Ecology, 84(11), 2003, pp. 3064–3077

BETTER INFERENCES FROM POPULATION-DYNAMICS EXPERIMENTS USING MONTE CARLO STATE-SPACE LIKELIHOOD METHODS

PERRY DE VALPINE[1]

National Center for Ecological Analysis and Synthesis, 735 State Street, Suite 300, Santa Barbara, California 93101 USA

Abstract. In experimental population ecology, there is often a gap between realistic models used to hypothesize about population dynamics and statistical models used to analyze data. Ecologists routinely conduct experiments where the data from each replicate are short time series of estimated population abundances structured by stage, species, and/or other information, and the conventional test for treatment effects uses a general linear model (GLM) such as analysis of variance (ANOVA). However, GLMs do not incorporate demographic relationships between abundances through time. An alternative is to use population-dynamics models as frameworks for statistical hypothesis testing. This approach requires general methods for fitting structured population models that can incorporate both process noise (stochastic dynamics) and observation error (inaccurate data). This paper presents such methods and compares them to GLMs for testing population-dynamics hypotheses from experiments. The methods are Monte Carlo state-space likelihood methods, including a basic Monte Carlo integration method and a recently developed Monte Carlo kernel likelihood method.

Three simulated examples of population-dynamics experiments were used to compare analysis with a population model to ANOVA, analysis of covariance (ANCOVA), and repeated-measures ANOVA. The examples considered manipulations of host-plant growth conditions, causing decreased survival and increased fecundity; predator addition to investigate a behaviorally mediated change in prey demography; and changed host-plant growth conditions with a more complex model for herbivore dynamics than the one used for analysis. For the first example, a population model gave much higher statistical power than any of the ANOVA methods and provides greater biological insight. For the second example, ANOVA models are not suited to test for the behavioral effect, but a population model detected it with high statistical power. The third example suggests that even incorrect biological structure can provide better inferences than omitting all biological structure. The likelihood methods presented here make analysis with structured population models feasible for a wide range of models incorporating process noise and observation error, thus offering higher statistical power and greater biological insight for population-dynamics experiments.

Key words: kernel density estimation; measurement error; Monte Carlo maximum likelihood; population dynamics; population model fitting; process error; stage-structured demography; state-space model, nonlinear; statistical power.

INTRODUCTION

In experimental population ecology there is often a gap between realistic models used to hypothesize about population dynamics and statistical models used to analyze data (Dennis et al. 1995, Kendall et al. 1999). Depending on the phenomena under study, realistic models might include stage-structured demography, nonlinear species interactions, movement, behavior, or other complicated processes (Metz and Diekmann 1986, Tuljapurkar and Caswell 1997, Gurney and Nisbet 1998). On the other hand, experiments produce data that might have the potential to distinguish between hypotheses about such processes, but they are analyzed with models—most commonly general linear models (GLMs) such as analysis of variance (ANOVA) and linear regression—that do not incorporate these processes. This disjunction limits ecologists' ability to investigate population dynamics using experiments.

Here I present a general method for fitting stage-structured population models to data and compare inference using population models to ANOVA, ANCOVA, and repeated-measures ANOVA. The idea of analyzing population data with population models has a long history (Kendall et al. 1999, Bjornstad and Grenfell 2001) but has been used almost exclusively for long observational time series rather than for replicated experimental time series, with a few exceptions such as by Dennis et al. (1995, 2001), Bjornstad et al. (1998, 2001), Ives et al. (1999), and Gibson et al. (1999). I examine the potential for population models to complement or replace analysis-of-variance models (and other non-dynamic models) as a framework for de-

Manuscript received 14 January 2002; revised 14 January 2003; accepted 19 January 2003; final version received 10 March 2003. Corresponding Editor: O. N. Bjornstad

[1] Present address: Department of Integrative Biology, VLSB 3060, University of California, Berkeley, California 94720-3140 USA. E-mail: perrydev@uclink.berkeley.edu

tecting treatment effects in experimental population time-series data.

The methods I present use a state-space model structure to incorporate both process noise (PN) and observation error (OE) in population models. ''Process noise'' refers to stochasticity in population dynamics, and ''observation error'' refers to inaccuracy in observations. The foremost obstacle to widespread use of population models as statistical frameworks is the lack of general methods for fitting complex models to data incorporating both PN and OE. A related difficulty is that realistic models often include unobserved variables, such as stage classes or species. Even for the simulated examples given here, incorporation of PN, OE, and incomplete observations was necessary for the examples to be realistic. Because state-space likelihood calculations require high-dimensional integration, Monte Carlo numerical integration methods are used for maximum-likelihood parameter estimation (and hence likelihood-ratio hypothesis testing).

In recent years, methods from the statistical literature for likelihood-based inference incorporating PN and OE have filtered into ecology, mostly for fisheries time series, with classical maximum likelihood for linear models (Mendelssohn 1988, Sullivan 1992, Pella 1993, Gudmundsson 1994, Schnute 1994, Freeman and Kirkwood 1995, Kimura et al. 1996), Bayesian analysis for relatively simple, discrete-time, nonlinear models (McAllister et al. 1994, Bjornstad et al. 1999, Meyer and Millar 1999, Quinn and Deriso 1999, Millar and Meyer 2000), and classical maximum likelihood for similar nonlinear models (de Valpine and Hastings 2002), but also for more complex models of plant–pathogen dynamics (Gibson and Renshaw 1998). The methods presented here give Monte Carlo approximate likelihood calculations for structured population models with PN and OE, which also naturally address the problem of incomplete observations. The models could include approximately continuous stage structure, multiple species, spatial structure, behavior, or almost any other process, subject to computational limitations that are not trivial. These methods are used here to compare between ANOVA and population models for testing experimental hypotheses from simulated data. The three examples include stage-structured prey dynamics, predation with a Type II functional response, and analysis of data generated from a more complicated model than the one used for fitting.

The next section introduces the differences between ANOVA models and population models as frameworks for hypothesis testing. Then I introduce the state-space framework for incorporating PN and OE and introduce methods for Monte Carlo maximum-likelihood estimation. Finally I give three examples comparing population models to GLMs in terms of statistical power and biologically meaningful inference.

ANOVA vs. Population Models

Consider a hypothetical experiment to ask whether populations of an herbivore have different dynamics in different growth conditions. Suppose an experiment is run with replicates of each set of conditions, and the data comprise surveys of different stage classes at several times. Conventionally, ANOVA or repeated-measures ANOVA might be used to analyze this experiment. The latter model includes a static effect of the growth conditions on abundance of each stage class, with time incorporated as if it were a treatment factor or a random effect. To test for a treatment effect, one fits a model to the data under null and alternative hypotheses and calculates the *P*-value significance level. The shortcoming of ANOVA here is that the model omits demographic relationships between numbers at one time and numbers at the previous time in the same or previous life stages. An alternative approach would be to use a stochastic, structured population model as the framework for hypothesis testing. A stochastic population model defines a distribution of possible observations, which may be difficult to calculate but can, in principle, be used for approximate likelihood-ratio hypothesis testing. In this way, a biological model can provide a tighter link between hypothesized processes and statistical tests to detect the patterns produced by those processes.

With a population model for this hypothetical experiment, an example null hypothesis is that demographic parameters are the same between treatment groups, with the alternative that parameters differ between groups. This can be tested by fitting models under the null and alternative hypotheses using maximum-likelihood methods and estimating a *P* value with asymptotic approximate likelihood-ratio distributions. With this approach, all of the separate analyses that might be conducted with ANOVA models can be united under one modeling framework, and hypotheses that are difficult to consider with ANOVA can be simple to express with a population model.

Maximum-likelihood estimation and likelihood-ratio testing are useful for a number of reasons (e.g., Dennis et al. 1995, Hilborn and Mangel 1997, Severini 2000). Roughly speaking, as the amount of data increases, maximum-likelihood parameter estimates (MLEs) become unbiased, normally distributed, and have the minimum variance possible among estimation schemes. MLEs may be thought of as ''optimal'' in these specific senses. ANOVA has a close relationship to likelihood-ratio tests because, as the amount of data increases, ANOVA converges to being identical with a likelihood-ratio test. Thus hypothesis testing is conceptually similar for an ANOVA model and a population model: both compare null and alternative hypotheses in ways that are rooted in asymptotic likelihood theory.

Many other types of population-dynamics hypotheses are inaccessible to ANOVA. For example, hypoth-

eses about behavioral effects on demographic rates such as when prey avoid predators (see *Examples: Example 2*, below; Schmitz et al. 1997, Peacor and Werner 2001), ontogenetic changes in species interactions (Polis and Strong 1996, Lundvall et al. 1999), movement and spatial dynamics (Kareiva 1987, Walde 1991, Ellner et al. 2001), intraguild predation (Polis 1991, Rosenheim et al. 1995), indirect effects (Strauss 1991, Wootton 1994), and emergent impacts of multiple predators (Sih et al. 1998) might all be studied with population-dynamics experiments. However, even when it is feasible to conduct experiments that might reveal effects of these processes, ANOVA is poorly suited to detect those effects because it omits the biological relationships between different species and life stages.

In summary, a key for conducting estimation and inference with general model structures is the ability to calculate likelihood values. Calculating likelihoods and finding maximum-likelihood parameters are complicated for structured population models that include both process noise and observation error, and they are the central technical challenges addressed by this paper.

State-Space and Monte Carlo Likelihoods

The central problem is to find maximum-likelihood parameter estimates for any population model that includes process noise (PN) and observation error (OE). First we must consider the form of the likelihood itself, which involves a state-space model or missing-data model framework (Harvey 1989, Schnute 1994, Robert and Casella 1999, de Valpine and Hastings 2002). The term ''state space'' refers to the mathematical space of all possible true population trajectories, which are jointly distributed with any set of observations.

There are three fundamental types of quantities involved: states, observations, and parameters. The states describe the true (but unknown) dynamics of the population. For example, the states may be the true stage-class abundances through time and space. The observations are estimates of some or all of the states (or of functions of the states). For example, the observations might be estimates of only adults at several times, or of eggs, juveniles, and adults, or of adult numbers and stage-class proportions. Parameters are the usual sorts of birth, growth, and survival rates of population models, and they are not treated as random variables (i.e., this paper does not take a Bayesian approach). Parameters will be written together in a vector, $\boldsymbol{\Theta}$.

For each replicate, let all of the population states through time be listed as a vector, $\mathbf{X}$, which includes all stage structure, all species, and all times. It is a full record of the trajectory of the system through time. Similarly, let all of the random PN values that affect the population trajectory through time be a vector, $\boldsymbol{\nu}$. Given parameters $\boldsymbol{\Theta}$, the model takes the PNs and produces the population states:

$$\mathbf{X} = F_{\boldsymbol{\Theta}}(\boldsymbol{\nu}). \tag{1}$$

To incorporate OE, let all of the random values that cause OE be a vector, $\boldsymbol{\varepsilon}$, and denote the model of how observations depend on states and OEs as:

$$\mathbf{Y} = G_{\boldsymbol{\Theta}}(\mathbf{X}, \boldsymbol{\varepsilon}). \tag{2}$$

The likelihood of the observations is defined as the probability density value of the observations, as a function of the parameters. States, $\mathbf{X}$, and observations, $\mathbf{Y}$, are jointly distributed random variables, which are functions of the random variables $\boldsymbol{\nu}$ and $\boldsymbol{\varepsilon}$ as well as the parameters, $\boldsymbol{\Theta}$. None of the elements of $\mathbf{X}$ and $\mathbf{Y}$ are necessarily independent, and only the $\mathbf{Y}$ are known. The likelihood of the observations is thus the marginal probability of the observations:

$$L(\boldsymbol{\Theta}) = \int P_{\boldsymbol{\Theta}}(\mathbf{Y}, \mathbf{X})\, d\mathbf{X} \tag{3}$$

where $P_{\boldsymbol{\Theta}}$ is the joint probability density of states and observations, which is defined by the models, Eqs. 1 and 2.

An alternative way to view the likelihood is to focus on PNs rather than population states. Since the PNs determine the population states, we could view the random variables of interest as $\boldsymbol{\nu}$ and $\mathbf{Y}$, rather than $\mathbf{X}$ and $\mathbf{Y}$. Then the likelihood can be written as

$$L(\boldsymbol{\Theta}) = \int P_{\boldsymbol{\Theta}}(\mathbf{Y}, \boldsymbol{\nu})\, d\boldsymbol{\nu} \tag{4}$$

where $P_{\boldsymbol{\Theta}}(\mathbf{Y}, \boldsymbol{\nu})$ is the joint probability of $\mathbf{Y}$ and $\boldsymbol{\nu}$. Where there is no room for confusion, I use P or $P_{\boldsymbol{\Theta}}$ in a generic way for probability density, with the arguments indicating what probability density is considered. Thus $P_{\boldsymbol{\Theta}}(\mathbf{Y}, \boldsymbol{\nu})$ is the probability density of noises and observations, while $P_{\boldsymbol{\Theta}}(\mathbf{Y}, \mathbf{X})$ is the probability density of states and observations. The two forms of the likelihood, Eqs. 3 and 4, are equal, but for the examples later Eq. 4 is useful, so I will use it from here on. (Note that it would also be possible to work with the space of random variables affecting individuals in an individual-based model, but this is typically very high dimensional.)

In an experimental situation with N replicates, the total likelihood is a product of integrals of the form in Eq. 4:

$$L(\boldsymbol{\Theta}) = \prod_{i=1}^{N} \int P_{\boldsymbol{\Theta}}(\mathbf{Y}_i, \boldsymbol{\nu}; i)\, d\boldsymbol{\nu} \tag{5}$$

where $\mathbf{Y}_i$ is the data vector for experimental unit i, and $P_{\boldsymbol{\Theta}}(\mathbf{Y}_i, \boldsymbol{\nu}; i)$ is the joint density of $\mathbf{Y}_i$ and $\boldsymbol{\nu}$ for that replicate. Typically this density function would be the same for different replicates of the same treatment group but different across treatment groups. If one can obtain maximum-likelihood estimates under null and alternative hypotheses, then P values can be estimated using the standard likelihood approximation that $-2\,(\log(L(\boldsymbol{\Theta}_0)/L(\boldsymbol{\Theta}_{\text{ALT}}))) \sim \chi^2_{\text{df}}$ (e.g., Hilborn and Man-

gel 1997, Severini 2000). Here $L(\Theta_0)$ and $L(\Theta_{ALT})$ are the maximum likelihoods under null and alternative hypotheses, respectively; df is the difference in the number of parameters between null and alternative hypotheses; and χ^2_{df} is the chi-squared distribution with df degrees of freedom.

The challenge is to calculate these likelihoods and find maximum-likelihood parameter estimates (MLEs). For nonlinear, non-Gaussian models, the integral may be high dimensional and have no closed-form solution. Approximating high-dimensional likelihood integrals numerically has been an intense focus in statistics research in recent years (e.g., Gilks et al. 1996, Robert and Casella 1999). In simple problems, one can use a discrete grid to approximate the range of values of the integration variables, but in high-dimensional problems grids become impractical. For state-space models that are Markov processes with a low-dimensional state space at each time step, the likelihood can be factored into sequential integrals that are feasible by grid methods (Kitagawa 1987, de Valpine and Hastings 2002), but this approach is impractical for higher dimensional state variables. The main alternatives to grid methods for high-dimensional integrations are Monte Carlo methods (Carlin et al. 1992, Geyer and Thompson 1992, Kitagawa 1996, 1998, Durbin and Koopman 1997, 2000, Shephard and Pitt 1997, Tanizaki and Mariano 1998, Pitt and Shephard 1999, Doucet et al. 2001). Next I describe basic Monte Carlo integration as well as a new method, the Monte Carlo Kernel Likelihood method, which, for stage-structured population models of experiments, is faster than several other potential methods (P. de Valpine, *unpublished manuscript*).

Monte Carlo Likelihoods and Likelihood Maximization

Basic Monte Carlo integration and importance sampling

The integral (Eq. 4) can be written as

$$L(\Theta) = \int P(\mathbf{Y}\,|\,\boldsymbol{\nu})P(\boldsymbol{\nu})\,d\boldsymbol{\nu}. \tag{6}$$

In basic Monte Carlo integration (Robert and Casella 1999), one obtains a large sample of points $\{\boldsymbol{\nu}_j\}$, $j = 1, \ldots, M$ from $P(\boldsymbol{\nu})$, and then uses the approximation

$$L(\Theta) \approx \frac{1}{M}\sum_{j=1}^{M} P(\mathbf{Y}\,|\,\boldsymbol{\nu}_j) \tag{7}$$

and, for multiple replicates,

$$L(\Theta) \approx \prod_{i=1}^{N}\frac{1}{M}\sum_{j=1}^{M} P(\mathbf{Y}_i\,|\,\boldsymbol{\nu}_j;\, i). \tag{8}$$

As the Monte Carlo sample size increases, the approximation increases in accuracy. I refer to this as the ''Monte Carlo direct'' (MCD) method.

The approximation (Expression 8) provides one option for calculating the likelihood (Eq. 5), which can then be maximized over Θ, and is used in the second example below. As an example, if the process noise is modeled with normal random variables, one can draw a large normal sample and use them in Expression 8 to approximate the likelihood. The same sample can be used for all vales of Θ (Robert and Casella 1999) so that the optimization surface can be smooth.

Unfortunately MCD is generally inefficient because most values of $\boldsymbol{\nu}$ give small $P(\mathbf{Y}_i\,|\,\boldsymbol{\nu})$, with rare $\boldsymbol{\nu}$ values giving large $P(\mathbf{Y}_i\,|\,\boldsymbol{\nu})$. There are many approaches to improving efficiency, and here I mention only one of the most useful—importance sampling. This uses the relationship

$$\int P(\mathbf{Y}\,|\,\boldsymbol{\nu})P(\boldsymbol{\nu})\,d\boldsymbol{\nu} = \int P(\mathbf{Y}\,|\,\boldsymbol{\nu})\frac{P(\boldsymbol{\nu})}{P_S(\boldsymbol{\nu})}P_S(\boldsymbol{\nu})\,d\boldsymbol{\nu} \tag{9}$$

where the density $P_S(\boldsymbol{\nu})$ is called the ''sampling'' density. Then the Monte Carlo approximation uses a sample of points $\{\boldsymbol{\nu}_j\}$, $j = 1, \ldots, M$ from $P_S(\boldsymbol{\nu})$ to calculate

$$L(\Theta) \approx \frac{1}{M}\sum_{j=1}^{M} P(\mathbf{Y}\,|\,\boldsymbol{\nu}_j)\frac{P(\boldsymbol{\nu}_j)}{P_S(\boldsymbol{\nu}_j)} \tag{10}$$

where the ratio $P(\boldsymbol{\nu}_j)/P_S(\boldsymbol{\nu}_i)$ is called an ''importance weight.'' Importance sampling can greatly improve efficiency if $P_S(\boldsymbol{\nu})$ is similar to $P(\boldsymbol{\nu}\,|\,\mathbf{Y})$, the distribution of states given observations. In fact, if $P_S(\boldsymbol{\nu}) = P(\boldsymbol{\nu}\,|\,\mathbf{Y})$, then every summand is just $P(\mathbf{Y})$, and the approximation is exact with only one sample, $M = 1$. The average (Expression 10) can behave very badly if the variance of the summands is infinite, which can happen if the tails of $P_S(\boldsymbol{\nu})$ are lighter than the tails of $P(\boldsymbol{\nu}\,|\,\mathbf{Y})$. Thus a goal of importance sampling is to find $P_S(\boldsymbol{\nu})$ close of $P(\boldsymbol{\nu}\,|\,\mathbf{Y})$ while avoiding too-light tails (Robert and Casella 1999).

Monte Carlo kernel likelihoods

Even with importance sampling and other improvements, basic Monte Carlo integration lacks efficiency for high-dimensional state-space problems. A more efficient method, the Monte Carlo kernel likelihood (MCKL) method (P. de Valpine, *unpublished manuscript*) works by obtaining a sample of points in the parameter space whose density is related to the likelihood surface, and this sample requires sampling from $P(\Theta, \boldsymbol{\nu}\,|\,\mathbf{Y})$ under different parameters. (This is not a Bayesian method, but it temporarily treats parameters as having prior and posterior density functions, which is mathematically similar to a Bayesian analysis.)

Suppose that the parameters have an initial distribution, $P(\Theta)$, which is wide and nearly flat (but still proper, i.e., it integrates to 1). This will temporarily play a role similar to a noninformative prior distribution in Bayesian analysis. Suppose that we can simulate a random point sample from the conditional distribution of parameters and noises given observations, $P(\Theta, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_N\,|\,\mathbf{Y}_1, \ldots, \mathbf{Y}_N)$, which is proportional

to $P(\boldsymbol{\Theta})\,\Pi_{i=1}^{N} P(\mathbf{Y}_i, \boldsymbol{\nu}_i \mid \boldsymbol{\Theta})$. Here the subscripts on $\boldsymbol{\nu}$ and $\mathbf{Y}$ index the N experimental replicates. The $\boldsymbol{\Theta}$ dimensions of the sample are a sample from

$$\frac{L(\boldsymbol{\Theta} \mid \mathbf{Y})P(\boldsymbol{\Theta})}{C_P} = \int \cdots \int P(\boldsymbol{\Theta}) \prod_{i=1}^{N} P(\mathbf{Y}_i, \boldsymbol{\nu}_i \mid \boldsymbol{\Theta})\, d\boldsymbol{\nu}_1 \cdots d\boldsymbol{\nu}_N \quad (11)$$

which is the usual Bayesian ''posterior'', $P(\boldsymbol{\Theta} \mid \mathbf{Y}) = L(\boldsymbol{\Theta} \mid \mathbf{Y})\, P(\boldsymbol{\Theta})/C_P$, and where $C_P = \int L(\boldsymbol{\Theta} \mid \mathbf{Y})\, P(\boldsymbol{\Theta})\, d\boldsymbol{\Theta}$ is an unknown normalizing constant. Label the sample by $\{\boldsymbol{\Theta}_j\}$, $j = 1, \ldots, M$.

There are two ways to estimate $L(\boldsymbol{\Theta} \mid \mathbf{Y})/C_P$ from the sample $\{\boldsymbol{\Theta}_j\}$ and $P(\boldsymbol{\Theta})$ using kernel density estimates. For notation, if $f(x)$ is a probability density, and $\{x_j\}$, $j = 1, \ldots, M$ is a sample from f, then a kernel density estimate of f is

$$\hat{f}(x) \approx \frac{1}{M} \sum_{j=1}^{M} K_h(x, x_j) \quad (12)$$

where K_h is a kernel function with bandwidth h (Silverman 1986, Scott 1992). For the common Gaussian kernel, K_h is a normal density function with mean x and standard deviation h, and the approach extends naturally to multivariate samples and kernels. Define $\tilde{f}_h(x) = \int K_h(x, x')\, f(x')\, dx'$. Then, for fixed h,

$$\lim_{M \to \infty} \hat{f}(x) = \tilde{f}_h(x). \quad (13)$$

In practice, as M increases, h is decreased so that $\tilde{f}(x)$ approaches $f(x)$.

To estimate $L(\boldsymbol{\Theta} \mid \mathbf{Y})/C_P$, take $P(\boldsymbol{\Theta} \mid \mathbf{Y})$ as an importance sampling density for the integral $\int K_h(\boldsymbol{\Theta}, \boldsymbol{\Theta}')L(\boldsymbol{\Theta}' \mid \mathbf{Y})\, d\boldsymbol{\Theta}'$:

$$\tilde{L}_h(\boldsymbol{\Theta} \mid \mathbf{Y}) = \int K_h(\boldsymbol{\Theta}, \boldsymbol{\Theta}') \frac{L(\boldsymbol{\Theta}' \mid \mathbf{Y})}{P(\boldsymbol{\Theta}' \mid \mathbf{Y})} P(\boldsymbol{\Theta}' \mid \mathbf{Y})\, d\boldsymbol{\Theta}' = C_P \int K_h(\boldsymbol{\Theta}, \boldsymbol{\Theta}') \frac{1}{P(\boldsymbol{\Theta}')} P(\boldsymbol{\Theta}' \mid \mathbf{Y})\, d\boldsymbol{\Theta}'. \quad (14)$$

Then a Monte Carlo estimate for $\tilde{L}_h(\boldsymbol{\Theta} \mid \mathbf{Y})/C_P$ given the sample $\{\boldsymbol{\Theta}_j\}$ from $P(\boldsymbol{\Theta} \mid \mathbf{Y})$ is

$$\frac{\hat{L}(\boldsymbol{\Theta} \mid \mathbf{Y})}{C_P} \approx \frac{1}{M} \sum_{j=1}^{M} \frac{K_h(\boldsymbol{\Theta}, \boldsymbol{\Theta}_j)}{P(\boldsymbol{\Theta}_j)}. \quad (15)$$

Even though C_P is unknown, we can use Expression 15 to estimate maximum-likelihood parameters and then use basic Monte Carlo integration to estimate the likelihood value itself. Given an initial estimate of the MLE, one can also use a focused prior with high density near the MLE to obtain more efficient, accurate estimates near the MLE than for a wide prior, but such ''zooming in'' requires more careful choice of h (P. de Valpine, *unpublished manuscript*). Here we use only wide priors and choice of h motivated by the asymptotically Gaussian shape of L near its maximum (Appendix).

The second way to estimate $L(\boldsymbol{\Theta} \mid \mathbf{Y})/C_P$ from $\{\boldsymbol{\Theta}_j\}$ and $P(\boldsymbol{\Theta})$ is by $\tilde{P}_h(\boldsymbol{\Theta} \mid \mathbf{Y})/P(\boldsymbol{\Theta})$ where $\tilde{P}_h(\boldsymbol{\Theta} \mid \mathbf{Y})$ can be estimated by a kernel density estimate of the posterior sample. This approach works well with wide priors, but not with ''zooming in'' efficiency gains similar to the first method, so it is less general and not used here.

Thus the problem is reduced to obtaining a sample of points from $P(\boldsymbol{\Theta}, \boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_N \mid \mathbf{Y}_1, \ldots, \mathbf{Y}_N)$ and using the $\boldsymbol{\Theta}$ dimensions to estimate likelihoods up to the unknown constant C_P. This approach also has the advantage that it delivers an approximation of the entire likelihood surface, not just the maximum-likelihood estimate (MLE). Once we obtain an MLE, its actual likelihood (i.e., not just up to an unknown constant) must be approximated by basic Monte Carlo integration or importance sampling (see Appendix). Obtaining the sample of states and parameters given data is not trivial because this can be very high dimensional, but the computational method of Markov chain Monte Carlo (MCMC) (Gilks et al. 1996, Robert and Casella 1999) is designed for such situations. In the Appendix, I summarize the MCMC algorithm used here.

Examples

I used the MCD (Monte Carlo direct) and MCKL (Monte Carlo kernel likelihood) methods on simulated experimental data to compare analysis with a population model to ANOVA, ANCOVA, and repeated-measures ANOVA. The *first example* used simple, density-independent population growth, with a treatment effect that increases fecundity but decreases survival. The *second example* used a treatment of adding a predator, with effects via direct consumption as well as a behavioral change in prey demographic rates. I focused on detecting the behavioral change. In these two examples, the comparison was idealized because I assumed that the correct population-model structure was known. Thus these examples compared the extreme ends of a spectrum: no demographic structure (ANOVA) vs. correct demographic structure. In the *third example*, the model that generated the data was more complex than the population model used for analysis. This example was similar to the first, but with data generated from a model that included density dependence, time-varying fecundity, and autocorrelated survival.

Each example used a stage-structured herbivore model, with age-cohorts separated by day and with no density dependence. The deterministic ''skeleton'' (sensu Tong 1990) of the model was

$$n(a, t) = n(a - 1, t - 1)S_{\mathrm{E}}(z_{\mathrm{E}}(t)) \quad (16)$$

$$1 < a \leq D_{\mathrm{E}}$$

$$n(D_{\mathrm{E}} + 1, t)_{\mathrm{E}} = n(D_{\mathrm{E}}, t - 1)h_{\mathrm{EE}}S_{\mathrm{E}}(z_{\mathrm{E}}(t)) \quad (17)$$

$$n(D_{\mathrm{E}} + 1, t)_{\mathrm{J}} = n(D_{\mathrm{E}}, t - 1)h_{\mathrm{EJ}}S_{\mathrm{J}}(z_{\mathrm{J}}(t)) \quad (18)$$

TABLE 1. Definitions of variables.

Symbol	Definition
General state-space model	
$\mathbf{Y}_j$	Observation vector for replicate j
$\mathbf{X}_j$	States of true population trajectory for replicate j
$\boldsymbol{\nu}_j$	Process noise vector for replicate j
$\boldsymbol{\varepsilon}_j$	Observation error vector for replicate j
$\boldsymbol{\Theta}$	Parameter vector
All example models: states	
$A(t)$	Adults at time t
$n(a, t)$	Number of age a individuals at time t, for $a \leq D_J + 1$
$n(a, t)_E$	Number of age a individuals at time t that are eggs, for $a \leq D_J + 1$
$n(a, t)_J$	Number of age a individuals at time t that are juveniles, for $a \leq D_J + 1$
All example models: basic parameters	
L_E	Length of egg stage
L_J	Length of juvenile stage
a_s	Logit transform of survival in average environment for stage s = E, J, or A
b_s	Standard deviation of logit transform of survival for stage s = E, J, or A
μ_β	Mean value of gamma distribution for β
σ_β	Standard deviation of gamma distribution β
All example models: process noises	
β	Per day birth rate of adults, gamma distributed across replicates
z_s	Process noise for survival of stage s = E, J, or A
$S_s(z_s)$	Per day survival of stage s = E, J, or A
All example models: bookkeeping aids	
$E(t)$	Total eggs at time t
$J(t)$	Total juveniles at time t
D_E	Largest integer $\leq L_E$
D_J	Largest integer $\leq L_E + L_J$
h_{EE}	$= L_E - D_E$, fraction of $n(D_E, t)$ that are eggs at time $t + 1$
h_{EJ}	$= 1 - h_{EE}$ fraction of $n(D_E, t)$ that are juveniles at time $t + 1$
h_{JJ}	$= (L_J + L_E) - D_J$, fraction of $n(D_J, t)$ that are juveniles at time $t + 1$
h_{JA}	$= 1 - h_{JJ}$, fraction of $n(D_J, t)$ that are adults at time $t + 1$
Predation-model parameters (Example 2)	
q_s	Inverse of the half saturation constant for predation on stage s = E or J
γ_s	γ_s/q_s is the maximum predation rate on stage s = E or J
More-complicated model parameters (Example 3)	
ρ	Autocorrelation of z_s for s = E, J, or A
ε_β	Process noise for time-variation in fecundity
τ_β	Standard deviation of ε_β
δ	Density dependence of adults and juveniles on fecundity

$$n(a, t) = n(a - 1, t - 1)S_J(z_J(t)) \qquad (19)$$
$$(D_E + 1) < a \leq D_J$$

$$n(D_J + 1, t)_J = n(D_J, t - 1)h_{JJ}S_J(z_J(t)) \qquad (20)$$

$$A(t) = [n(D_J + 1, t - 1)_J + n(D_J, t - 1)h_{JA} + A(t - 1)]S_A(z_A(t)) \qquad (21)$$

$$n(1, t) = \beta A(t - 1)S_E(z_E(t)). \qquad (22)$$

Definitions of all variables are given in Table 1. Equation 16 gives growth and mortality for the egg-only day classes. Eqs. 17 and 18 give the egg and juvenile portions of the egg-to-juvenile day class. Eq. 19 gives growth and survival of juveniles, and Eq. 20 gives the juvenile portion of the juvenile-to-adult day class. Eqs. 21 and 22 give adult dynamics and births, respectively.

The reason for allowing fractional maturation times is that the Monte Carlo likelihood methods require parameters with continuous ranges of values, not just integers. To allow a continuous range of the stage-duration parameters, I assumed that age cohorts are distinct by day but well mixed within days. This means I defined an egg stage of 4.6 days to mean that 60% of individuals have an egg stage of 4 days and the remaining 40% have an egg stage of 5 days. This would be realistic if, for example, molting occurs only during a fixed time of day, so that individuals either molt at that time one day or wait until that time the next day. Note that for $1 \leq a \leq D_E$, $n(a, t)_E = n(a, t)$; the subscripts "E" and "J" (or "J" and "A") are important only for the day classes with both eggs and juveniles (or juveniles and adults).

Process noise

For the examples here, I considered reproductive rates that differ randomly between replicates but not over time and survival rates that differ between replicates and over time. For the survival rates, let S_E, S_J,

TABLE 2. Simulation parameter values for all three examples comparing a population model to ANOVA, ANCOVA, and repeated-measures ANOVA.

Parameter	T_0	T_1	T_2	P	P_B	$I_{M,0}$	$I_{M,1}$	$I_{S,0}$	$I_{S,1}$
μ_F	4.0	6.0	8.0	4.0	4.0	4.0	6.0	4.0	6.0
σ_F^2	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
$S_E(0)$	0.9	0.877	0.86	0.9	0.88	0.9	0.877	0.9	0.877
$S_J(0)$	0.8	0.777	0.76	0.8	0.8	0.8	0.777	0.8	0.777
$S_A(0)$	0.7	0.677	0.66	0.7	0.7	0.7	0.677	0.7	0.677
b	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
L_E	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
L_J	6.0	6.0	6.0	6.0	6.5	6.0	6.0	6.0	6.0
q_E	⋯	⋯	⋯	0.025	0.0125	⋯	⋯	⋯	⋯
q_J	⋯	⋯	⋯	0.0125	0.00625	⋯	⋯	⋯	⋯
γ_E	⋯	⋯	⋯	9.0	9.0	⋯	⋯	⋯	⋯
γ_J	⋯	⋯	⋯	3.0	3.0	⋯	⋯	⋯	⋯
ρ	⋯	⋯	⋯	⋯	⋯	0.6	0.6	0.6	0.6
δ	⋯	⋯	⋯	⋯	⋯	0.01	0.01	0.05	0.05
τ_β	⋯	⋯	⋯	⋯	⋯	0.1	0.1	0.1	0.1

Notes: Example 1 uses T_0 (control, in all cases), T_1 (small treatment effect), and T_2 (large treatment effect). Example 2 uses T_0, P (predator without behavior effect) and P_B (predator with behavior effect) parameters. Example 3 uses mild ($I_{M,0}$, $I_{M,1}$) and strong ($I_{S,0}$, $I_{S,1}$) deviations from the model (where I = incorrect model being used to fit data).

and S_A be inverse logit transformations of normal deviates that change every two days:

$$S_s[z_s(t)] = \frac{e^{a_s + b_s z_s(t)}}{1 + e^{a_s + b_s z_s(t)}} \tag{23}$$

$$z_s(t) \sim \mathcal{N}(0, 1) \qquad t \text{ odd} \tag{24}$$

$$z_s(t) = z_s(t - 1) \qquad t \text{ even} \tag{25}$$

where s = E, J, or A. The two-day periods reflect an assumed time scale of variation of environmental conditions. One could also let environmental variation be autocorrelated, and/or have a time scale that is estimated. For individual variation in fecundity, β for each replicate is drawn from a gamma distribution. I let the simulated experiments run for 40 days, so $\boldsymbol{\nu}$ is 61-dimensional for each replicate—one dimension for β and three dimensions for $\mathbf{z} = (z_E, z_J, z_A)$ for each of 20 two-day stretches.

Observation error

I assumed that all experiments started with known initial conditions and that observations were made every 10 days for 40 days. Observations were grouped by stage class, and for each stage class the observed values were normally distributed with means equal to the true number and standard deviations equal to 0.1 × (true number) + 0.01:

$$E(t) = \sum_{a=1}^{D_E+1} n(a, t)_E \tag{26}$$

$$J(t) = \sum_{a=D_E+1}^{D_J+1} n(a, t)_J \tag{27}$$

$$Y_{s,t} \sim \mathcal{N}(s, 0.1s + 0.01) \tag{28}$$

where s = E, J, or A and $Y_{s,t}$ is the estimate of stage s at time t. This was motivated by the situation where only a fraction of each replicate is sampled, and then the sample is multiplied by the inverse of the fraction to obtain an estimate of the total number, so the standard deviation scales with the mean. In addition to being inaccurate, these observations represent incomplete information because they do not include within-stage age structure, which is an essential part of the true dynamics. The 0.01 term adds numerical regularity to the algorithms when the true number is very small, and can be interpreted as a tiny chance that even with no individuals, an individual will mistakenly be counted. I also assumed that, aside from the experiment of interest, the investigators had studied their own observation process (i.e., by making replicate subsample counts from the type of enclosures used in the experiment) and know the model (Expression 28) for the distribution of observations. Finally, the abundances here are unscaled (e.g., the units could be 10^6 individuals).

Example 1: Plant effects on herbivore demography

First I assumed that the experimental treatment was to change host-plant growth conditions, such as by watering, fertilizing, or inducing secondary chemical defenses, and that the effect was a shift in the herbivore's life-history pattern, with higher fecundity and lower survival. This is a difficult situation for ANOVA because the number of adults at the end of the experiment may be similar between treatment and control even though the demographic trajectory that produced those adults is different. I used 20 experimental units, with 10 each of control and treatment.

All examples used a Type I error rate of $\alpha = 0.05$. Table 2 gives parameters for no effect (T_0, = control in all cases), small treatment effect (T_1), and large treatment effect (T_2). Relative to the control, the log number of adults at the end of the experiment in both treatment

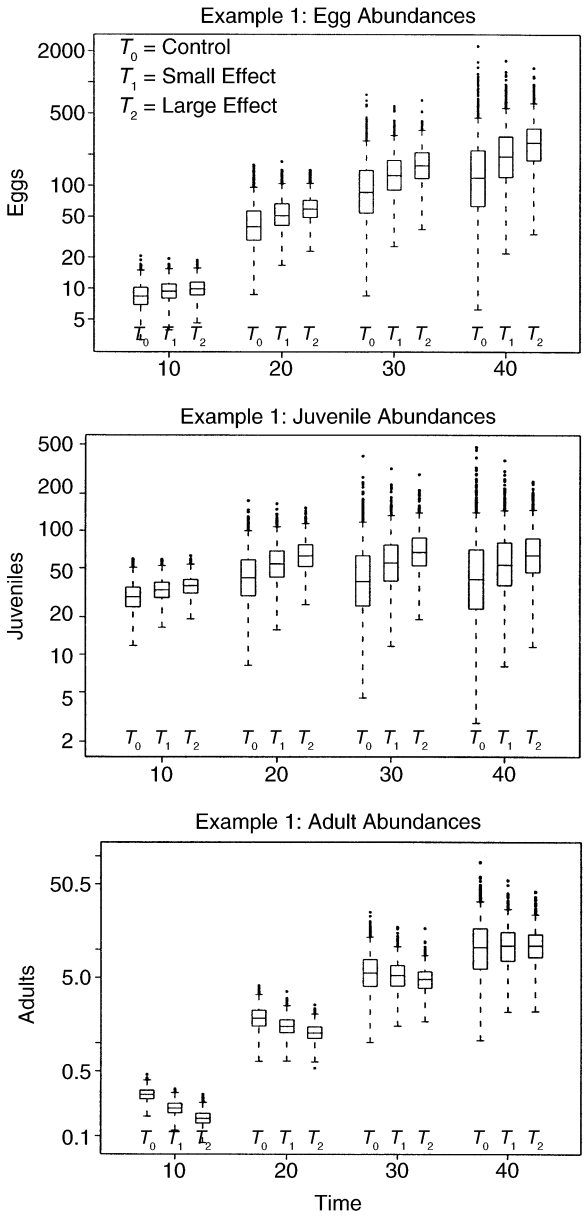

FIG. 1. Box-and-whisker plots on a log scale of observations of eggs, juveniles, and adults for Example 1 under control (T_0), small (T_1), and large (T_2) treatment effects. Distributions for T_0 and T_2 are horizontally offset for visual clarity. Plots are from 1000 simulated trajectories, although for each simulated experiment there are only 10 replicates. For each box-and-whisker diagram, dashed horizontal lines of the box show the 25th, 50th, and 75th percentiles, dashed ''whisker'' lines extend to the most extreme value within 1.5 times the interquartile range (i.e., difference between 25th and 75th percentiles) above and below the 25th and 75th percentiles, and dots show individual values beyond this range.

conditions had a similar mean but lower variance (Fig. 1). All examples used 100 simulated data sets of each hypothetical experiment.

I considered three types of ANOVA, all with log-transformed abundances and all for eggs, juveniles, or adults separately: ANOVA on abundance at the end of the experiment, repeated-measures ANOVA, and ANCOVA with time as a covariate. The first ANOVA amounts to a *t* test, and since the variances were generally unequal I used a Welch correction. For repeated-measures ANOVA, the assumptions were violated because the variance increased through time, but I report results anyway for comparative purposes. For ANCOVA the assumptions were also not met, since the log abundances did not increase linearly (over long times they would, but these simulations do not reach stable age structure) and the variances changed through time. Also, in ANCOVA one typically first tests whether slopes are not different among groups and then tests for significantly different intercepts assuming homogenous slopes. I considered an investigator seeking any significant result, whether for different slopes or different intercepts if the slopes are similar, so I report the proportion of data sets where either hypothesis was rejected.

For analysis with the population model, I used as a null hypothesis that all parameters were equal between control and treatment, and as an alternative that μ_β, a_E, a_J, and a_A may vary between control and treatment. For Example 1 the MLEs were obtained using MCKL and the Nelder-Mead simplex algorithm for maximization (Press et al. 1992).

Example 2: Predator effects on herbivore demography via behavior

Next I considered a predator–prey experiment, with two questions of interest: whether (and how much) the predator eats the herbivores, and whether the presence of the predator causes behavioral changes in prey demography, such as in behavioral trait-mediated indirect effects. I focused on the second question because ANOVA is wholly unsuited to it. The predator was modeled with a Type II functional response and with a constant population size for simplicity. Mortality caused by the predator was modeled by multiplying egg survival by

$$S_{E,P} = \exp\left(\frac{-\gamma_E}{1 + q_E E + q_J J}\right) \tag{29}$$

and juvenile survival by

$$S_{J,P} = \exp\left(\frac{-\gamma_J}{1 + q_E E + q_J J}\right). \tag{30}$$

Under a Type II functional response, the egg mortality rate from predation would be $(\gamma_E)/(1 + q_E E + q_J J)$, so that after a day the fraction surviving would be Eq. 29, and a similar explanation applies for juveniles (Eq. 30). Usually there is a predator abundance in the numerator of a Type II functional response, but since I assumed constant predator numbers, their abundance was incorporated in γ_E and γ_J.

Table 2 lists control (T_0), no-behavior-predation (P), and behavior-predation (P_B) parameters. For P_B, the

parameters reflect increased juvenile movement to avoid predators and increased adult movement to disperse eggs, with the costs of 0.5-day-longer juvenile development time and 0.02 decrease in the adult-survival parameter and with the consequence of slower saturation of predation rates as prey density increases. Each experiment consisted of T_0 and either P or P_B.

For analysis with the population model to detect the effect on prey behavior, I used a null hypothesis that all parameters were equal between treatment and control except for the presence of the predator in the treatment (no predator corresponds to $\gamma_E = \gamma_J = 0$) and an alternative that μ_β, a_E, a_J, a_A, L_E, and L_J may vary between treatments. The likelihoods were approximated with basic Monte Carlo integration and maximized with the Nelder-Mead simplex algorithm (Press et al. 1992). A Monte Carlo (MC) sample size of 5000 was used for an initial maximization, followed by a second maximization with MC sample size of 20 000. Diagnostic runs showed this protocol to be accurate.

Example 3: Plant effects on herbivore demography with a more complex model

The final example used an experiment similar to Example 1, but with data generated by a more complex model than the population model used for analysis. Two levels of deviation from model assumptions were used. The more complex model included autocorrelation of mortalities, fecundity that changed randomly every two days, and density dependence for fecundity. Autocorrelation of mortalities was implemented with a first-order autoregressive model for z_i, i = E, J, or A:

$$z_i(t + 1) = \rho z_i(t) + \varepsilon_i(t) \tag{31}$$

where ρ is the autocorrelation and $\varepsilon_i(t)$ is normally distributed and sequentially independent noise. To keep variance of $z_i(t)$ equal to 1, as before, I set $\sigma_\varepsilon^2 = 1 - \rho^2$. Time variation and density dependence of fecundity were implemented by setting

$$\beta(t) = [\overline{\beta} + \varepsilon_\beta(t)]e^{-\delta[J(t)+A(t)]} \tag{32}$$

where $\varepsilon_\beta(t)$ is normal with mean zero and standard deviation τ_β; δ is a density-dependence parameter; and $\overline{\beta}$ plays the previous role of β: it is gamma-distributed across replicates. Model fitting was the same as for Example 1. Parameter values for mild ($I_{M,0}$, $I_{M,1}$) and strong ($I_{S,0}$, $I_{S,1}$) deviations from the model used for fitting are given in Table 2.

RESULTS

Parameter estimates

Parameter estimates showed good estimation properties (Fig. 2). For example, for the six nonvariance parameters of the control group (T_0) of Example 1, there was only mild bias in the parameters, such as an average μ_β estimate of 4.3 instead of the true value of 4.0. In addition, parameter estimates were correlated in biologically intuitive ways, with higher mean fecundity values associated with lower mean egg survival, higher mean juvenile survival, lower egg stage length, and higher juvenile stage length. These correlations make sense because, for example, similar egg abundances might be produced by higher fecundity and lower egg survival or lower fecundity and higher egg survival, so that across a range of data sets, some will tend to estimate the former and others the latter. Alternating correlations between estimates of fecundity and successive survival rates were discussed by Wood (1994, 1997).

Statistical power

In all three examples the population-model analysis gave higher power and more insightful inferences than ANOVA. In Example 1, the distributions of log abundance of eggs, juveniles, and adults overlapped substantially (Fig. 1), so ANOVA with the final 10 abundances from control (T_0) and treatment groups (T_1 or T_2) had low power to detect the different means, with highest power for the egg distributions (Fig. 3a and b). Repeated-measures ANOVA improved power for analyzing juveniles or adults, but the maximum power was still only ~40%. ANCOVA produced higher power but had the undesirable feature of inflated Type I error rate, shown by ~20% rejection of the null when it was true.

The population model had very high power, with rejection of the null in all 100 simulations for both treatment magnitudes. In fact, the P values for the population model were very small, much less that 10^{-6} even for the small treatment effect. This is encouraging because even if we had allowed all of the parameters to vary between treatment and control, we would have obtained 100% power; $L(\boldsymbol{\Theta}_{ALT})$ would be greater than or equal to the value we obtained with fewer parameters, but the likelihood ratios we obtained were large enough to reject under a chi-squared distribution with df = 10.

In Example 2, the distributions of eggs, juveniles, and adults reflected the presence of the predator (Fig. 4). The direct effect of reducing population sizes was strong—ANOVA could probably detect this with reasonable power—but the additional behavior effect was difficult to see intuitively. Comparing P with P_B shows that with the behavior effect there were slightly higher abundances with lower variances than without the behavior effect. Each experiment, however, included 10 points from T_0 and either P or P_B, so the foregoing comparison would not be possible. Loosely speaking, the population model analysis ''controls for'' predation and estimates any remaining demographic changes in the prey. With no behavior effect, only three out of 100 null hypotheses were rejected, close to the expected Type I error rate of 5%. With the behavior effect, the null hypothesis was rejected 81 out of 100 times. Thus the population model had high power to detect

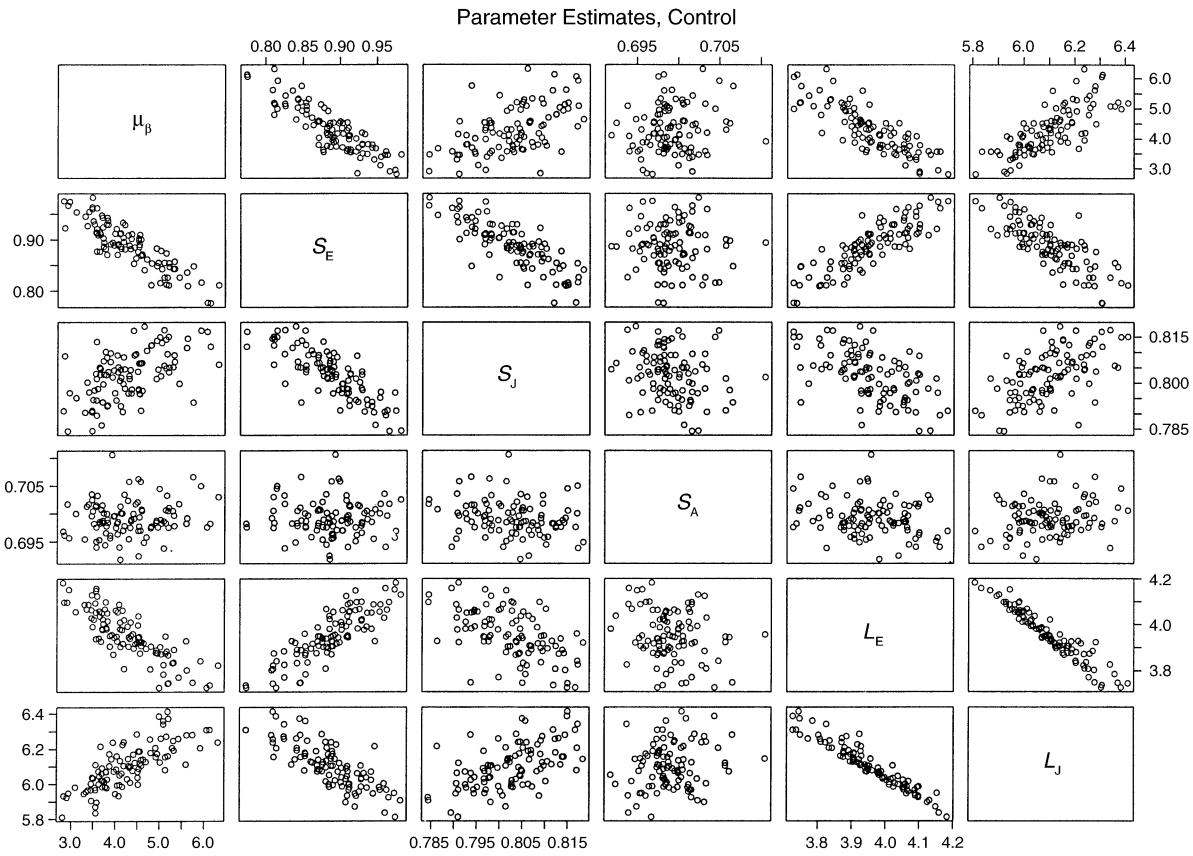

FIG. 2. Maximum-likelihood parameter estimates for the six nonvariance parameters of the control group from 100 simulated experiments of Example 1 (parameters T_0). Each subplot shows the 100 maximum-likelihood parameters plotted in two of the six dimensions. Each diagonal entry labels the horizontal axis of its column and the vertical axis of its row.

changes in demographic rates that were not intuitively obvious. The power levels of these examples are a function of the magnitude of the simulated effects.

In Example 3, the abundance distributions decreased nonlinearly, relative to Example 1, due to density dependence. As in Example 1, analysis with a population model provided higher power than the three varieties of ANOVA (Table 3), and ANCOVA had inflated Type I Error rate when the null ($I_{M,0}$ or $I_{S,0}$) is true. For the mildly incorrect model ($I_{M,0}$ $I_{M,1}$), the population model performed as in Example 1. For the strongly incorrect model ($I_{S,0}$, $I_{S,1}$), power of the population model remained nearly perfect when there was a true difference ($I_{S,1}$), but Type I error rate was inflated to 10% when there was no true difference ($I_{S,0}$). Diagnostic analyses could be useful for evaluating and comparing different model fits, but here the main point is that even approximately correct demographic structure can provide higher power than omitting demographic structure—but as model structure becomes more incorrect ($I_{S,0}$, $I_{S,1}$), Type I error rate, power, or both, may degrade.

DISCUSSION

Why are population models so much more powerful for detecting treatment effects than ANOVA models? ANOVA does not generally incorporate dynamics or multiple stages or species in a way that corresponds closely with biologically motivated, demographic hypotheses. In analysis with a population model, the distribution of trajectories from a stochastic demographic hypothesis is used to calculate likelihoods, so structured data enter naturally into a single analysis. The distributions of stage classes through time under stochastic population models with process noise (PN) and observation error (OE) are not easy to calculate analytically, but they can be calculated numerically and provide better evaluation of hypotheses with data than ANOVA models.

Although the majority of population-dynamics experiments are not analyzed with population models, a relative handful of studies have been. These highlight the value of the approach but offer less general methods than those presented here. For example, Dennis et al. (1995, 2001) used approximate likelihood-ratio tests for experimental data where the likelihoods are relatively easy to calculate because the data really have no OE. Ives et al. (1999) and Klug et al. (2000) used linear time-series models to investigate experimental-lake time series. Bjornstad et al. (1998, 2001) fit time-lag models to replicated time series of Indian meal moth

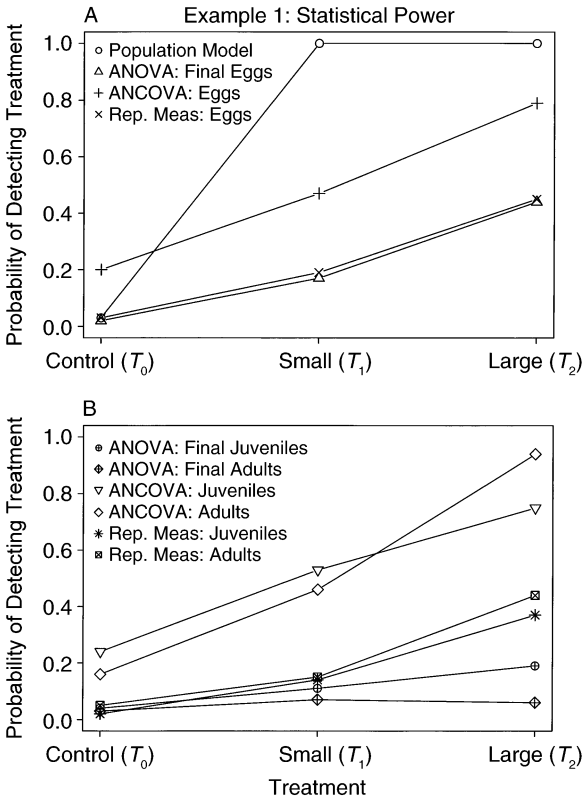

FIG. 3. Statistical power for Example 1, showing probability of detecting treatment effect as a function of true effect size. (A) Statistical power for rejecting the null hypothesis of no treatment effect for Example 1 using the population model or ANOVA, ANCOVA, or repeated-measures ANOVA for log final abundance of eggs. (B) Statistical power for Example 1 using ANOVA, ANCOVA, and repeated-measures ANOVA for log final abundances of juveniles and adults.

(*Plodia interpunctella*) populations with or without a virus and with or without a parasitoid, but they stopped short of a classical hypothesis test. Wood (1994, 1997) developed a method for fitting stage-structured models with splines of age–time surfaces in a nonlikelihood way, assuming a priori knowledge of stage lengths. Finally, Gibson and Renshaw (1998) and Gibson et al. (1999) give methods related to this paper in using Bayesian posteriors to estimate frequentist conclusions.

Several implementation issues merit discussion. The choice to work with population states, PN's, or individuals has ramifications for choice of model and Monte Carlo method. The examples here tracked a 10-dimensional age vector for 40 days. If the model had been defined in terms of each population state (i.e., age cohort), there would have been 400 dimensions for the likelihood integral. One could design a Markov chain Monte Carlo (MCMC) algorithm to sample these states for each replicate (given observations) along with parameters. For most states (i.e., age-time abundances), Metropolis-Hastings steps would involve at least the probability of obtaining a particular state (e.g., by survival from the previous age at the previous time) and the probability of obtaining the next state from that state. Thus, there would be many state dimensions but relatively low computational burden for MCMC moves for each one. For models with density dependence, changing a state could affect other transition probabilities, which would increase the computational cost of each MCMC move.

In contrast, by working with the space of process noises, one can decrease the dimensionality but increase the computational burden for MCMC moves in each dimension. In the examples here, 61 PN dimen-

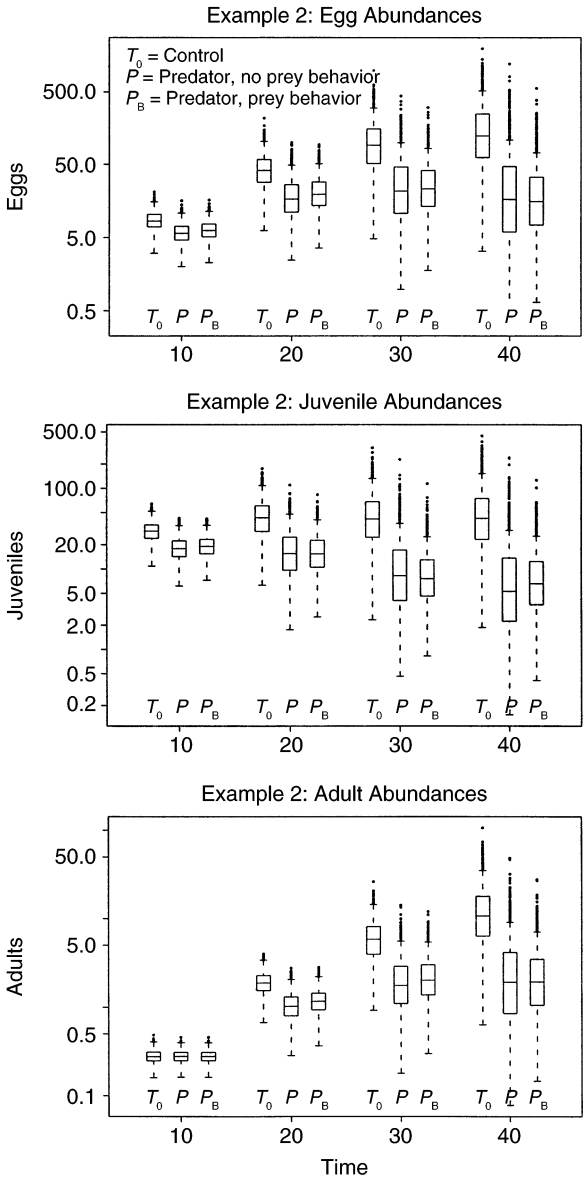

FIG. 4. Box-and-whisker plots on a log scale of observations of eggs, juveniles, and adults for Example 2 under control (T_0), predator without behavior effect (P), and predator with behavior effect (P_B). The population model rejected the no-behavior hypothesis in 81% of cases when it was false.

TABLE 3. Probability of detecting treatment effect for example 3.

Model	Probability			
	$I_{M,0}$	$I_{M,1}$	$I_{S,0}$	$I_{S,1}$
Population model	0.05	1.00	0.10	0.99
ANOVA				
final eggs	0.05	0.23	0.05	0.30
final juveniles	0.05	0.13	0.04	0.06
final adults	0.05	0.05	0.04	0.05
ANCOVA				
eggs	0.21	0.56	0.16	0.54
juveniles	0.27	0.61	0.26	0.61
adults	0.12	0.43	0.12	0.42
Repeated measures				
eggs	0.05	0.24	0.05	0.21
juveniles	0.05	0.19	0.04	0.18
adults	0.05	0.12	0.05	0.12

Notes: $I_{M,0}$ and $I_{M,1}$ use mildly incorrect model structure, while $I_{S,0}$ and $I_{S,1}$ use strongly incorrect model structure. Each number is the fraction of 100 simulated experiments for which the null hypothesis was rejected using a particular model. In $I_{M,0}$ and $I_{S,0}$ the null hypothesis is true, while in $I_{M,1}$ and $I_{S,1}$ it is false.

sions determined the trajectory of a replicate. This gave fewer dimensions (than 400), but a change in a single noise value required recalculation of the entire population trajectory after the time of that noise. The tradeoff between dimensionality and computation effort for each dimension is connected to the type of model used. The 61-dimensional approach here represented environmental stochasticity, while the 400-dimensional approach could also include demographic stochasticity. Both approaches have merit in the types of models they can represent and the implementation and computational tradeoffs involved.

Choice of model, state, and noise vectors affects efficiency of Monte Carlo (MC) likelihood methods. In general Monte Carlo kernel likelihood (MCKL) is more efficient than basic MC integration, but efficiency of MCKL depends on efficiency of its posterior-sampling algorithm. The MCMC approach used here is the tip of an iceberg of posterior-sampling possibilities. In addition to the many available improvements to MCMC (Robert and Casella 1999), another promising approach is joint parameter-state particle filtering (Doucet et al. 2001). Relative efficiency of basic MC integration increases with the number of experimental replicates because the same state trajectories—the most computationally expensive step in calculating a summand of Expression 10—can be used for every replicate in a treatment group.

The example models here may be viewed as discretizations of the type of continuous time and age model described by Gurney et al. (1983). Monte Carlo likelihood methods could be used to fit more nearly continuous models, i.e., with shorter stage classes, but several challenges would arise. Numerical calculation of state trajectories would be slower, and dimensionality of PN could be much higher. Since many thousands of state trajectories are required for Monte Carlo methods, these changes could create significant computational challenges. In many cases a mildly discretized model may be as biologically justified as a continuous-time model.

A major question for further investigation is how much biological structure must be correctly known to provide a useful model for inference. This study has focused primarily on two extremes: ANOVA, which has no biological structure, and a population model with the correct structure. In analysis of real data, model structures may be at best roughly correct. The high power of analysis with a population model, along with Example 3 (see *Examples*, above), indicate that using some population structure can be better than none, but the boundaries of how much structure is enough remain to be explored. One promising option is to use the ''semi-mechanistic'' approach of using flexible functions in key parts of a model for which mechanistic structure can not be assumed (Ellner et al. 1998, Kendall et al. 1999).

This study has not discussed several aspects of statistical analysis that would be critical with real data, including diagnosis of model fit and analysis of sensitivity to model structure. Diagnosis of model fit can use residual analysis, but with a state-space model there are no point residuals; instead there are distributions of states (or process noises) given observations. In place of residuals, one could use the means or full distributions of states given observations to look for violations of assumptions, such as lack of fit to the assumed distribution or lack of temporal independence. Finally, although only point estimates and likelihood-ratio tests were considered here, one could also use Monte Carlo likelihood approximations to estimate confidence regions in standard ways (Stuart and Ord 1991, Severini 2000).

Using population models to analyze population data is part of the larger trend in science of using computational methods to analyze data in the framework of mechanistic models (Smith 1992). The theoretical argument for this approach is strong, but application has been limited by computational methods. The computational limitations are not trivial, but with continuing advances in algorithms and computational power, the possibility of improved experimental study of complex ecological systems is promising.

ACKNOWLEDGMENTS

I thank R. Karban for discussions about experimental population ecology that helped spark this work. S. Ellner for discussions about model fitting and inference, and N. Underwood, J. Schmidt, O. Bjornstad, and two anonymous reviewers for helpful manuscript comments. This work was conducted while the author was a Postdoctoral Associate at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant DEB-0072909), the University of California, and the Santa Barbara campus.

Literature Cited

Bjornstad, O. N., M. Begon, N. C. Stenseth, W. Falck, S. M. Sait, and D. J. Thompson. 1998. Population dynamics of the indian meal moth: demographic stochasticity and delayed regulatory mechanisms. Journal of Animal Ecology **67**:110–126.

Bjornstad, O. N., J.-M. Fromentin, N. C. Stenseth, and J. Gjosaeter. 1999. Cycles and trends in cod populations. Proceedings of the National Academy of Sciences (USA) **96**: 5066–5071.

Bjornstad, O. N., and B. T. Grenfell. 2001. Noisy clockwork: time series analysis of population fluctuations in animals. Science **193**:638–643.

Bjornstad, O. N., S. M. Sait, N. C. Stenseth, D. J. Thompson, and M. Begon. 2001. The impact of specialized enemies on the dimensionality of host dynamics. Nature **409**:1001–1006.

Carlin, B. P., N. G. Polson, and D. S. Stoffer. 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. Journal of the American Statistical Association **87**:493–500.

de Valpine, P., and A. Hastings. 2002. Fitting population models with process noise and observation error. Ecological Monographs **72**:57–76.

Dennis, B., R. A. Desharnais, J. M. Cushing, and R. Costantino. 1995. Nonlinear demographic dynamics: mathematical models, statistical methods, and biological experiments. Ecological Monographs **65**:261–281.

Dennis, B., R. A. Desharnais, J. M. Cushing, S. M. Henson, and R. F. Costantino. 2001. Estimating chaos and complex dynamics in an insect population. Ecological Monographs **71**:277–303.

Doucet, A., N. de Freitas, and N. Gordon, editors. 2001. Sequential Monte Carlo methods in practice. Springer-Verlag, New York, New York, USA.

Durbin, J., and S. J. Koopman. 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika **84**:669–684.

Durbin, J., and S. J. Koopman. 2000. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. Journal of the Royal Statistical Society, B **62**:3–56.

Ellner, S. P., B. A. Bailey, G. V. Bobashev, A. R. Gallant, B. T. Grenfell, and D. W. Nychka. 1998. Noise and nonlinearity in measles epidemics: combining mechanistic and statistical approaches to population modeling. American Naturalist **15**:425–440.

Ellner, S. P., et al. 2001. Habitat structure and population persistence in an experimental community. Nature **412**: 538–543.

Freeman, S. N., and G. P. Kirkwood. 1995. On a structural time series method for estimating stock biomass and recruitment from catch and effort data. Fisheries Research **22**:77–98.

Geyer, C. J., and E. A. Thompson. 1992. Constrained Monte Carlo maximum likelihood for dependent data. Journal of the Royal Statistical Society, B **54**:657–699.

Gibson, G., C. Gilligan, and A. Kleczkowski. 1999. Predicting variability in biological control of a plant-pathogen system using stochastic models. Proceedings of the Royal Society of London B **266**:1743–1753.

Gibson, G. J., and E. Renshaw. 1998. Estimating parameters in stochastic compartmental models using Markov chain methods. IMA Journal of Mathematics Applied to Medicine and Biology **15**:19–40.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter, editors. 1996. Markov Chain Monte Carlo in practice. Chapman and Hall, New York, New York, USA.

Gudmundsson, G. 1994. Time series analysis of catch-at-age observations. Applied Statistics **43**:117–126.

Gurney, W., and R. Nisbet. 1998. Ecological dynamics. Oxford University Press, New York, New York, USA.

Gurney, W., R. Nisbet, and J. Lawton. 1983. The systematic formulation of tractable single-species population models incorporating age structure. Journal of Animal Ecology **52**: 479–495.

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

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

Ives, A. R., S. R. Carpenter, and B. Dennis. 1999. Community interaction webs and zooplankton responses to planktivory manipulations. Ecology **80**:1405–1421.

Kareiva, P. 1987. Habitat fragmentation and the stability of predator–prey interactions. Nature **326**:388–390.

Kendall, B. E., C. J. Briggs, W. W. Murdoch, P. Turchin, S. P. Ellner, E. McCauley, R. M. Nisbet, and S. N. Wood. 1999. Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches. Ecology **80**:1789–1805.

Kimura, D. K., J. W. Balsiger, and D. H. Ito. 1996. Kalman filtering the delay-difference equation: practical approaches and simulations. Fishery Bulletin **94**:678–691.

Kitagawa, G. 1987. Non-Gaussian state-space modeling of nonstationary time series (with discussion). Journal of the American Statistical Association **82**:1032–1063.

Kitagawa, G. 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. Journal of Computational and Graphical Statistics **5**:1–25.

Kitagawa, G. 1998. A self-organizing state-space model. Journal of the American Statistical Association **93**:1203–1215.

Klug, J. L., J. Fischer, A. Ives, and B. Dennis. 2000. Compensatory dynamics in planktonic community responses to pH perturbations. Ecology **81**:387–398.

Lundvall, D., R. Svanback, L. Persson, and P. Bystrom. 1999. Size-dependent predation in piscivores: interactions between predator foraging and prey avoidance abilities. Canadian Journal of Fisheries and Aquatic Sciences **56**:1285–1292.

McAllister, M. K., E. K. Pikitch, A. E. Punt, and R. Hilborn. 1994. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. Canadian Journal of Fisheries and Aquatic Sciences **51**:2673–2687.

Mendelssohn, R. 1988. Some problems in estimating population sizes from catch-at-age data. Fishery Bulletin **86**: 617–630.

Metz, J., and O. Diekmann, editors. 1986. The dynamics of physiologically structured populations. Number 68 in Lecture Notes in Biomathematics. Springer-Verlag, Berlin, Germany.

Meyer, R., and R. B. Millar. 1999. Bayesian stock assessment using a state-space implementation of the delay difference model. Canadian Journal of Fisheries and Aquatic Sciences **56**:37–52.

Millar, R. B., and R. Meyer. 2000. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. Applied Statistics **49**: 327–342.

Peacor, S. D., and E. E. Werner. 2001. The contribution of trait-mediated indirect effects to the net effects of a predator. Proceedings of the National Academy of Sciences (USA) **98**:3904–3908.

Pella, J. J. 1993. Utility of structural time series models and the Kalman Filter for predicting consequences of fishery actions. Pages 571–593 in G. Kruse, D. Eggers, R. Marasco, C. Pautzke, and T. J. Quinn II, editors. Proceedings of the

International Symposium on Management Strategies for Exploited Fish Population. Alaska Sea Grant College Program Report Number 93-02. University of Alaska, Fairbanks, Alaska, USA.

Pitt, M. K., and N. Shephard. 1999. Filtering via simulation: auxiliary particle filters. Journal of the American Statistical Association **94**:590–599.

Polis, G. 1991. Complex trophic interactions in deserts: an empirical critique of food web theory. American Naturalist **128**:123–155.

Polis, G. A., and D. R. Strong. 1996. Food web complexity and community dynamics. American Naturalist **147**:813–846.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical recipes in C: the art of scientific computing. Second edition. Cambridge University Press, Cambridge, UK.

Quinn, T. J., II, and R. B. Deriso. 1999. Quantitative fish dynamics. Oxford University Press, New York, New York, USA.

Robert, C. P., and G. Casella. 1999. Monte Carlo statistical methods. Springer-Verlag, New York, New York, USA.

Rosenheim, J. A., H. K. Kaya, L. E. Ehler, J. J. Marois, and B. A. Jaffee. 1995. Intraguild predation among biological-control agents: theory and evidence. Biological Control **5**: 303–335.

Schmitz, O. J., A. P. Beckerman, and K. M. O'Brien. 1997. Behaviorally mediated trophic cascades: effects of predation risk on food web interactions. Ecology **78**:1388–1399.

Schnute, J. T. 1994. A general framework for developing sequential fisheries models. Canadian Journal of Fisheries and Aquatic Sciences **51**:1676–1688.

Scott, D. 1992. Multivariate density estimation. John Wiley and Sons, New York, New York, USA.

Severini, T. A. 2000. Likelihood methods in statistics. Oxford University Press, New York, New York, USA.

Shephard, N., and M. K. Pitt. 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika **84**: 653–667.

Sih, A., G. Englund, and D. Wooster. 1998. Emergent impacts of multiple predators on prey. Trends in Ecology and Evolution **13**:350–355.

Silverman, B. 1986. Density estimation for statistics and data analysis. Chapman and Hall, New York, New York, USA.

Smith, A. 1992. Discussion of the paper by Geyer and Thompson. Journal of the Royal Statistical Society, B **54**: 684–686.

Strauss, S. Y. 1991. Indirect effects in community ecology—their definition, study and importance. Trends in Ecology and Evolution **6**:206–210.

Stuart, A., and J. K. Ord. 1991. Kendall's advanced theory of statistics. Volume 2. Classical inference and relationship. Fifth edition. Oxford University Press, New York, New York, USA.

Sullivan, P. J. 1992. A Kalman filter approach to catch-at-length analysis. Biometrics **48**:237–257.

Tanizaki, H., and R. S. Mariano. 1998. Nonlinear and non-Gaussian state space modeling with Monte Carlo simulations. Journal of Econometrics **83**:263–290.

Tong, H. 1990. Non-linear time series: a dynamical system approach. Oxford University Press, New York, New York, USA.

Tuljapurkar, S., and H. Caswell, editors. 1997. Structured-population models in marine, terrestrial, and freshwater systems. Chapman and Hall, New York, New York, USA.

Walde, S. J. 1991. Patch dynamics of a phytophagous mite population—effect of number of subpopulations. Ecology **72**:1591–1598.

Wood, S. N. 1994. Obtaining birth and mortality patterns from structured population trajectories. Ecological Monographs **64**:23–44.

Wood, S. N. 1997. Inverse problems and structured-population dynamics. Pages 555–586 *in* S. Tuljapurkar and H. Caswell, editors. Structured-population models in marine, terrestrial, and freshwater systems. Chapman and Hall, New York, New York, USA.

Wootton, J. T. 1994. The nature and consequences of indirect effects in ecological communities. Annual Review of Ecology and Systematics **25**:443–466.

APPENDIX

A summary of the Markov chain Monte Carlo algorithms and kernel density estimates used in this paper is available in ESA's Electronic Data Archive: *Ecological Archives* E084-081-A1.