Density-Structured Models for Plant Population Dynamics.
Author(s): Robert P. Freckleton, William J. Sutherland, Andrew R. Watkinson, Simon A. Queenborough
Source: *The American Naturalist*, Vol. 177, No. 1 (January 2011), pp. 1-17
Published by: The University of Chicago Press for The American Society of Naturalists
Stable URL: http://www.jstor.org/stable/10.1086/657621
Accessed: 06/06/2011 14:24

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=ucpress.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

The University of Chicago Press and *The American Society of Naturalists* are collaborating with JSTOR to digitize, preserve and extend access to *The American Naturalist.*

http://www.jstor.org

VOL. 177, NO. 1 THE AMERICAN NATURALIST JANUARY 2011 Ⓔ

Density-Structured Models for Plant Population Dynamics

Robert P. Freckleton,[1,*] William J. Sutherland,[2] Andrew R. Watkinson,[3] and Simon A. Queenborough[4]

1. Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom; 2. Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, United Kingdom; 3. School of Environmental Science, University of East Anglia, Norwich NR4 7TJ, United Kingdom 4. National Center for Ecological Analysis and Synthesis, Santa Barbara, California 93106

Submitted April 8, 2010; Accepted October 1, 2010; Electronically published December 2, 2010

Online enhancements: zip file.

ABSTRACT: Density-structured models are structured population models in which the state variable is the proportion of populations or sites in a small number of discrete density states. Although such models have rarely been used, they have the advantage that they are straightforward to parameterize, make few assumptions about population dynamics, and permit rapid data collection using coarse density assessment. In this article, we highlight their use in relating population dynamics to environmental variation and their robustness to measurement error. We show that density-structured models are able to accurately represent population dynamics under a wide range of conditions. We look at the effects of including a persistent seedbank and describe numerical approximations for the mean and variance of population size. For simulated data, we determine the extent to which the underlying continuous process may be inferred from density-structured data. Finally, we discuss issues of parameter estimation and applications for which these types of models may be useful.

Keywords: population dynamics, structured model, measurement error, annual plants.

Introduction

Population models are key tools in ecology (Maynard Smith 1973; May 1974, 1989; Sutherland 1996; Hilborn and Mangel 1997; Lande et al. 2003; Bolker 2008). Models serve a range of functions, including predicting whether populations of conservation concern are able to persist (Lande et al. 2003), whether control measures will eradicate pest species (e.g., Firbank and Watkinson 1986; Rees and Paynter 1997), or how populations might respond to changes in the environment (Ranta et al. 2006). A variety of data sources may be used to generate population models, ranging from data on individual performance and fitness (e.g., Pacala et al. 1996; Sutherland 1996) to long-term census data (Dennis and Taper 1994; Rees et al. 1996; Sibly et al. 2005, 2007).

Although modeling is the focus of many studies in applied ecology, a critical limitation in many, if not most, applications is that it is difficult to generate estimates of the spatiotemporal variation in key parameters (May 1989; Freckleton and Watkinson 1998; Sæther et al. 2007; Grøtan et al. 2008), and consequently it is difficult to apply models at large scales. Developing methodologies that allow us to deal with these problems is an important challenge (Lande et al. 2003; Clark and Bjørnstad 2004).

Several types of population model are commonly employed in the literature. The most commonly used models are simple differential and difference equations predicting changes in population size as a function of a small number of parameters (e.g., see Hassell 1975; Watkinson 1980; Royama 1992; Lande et al. 2003). Such models can be used as tools to explore population dynamics in theoretical analyses (e.g., May 1974), but they can also be parameterized with field data and used to describe or predict population dynamics (e.g., Rees et al. 1996; Freckleton et al. 2000; Sibly et al. 2005). More complex models include structured models, such as Leslie and Lefkovitch matrix models for age- or stage-structured populations (e.g., Caswell 2000). More recently, integral projection models have been developed for modeling populations in which growth and survival are size dependent (Easterling et al. 2000; Ellner and Rees 2007).

When detailed data on individual performance and survival are available, it is possible to build up models based on a detailed understanding of the various processes that occur in the life cycle and drive population dynamics (e.g., Pacala et al. 1996; Sutherland 1996; Stephens et al. 2005). The drawback is that frequently the information is not available for such modeling. It is more common that annual census data have been collected, and these can be used as the basis for parameterizing regression models with less detail but at longer and larger scales.

---

[*] Corresponding author; e-mail: r.freckleton@sheffield.ac.uk.

Am. Nat. 2011. Vol. 177, pp. 1–17.
DOI: 10.1086/657621

As noted above, whatever approach is used, a key limitation in developing models for population dynamics is that more often than not, insufficient data are available for estimating the variance in model parameters. Large-scale census data are important in this regard, as such data sets potentially allow the spatiotemporal variation in population dynamics to be quantified and related to environmental drivers (Clark and Bjørnstad 2004; Sæther et al. 2007). However, there are complications with this approach: measurement error can be a significant component of the variation in long-term ecological data sets. Specifically, errors in estimates of population density can lead to analyses yielding biased estimates of model parameters. For example, the strength of density dependence may be overestimated, resulting in erroneous predictions of mean population size (Shenk et al. 1998; Freckleton et al. 2006; Green 2008) or possible underestimates of the effects of environmental predictors (e.g., Pablo Almaraz et al., unpublished manuscript).

The considerations described above are extremely important for the design of long-term monitoring programs. There is, unsurprisingly, a trade-off to be considered between balancing the extent and amount of data collected and the precision or resolution of data collection. More detailed data—for instance, on the fates of individuals and measuring the various demographic processes determining population growth with quantified errors—will yield more reliable models. However, such data are more expensive and time-consuming to collect. On the other hand, rapidly collected data with less precision or resolution can achieve a broader coverage but will yield less resolved or detailed predictions.

In a study attempting to model the population growth and control of an alien species, Taylor and Hastings (2004) used a "density-structured" model. This model was somewhat of a departure from previous models for population dynamics, in that the state variable in this model was a discrete density state. In their model, this was a simple assignment of the state of a site as "low," "medium," or "high." The model then simply modeled the rate at which sites move between these density states, with the parameters of the model implicitly being a matrix of transition probabilities (see also Moore and Noble 1990; Bogich and Shea 2008). In previous applications, such models have used either these three density states or presence/absence as the state variables.

This might, at first sight, appear to be a somewhat crude approach to modeling, in that population dynamics are not modeled explicitly but rather are summarized by coarse transition probabilities. However, this modeling framework offers some empirical and analytical advantages. In empirical terms, there are two key advantages: first, the data for parameterizing a density-structured model should be relatively easy to collect, because the density states should be easier and quicker to assign than detailed enumeration of population densities; and second, census errors in the data are straightforward to assess via replicated surveys, which should be relatively quick and easy to perform. In analytical terms, these models offer three advantages: (1) the transition probabilities should be statistically straightforward to estimate; (2) analysis is simplified if it is not necessary to specify an underlying model for population dynamics (e.g., Ellner et al. 2002): in this case, because the model is defined purely by the estimated probabilities, it is not necessarily to specify any functions describing the underlying mechanisms driving population dynamics; and (3) the model is a simple linear one, which means that it has well-understood properties because it is a special case of a structured model, for which a great deal of theory exists (Caswell 2000).

Apart from the simplification of population dynamics, there are potential drawbacks to this approach. One obvious issue is whether population dynamics can be approximated accurately enough for predictions to be useful. For instance, when considering size-structured models, Easterling et al. (2000) pointed out that matrix models with small numbers of discrete states could potentially provide a misleading description of population dynamics because over several generations individuals are predicted to move too quickly between successive states. This is an artifact of the classification into discrete states and can be dealt with by using an alternative approach that uses a continuous state variable (Easterling et al. 2000; Ellner and Rees 2006, 2007). It seems possible that such artifacts could occur in models structured by density; however, the extent of such effects is uncertain.

The density-structured approach to modeling has not been used a great deal in modeling to date, which seems surprising, given its potential. In this article, we highlight density-structured models and their possible application to problems in population modeling. We show that they are not crude approximations but, under reasonable assumptions, accurate descriptions of population dynamics and population structure. They can be used to provide accurate estimates of the mean and variance of population sizes, and even for models with a few states, the underlying stationary distribution of population size can be retrieved. Finally, we look at the degree to which the underlying continuous dynamics can be retrieved from discrete density state data.

Methods

We begin by outlining a simple demographic model for the population dynamics of an annual plant in a stochastic environment both without and with a seedbank. Using this model as an example, we then describe the formulation of a density-structured model from a continuous model and describe how we derive numerical approximations for the stationary distribution of population size, mean and variance of population size, and population growth rates. Finally, we ask, given a density-structured data set or model, can we expect to be able to recover estimates of the parameters of the continuous population model?

Difference Equation Model

The starting point is a conventional model for population dynamics based on a difference equation relating population size (measured as a continuous variable) in one year to that in the previous year. The general model is

$$N(t+1) = \lambda(t)N(t)f(N(t)) \tag{1}$$

(e.g., Maynard Smith 1973; May 1974; Hassell 1975; May and Oster 1976; Watkinson 1980). Here N is population size measured at times t and $t+1$, λ is the finite rate of population increase from low densities, and $f()$ is a function describing the effects of density dependence on population growth. In this equation, λ is written as a function of t to denote that it is a stochastic variable. To model density dependence, we employed the following form for $f()$:

$$f(N) = (1 + aN)^{-b} \tag{2}$$

(Hassell 1975; Watkinson 1980), where a is a parameter scaling the strength of density dependence and b characterizes whether the density response is undercompensating ($b < 1$), compensating ($b = 1$), or overcompensating ($b > 1$).

In many plant species, there is a persistent seedbank. In this case, the model for population dynamics is more complex. A simple model, assuming no age dependence in the seedbank, is

$$\begin{aligned} s(t+1) &= q(t)(1-g(t))s(t) + s_m N(t)f(N(t)), \\ N(t+1) &= g(t+1)s(t+1) \\ &= q(t)g(t+1)(1-g(t))s(t) \\ &\quad + g(t+1)s_m N(t)f(N(t)) \end{aligned} \tag{3}$$

(MacDonald and Watkinson 1981; Ellner 1984), where s is the density of seed, s_m is the maximal mean production of seed by an isolated plant, g is the fraction of seed germinating in a given year, and q is the annual survival of seed. In the current context, it is useful to note that this equation reduces to a single difference equation for the density of mature plants:

$$\begin{aligned} N(t+1) &= q(t)\frac{g(t+1)}{g(t)}(1-g(t))N(t) \\ &\quad + g(t+1)s_m N(t)f(N(t)) \\ &= \gamma(t)N(t) + \lambda(t)N(t)f(N(t)). \end{aligned} \tag{4}$$

(MacDonald and Watkinson 1981; Ellner 1984). This equation has two stochastic components, γ and λ. These model recruitment from the seedbank and from newly produced seeds, respectively. The new parameter, $\gamma(t) = q(t)(g(t+1)/g(t))(1-g(t))$, is a composite of the parameters driving the seedbank dynamics (it is assumed that the correlation between γ and λ is negligible; this is a reasonable assumption because $g(t)$ is small and the dynamics of γ are driven by variation in q in addition to that in g). In a constant environment, this term simplifies to $\gamma = q(1-g)$. In a stochastic environment, the mean of γ will lie between 0 and 1; however, individual values of γ may exceed 1 if $g(t)$ is small enough.

In numerical analyses, we assumed that both λ and γ were lognormally distributed, with variances σ_λ^2 and σ_γ^2, respectively. Equation (9) was applied across the full $k \times k$ matrix of possible density state transitions to derive a state-transition matrix.

Density-Structured Dynamics

In a density-structured model, we define $\mathbf{s}$ to be a vector of states (s) of the population. In general, there are k ordered states, defined as a series of density intervals. An interval i is defined by lower limit N_i^-, upper limit N_i^+, and midpoint N_i^{mid}.

The entries of $\mathbf{s}$ are $s_1, s_2, \ldots, s_k$ and are the proportions of the population in each state. These proportions can be viewed in two equivalent ways: first, s_j could represent the proportion of a given set of sites that are in state j; second, s_j could measure the probability that a given site is in state j. In the first case, if we modeled the population dynamics of n sites, $s_j n$ would be the proportion of sites expected to be in state j. In the second case, if we studied the dynamics of a single site, s_j would be the probability that the site would be in state j at equilibrium.

The model relating the state of the population at time $t+1$ to that at time t is

$$\mathbf{s}(t+1) = \mathbf{T} \cdot \mathbf{s}(t), \tag{5}$$

where $\mathbf{T}$ is a $k \times k$ transition matrix with the following form:

$$\mathbf{T} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & \cdots & p_{1k} \\ p_{21} & p_{22} & p_{23} & p_{24} & \cdots & p_{2k} \\ p_{31} & p_{32} & p_{33} & p_{34} & \cdots & p_{3k} \\ p_{41} & p_{42} & p_{43} & p_{44} & \cdots & p_{4k} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & p_{k3} & p_{k4} & \cdots & p_{kk} \end{bmatrix}. \tag{6}$$

The diagonal entries of **T** are the probabilities (p) that a patch in a given state will remain in that state for the next year; that is, p_{jj} is the probability that a patch in state j will remain in that state. The off-diagonal entries of **T** model the transitions between density states. For example, $p_{(j+1)j}$ is the probability that a patch in state j will be in state $j+1$ in the next year, whereas $p_{j(j-1)}$ is the probability that the reverse transition will occur. The model defined by equation (6) is a Markov chain model. These models are described in detail in numerous texts (see, e.g., Taylor and Karlin 1998). So long as the probabilities are constant and the model obeys certain assumptions (irreducibility, i.e., that all states are reachable; and ergodicity, i.e., that the eventual state is the same irrespective of the starting point; see Stott et al. 2010), then the model can be used to predict population sizes.

Because the entries of **T** are probabilities, all of the entries within a column of **T** must sum to 1. Because all of the rows of **T** sum to 1, it is always the case that $\sum s_i(t) = \sum s_j(t+1) = 1$, and consequently, the dominant eigenvalue of **T** is 1.

Calculation of Transition Probabilities

The probability density of a site at density x being at density y in the next year is

$$P(y|x) = G(y;\ F(x), \sigma_\lambda^2), \tag{7}$$

where $F()$ is the mean population size at time $t+1$ predicted by the continuous population model and $G()$ is the density function for the distribution of population sizes about the mean. For a site that is at density x in year t, the probability that it will be in state j in year $t+1$ is then

$$P(j|x) = \int_{N_j^-}^{N^+} G(y;\ F(x), \sigma_\lambda^2)dy. \tag{8}$$

The probability that a site in state i at time t will be in state j at time $t+1$ is then given by the following integral:

$$P(j|i) = \frac{\int_{N_i^-}^{N^+}\int_{N_j^-}^{N^+} G(y;\ F(x), \sigma_\lambda^2)P(x)dydx}{\int_{N_j^-}^{N^-} P(x)dx}. \tag{9}$$

In equation (9), $P(x)dx$ is the probability that a site is at density x, that is, $P(x)$ is the probability density of x in the quasi-stationary distribution of population sizes. This is important to include in the calculation of density state transition probabilities because variation in initial density will affect the overall probability of transition because of the nonlinear relationship between densities in successive years. Specifically, equation (8) has to be integrated across the stationary distribution of population sizes. If this is not known, it will yield only approximate estimates of the transition probabilities. Equation (9), however, is exact. The assumption in equation (9) is that empirical estimates of density state transitions are measured on an equilibrium population distributed according to the quasi-stationary distribution of population sizes.

A simple expression for the quasi-stationary distribution of population sizes predicted by the model defined by equations (1) and (2) does not exist except in special cases (Dennis and Patil 1984; Dennis and Costantino 1988). For generating theoretical transition matrices according to equation (9), we therefore used a numerical approach to estimate $P(x)$ in equation (9). For each set of parameter values, we first simulated 10^6 generations by using the stochastic model, iterating initially for 1,000 generations to ensure that the quasi-stationary distribution was achieved. We then approximated the distribution of population sizes by using kernel density estimation with the density function in R (R Development Core Team 2010). For generating estimates of the mean and variance of population size from density-structured data, we explore the use of gamma distributions (below).

In the model with no seedbank, $G()$ is simply a lognormal probability distribution. It is straightforward to modify equation (9) to account for the presence of a seedbank, simply by replacing the function $F()$ with the predicted density according to equation (4). However, the complication here is that the variance about the expected density is given by a probability distribution that is the sum of the two lognormal random variates λ and γ. This is not a straightforward distribution and has no closed form; therefore, we used a numerical approach: 10^6 generations were simulated from the initial density x, and kernel density estimation was used to approximate $G(x)$ by means of the density function in R (R Development Core Team 2010).

Determining the Density States

The density states may be set in one of a number of ways; however, a practically relevant approach is that density classes are set a priori in order to reflect the range of dynamics observed across a range of environmental conditions (e.g., Queenborough et al., forthcoming). In the results, we describe how the number of density states was set at 10, five, and three for examining bias and at five and three for exploring precision. The largest value was chosen as an unrealistically high number intended to more closely approximate the underlying stochastic difference-equation model. The two smaller values were chosen to represent values that might more usually be employed in real applications. In the appendix tables, available in a zip file in the online edition of the *American Naturalist*, we also report results from $k = 2$–10 density states in exploring model bias (appendix tables 1–4) and from $k = 2$–6 density states when exploring precision (appendix tables 5–7). A smaller range was used in exploring precision because for matrices with more than six density states, the number of parameters to be estimated becomes impracticably large relative to the number of transitions simulated.

Approximating the Stationary Distribution of Population Size

As noted above, the stationary distribution of population sizes $P(x)$ does not have an exact analytical form except under certain circumstances. However, for populations with a stable equilibrium, the gamma distribution can be expected to provide an reasonable fit (May 1974; Dennis and Patil 1984; Dennis and Costantino 1988). We therefore explored the potential for the gamma distribution to approximate the mean and variance of the logarithm of population size on the basis of the stable state distribution (we also explored the use of the gamma and lognormal distributions; however, these performed slightly worse in simulations).

The gamma distribution is defined by two parameters, the shape (α) and rate (β):

$$\text{gamma}(x; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(a)} x^{\alpha-1} e^{-\beta x}. \tag{10}$$

The mean is α/β, and the variance is α/β^2. In the case of the log-gamma distribution, $x = \log(N)$ is the logarithm of the variable of interest.

If $\mathbf{s}^* = \{s_1^*, s_2^*, \ldots, s_k^*\}$ was the stable distribution of density states predicted by the density-structured model, we approximated the stable distribution of population sizes by using nonlinear modeling to find values of α and β that yielded the gamma distribution with the best fit to $\mathbf{s}^*$. We did this by using nonlinear modeling (using the optim function in R; R Development Core Team 2010) to find the best-fitting values according to least squares minimization. From the fitted log-gamma distributions, we estimated the mean and variance in population sizes. This was done for a range of λ between 1.5 and 100 and a range of σ_λ between 0.1 and 1.0. Note that because λ is assumed to be lognormally distributed, this upper value of σ_λ is extremely high. We included this high value to incorporate the extreme possibilities; however, in reality we would expect values of σ_λ to lie between 0 and 0.5 (see appendix table 8).

Populations were simulated with no seedbank ($\gamma = 0$) or with a moderate ($\gamma = 0.2$; $\sigma_\gamma = 0.2$) or highly persistent ($\gamma = 0.8$; $\sigma_\gamma = 0.2$) seedbank. Finally, we simulated populations in which density dependence was overcompensating ($b = 4$). We included overcompensating density dependence for the sake of completeness; however, we note that such dynamics are expected to be rare (Rees and Crawley 1991; Freckleton and Watkinson 2002). Below, we report analyses of distributions of population size for $k = 3$, 5, and 10 density states (see appendix tables 1–4 for expanded results for $k = 2$–10 states). Each parameter combination was replicated 100 times.

For each parameter combination, we also estimated the precision of the estimates, using a simple randomization procedure. For each parameter combination, we generated T independent transitions from $N(t)$ to $N(t+1)$. We used these to generate empirically observed transition matrices by calculating the rates of transitions between density states. These matrices were then iterated to generate predicted stable state distributions, which were used in turn to predict the mean and standard deviation of population size using the approach described above. Precision was measured as the coefficient of variation in parameter estimate in 100 simulations; this corresponds to the usual definition of sampling variance for statistical estimators, that is, the sampling variance about the estimated, but not necessarily true, value.

Here we report results for $T = 500$ transitions. This number is equivalent to the number of plots recensused in a study: for example, 500 transitions could represent 500 plots resurveyed once or 250 plots resurveyed twice. This is a number comparable with sample sizes used in previous detailed demographic modeling (e.g., Rees et al. [1996] used up to 700 transitions; Watkinson et al. [2000] used between 300 and 800 transitions; Freckleton et al. [2000] used 480 transitions). Note that usually we would expect that the ease of collection of density-structured data would permit much larger sample sizes than using conventional demographic approaches. In the appendix, we report expanded results for 200, 500, and 1,000 transitions, and for between two and six density states.

In analyzing bias and precision, we concentrate on the standard deviation of population size. To interpret these, we note that a bias (or precision) of +0.05 will imply that the 95% upper limit for population size will be 5% too high or that the standard error for mean population size will be 5% too high. To put this degree of error in context, if 1.96 is the multiplier for the usual 0.975 tail of a normal distribution, with a 5% error the effective multiplier becomes 2.058. This corresponds to a percentile of 0.98; that is, it leads to only a 0.5% difference in probability in this tail, which is relatively minor. For instance, in a two-tailed test, this would change a "true" $p = 0.04$ to 0.05. In terms of interpretation, we would argue that a 5% bias is only small, and we suggest that 10% is a bias that would start to become notable, although not necessarily fatal. In tables 1 and 2, we highlight all estimates of bias or precision that are greater than 10%.

Recovering the Continuous Model

As we noted above, in a density-structured model it is not necessary to explicitly specify a model for density dependence or detailed population dynamics. Simply by using estimated probabilities to form a transition matrix and iterating this model, it is possible to generate predictions of population size. In practical applications, however, it may be desirable to be able to link density-structured models to continuous demographic models. This may be useful, for example, if other information is available or if the aim of modeling is to simulate populations under new conditions. In this section, we therefore ask whether the underlying continuous model can be recovered from a density-structured model. More generally, the simulations we report demonstrate the degree of equivalence of the predictions of the two types of model.

Specifically, we asked, given a matrix $\mathbf{T}$ generated according to equation (9), could the underlying deterministic model be recovered in the absence of any other information? We did this by using a numerical procedure utilizing the log-gamma approximation of the stationary distribution of population sizes. The analysis proceeded in the following stages: (1) The matrix $\mathbf{T}$ was solved to yield the stable distribution of density states, $\mathbf{s}^* = \{s_1^*, s_2^*, \ldots, s_k^*\}$. (2) Via the log-gamma approximation, $\mathbf{s}^*$ was used to approximate $P(x)$, the stationary distribution of population sizes for the continuous population model using the algorithm described above. (3) Values of λ, a, and σ_λ were chosen (we restricted this analysis to the simpler model of a population with no seedbank and stable dynamics, i.e., $b = 1$, as the more complex model requires an additional computationally relatively intensive step to numerically estimate the distribution of population sizes). (4) Given the population parameters in step 3 and the approximation of the stable distribution of population sizes in step 2, a new matrix $\mathbf{T}_{\mathrm{rat}}$ was formed according to equation (9) and the methods described above to estimate $\mathbf{T}$. (5) The elements of $\mathbf{T}_{\mathrm{rat}}$ were regressed on the corresponding elements of $\mathbf{T}$, via regression through the origin. The residual variance of this regression was used as a measure of goodness of fit (preliminary analysis indicated that this metric yielded the fastest and most reliable convergence). (6) Steps 3–5 were repeated until the best-fitting estimates of λ, a, and σ_λ were found. The optim function in R (R Development Core Team 2010) was used to find the best-fitting model parameters.

This procedure was used as a simple method for finding the best-fitting estimates of the model parameters in the numerical analyses. This is not intended as a statistical method for analyzing data in real applications in which there will be error in the estimates of the elements of $\mathbf{T}$, among other issues. The aim here is simply to determine whether the underlying parameters of the continuous model can in principle be recovered and hence whether the information contained in the density-structured model is comparable with that contained in a full continuous-density model. This analysis was performed for a range of values of λ between 1.5 and 100 and values of σ_λ between 0.1 and 0.5, and we used $k = 3$, 5, and 10 density states.

Model Results

Examples of Discretized Population Dynamics

Figure 1 shows examples of three population models that differ in several respects. Figure 1*A* is a model of population dynamics in which there is no seedbank and population dynamics are determined by a compensatory density response. In figure 1*B*, there is a persistent bank of seed, while in figure 1*C*, density dependence is overcompensating and populations show persistent two-point cycles.

Figure 2 shows the discretized versions of the models in figure 1, along with the predictions of the stable distribution of density states. The main points to be made from this figure are that (1) the discretized versions of the models still contain information on population dynamics and the variance in parameters between models; (2) as anticipated from equation (9), the stable distribution of density states is able to exactly match the distribution from the continuous model; and (3) although the description of dynamics based on three states is relatively coarse, with as few as five density states the models and predictions provide a reasonable level of resolution.

Because the transition probabilities calculated according to equation (9) are exact, so long as the stable distribution of population sizes exists, the density-structured model is able to provide an accurate characterization of population dynamics irrespective of the underlying continuous model. Thus, the density-structured models are able to accurately

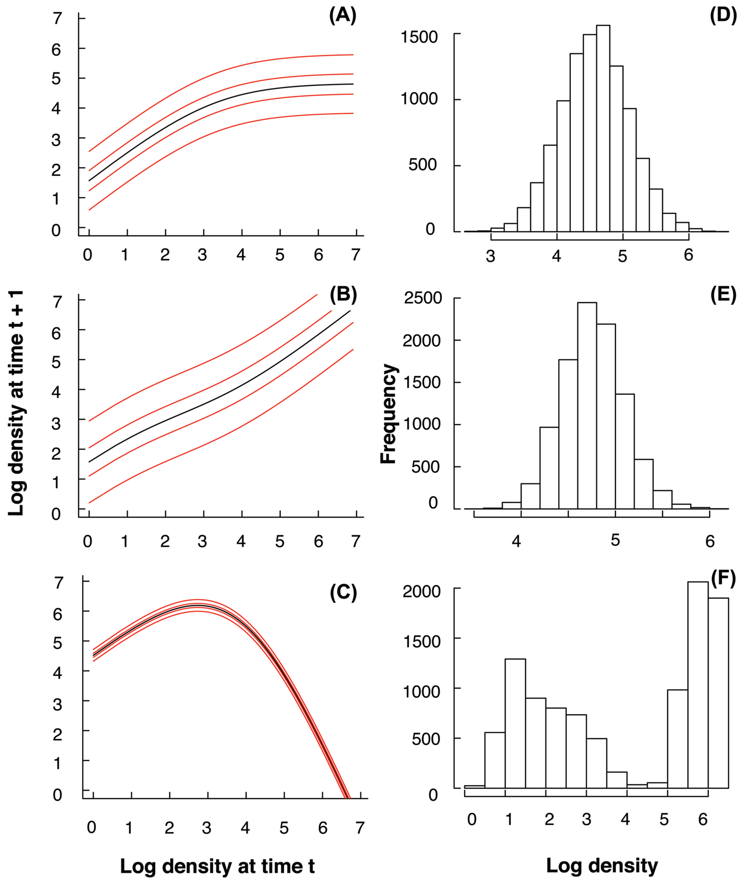

Figure 1: Examples of stochastic population models and corresponding stationary distributions of population sizes. *A*, Compensating density dependence with no seedbank: $\lambda = 10$, $\sigma_\lambda^2 = 0.3$, $N_{eq} = 1{,}000$, $b = 1$, $\gamma = 0$, $\sigma_\gamma = 0$. *B*, Compensating density dependence with a persistent seedbank: $\lambda = 5$, $\sigma_\lambda^2 = 0.5$, $N_{eq} = 100$, $b = 1$, $\gamma = 0.8$, $\sigma_\gamma = 0.2$. *C*, Overcompensating density dependence with no seedbank: $\lambda = 100$, $\sigma_\lambda^2 = 0.1$, $N_{eq} = 100$, $b = 4$, $\gamma = 0$, $\sigma_\gamma = 0$. The black line represents the mean response, and the red lines represent the fiftieth and ninety-fifth percentiles. *D–F* show the corresponding stationary distributions of population sizes.

summarize the model with a persistent seedbank and the one with overcompensating density dependence as well as the one with compensating density dependence.

Accuracy of Approximation of Mean and Variance

The approximations of the stable distributions of population sizes based on the log-gamma distribution provide generally accurate estimates of the mean and variance of population sizes under a range of conditions (see fig. 3 for graphical examples). It is perhaps unsurprising that the estimate of the mean is usually accurate (to within 1% in all simulations), as assignment of density states assumes that the density states encompass the range of densities expected to occur. Table 1, however, summarizes how accurately the log-gamma approximations estimate the standard deviation of the stationary distribution of population sizes (see also appendix tables 1–4).

The log-gamma approximations are very accurate for populations with compensatory density dependence, no seedbank, and varying levels of stochasticity; the only notable deviation occurs when the variance in λ is very high ($\sigma_\lambda = 1$), and λ is low (1.5). As shown in figure 3, the reason why the log-gamma approximations begin to fail is that the distribution of population sizes has a long lower tail when the variance in population growth rate is high and λ is close to 1. The gamma distribution cannot accommodate this fat lower tail, and hence the approximation breaks down. Otherwise the approximation works well, with little bias.

The approximation of population size is also very effective when there is a moderate seedbank. There is vir-

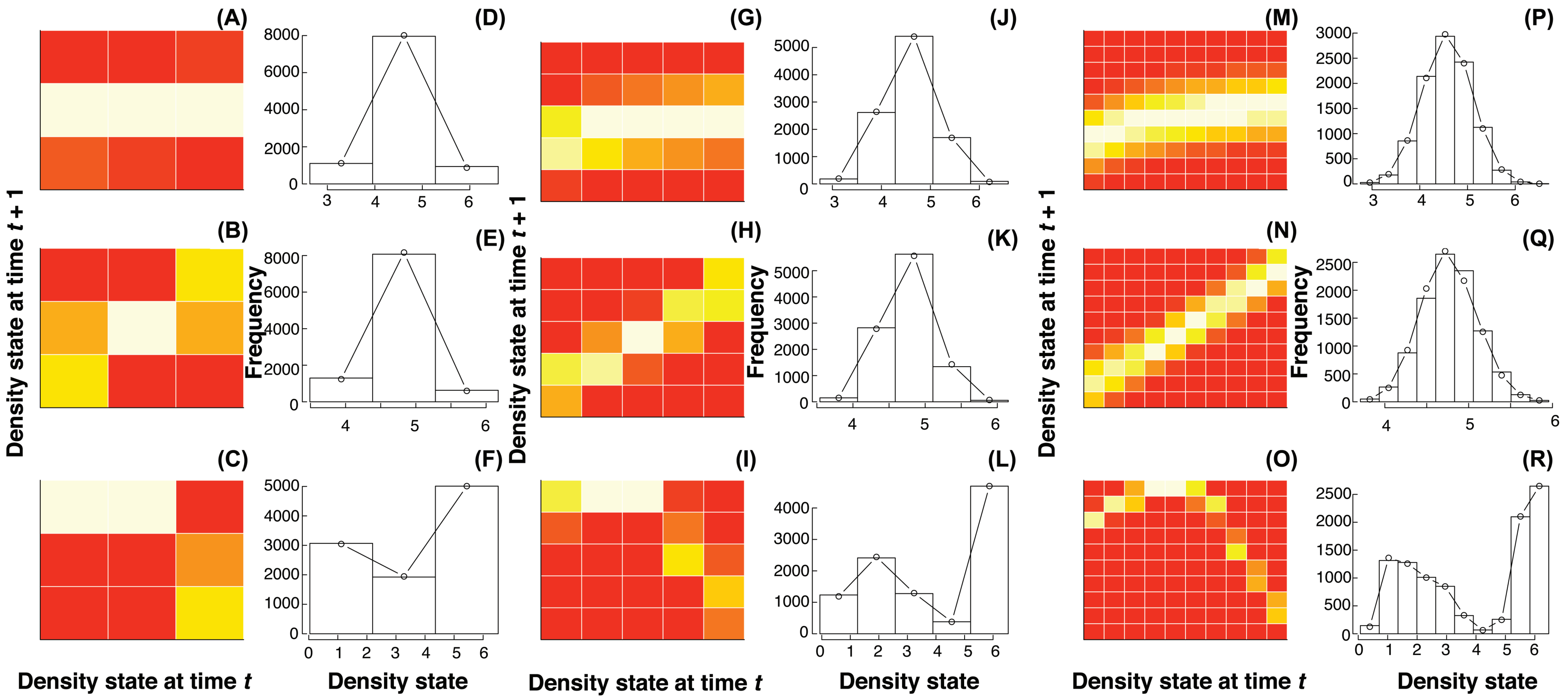

Figure 2: Discretized versions of the models shown in figure 1. The matrices shown are density-structured matrices for population dynamics (*red* for low probability, *yellow* for high probability) calculated according to equation (9). The models correspond to those in figure 1: *A*, *G*, and *M* are discretized versions of figure 1*A*; *B*, *H*, and *N* are discretized versions of figure 1*B*; *C*, *I*, and *O* are discretized versions of figure 1*C*. The number of density states was set at 3 (*A*–*C*), 5 (*G*–*I*), or 10 (*M*–*O*). Also shown for each model are the stable state distributions predicted for each model (*circles connected by lines*), together with the proportions calculated from the stationary distribution of population sizes (*bars*).

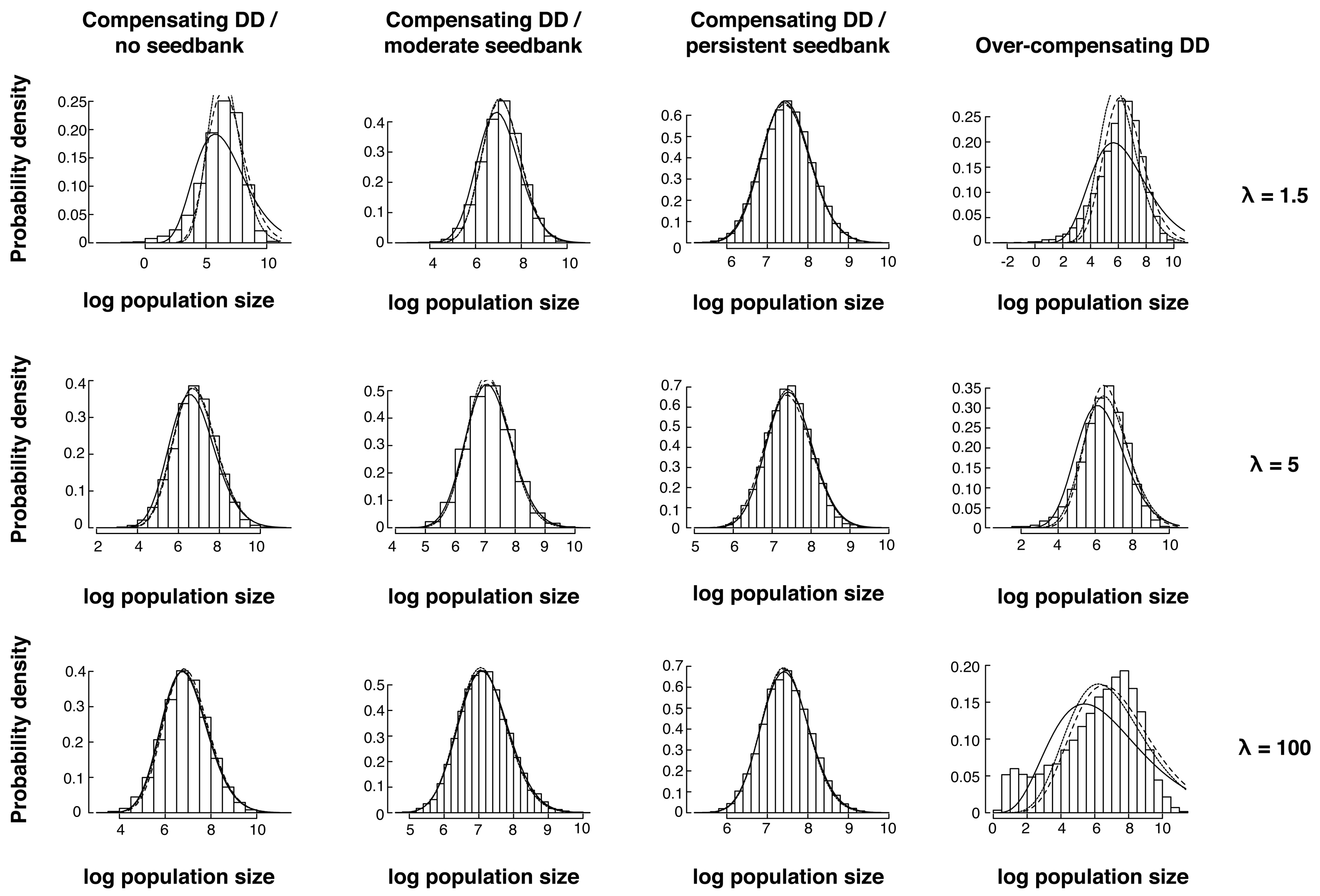

Figure 3: Log-gamma approximations of the stationary distribution of population size using density states. The bars represent the simulated data from a continuous model. The solid lines represent fitted models based on three states, dotted lines those based on five states, and dashed lines those based on 10 states.

Table 1: Bias in estimates of the standard deviation of population size using log-gamma approximations

	Compensating DD/ no seedbank			Compensating DD/ moderate seedbank			Compensating DD/highly persistent seedbank			Overcompensating DD/ no seedbank		
λ, σ_λ	3 states	5 states	10 states	3 states	5 states	10 states	3 states	5 states	10 states	3 states	5 states	10 states
λ = 1.5:												
.1	.00	.00	.00	.00	.00	.00	−.05	.01	.01	.00	.00	.00
.2	−.01	.00	.00	.00	.00	.00	−.04	.00	.02	−.01	−.01	.01
.3	−.03	.00	.01	.00	.00	.01	−.02	.00	.01	−.03	−.01	.01
.4	−.04	.00	.02	.00	.00	.01	−.01	.00	.01	−.06	.01	.02
.5	−.08	.01	.03	.00	.01	.01	.00	.01	.01	−.09	.03	.03
1.0	**−.13**	**.14**	.10	−.01	.02	.02	.00	.01	.01	**−.12**	**.14**	**.12**
λ = 5:												
.1	.00	.00	.00	.00	.00	.00	−.07	.03	.02	.00	.00	.00
.2	.00	.00	.00	.00	.00	.00	−.05	.01	.02	.00	.00	.00
.3	.00	.00	.00	.00	.00	.00	−.03	.00	.01	−.01	.00	.01
.4	.00	.00	.00	.00	.00	.00	−.01	.00	.01	−.03	.00	.01
.5	.00	.00	.00	.00	.00	.01	.00	.01	.01	−.05	.00	.02
1.0	−.01	−.01	.00	−.01	.00	.02	.00	.01	.01	−.09	.02	.06
λ = 10:												
.1	.00	.00	.00	.00	.00	.00	−.07	.02	.02	−.02	−.01	.00
.2	.00	.00	.00	.00	.00	.00	−.05	.01	.02	−.06	.00	.02
.3	.00	.00	.00	.00	.00	.00	−.03	.01	.02	**−.11**	.04	.03
.4	.00	.00	.00	−.01	.00	.01	−.01	.01	.01	**−.15**	.08	.05
.5	.00	.00	.00	−.01	.00	.01	.00	.00	.01	**−.18**	**.10**	.06
1.0	−.01	.00	.00	−.02	.00	.02	.00	.01	.01	**−.11**	.04	.10
λ =20:												
.1	.00	.00	.00	.00	.00	.00	−.07	.02	.02	**−.24**	**−.43**	**−.37**
.2	.00	.00	.00	.00	.00	.00	−.05	.01	.02	**−.11**	−.09	**−.14**
.3	.00	.00	.00	.00	.00	.00	−.03	.01	.02	**−.12**	.06	−.02
.4	.00	.00	.00	−.01	.00	.01	−.01	.00	.01	**−.12**	**.12**	.05
.5	.00	.00	.00	−.01	.00	.01	.00	.01	.01	**−.13**	**.14**	.08
1.0	.00	.00	.00	−.03	.00	.02	.00	.01	.01	**−.11**	.05	**.13**
λ =100:												
.1	.00	.00	.00	.00	.01	.00	−.07	.02	.02	**−.24**	**.64**	**.43**
.2	.00	.00	.00	.00	.00	.00	−.05	.01	.02	−.01	**−.39**	**−.37**
.3	.00	.00	.00	.00	.00	.00	−.03	.01	.02	−.04	**−.35**	**−.28**
.4	.00	.00	.00	−.01	.00	.01	−.01	.00	.01	**−.13**	**−.26**	**−.20**
.5	.00	.00	.00	−.01	.00	.01	.00	.01	.01	**−.24**	**−.13**	**−.12**
1.0	.00	.00	.00	−.03	.01	.02	.00	.01	.01	**−.11**	.00	.05

Note: Estimates of bias of more than 10% greater or less than the true value are shown in boldface. λ is the finite rate of population increase; σ_λ is the standard deviation of λ. DD = density dependence.

tually no bias in the estimates of the standard deviation of population size. This is clear in figure 3, in which the approximation of the distribution of population sizes is very close indeed.

When a persistent seedbank is present, the approximation is generally also quite reasonable. When the number of density states is low ($k = 3$) and the variance in λ is low (0.1), the bias in the estimate of the variance in population size can be as high as 10%. However, as the variance in λ increases or the number of density states is increased to $k = 5$ or 10, this bias declines considerably.

When population dynamics are driven by overcompensating density dependence, the log-gamma approximation performs less well, especially as the mean value of λ increases and population dynamics become intrinsically less stable. For λ = 20 or 100, the approximation is unreliable for any number of density states, although for lower values the approximation is generally reasonable.

Table 2 summarizes the precision of estimates of the standard deviation of population size for models using three and five density states. We did not estimate precision for 10 density states because the 500 transitions employed is too low a number to estimate the large number of density-state transitions (as noted above, 10-state models were included in the previous analysis, although this was anticipated to be too many density states to be practicable).

Table 2: Precision of estimates of the standard deviation of population size using log-gamma approximations

	Compensating DD/ no seedbank		Compensating DD/ moderate seedbank		Compensating DD/highly persistent seedbank		Overcompensating DD/ no seedbank	
λ, σ_λ	3 states	5 states	3 states	5 states	3 states	5 states	3 states	5 states
$\lambda = 1.5$:								
.1	.06	.07	.05	.06	.12	.08	.06	.07
.2	.07	.07	.06	.07	.10	.09	.05	.07
.3	.07	.08	.06	.07	.08	.07	.08	.06
.4	.09	.08	.07	.07	.08	.08	.08	.08
.5	.10	.07	.08	.07	.10	.09	**.11**	.07
1.0	**.23**	**.11**	**.11**	.07	.09	.09	**.25**	**.11**
$\lambda = 5$:								
.1	.04	.06	.05	.05	**.11**	.08	.05	.05
.2	.05	.05	.05	.06	.10	.07	.05	.06
.3	.05	.05	.05	.05	.09	.08	.05	.05
.4	.05	.05	.05	.05	.08	.09	.05	.05
.5	.05	.05	.05	.06	.08	.09	.05	.06
1.0	.06	.06	.05	.06	.10	**.11**	.08	.06
$\lambda = 10$:								
.1	.04	.05	.04	.05	**.11**	.08	.06	.07
.2	.05	.05	.05	.05	.09	.09	.07	.06
.3	.05	.05	.04	.05	.09	.08	.06	.06
.4	.05	.05	.05	.06	**.11**	.09	.07	.07
.5	.04	.06	.05	.06	.07	.08	.07	.07
1.0	.05	.05	.05	.04	.07	.08	.08	.08
$\lambda = 20$:								
.1	.05	.05	.04	.05	.09	.08	NA	.05
.2	.04	.05	.04	.05	.10	.09	.10	.10
.3	.05	.05	.04	.05	.09	.07	.07	.09
.4	.05	.05	.04	.05	.07	.08	.09	.08
.5	.05	.05	.05	.05	.09	.08	.09	.08
1.0	.05	.06	.05	.04	.08	.09	.07	.09
$\lambda = 100$:								
.1	.04	.05	.04	.05	**.11**	.08	NA	NA
.2	.04	.05	.05	.05	.09	.08	NA	.03
.3	.04	.05	.05	.05	.09	.07	.05	.06
.4	.04	.05	.05	.06	.08	.09	.08	.06
.5	.05	.04	.05	.05	.08	.08	**.11**	.07
1.0	.06	.05	.05	.05	.08	.09	.09	.10

Note: Estimates greater than 0.1 (i.e., coefficient of variation in estimates is greater than 10%) are shown in boldface. λ is the finite rate of population increase; σ_λ is the standard deviation of λ. DD = density dependence; NA = not available.

Precision is generally high (i.e., the values of the coefficient of variation are typically low), with the coefficient of variation typically being less than 0.1 and usually less than 0.05. The only exceptions, unsurprisingly, are when population dynamics are highly nonlinear or extremely stochastic ($\sigma_\lambda = 1$). In the former case, it is often not possible to fit the gamma distribution at all for small or moderate sample sizes (the "NA" values in table 2 and appendix tables 5–7).

Precision is frequently the same or even marginally worse as the number of states is increased (table 2; see also appendix tables 5–7). This is because, although increasing the number of states improves the description of the distribution of population size, for the same size of data set, the effective sample size for estimating the transitions between a large number of states is lower than that for a small number of states. For example, with 500 transitions, there are effectively $500/(3 \times 3) = 55.6$ observations per transition in a three-state model, whereas there are effectively $500/(5 \times 5) = 20$ observations per transition in a five-state model.

Recovering the Continuous Process

As shown in table 3, the continuous model is readily recovered from transition matrices, with very little intrinsic

Table 3: Estimates of continuous model parameters from transition matrices

True values of	3 states			5 states			10 states		
λ, σ^2_λ	λ	σ^2_λ	N_{eq}	λ	σ^2_λ	N_{eq}	λ	σ^2_λ	N_{eq}
$\lambda = 1.5$:									
.1	1.50	.10	98.74	1.50	.10	99.62923	1.47	.10	99.42
.2	1.49	.20	97.17	1.44	.20	97.62541	1.47	.20	98.37
.3	1.43	.32	92.19	1.47	.30	95.6521	1.48	.30	99.27
.4	1.42	.44	90.10	1.41	.41	91.39677	1.45	.40	95.13
.5	1.37	.56	80.76	1.38	.51	86.63374	1.39	.51	89.19
1.0	NA	NA	NA	NA	NA	NA	NA	NA	NA
$\lambda = 5$:									
.1	5.07	.10	100.04	4.90	.10	99.96023	5.01	.10	100.23
.2	4.98	.20	100.28	4.87	.20	100.0865	4.83	.20	100.26
.3	4.93	.30	99.63	4.93	.30	99.7484	4.89	.30	100.09
.4	4.89	.40	99.39	4.83	.40	99.7398	4.90	.40	100.03
.5	4.74	.50	99.16	4.79	.50	99.79655	4.36	.50	98.93
1.0	4.00	.98	99.48	4.22	1.00	99.87	4.33	1.00	97.69
$\lambda = 10$:									
.1	9.92	.10	99.92	10.09	.10	100.3133	9.66	.10	99.55
.2	9.97	.20	100.08	9.97	.20	99.66012	9.78	.20	100.32
.3	9.79	.30	99.96	9.96	.30	99.98	9.97	.30	99.73
.4	9.65	.40	99.69	9.76	.40	99.68	9.87	.40	100.02
.5	9.54	.50	99.80	9.77	.50	99.88955	9.71	.49	99.54
1.0	8.63	1.00	99.19	8.38	1.01	98.99	8.95	1.00	100.15
$\lambda = 20$:									
.1	19.80	.10	99.90	20.68	.10	99.70	20.36	.10	100.09
.2	19.77	.20	99.91	20.40	.20	99.96	19.51	.20	99.85
.3	19.72	.30	99.83	19.50	.30	100.07	17.23	.30	100.12
.4	19.49	.40	99.79	19.83	.40	99.96	20.44	.40	100.45
.5	19.27	.50	100.11	19.65	.50	99.92	19.51	.50	100.10
1.0	16.95	1.00	99.60	17.60	1.00	99.91	17.73	1.00	99.83
$\lambda = 100$:									
.1	100.30	.10	100.09	102.31	.10	100.40	98.04	.10	99.70
.2	100.17	.20	100.02	99.63	.20	99.97	97.98	.20	100.17
.3	96.32	.30	99.98	98.30	.30	99.89	101.66	.30	100.03
.4	102.95	.40	100.05	96.56	.40	100.04	94.49	.40	99.62
.5	92.31	.50	99.82	95.27	.50	100.42	96.77	.50	99.86
1.0	82.58	1.00	100.05	84.16	1.00	99.85	89.54	1.00	99.93

Note: λ is the finite rate of population increase; σ_λ is the standard deviation of λ; N_{eq} is the equilibrium population size (assumed to be 100 throughout). NA = not available.

bias. The only exception occurs when the value of λ is low and its variance very high (1). There is a degree of underestimation of λ when the variance in λ is high; however, the extent of the bias is generally low. There is also a small bias in the estimate of N_{eq} when λ is low and variance in λ is high. Estimates of the variance in λ are typically very close indeed to the true values. As emphasized above, the aim of the analysis summarized in table 2 is to show that the parameters of the continuous model can, in principle, be recovered accurately from the discrete transition matrix. However, estimation from real data would require additional steps, specifically to account for statistical uncertainty in estimates of the elements of **T** and to generate variances for the estimated parameters.

Discussion

The density-structured modeling approach is a fundamentally empirical one. The aim is to use large amounts of rapidly collected data to generate models that encapsulate and summarize population dynamics. The empirically generated transition matrices and their projections then become tools for asking how different factors affect population dynamics and for simulating and projecting population sizes. Using simulations, we have shown that potentially, this approach can very closely approximate distributions of population size.

Of course, as with any modeling approach, there are decisions and trade-offs that have to be made. Density-structured approaches have the advantage that they require less intensive data to parameterize. The disadvantages are that they are not explicitly formulated in terms of parameters describing the processes generating population dynamics and that they require reasonable amounts of data to parameterize. No single modeling approach fits all purposes: our aim in this article has been to highlight that for studies that primarily aim to characterize population variability and to account for it in predicting population size, density-structured approaches may offer advantages in that they permit faster data collection. This may permit robust modeling in circumstances when other approaches may be compromised by lack of data (e.g., see Freckleton et al. 2008).

Model Structure

Density-structured models make few assumptions about the processes driving population dynamics. Like other purely empirical approaches, such as Leslie and Lefkovitch matrices (e.g., Caswell 2000), or empirical approaches, such as semiparametric modeling (Ellner et al. 2002), the model is itself a tool for summarizing the data and the underlying dynamics as well as for modeling and population projections.

The advantage of using an empirical approach for modeling is that it is not necessary to make assumptions about the way that the system is structured or to impose particular function for the relationships between key variables (e.g., Ellner et al. 2002). This avoids the problem that misspecifying either of these elements can lead to potentially misleading predictions (e.g., Freckleton et al. 2008). The downside of this empirical focus is that detail may be lost. For instance, using a density-structured approach to model annual census data ignores details of within-year population flux or details of how other processes affect demographic rates. Frequently, only annual census data are available, however, and the potential for such analyses is limited in any case.

Density-structured models are inherently density dependent if density dependence exists in the population. This is because the transition probabilities are integrated across the density-dependent function (fig. 1). This is different from other structured models, such as stage- or age-structured or integral projection models, in which density dependence has to be explicitly incorporated into the transition matrix by allowing the transition probabilities to vary with density (e.g., Alvarez-Bullya 1994; Silva Matos et al. 1999). This is an important advantage, as measuring and modeling density dependence are well known to constitute a difficult part of the process of developing and fitting models to continuous density data (Bulmer 1975; Dennis and Taper 1994; Dennis et al. 2006; Freckleton et al. 2006).

In the analyses presented above, we considered annual plant populations because they present relatively simple life histories. We have shown that the approach is robust to the inclusion of the additional complexity of a long-lived bank of seed. This robustness results from the population achieving a stable stage distribution, for example, also permitting the model for the population to be written as a single difference equation (eq. [4]). We anticipate that the approach could be extended to perennial plants, on the assumptions that stage distributions are relatively stable and that density states are not too heterogeneous. Certainly, the method should be highly applicable for biennials and short-lived perennials.

The idea of taking a coarsened approach to measuring and modeling population data runs counter to trends in the analysis of demographic data on size- or stage-structured populations, in which integral projection models are becoming increasingly commonly used (Easterling et al. 2000; Ellner and Rees 2006, 2007). This approach has been developed because in stage-structured models in which flux between states is modeled by probabilities of growth and stasis, predictions of growth between stages can be too high, with individuals moving from the smallest to the largest states too quickly. This problem can be dealt with by using a model with a continuous state variable (or approximating this by using a very large number of states) and integrating growth and survival functions to generate growth transition probabilities.

Our simulations showed that there is potential for recovery of the continuous process from the discrete transition matrix. Although density-structured models can be used as modeling tools in their own right, this is potentially an extremely valuable application of this approach. What is particularly important is that density-structured data can be collected very much more rapidly than can detailed demographic data. Thus, it may be more economical to attempt to fit demographic models using initially coarse-grained data.

Scale

In the theoretical analysis, we have not been explicit about the scale at which the model operates. In reality, these models can be applied at a range of scales. For instance, Taylor and Hastings (2004) used individual discrete populations as the level of analysis. The proximate application of this approach, however, probably lies in modeling mesoscale dynamics. The typical application of a model would be to a single-site population in which quadrats or patches are delimited and then recensused annually. The spatial variation in density is therefore used as the basis for constructing the transition matrix.

This approach is, in principle, the same as that commonly used for monitoring and modeling plant populations (e.g., Freckleton and Watkinson 2002). The difference is that density-structured data can be collected very rapidly. As an example, in an ongoing study, ∼10,000 quadrats are resurveyed annually by a single researcher (S. A. Queenborough and R. P. Freckleton, unpublished data).

The rapidity of data collection at single sites should facilitate data collection across larger scales; it should be more feasible to collect replicate data at different sites. The data collected could be treated individually at each site, permitting intersite variability to be quantified. As we highlight below (see "Limitations and How to Recognize Them"), preliminary data will be required to set the appropriate spatial scale of modeling as well as to decide the density states and how many of these will be used.

Robustness

As noted above, a density-structured model is not an approximation at equilibrium because the probability in equation (9) is exact. This result is, however, dependent on the denominator of equation (9) being constant and not changing with time or density. If this does change, for example, because the distribution of population sizes is changing, then the model predictions may not be robust (see below).

There has been a great deal of discussion of the importance of accounting for measurement error in models of population dynamics, and numerous techniques have been suggested for dealing with this (Bulmer 1975; Clark and Bjørnstad 2004; Freckleton et al. 2006). Accounting for measurement error is necessary because ignoring it can lead to bias in estimates of the density-dependent components of models (Shenk et al. 1998; Freckleton et al. 2006) as well as misestimation of the effects of environmental effects (Pablo Almaraz et al., unpublished manuscript). Although techniques exist for estimating models in the presence of unknown levels of census error (e.g., Dennis et al. 2006), the most powerful approaches are possible only when census error can be estimated (reviewed by Freckleton et al. 2006).

Through the use of density-structured models, it should be reasonably straightforward to include measurement error in the modeling process, as the fitting of these models is based on vector-generalized linear models, which are closely related to conventional linear modeling approaches. Methods such as SIMEX (simulation-extrapolation; Cook and Stefanski 1994) are straightforward to implement and can be used to correct for bias, and for misclassification of discrete states a method called MC-SIMEX (misclassification simulation-extrapolation) has been developed (Küchenhoff et al. 2006). Alternatively, state-space models (Harvey 1989) allow a combined model for the process of generating the true data as well as the error to be specified and fitted simultaneously.

Dealing with Density-Independent Populations

One of the assumptions of the approach is that the population is at or near a stable state distribution and that an equilibrium exists with a quasi-stationary distribution of population sizes. If this is not true, then the equilibrium population size and its variance, by definition, do not exist. To obtain insight into how this affects the analysis of data from density-structured models, note that if population dynamics are density independent, then equation (9) becomes

$$P(j,\, i \,|\, P_1(x)) = \frac{\int_{N_i^-}^{N_i^+}\int_{N_j^-}^{N_i^+} G(y;\, \lambda x,\, \sigma_\lambda^2) P_1(x)\, dy\, dx}{\int_{N_j^-}^{N_i^+} P_1(x)\, dx}. \qquad (11)$$

The estimated transition probabilities are dependent on $P_t(x)$, which is the distribution of population sizes at time t. This will change with t; that is, it is not stationary. However, two important conclusions can be drawn from equation (11). First, for that time period, transition probabilities calculated according to equation (11) are exact and not approximations. Thus, interannual changes will be measured correctly. Second, if an approximation for the distribution of population sizes can be obtained, density-structured data could be used to estimate rates of population increase or decrease. The algorithm for doing this would follow that used above to fit the model for density-dependent population dynamics.

Another approach is possible when a long time series of data is available: it should be possible to generate annual estimates of transition probabilities and to relate these to the average or variance in density. This can be used to generate a model for any changes in transition probabilities that result from changes in the distribution of population sizes; this model can be used as a heuristic adjustment. In addition, annual matrices can be used to generate annual predictions of population sizes and population growth rates. These can be used to generate estimates of how sensitive the model predictions are to violation of the model assumptions.

Practical Considerations

One of the prime reasons for using a density-structured modeling approach in a field program is that it should permit census data to be collected quickly and relatively inexpensively. However, there are several important considerations in the design of such a study. The key first steps involve deciding what the density states should be and training observers to estimate these. The exact way that these are done obviously depends precisely on the study system. We have used density-structured data as the basis for designing large-scale field monitors of arable weeds and sand dune annuals, and the details of how we designed these programs are described elsewhere (Queenborough et al., forthcoming). In doing this, we have come to the following broad conclusions: (1) initial data are invaluable in designating the data states; however, for many systems such data exist in the literature or can be readily collected in pilot analyses; and (2) it is relatively straightforward to train observers to monitor and estimate density states. This can be done either by having fixed criteria for visual assignment of density or by using small subsamples, within the main sampling units, that are used to assign states on a numerical basis.

Dealing with the issue of measurement error is an important consideration in the design of monitoring programs that employ a density-structured approach (Queenborough et al., forthcoming). It is relatively straightforward to conduct replicated surveys to measure within-observer variability as well as the variability between observers. This may be done by having an observer conduct repeat surveys of a subset of populations or by surveying a single population several times using different observers. This should not add significantly to the workload in a field program; however, the advantages of doing so are huge (Dennis et al. 2009).

One of the downsides of using transition matrices is that each matrix contains k^2 entries, each of which has to be estimated. It might therefore be argued that the data requirements would be prohibitive for accurate parameterization. On the other hand, the rapidity with which data can be collected should enable large data sets to be built up that encapsulate all commonly observed transitions. Moreover, as shown in the examples in figure 2, a high proportion of transition probabilities might be expected to be 0 in any case, which would reduce the data requirements. The simulation results showed that reasonable results can be expected for between 200 and 500 samples, depending on the number of density states and the nature of population dynamics. As noted above, this is not unreasonable, compared with sample sizes in previous detailed demographic modeling analyses.

Limitations and How to Recognize Them

Our intention in this article has been to highlight the possible utility of density-structured models and data in plant ecology. One single modeling approach cannot deal with all eventualities, and just as density-structured models will be preferable under some circumstances, there are conditions under which they will not be suitable. Here we outline what such conditions might be.

First, as outlined above, if population dynamics are nonequilibrium or there is no quasi-stationary distribution of population sizes, it is not possible to capture population dynamics and predict population sizes with a Markov model. As suggested above, density-structured models and data can have some value under such circumstances; however, they would have to be carefully analyzed.

Second, if there are complex life-history divisions within the population, then this can generate complexity that cannot be averaged over in a similar manner. For example, the convenient approximation in equation (4) allows the complexity of a seedbank to be incorporated in a model of little more complexity (i.e., a single difference equation) than one without a seedbank. Complex age or stage dependence would complicate matters, however. This would require additional states to be included within the model, with the consequence that the number of parameters would be inflated. If this complexity were ignored, the model predictions could be compromised.

Finally, the simulation results highlight conditions under which the performance of the approach may suffer. This is particularly the case when stochasticity is extreme or population dynamics are highly nonlinear. It should be pointed out that the density-structured approach is not inherently biased or flawed under such conditions (e.g., fig. 2; eq. [9]). Rather, it may prove difficult to obtain reliable approximations of population size distributions under such circumstances (tables 1, 2; fig. 3). The main conclusion here is that for populations in which stochasticity is extreme or population dynamics are nonlinear, large amounts of data and relatively high numbers of density states will be required.

These considerations all underline, as emphasized above, that before a practical program of data collection can be embarked upon, a certain amount of preliminary information is required. This includes a qualitative understanding of the population ecology of the species to be studied and some information on typical ranges of density, as well as the likelihood of extreme variations in density. Usually, a good proportion of such information is available from the literature or from pilot studies.

In our analyses, we have used one approach to generating predictions of the mean and variance of population size from density-structured models and data, namely, to treat the distribution of predicted density states as quadrature points and use these to fit a gamma distribution. In general, numerical quadrature of this type is extremely accurate if the data are taken from the assumed distribution (e.g., as in Gauss quadrature; Abramowitz and Stegun 1972). As we show in the appendix, because the distribution of log population sizes is often very well approximated by a gamma distribution, even with only two density states it is often possible to obtain an unbiased estimate of the standard deviation of population size.

In practical applications, it should be straightforward to diagnose failure of the gamma model. The fit of the gamma model will be poor, and goodness-of-fit tests (e.g., χ^2 test) can be used to test whether a fitted gamma model is adequate. If the gamma model does not fit, then the simulation results (fig. 3) suggest two immediate modifications. First, if it suspected that the dynamics are nonlinear, then a multimodal distribution could be fitted. This could be done by fitting mulitmodal gamma distributions (e.g., Dennis and Patil 1984). In other cases, the lack of fit could be related to a long lower tail. Options that could be explored here include the Adès distribution (e.g., Perry and Taylor 1985), which is a power transformation of a gamma distribution.

It is important to point out that the performance of approximations should be regarded as a separate issue from the intrinsic bias of density-structured models. Used by themselves, the predictions of density-structured models should be unbiased. Predictions of rates of transitions between states of the proportions of sites in each density state should be the same as would be obtained by discretizing the predictions of a continuous model.

Concluding Remarks

In the modeling and analysis presented above, we explored how density-structured models can be applied to an important problem, namely, predicting how population dynamics can be modeled and predicted while accounting for temporal variability. The key elements that the models included are density dependence and temporal variability that leads to a distribution of population sizes around the mean.

Density-structured models can offer a useful starting point for analysis of population dynamics. They can be used to simplify the data and modeling requirements, especially in terms of mathematical and computational intensity. This can allow the data to be explored and initial hypotheses to be tested and refined. For designing large-scale monitoring programs and in using data from these to generate forecasting tools, density-structured models should be an extremely powerful approach.

Acknowledgments

R.P.F. is a Royal Society University Research Fellow. Thanks to I. Cook and M. Rees for comments on an earlier draft of this manuscript. This work was funded in part by a Leverhulme Trust Research Leadership Award to R.P.F.

Literature Cited

Abramowitz, M., and I. A. Stegun. 1972. Handbook of mathematical functions with formulas, graphs and mathematical tables. Dover, New York.

Alvarez-Buylla, E. R. 1994. Density dependence and patch dynamics in tropical rain forests: matrix models and applications to a tree species. American Naturalist 143:155–191.

Bogich, T., and K. Shea. 2008. A state-dependent model for the optimal management of an invasive metapopulation. Ecological Applications 18:748–761.

Bolker, B. M. 2008. Ecological models and data in R. Princeton University Press, Princeton, NJ.

Bulmer, M. G. 1975. The statistical analysis of density dependence. Biometrics 31:901–911.

Caswell, H. 2000. Matrix population models. Sinauer, Sunderland, MA.

Clark, J. S., and O. Bjørnstad. 2004. Population time series: process variability, observation error, missing values, lags, and hidden states. Ecology 85:3140–3150.

Cook, J. R., and L. A. Stefanski. 1994. Simulation-extrapolation estimation in parametric measurement error models. Journal of the American Statistical Association 89:1314–1328.

Dennis, B., and R. F. Costantino. 1988. Analyses of steady-state populations with the gamma abundance model: application to *Tribolium*. Ecology 69:1200–1213.

Dennis, B., and G. P. Patil. 1984. The gamma distribution and weighted multimodal gamma distributions as models of population abundance. Mathematical Biosciences 68:187–212.

Dennis, B., and M. L. Taper. 1994. Density dependence in time series of observations of natural populations: estimation and testing. Ecological Monographs 64:205–224.

Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples. 2006. Estimating density dependence, process noise and observation error. Ecological Monographs 76:323–341.

Dennis, B., J. M. Ponciano, and M. L. Taper. 2009. Replicated sampling increases efficiency in monitoring biological populations. Ecology 91:610–620.

Easterling, M. B., S. P. Ellner, and P. M. Dixon. 2000. Size-specific sensitivity: applying a new structured population model. Ecology 81:694–708.

Ellner, S. 1984. Asymptotic behavior of some stochastic difference equation population models. Journal of Mathematical Biology 19: 169–200.

Ellner, S. P., and M. Rees. 2006. Integral projection models for species with complex demography. American Naturalist 167:410–428.

———. 2007. Stochastic stable population growth in integral projection models: theory and application. Journal of Mathematical Biology 54:227–256.

Ellner, S. P., Y. Seifu, and R. H. Smith. 2002. Fitting population dynamic models to time series data by gradient matching. Ecology 83:2256–2270.

Firbank, L. G., and A. R. Watkinson. 1986. Modelling the population dynamics of an arable weed and its effects upon crop yield. Journal of Applied Ecology 23:147–159.

Freckleton, R. P., and A. R. Watkinson. 1998. How does temporal variability affect predictions of weed population numbers? Journal of Applied Ecology 35:340–344.

———. 2002. Are weed population dynamics chaotic? Journal of Applied Ecology 39:699–707.

Freckleton, R. P., A. R. Watkinson, P. M. Dowling, and A. R. Leys. 2000. Determinants of the abundance of invasive annual weeds: community structure and non-equilibrium dynamics. Proceedings of the Royal Society B: Biological Sciences 267:1153–1161.

Freckleton, R. P., W. J. Sutherland, R. E. Green, and A. R. Watkinson. 2006. Census error and the detection of density dependence. Journal of Animal Ecology 75:837–851.

Freckleton, R. P., W. J. Sutherland, A. R. Watkinson, and P. A. Stephens. 2008. Modelling the effects of management on population dynamics: some lessons from annual weeds. Journal of Applied Ecology 45:1050–1058.

Green, R. E. 2008. Assessing the impact of culling on population size in the presence of uncertain density dependence: lessons from a great cormorant population. Journal of Applied Ecology 45:1683–1688.

Grøtan, V., B.-E. Sæther, F. Filli, and S. Engen. 2008. Effects of climate on population fluctuations of ibex. Global Change Biology 14: 218–228.

Harvey, A. C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.

Hassell, M. P. 1975. Density-dependence in single-species populations. Journal of Animal Ecology 44:283–295.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, NJ.

Küchenhoff, H., S. M. Mwalli, and E. Lesaffre. 2006. A general method for dealing with misclassification in regression: the misclassification SIMEX. Biometrics 62:85–96.

Lande, R., B.-E. Sæther, and S. Engen. 2003. Stochastic population dynamics in ecology and conservation. Oxford University Press, Oxford.

MacDonald, N., and A. R. Watkinson. 1981. Models of an annual plant population with a seedbank. Journal of Theoretical Biology 93:643–653.

May, R. M. 1974. Biological populations obeying difference equations: stable points, stable cycles, and chaos. Science 186:645–647.

———. 1989. Levels of organization in ecology. Pages 339–363 *in* J. M. Cherrett, ed. Ecological concepts: the contribution of ecology to an understanding of the natural world. Blackwell, Oxford.

May, R. M., and G. F. Oster. 1976. Bifurcations and dynamic complexity in simple ecological models. American Naturalist 110:573–599.

Maynard Smith, J. 1973. Models in ecology. Cambridge University Press, Cambridge.

Moore, A. D., and I. R. Noble. 1990. An individualistic model of vegetation stand dynamics. Journal of Environmental Management 31:61–81.

Pacala, S. W., C. D. Canham, J. Saponara, J. A. Silander Jr., R. K. Kobe, and E. Ribbens. 1996. Forest models defined by field measurements: estimation, error analysis and dynamics. Ecological Monographs 66:1–43.

Perry, J. N., and L. R. Taylor. 1985. Adès: new ecological families of species-specific frequency distributions that describe repeated spatial samples with an intrinsic power-law variance-mean property. Journal of Animal Ecology 54:931–953.

Queenborough, S. A., K. M. Burnet, W. J. Sutherland, A. R. Watkinson, and R. P. Freckleton. Forthcoming. From meso- to macroscale population dynamics: a new density-structured approach. Methods in Ecology and Evolution, doi:10.1111/j.2041-210X.2010.00075.x.

Ranta, E., P. Lundberg, and V. Kaitala. 2006. Ecology of populations. Cambridge University Press, Cambridge.

R Development Core Team. 2010. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Rees, M., and M. J. Crawley. 1991. Do plant populations cycle? Functional Ecology 5:580–582.

Rees, M., and Q. Paynter. 1997. Biological control of Scotch broom: modelling the determinants of abundance and the potential impact of introduced insect herbivores. Journal of Applied Ecology 34: 1203–1221.

Rees, M., P. J. Grubb, and D. Kelly. 1996. Quantifying the impact of competition and spatial heterogeneity on the structure and dynamics of a four-species guild of winter annuals. American Naturalist 147:1–32.

Royama, T. 1992. Analytical population dynamics. Chapman & Hall, London.

Sæther, B.-E., S. Engen, V. Grøtan, W. Fielder, E. Matthysen, M. E. Visser, J. Wright, et al. 2007. The extended Moran effect and large-scale synchronous fluctuations in the size of great tit and blue tit populations. Journal of Animal Ecology 76:315–325.

Shenk, T. M., G. C. White, and K. P. Burnham. 1998. Sampling variance effects on detecting density dependence from temporal trends in natural populations. Ecological Monographs 68:445–463.

Sibly, R. M., D. Barker, M. C. Denham, J. Hone, and M. Pagel. 2005. On the regulation of populations of mammals, birds, fish and insects. Science 309:607–610.

Sibly, R. M., D. Barker, J. Hone, and M. Pagel. 2007. On the stability of populations of mammals, birds, fish and insects. Ecology Letters 10:970–976.

Silva Matos, D. M., R. P. Freckleton, and A. R. Watkinson. 1999. The role of density dependence in the population dynamics of a tropical palm. Ecology 80:2635–2650.

Stephens, P. A., A. F. Russell, A. J. Young, W. J. Sutherland, and T. H. Clutton-Brock. 2005. Dispersal, eviction, and conflict in meerkats (*Suricata suricatta*): an evolutionarily stable strategy model. American Naturalist 165:120–135.

Stott, I., S. Townley, D. Carslake, and D. J. Hodgson. 2010. On reducibility and ergodicity of population projection models. Methods in Ecology and Evolution 1:242–252.

Sutherland, W. J. 1996. From individual behaviour to population ecology. Blackwell, Oxford.

Taylor, C. M., and A. Hastings. 2004. Finding optimal control strategies for invasive species: a density-structured model for *Spartina alterniflora*. Journal of Applied Ecology 41:1049–1057.

Taylor, H. M., and S. Karlin. 1998. An introduction to stochastic modelling. Academic Press, New York.

Watkinson, A. R. 1980. Density-dependence in single-species populations of plants. Journal of Theoretical Biology 83:345–357.

Watkinson, A. R., R. P. Freckleton, and L. Forrester. 2000. Population dynamics of *Vulpia ciliata*: regional, patch and local dynamics. Journal of Ecology 88:1012–1029.

Associate Editor: Tim Coulson
Editor: Mark A. McPeek