 0 & 0.33 & 0.67 \end{bmatrix}.$$

Combining the above phase-type distributions with the appropriate survival and fecundity values yields following population matrix that can now be used with Eq. 1 to produce a population trajectory for this organism:

$$\mathbf{A} = \begin{bmatrix} 0.75 & 0 & 0 & 1.04 & 1.04 & 1.04 \\ 0.25 & 0.75 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.105 & 0.67 & 0 & 0 \\ 0 & 0 & 0.195 & 0 & 0.67 & 0 \\ 0 & 0 & 0 & 0 & 0.33 & 0.67 \end{bmatrix}.$$

NUMERICAL ANALYSES

In this section, we use the techniques outlined in the preceding sections to verify the integrity of the approach and to highlight the importance of correctly handling variance in the development rates/times of organisms. To simplify our analyses we consider a hypothetical organism with two stages: juvenile and adult. Our goal is to use our model to illustrate how the population rate of increase is affected when the mean of the development times is held constant and the variance changes. (The rate of increase for the population is a time invariant measure of a population's growth rate and is frequently used as a simple index of population performance, e.g., Caswell 2001.) We shall also explore the interaction between variance in development times and other life-history parameters (fecundity, adult longevity, and juvenile survival). The base set of parameters for our hypothetical organism is detailed in Table 1. The coefficient of variation (CV) of development times is systematically varied with either fecundity, adult longevity, or juvenile survival through the ranges specified in Table 1. For each pair of manipulated parameter values, we calculate the rate of increase of the modeled population and use this as a measure of population performance.

Figs. 1–3 show the results of these analyses, presented as three-dimensional surfaces. The graphs display the sensitivity of population growth rate (rate of increase) to the manipulated parameter values. In all cases, it can be seen that the rate of increase predicted is sensitive to the variance of development times for the (hypothetical) organism. It can also be seen that the size of this effect is related to the specific life-history of the organism—for example the rate of population increase is highly sensitive to the variance in development times when fecundity, adult longevity, and juvenile survival are high.

TABLE 1. Values used in sensitivity studies.

Parameter	Value	Range
Juvenile development time	always 20 d	fixed
Juvenile development time CV	always varied	0–97.46%
Fecundity	2 offspring·female ⁻¹ ·d ⁻¹	1–10 offspring·female ⁻¹ ·d ⁻¹
Proportion of juveniles surviving	0.9	0.1–0.9
Mean adult longevity	5 d	1–10 d

It might also be noted that the maximum CV of juvenile development time is 0.9746 because this is the CV of a geometric distribution that has a mean of 20.

There are a couple of important implications of these graphs. The first alludes to the fact that population dynamics are sensitive to both the mean and the variance of development or maturation times. The second implication of these analyses is that since models are used to simplify complex systems, modelers must make informed decisions about what details to strip away and the consequences of these simplifications considering their specific modeling goals. The information presented here suggests that measuring and correctly representing the variance of development rates are crucial for obtaining accurate and reliable model outputs.

To show the validity of our matrix model, an individual-based simulation model was developed in

which individual organisms were created, developed, produced new organisms, and died according to random variates that were generated according to the distributions being modeled. The individual-based simulation approach allows for the construction of discrete individuals within the computer that follow specific sets of rules described by formulas of almost any complexity (see the Appendix for a description of the methodology used). As a result, complex life-history and behavior can easily be incorporated into the simulation. Fig. 4 shows a comparison of rates of increase calculated using both the matrix approach outlined in this paper and the individual-based simulation. It can be seen that over the majority of the range of the simulations, the answers given by the simulation model and the matrix approach presented here are similar, and differ only by amounts that are explained by the sample error associated with obtaining finite rates of increase from a non-analytical

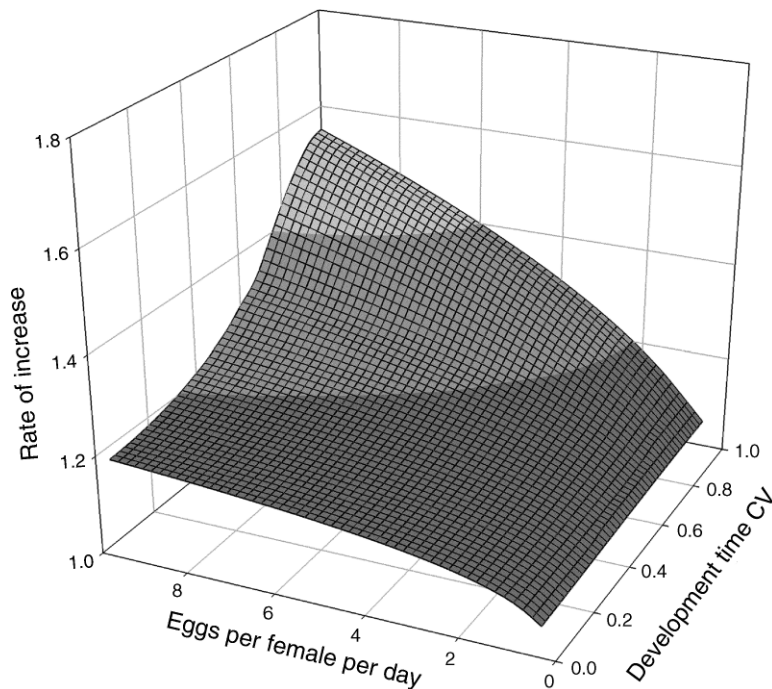


FIG. 1. The sensitivity of the rate of increase to fecundity and intrapopulation variability of development time. The rate of increase is a time-invariant measure of the growth of a population and is commonly derived from matrix models. The graph shows rates of increase calculated for hypothetical populations with different fecundities (eggs per female per day) and different levels of intrapopulation variability (but with a fixed mean of 20 days). Intrapopulation variation in development time is expressed as the coefficient of variation (CV).

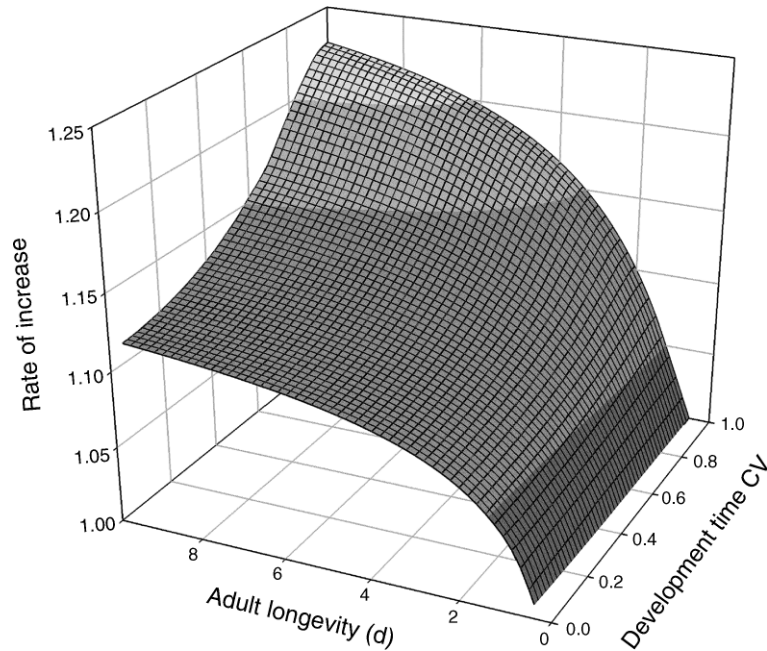


FIG. 2. The sensitivity of the rate of increase to adult longevity and intrapopulation variability of development time. The rate of increase is a time-invariant measure of the growth of a population and is commonly derived from matrix models. The graph shows rates of increase calculated for hypothetical populations with different adult longevitys and different levels of intrapopulation variability (but with a fixed mean of 20 days). Intrapopulation variation in development time is expressed as the coefficient of variation (CV).

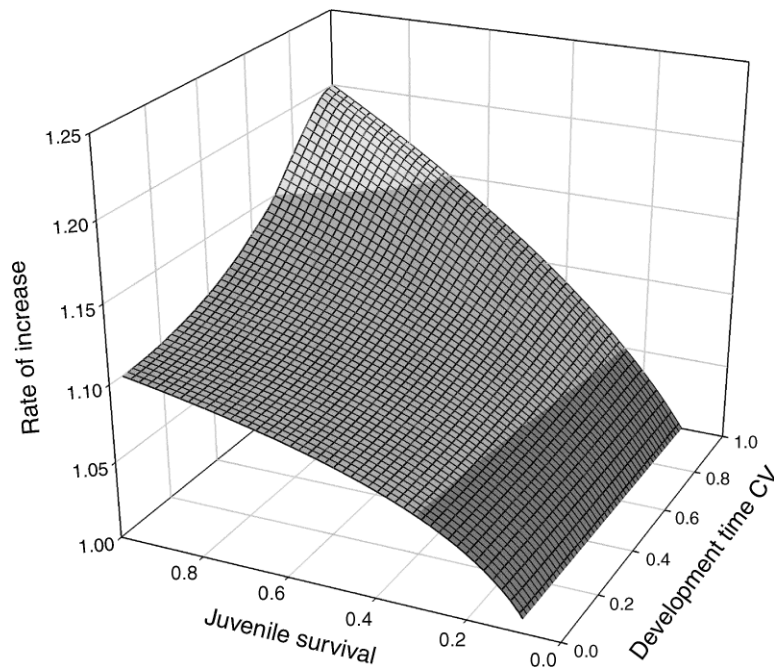


FIG. 3. The sensitivity of the rate of increase to juvenile survival and intrapopulation variability of development time. The rate of increase is a time-invariant measure of the growth of a population and is commonly derived from matrix models. The graph shows rates of increase calculated for hypothetical populations with different juvenile survival (expressed as daily probability of survival) and different amounts of intrapopulation variability (but with a fixed mean of 20 days). Intrapopulation variation in development time is expressed as the coefficient of variation (CV).

simulation. Notice that the CV of the geometric distribution is slightly less than 1.0 so that the maximum figure in the Fig. 4 is 0.95 instead of 1.0.

DISCUSSION

This paper describes some mathematical foundations that enable matrix population models to accurately represent intrapopulation variability in the development rates of organisms. Using these techniques, model outputs were practically identical (differences were attributable to rounding error) to those from a more complex, computationally intensive model. However the matrix approach benefited from massive computational savings. These methods are not the only option for representing variability in the development rates of populations (for a general review see Schaalje and Vaart 1989 and Caswell 2001 for matrix model methods). However, matrix population models offer a number of advantages over other techniques. These include an extensive body of research, a suite of effective analytical tools (e.g., Tuljapurkar and Caswell 1997, Caswell 2001), conceptually simple mathematical and ecological representations and accessible methods and software for simulation (e.g., standard matrix mathematics and MATLAB).

Without modification (such as those detailed here or the pseudostage approach adopted by Caswell 2001), stage-structured matrix models represent geometrically distributed development times. A geometric distribution constructed so as to yield a mean development time of \bar{t} will have a CV equal to $\sqrt{1 - 1/\bar{t}}$, which yields significantly more variability than many organisms. Table 2 reviews development times for a variety of organisms. Our methodology suggests that the sensitivity of the model to variations in development rate is of the same order of magnitude as fecundity, survival or mean development times. Put simply, variations in the development times of individuals in a population are usually not geometrically distributed (they have a much smaller CV). Given the sensitivity of populations to this phenomenon, stage-structured matrix models that do not account for realistic intrapopulation variability in development are likely to yield erroneous results.

Other researchers have highlighted the importance of intrapopulation variability and population the output of population models. For example Bellows (1986) assessed the impact of developmental variance on the growth of a hypothetical insect population and found a negative relationship between a CV of 0–60% and a positive one thereafter (60–200%). In contrast, our findings suggest a consistent positive relationship between population growth rate and intrapopulation development time variability. These differences may stem from difference in the modeling methodologies used. Bellows used a truncated normal distribution to describe distributions in development rates. The distributions we use are from a family formed from sums of geometric random variables. We attribute the contrasting results to these

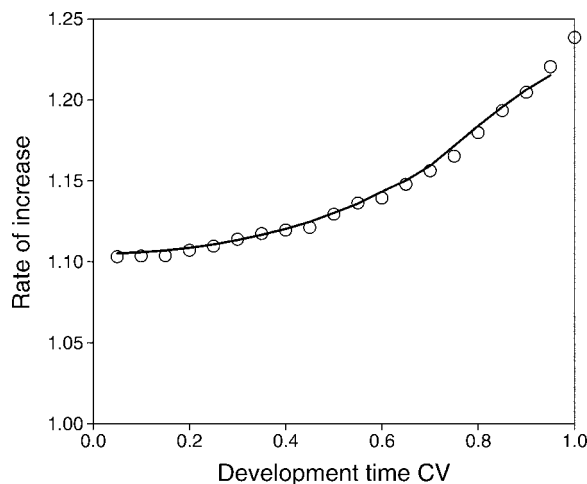


FIG. 4. Equivalence of outputs from our matrix model (solid line) and an individual-based simulation model (open circles) parameterized to represent a range of intrapopulation variability in development times (CV of development time). The graph shows that the rates of increase estimated using our matrix methods estimate are equivalent to estimates from a more complex, computationally intensive individual-based simulation model. In all cases, mean development time is fixed (20 days) showing that intrapopulation development time has a large effect on the rate of increase. Since the maximum CV for a geometric distribution with mean 20 is 0.975, the results from the matrix models are valid only up to this value.

differences in the shape of the distributions used to represent development times. In particular, the formulation used by Bellows (1986) results in large changes in the shape of the function representing distributions of development times such that (in his hypothetical population) the modal time of development is approximately 1 day later for a CV of 0–60% than 60–200%. This illustrates that both the shape of the developmental distributions (early developers are particularly important for population growth) and the relative timing of life-history events are important components of population models.

Although it can be shown that intrapopulation variability in development time plays an important role in the ecology of species, a survey of the literature (Table 2) shows that authors do not generally report these data. Instead, studies tend to focus on mean development times, usually reporting the standard error as an indicator of the accuracy of this measurement. In some cases (especially for controlled experiments), this is despite the fact that mean development times are calculated by measuring the development times for each individual in a cohort (i.e., variation is measured by default). There may be two reasons for this. The first is that the importance of variability in development rate is generally not fully appreciated by ecologists, a point that this paper aims to address. Second, many studies that involve the measurement of development rates are focused on hypothesis

TABLE 2. Variability in CV among organisms.

Species	Stages	Temperature (°C)	Mean	CV (%)	Reference
Insect					
<i>Aglais urticae</i>	egg to adult	11.8	91.7	6.59	Bryant et al. (1997)
<i>Aglais urticae</i>	egg to adult	32.8	13.6	4.41	
<i>Aglais urticae</i>	egg to adult	34	15.2	3.22	
<i>Inachis io</i>	egg to adult	12.2	118	1.90	
<i>Inachis io</i>	egg to adult	31.9	18.9	1.40	
<i>Inachis io</i>	egg to adult	33.7	25.3	4.18	
<i>Polygonia c-album</i>	egg to adult	12.1	69.6	7.62	
<i>Polygonia c-album</i>	egg to adult	32.9	19.1	4.16	
<i>Vanessa atalanta</i>	egg to adult	11.8	99	1.98	
<i>Vanessa atalanta</i>	egg to adult	32	17.3	4.33	
<i>Vanessa atalanta</i>	egg to adult	33.5	19.3	4.92	
<i>Tribolium confusum</i>	egg to adult	27	~149	~3.10	Bellows (1986)
<i>Scathophaga stercoraria</i> (yellow dung fly)	egg to adult	10	79.51	5.49	Blanckenhorn (1997)
<i>Scathophaga stercoraria</i>	egg to adult	15	36.42	5.14	
<i>Scathophaga stercoraria</i>	egg to adult	20	23.22	3.58	
<i>Scathophaga stercoraria</i>	egg to adult	25	21.2	4.90	
<i>Aphis gossypii</i>	birth to adult	15	12	9.12	Kersting et al. (1999)
<i>Aphis gossypii</i>	birth to adult	20	8.1	16.00	
<i>Aphis gossypii</i>	birth to adult	25	5.7	10.47	
<i>Aphis gossypii</i>	birth to adult	30	4.5	12.81	
<i>Dendroctonus frontalis</i>	larval	15	44.83	26.50	Gagne (1980)
	larval	20	23.08	15.40	
	larval	25	17.12	23.26	
	larval	27	27.69	25.20	
<i>Myzus persicae</i>	apterous (instars)	6.2	44.5	9.91	Liu and Meng (1999)
	apterous (instars)	11.3	17.8	8.17	
	apterous (instars)	14.3	11.7	5.80	
	apterous (instars)	20.1	7.0	10.52	
	apterous (instars)	24.7	5.5	12.01	
	apterous (instars)	30	6.2	10.33	
Copepod					
<i>Pseudocalanus minutus</i>	egg to emergence	0	10.89	5.51	McLaren (1966)
	egg to emergence	3.18	6.61	3.93	
	egg to emergence	4.6	5.91	4.57	
	egg to emergence	13.13	2.78	5.04	
Amphibian					
<i>Scaphiopus couchii</i>	to metamorphosis	unknown	14.57	4.30	Newman (1988)
<i>Hyla pseudopuma</i>	to metamorphosis	unknown	25.7	37.74	Crump (1989)
	to metamorphosis	unknown	25.0	33.26	
Fish					
<i>Salvelinus fontinalis</i> (Brook trout)	egg	7.5	75.8	2.77	Hutchings (1991)

testing rather than the integration of findings into population level effects via population models.

For example, Bridges (2000) presents a toxicity study assessing the effects of the pesticide Carbaryl to different life-stages of southern leopard frog and found that although the mean development time (to metamorphosis) was not significantly different between treatment groups, some treatments led to changes in the number of individuals displaying rapid growth rates. Bridges concludes that since leopard frogs often breed in temporary pools, individuals that develop before ponds desiccate may contribute considerably to population growth and therefore changes in the distribution of development times may be ecologically significant. However, our results also suggest that changes in the intrapopulation variability in development times can affect time invariant measures of population growth and

may thus provide a more straightforward interpretation of such results.

The variability of intrapopulation development times is often also studied as an explicable and adaptive ecological phenomenon. Werner (1988) estimated that approximately 80% of all animal species have complex life-cycles and theoretical ecologists in particular are interested in understanding the fitness consequences of differences in the timing and magnitude of the life-history events that define this complexity. The majority of these studies have focused on understanding why distinct populations of the same species exhibit differences in the timing of life-history events. Examples include the trade-offs between growth and predation in guppies (Reznick and Endler 1982) and mayfly larvae (Peckarsky et al. 2001). However a number of researchers have also attempted to explain variation in development rates in organisms occurring sympatrically. For example, Mangel and Stamps (2001)

present a review of the trade-off between growth and survival and provide mathematical arguments for the maintenance of intrapopulation variability in individual development times. Other researchers believe that for organisms that live in fluctuating and uncertain environments, variation in development times (or development plasticity) is an adaptive response to these uncertainties. The methods presented here allow effective control over the distribution of development times amongst individuals and may therefore provide useful tools for further research in this area.

ACKNOWLEDGMENTS

This research was supported in part by the USDA Forest Service Cooperative agreement SRS 06-CA-11330124-196 and through the Knowledge Engineering Laboratory, College of Agriculture, Texas A&M University.

LITERATURE CITED

- Bellows, T. S. 1986. Impact of developmental variance on behavior of models for insect populations I. Models for populations with unrestricted growth. *Researches on Population Ecology* 28:53–62.
- Blanckenhorn, W. U. 1997. Effects of temperature on growth, development and diapause in the yellow dung fly: against all the rules? *Oecologia* 111:318–324.
- Blanckenhorn, W. U. 1998. Adaptive phenotypic plasticity in growth, development, and body size in the yellow dung fly. *Evolution* 52:1394–1407.
- Bridges, C. M. 2000. Long-term effects of pesticide exposure at various life stages of the southern leopard frog (*Rana sphenoccephala*). *Archives of Environmental Contamination and Toxicology* 39:91–96.
- Bryant, S. R., C. D. Thomas, and J. S. Bale. 1997. Nettle-feeding nymphalid butterflies: temperature, development and distribution. *Ecological Entomology* 22:390–398.
- Caswell, H. 2001. *Matrix population models*. Second edition. Sinauer Associates, Sunderland, Massachusetts, USA.
- Crump, M. L. 1989. Effect of habitat drying on developmental time and size at metamorphosis in *Hyla pseudopuma*. *Copeia* 1989:794–797.
- Davidson, J. 1944. On the relationship between temperature and rate of development of insects at constant temperatures. *Journal of Animal Ecology* 13:26–38.
- Fischer, K., and K. Fiedler. 2000. Response of the copper butterfly *Lycaena tityrus* to increased leaf nitrogen in natural food plants: evidence against the nitrogen limitation hypothesis. *Oecologia* 124:235–241.
- Gagne, J. 1980. The effects of temperature on population processes of the southern pine beetle *Dendroctonus frontalis Zimmerman*. Dissertation. Texas A&M, College Station, Texas, USA.
- Haukioja, E., E. Pakarinen, P. Niemelä, and L. Iso-Iivari. 1988. Crowding-triggered phenotypic responses alleviate consequences of crowding in *Epirrita autumnata* (Lep., Geometridae). *Oecologia* 75:549–558.
- Hutchings, J. A. 1991. Fitness consequences of variation in egg size and food abundance in brook trout *Salvelinus fontinalis*. *Evolution* 45:1162–1168.
- Kersting, U., S. Satar, and N. Uygun. 1999. Effect of temperature on development rate and fecundity of apterous *Aphis gossypii* Glover (Hom., Aphididae) reared on *Gossypium hirsutum* L. *Journal of Applied Entomology* 123:23–27.
- Leslie, P. H. 1945. The use of matrices in certain population. *Mathematics. Biometrika* 33:183–212.
- Leslie, P. H. 1948. Some further notes on the use of matrices in population mathematics. *Biometrika* 35:213–245.
- Lewis, E. G. 1942. On the generation and growth of a population. *Sankhya* 6:93–96.
- Liu, S. S., and X. D. Meng. 1999. Modelling development time of *Myzus persicae* (Hemiptera: Aphididae) at constant and natural temperatures. *Bulletin of Entomological Research* 89: 53–63.
- Mangel, M., and J. Stamps. 2001. Trade-offs between growth and mortality and the maintenance of individual variation in growth. *Evolutionary Ecology Research* 3:583–593.
- McLaren, I. A. 1966. Predicting development rate of copepod eggs. *Biological Bulletin* 131:457–469.
- Michaud, J. P., and A. K. Grant. 2003. Sub-lethal effects of a copper sulfate fungicide on development and reproduction in three coccinellid species. *Journal of Insect Science* 3:1–6.
- Neuts, M. F. 1981. Matrix-geometric solutions in stochastic models: an algorithmic approach. The Johns Hopkins University Press, Baltimore, Maryland, USA.
- Newman, R. A. 1988. Genetic variation for larval anuran (*Scaphiopus couchii*) development time in an uncertain environment. *Evolution* 42:763–773.
- Peacor, S. D., J. R. Bence, and C. A. Pfister. 2007. The effect of size-dependent growth and environmental factors on animal size variability. *Theoretical Population Biology* 71:80–94.
- Peckarsky, B. L., B. W. Taylor, A. R. McIntosh, M. A. McPeck, and D. A. Lytle. 2001. Variation in mayfly size at metamorphosis as a developmental response to risk of predation. *Ecology* 82:740–757.
- Pipe, R. K., and P. Walker. 1987. The effect of temperature on development and hatching of Scad, *Trachurus trachurus* L., eggs. *Journal of Fish Biology* 31:675–682.
- Reznick, D. A., H. Bryga, and J. A. Endler. 1990. Experimentally induced life-history evolution in a natural population. *Nature* 346:357–359.
- Reznick, D., and J. A. Endler. 1982. The impact of predation on life history evolution in Trinidadian guppies (*Poecilia reticulata*). *Evolution* 36:160–177.
- Ruohomaki, K., T. Klemola, P. Kaitaniemi, and M. Kaar. 2003. Crowding-induced responses in a geometrid moth revisited: a field experiment. *Oikos* 103:489–496.
- Ryer, C. H., and B. L. Olla. 1995. The influence of food distribution upon the development of aggressive and competitive behaviour in juvenile chum salmon, *Oncorhynchus keta*. *Journal of Fish Biology* 46:264–272.
- Schaalje, G. B., and H. R. Vaart. 1989. Relationships among recent models for insect population dynamics with variable rates of development. *Journal of Mathematical Biology* 27: 399–428.
- Schneider, P., W. Takken, and P. J. McCall. 2000. Interspecific competition between sibling species larvae of *Anopheles arabiensis* and *An. gambiae*. *Medical and Veterinary Entomology* 14:165–170.
- Taylor, F. 1981. Ecology and evolution of physiological time in insects. *American Naturalist* 117:1–23.
- Teng, H. J., and C. S. Apperson. 2000. Development and survival of immature *Aedes albopictus* and *Aedes triseriatus* (Diptera: Culicidae) in the laboratory: effects of density, food, and competition on response to temperature. *Journal of Medical Entomology* 37:40–52.
- Tuljapurkar, S., and H. Caswell. 1997. *Structured-population models in marine, terrestrial, and freshwater systems*. Chapman and Hall, New York New York, USA.
- Twombly, S. 1993. Inter- and intrapopulation variation in time to metamorphosis in a freshwater copepod. *Freshwater Biology* 30:105–118.
- van Tienderen, P. H. 1995. Life cycle trade-offs in matrix population models. *Ecology* 76:2482–2489.
- Wagner, T. L., H. I. Wu, R. M. Feldman, P. J. H. Sharpe, and R. N. Coulson. 1985. Multiple-cohort approach for simulating development of insect populations under variable temperatures. *Annals of the Entomological Society of America* 78:691–704.

- Wang, J. J., Z. M. Zhao, and J. P. Zhang. 2004. The host plant-mediated impact of simulated acid rain on the development and reproduction of *Tetranychus cinnabarinus* (Acari, Tetranychidae). *Journal of Applied Entomology* 128:397–402.
- Werner, E. E. 1988. Size, scaling, and the evolution of complex life cycles. Pages 60–81 in B. Ebenman and L. Persson, editors. *Size-structured populations: ecology and evolution*. Springer-Verlag, Berlin, Germany.

APPENDIX

A proof and discussion regarding the importance of not using a survival term within the transition probabilities of Eqs. 3–5 (*Ecological Archives* E090-004-A1).