Ecological Monographs, 72(1), 2002, pp. 57–76

FITTING POPULATION MODELS INCORPORATING PROCESS NOISE AND OBSERVATION ERROR

PERRY DE VALPINE[1] AND ALAN HASTINGS

Department of Environmental Science and Policy, Center for Population Biology and Institute for Theoretical Dynamics, University of California, One Shields Avenue, Davis, California 95616 USA

Abstract. We evaluate a method for fitting models to time series of population abundances that incorporates both process noise and observation error in a likelihood framework. The method follows the probability logic of the Kalman filter, but whereas the Kalman filter applies to linear, Gaussian systems, we implement the full probability calculations numerically so that any nonlinear, non-Gaussian model can be used. We refer to the method as the "numerically integrated state-space (NISS) method" and compare it to two common methods used to analyze nonlinear time series in ecology: least squares with only process noise (LSPN) and least squares with only observation error (LSOE). We compare all three methods by fitting Beverton-Holt and Ricker models to many replicate model-generated time series of length 20 with several parameter choices. For the Ricker model we chose parameters for which the deterministic part of the model produces a stable equilibrium, a two-cycle, or a four-cycle. For each set of parameters we used three process-noise and observation-error scenarios: large standard deviation (0.2) for both, and large for one but small (0.05) for the other. The NISS method had lower estimator bias and variance than the other methods in nearly all cases. The only exceptions were for the Ricker model with stable-equilibrium parameters, in which case the LSPN and LSOE methods has lower bias when noise variances most closely met their assumptions. For the Beverton-Holt model, the NISS method was much less biased and more precise than the other methods.

We also evaluated the utility of each method for model selection by fitting simulated data to both models and using information criteria for selection. The NISS and LSOE methods showed a strong bias toward selecting the Ricker over the Beverton-Holt, even when data were generated with the Beverton-Holt. It remains unclear whether the LSPN method is generally superior for model selection or has fortuitously better biases in this particular case. These results suggest that information criteria are best used with caution for nonlinear population models with short time series.

Finally we evaluated the convergence of likelihood ratios to theoretical asymptotic distributions. Agreement with asymptotic distributions was very good for stable-point Ricker parameters, less accurate for two-cycle and four-cycle Ricker parameters, and least accurate for the Beverton-Holt model. The numerically integrated state-space method has a number of advantages over least squares methods and offers a useful tool for connecting models and data and ecology.

Key words: Beverton-Holt model; Kalman filter; least-squares cf. state-space models; model-fitting; observation error; parameter estimation; population models; process noise; Ricker model; time series.

INTRODUCTION

Relating population models statistically to data is central to answering many important questions in ecology. Does a population exhibit direct, delayed, and/or joint density dependence (Pollard et al. 1987, Turchin 1990, Turchin and Taylor 1992, Hanski et al. 1993, Kemp and Dennis 1993, Dennis and Taper 1994, Wolda et al. 1994, Dennis et al. 1998, Zeng et al. 1998, Saitoh et al. 1999)? What are the relative influences of exogenous and endogenous forces in generating observed dynamics of populations (Hastings et al. 1993, Dennis et al. 1995, Ellner and Turchin 1995, Costantino et al. 1997, Higgins et al. 1997, Ellner et al. 1998, Stenseth et al. 1998, Bjornstad et al. 1999*a*, Kendall et al. 1999)? Are population dynamics spatially synchronous (Grenfell et al. 1998, Bjornstad et al. 1999*b*)? Do two or more species interact (Ives et al. 1999)? These are the types of questions that can be addressed by fitting population models to data—either experimental or observational—and testing statistical hypotheses.

Using population models to analyze data is conceptually similar to using analysis-of-variance models to analyze data: start with a model structure that includes deterministic and stochastic components, estimate parameters under different hypotheses, make probabilistic statements about the relationship between model

Manuscript received 4 February 2000; revised 20 February 2001; accepted 23 February 2001.

[1] Present address: National Center for Ecological Analysis and Synthesis, University of California, 735 State Street, Suite 300, Santa Barbara, California 93101 USA.
E-mail: devalpin@nceas.ucsb.edu

and data that compare hypotheses—such as P values or Type I and Type II error rates—and consider the biological implications of the parameterized models. However, for population models and data the problem is more complicated because the data are related through time, many relevant models are nonlinear, and there is a wide range of possible model structures. Most importantly, there is noise in both the underlying demographic process as well as our observations of population sizes.

The problem of simultaneously including two types of noise has been vexing. The most common approaches have been to assume either that the observations are perfect or that the process is purely deterministic, but these assumptions can lead to biases in parameter estimation and hypothesis testing (Hilborn 1979, Uhler 1980, Walters and Ludwig 1981, Ludwig et al. 1988, Ludwig and Walters 1989, Polacheck et al. 1993, Schnute and Richards 1995, Hilborn and Mangel 1997, Quinn and Deriso 1999). Other approaches have been explored (Ludwig and Walters 1981, Collie and Sissenwine 1983, Ludwig et al. 1988, Ludwig and Walters 1989, Carpenter et al. 1994, Schnute and Richards 1995) but have the same basic shortcoming that they do not fully handle the probability structure created by the joint presence of process noise and observation error.

In this paper we test a method to fit population models to data simultaneously incorporating both process noise and observation error. The method uses state-space models and an extension of the Kalman filter that were developed over the past 40 yr in the engineering and statistics literature (e.g., Kalman 1960, Meinhold and Singpurwalla 1983, Kitagawa 1987, Harvey 1989, 1993, Tong 1990, Shumway and Stoffer 2000). A state-space model includes two models, one for the true dynamics of the system with process noise, the other for our observations of the true states of the system with observation error. The Kalman filter is most commonly used in engineering for estimating the state of a system when the governing model is known, but it can also be used to calculate the likelihood that a set of parameters produced a time series of data, making it useful for maximum-likelihood parameter estimation. In its standard uses, the Kalman filter applies to linear systems with Gaussian noises, and "extended" Kalman filters use Taylor series to approximate the same calculations for nonlinear systems. We extend the probability logic of the linear Kalman filter to nonlinear, non-Gaussian systems by numerically estimating all of the relevant probability distributions and calculations. We refer to this as the "numerically integrated state-space" method. This approach was used by Kitigawa (1987) for the problem of non-Gaussian noise in a non-stationary time series, and the potential for this approach was recognized before advances in computer technology made it practical (Kitigawa 1987, Kohn and Ansley 1987).

Approaches related to ours have recently begun to appear in ecology, mostly in the fisheries literature (Mendelssohn 1988, Sullivan 1992, Pella 1993, Gudmundsson 1994, Schnute 1994, Freeman and Kirkwood 1995, Kimura et al. 1996, Reed and Simmons 1996, Newman 1998, Bjornstad et al. 1999*a*, Meyer and Millar 1999). Most of these papers used Kalman filters for linear models or linear approximations of nonlinear models. Meyer and Millar (1999) and Bjornstad et al. (1999*a*) used the Bayesian Markov chain Monte Carlo method (Gilks et al. 1996). Like our method, this produces likelihood calculations that can incorporate nonlinear, non-Gaussian structure. However, it differs from our approach by taking a Bayesian view of parameters and using different—stochastically approximated—numerical methods. Schnute (1994) provides a good overview of different methods, contrasting especially the "errors in variables" approach (Ludwig and Walters 1981, 1989, Ludwig et al. 1988) with the linear and extended Kalman-filter approach. None of these studies systematically evaluated and compared the properties of Kalman-filter-related methods and other methods. Zeng et al. (1998) applied a linear Kalman filter to insect population data, but they used it to incorporate autocorrelated model parameters rather than observation noise.

We test the numerically integrated state-space (NISS) method by generating noisy data with Ricker and Beverton-Holt models and then fitting the generated data with these models. Ricker and related models are popular for detecting density dependence and have thus been a focus for model-fitting and inference methods in ecology (e.g., Turchin 1990, Turchin and Taylor 1992, Dennis and Taper 1994). We compare the bias and precision of the NISS method to two least-squares methods: least squares with only process noise and least squares with only observation error. The words "noise" and "error" are interchangeable in this context, but we tend toward using the former for process randomness and the latter for observation randomness. Since the NISS method produces likelihood calculations, we also evaluate convergence to asymptotic likelihood-ratio distributions, which would be useful for constructing confidence intervals, and power of information criteria to identify correct model structure.

We first introduce the mathematics and statistics of the method. This section can be skipped by readers interested only in how the method performs. We then describe a test of the method, including details of our implementation of the calculations, which can also be skipped. The *Results* section on the performance of the method is self-contained and can be read by itself.

Numerically Integrated State-Space Method

General framework

This introduction to state-space models follows Kitigawa (1987), Harvey (1989, 1993), and Schnute

(1994), but only Kitigawa (1987) gives the general nonlinear non-Gaussian equations, which we give here in more explicit detail. Suppose that a population is governed by the stochastic process

$$n_t = F(n_{t-1}, v_{t-1}) \quad (1)$$

where n_t is the population state at time t, v_t is a random variable for the process noise at time t, and F determines the new population state as a function of the old population state and the process noise. In the simplest case, the population state n_t would be population size, but in more complicated cases it could be a vector of population sizes of different age or stage classes, or any other vector of population information. Similarly, v_t may be a single number or a vector of more than one noise input to the system.

Suppose that our observation, y_t, at time t of n_t is governed by the stochastic observation model

$$y_t = G(n_t, \varepsilon_t) \quad (2)$$

where ε_t is a random variable for the observation noise at time t, and G determines the process of observing n_t with error ε_t. An observation y_t may be a single estimate of population size or a vector of estimates of age or stage classes or other population information. An observation error ε_t may be a single number or a vector. There are parameters to estimate contained in the functions F and G as well as in the distributions of values of v_t and ε_t, and we call the vector of these parameters Θ. In the examples in this paper, Θ is constant over time.

In our exposition, we assume that values of v_t and ε_t are each serially independent, independent from each other, and have constant distributions. In time-series terminology, the population model (Eq. 1) is a nonlinear, first-order, autoregressive process (Tong 1990). However, the state-space model structure (Eqs. 1 and 2) can be extended to more complicated cases, including models of multiple species, multiple stage classes, and/or time lags, and to models where some dimensions of the state variable are not estimated by the observations. The use of separate equations for the true underlying state of the system and observations of that state is what makes this a "state-space" model.

In any particular application, the observation model (Eq. 2) would reflect the method used to obtain the data y_t. In a simple case data may be obtained from direct but inaccurate estimation of all of the stage classes of n_t. However, in many cases data may be estimates of only one of several components of n_t, as when only adults can be surveyed but juveniles are built into the population model F (e.g., Higgins et al. 1997), or estimates of a combination of components of n_t, as when stage classes must be lumped for estimation but modeled separately in F. Another example of a stochastic observation model would be a model for mark–recapture data. In each of these cases, given a true unknown population state, there can be an observation error that is independent of the process noise that led to that state. Coarse observations are obviously less desirable than detailed observations, but it is useful that the state-space framework naturally accommodates either situation.

Likelihoods

We begin by defining likelihoods, which are a common foundation for statistical inference (e.g., Stuart and Ord 1991, Edwards 1992, Dennis et al. 1995, Hilborn and Mangel 1997). Suppose we have a series of T observations $y_1, \ldots, y_T$. We label the first t observations as $\mathbf{y}^t$, so subscript t denotes the particular observation of y at time t, and superscript t denotes the series of observations up to and including time t, $(y_1, \ldots, y_t)$ (following Schnute 1994). The entire series is $\mathbf{y}^T$. We first explain how to calculate the likelihood that a fixed set of parameters, Θ, produced the data $\mathbf{y}^T$. Then we can maximize the likelihood over the parameter space and compare it to the likelihood under a null hypothesis.

The likelihood of a possible choice of parameters Θ is defined as the probability that if those were the real parameters, they would have produced the observed data. This is written as $L(\Theta \mid \mathbf{y}^T) = P(\mathbf{y}^T \mid \Theta)$. Here "$L$" stands for likelihood and "P" for probability. The difference between them is that L is viewed as a function of the parameters given the data and P is viewed as a function of the data given the parameters. For a time series, the likelihood can be conveniently expressed recursively as

$$L(\Theta \mid \mathbf{y}^T) = P(y_1 \mid \Theta) \prod_{t=2}^{T} P(y_t \mid \mathbf{y}^{t-1}, \Theta). \quad (3)$$

This equation says that the probability of the entire sequence of observations is the probability of the first observation times the probability of the second given the first times the probability of the third given the first two, and so on. The form of this equation suggests an iterative calculation procedure.

The discrete case

We now consider calculation of the likelihood, Eq. 3, using the model's components (F, G, and the distributions of v_t and ε_t) for some choice of parameters, Θ. We introduce the calculations as if the states and observations take only discrete values, such as integers, before giving the full equations for states and observations that can take a continuous range of values. This allows simplification of the equations and emphasis on the concepts.

The likelihood can be calculated iteratively forward in time: compute $P(y_1 \mid \Theta)$ first, then use this to compute $P(y_2 \mid \mathbf{y}^1, \Theta)$ (note $\mathbf{y}^1 = y_1$), and in general use the results of step $t - 1$ to compute $P(y_t \mid \mathbf{y}^{t-1}, \Theta)$. We violate this pattern in one respect: $P(y_1 \mid \Theta)$ depends on $P(n_1 \mid \Theta)$, but for explanation purposes we assume we know $P(n_1 \mid \Theta)$ for all values of n_1. Later we use the stationary

distribution of the population model (Eq. 1) for this initial state distribution and describe how this is obtained. The calculations are for a specific set of parameter values, so the dependence on Θ will be omitted from here on.

Step 1.—For the probability of the first observation, we get

$$P(y_1) = \sum_{n_1} P(n_1)P(y_1|n_1) \quad (4)$$

where the sum is over all possible values of n_1. The interpretation of this equation is that one way to obtain observation y_1 is to have state n_1 and then observation y_1 given n_1, and we must sum the probabilities of all such combinations to get the total probability of y_1. The probability of an observation given a state, $P(y_1|n_1)$ is defined by Eq. 2. For example if the observation error is additive ($y_1 = G(n_1, \varepsilon_1) = n_1 + \varepsilon_1$), then $P(y_1|n_1) = P(\varepsilon_1 = y_1 - n_1)$.

Our final use of the first observation is to adjust the distribution of states to reflect the information contained in the first observation. Mathematically, the state n_1 and the observation y_1 are jointly distributed random variables, and we want the conditional distribution of n_1 given y_1. This is given by

$$P(n_1|\mathbf{y}^1) = \frac{P(n_1)P(y_1|n_1)}{P(y_1)}. \quad (5)$$

This equation will be used in Step 2.

Step 2.—For the probability of the second observation, we get

$$P(y_2|\mathbf{y}^1) = \sum_{n_2} P(n_2|\mathbf{y}^1)P(y_2|n_2) \quad (6)$$

where the sum is over all possible values of n_2. This is similar to Eq. 4, but now we must deal with $P(n_2|\mathbf{y}^1)$, the probability of state n_2 given the information up to the previous observation. This is given by

$$P(n_2|\mathbf{y}^1) = \sum_{n_1} P(n_1|\mathbf{y}^1)P(n_2|n_1) \quad (7)$$

where the sum is over all possible values of n_1. The interpretation of this equation is that one way to have state n_2 given observation $\mathbf{y}^1$ is to have state n_1 given $\mathbf{y}^1$ and state n_2 given n_1, and we must sum the probabilities of all such combinations. Note that $P(n_2|n_1)$ is defined by the stochastic population model (Eq. 1). We write $P(n_2|n_1)$ rather than $P(n_2|n_1, \mathbf{y}^1)$ because $\mathbf{y}^1$ is uninformative about n_2 if n_1 is known. Combining Eqs. 6 and 7, we get

$$P(y_2|\mathbf{y}^1) = \sum_{n_1} \sum_{n_2} P(n_1|\mathbf{y}^1)P(n_2|n_1)P(y_2|n_2) \quad (8)$$

where the sums are over all possible values of n_1 and n_2 that could have led to observation y_2.

Parallel to step 1, our final use of the second observation is to adjust the distribution of states at time 2 to reflect the information contained in the second observation. Here n_2 (given $\mathbf{y}^1$) and y_2 are jointly distributed, and the conditional distribution of n_2 given y_2 and $\mathbf{y}^1$ is

$$P(n_2|\mathbf{y}^2) = P(n_2|y_2, \mathbf{y}^1) = \frac{P(n_2|\mathbf{y}^1)P(y_2|n_2)}{P(y_2|\mathbf{y}^1)}. \quad (9)$$

Note that whereas in Eq. 6 we are interested in only a particular value of y_2—our observation—in Eq. 9 we are interested in all values of n_2, and the numerator of Eq. 9 for each value of n_2 corresponds to one term in the sum of Eq. 6.

Later steps.—All subsequent steps are similar to step 2. For step t, we begin with $P(n_{t-1}|\mathbf{y}^{t-1})$ from the previous step, calculate $P(y_t|\mathbf{y}^{t-1})$, and update the state distribution to be $P(n_t|\mathbf{y}^t)$. The formulae for $P(y_t|\mathbf{y}^{t-1})$, $P(n_t|\mathbf{y}^{t-1})$, and $P(n_t|\mathbf{y}^t)$ are the same as Eqs. 6, 7, and 9, respectively, but with "$t - 1$" in place of "1" and "t" in place of "2."

It is worth commenting on the meaning of the probability of state n_t given the observations up to time $t - 1$. One can imagine an infinite set of repetitions of the process up to time t (i.e., each repetition gives a series of length t, and there are infinitely many such repetitions). If we take the subset of those repetitions that have exactly the same observations as ours up to time $t - 1$, then $P(n_t|\mathbf{y}^{t-1})$ is the proportion of those that have state n_t at time t.

A comment is also due about the conditional probability equations (Eqs. 5 and 9). Each of these equations is Bayes' law, but this is not a philosophically Bayesian analysis. In Eqs. 5 and 9 the conditional probability involves probability distributions for true random variables – the states and observations of the system, and parameters are viewed as fixed. Philosophically Bayesian analyses view parameters as random variables and reinterpret their probability distribution in terms of "degree of belief" or some similar concept, but this is different from our mathematical use of Bayes' law (Edwards 1992, Dennis 1996, Hilborn and Mangel 1997).

The continuous case

We now introduce the calculations for continuous probability distributions, and this requires more careful notation. The procedure involves four related probability density functions. A probability density function (pdf) describes the probability with which any range of values of a continuous random variable occurs. The first pdf is for the distribution of states n_t given observations up to and including that time, $\mathbf{y}^t$, which is written as $f_{N_t|\mathbf{y}^t}(n_t)$. In this notation for pdf's, the subscript denotes the random variable in question, which is written as a capitalized version of the notation for a particular realization of that variable. The argument of the pdf is a particular possible value of the random variable. For example, $f_{N_t|\mathbf{y}^t}(n_t)$is the probability density for the particular value n_t from the distribution of all possible values of n_t given $\mathbf{y}^t$. This notation is a bit

cumbersome, and in other situations one could suppress the subscripts when they are obvious from the function arguments. However, in what follows we sometimes have function arguments that do not reveal what distribution is being considered, so the subscripts on f are necessary. The other important pdf's are $f_{N_t|\mathbf{y}^{t-1}}(n_t)$, the probability density of n_t given the observations up to time $t-1$; $f_{Y_t|\mathbf{y}^{t-1}}(y_t)$, the probability density of y_t given observations up to time $t-1$; and $f_{Y_t|n_t}(y_t)$, the probability density of y_t given that the current state is n_t. This last pdf is defined entirely by the function G and the pdf of ε.

The likelihood equation, equivalent to Eq. 3, is now

$$L(\boldsymbol{\Theta}|\mathbf{y}^T) = f_{Y_1}(y_1|\boldsymbol{\Theta}) \prod_{t=2}^{T} f_{Y_t|\mathbf{y}^{t-1}}(y_t|\boldsymbol{\Theta}). \quad (10)$$

Step 1.—As in the discrete case, we begin with y_1 and n_1 and omit $\boldsymbol{\Theta}$ from the notation. We again start with the stationary distribution of the stochastic process model (Eq. 1) as the distribution of n_1 and put off discussion of obtaining the stationary distribution.

The continuous equivalents of Eqs. 4 and 5 are

$$f_{Y_1}(y_1) = \int f_{N_1}(n_1) f_{Y_1|n_1}(y_1)\, dn_1 \quad (11)$$

$$f_{N_1|\mathbf{y}^1}(n_1) = \frac{f_{N_1}(n_1) f_{Y_1|n_1}(y_1)}{f_{Y_1}(y_1)} \quad (12)$$

respectively. Integrals in the continuous case correspond to sums in the discrete case.

Step 2.—Continuing to parallel the discrete case, we have the equivalent of Eq. 6 for the pdf of the second observation:

$$f_{Y_2|\mathbf{y}^1}(y_2) = \int f_{N_2|\mathbf{y}^1}(n_2) f_{Y_2|n_2}(y_2)\, dn_2. \quad (13)$$

As before, this requires that we calculate the pdf of states at time 2 given the first observation. This corresponds to Eq. 7 and is given for the continuous case by

$$f_{N_2|\mathbf{y}^1}(n_2) = \int f_{N_1|\mathbf{y}^1}(n_1) f_{N_2|n_1}(n_2)\, dn_1. \quad (14)$$

Finally, the conditional distribution of n_2 given y_2, the equivalent of Eq. 9, is

$$f_{N_2|\mathbf{y}^2}(n_2) = \frac{f_{N_2|\mathbf{y}^1}(n_2) f_{Y_2|n_2}(y_2)}{f_{Y_2|\mathbf{y}^1}(y_2)}. \quad (15)$$

Later steps.—Again we can generalize that for step t, we start with $f_{N_{t-1}|\mathbf{y}^{t-1}}(n_{t-1})$ and calculate the pdf of observation y_t:

$$f_{Y_t|\mathbf{y}^{t-1}}(y_t) = \int f_{N_t|\mathbf{y}^{t-1}}(n_t) f_{Y_t|n_t}(y_t)\, dn_t. \quad (16)$$

This requires the pdf of state n_t:

$$f_{N_t|\mathbf{y}^{t-1}}(n_t) = \int f_{N_{t-1}|\mathbf{y}^{t-1}}(n_{t-1}) f_{N_t|n_{t-1}}(n_t)\, dn_{t-1}. \quad (17)$$

To calculate the conditional distribution of states given the latest observation, the general equation is

$$f_{N_t|\mathbf{y}^t}(n_t) = \frac{f_{N_t|\mathbf{y}^{t-1}}(n_t) f_{Y_t|n_t}(y_t)}{f_{Y_t|\mathbf{y}^{t-1}}(y_t)}. \quad (18)$$

The stationary distribution for $P(n_1)$.—We now return to an issue deferred earlier: what to use for the distribution of $f_{N_1}(n_1)$, which is the distribution of states before we have conditioned on any observations. It turns out that under a wide range of circumstances, a Markov chain model such as Eq. 1 produces a stationary distribution of population states. A stationary distribution can be thought of as a stochastic equilibrium (e.g., Turchin 1995). If the system has been in operation for a long time and we start observing it at a random time, the distribution of population states will be the stationary distribution. Most stochastic population models with some type of regulation and vanishingly small chance of extinction will have a stationary distribution. In contrast, models that allow arbitrarily large excursions of population size, such as random walks, do not have stationary distributions.

Tong (1990:122–126) more formally discusses conditions in which stationary distributions exist and connects them to the idea of "stochastic stability." Roughly speaking, a stationary distribution will exist for models that lack fixed periodicity, allow all states to eventually be reached from all other states (i.e., lack subsets of states that trap the system), and tend to return from large excursions away from common states. Stationary distributions for stochastic models often exist if the underlying deterministic model has a basin of attraction, even if the deterministic dynamics converge to limit cycles or chaotic attractors. For the examples we use in this paper, stationary distributions exist for all model-parameter combinations we study. There are parameters for these models that lack density-dependence, and hence allow arbitrarily large excursions and lack a stationary distribution (e.g., the Ricker with no density-dependence, Dennis and Taper 1994), but none of our parameter estimates approach these parameter values.

A stationary distribution for (1) can be calculated by iterating the projection equation for the pdf of population states,

$$f_{N_t}(n_t) = \int f_{N_{t-1}}(n_{t-1}) f_{N_t|n_{t-1}}(n_t)\, dn_{t-1} \quad (19)$$

until $f_{N_t}(n)$ is the same distribution as $f_{N_{t-1}}(n)$ (cf. Tong 1990, eq. 4.11). For example, one can start with any distribution for $f_{N_t}(n)$, set $f_{N_{t-1}}(n) = f_{N_t}(n)$, calculate $f_{N_t}(n)$, and repeat until the distributions are virtually identical. By analogy with a deterministic system, $n_t = h(n_{t-1})$, one could estimate a stable equilibrium of h

by starting with any value n_t, setting $n_{t-1} = n_t$, calculating n_t, and repeating until $n_t \approx n_{t-1}$. Note that Eq. 19 is the same as (17) without any conditioning on observations. As with all of our probability calculations, we iterate (19) numerically.

Calculating pdf's from functions of random variables.—We are still one step shy of giving usable formulae for the likelihood (10); we need to relate Eq. 11–19 to the components of the process and observations models (1) and (2), namely F, G, and the distributions of values of ν_t and ε_t. We consider the equations in the form (16–18). The pdf's of ν_t and ε_t will be written $f_\nu(\nu_t)$ and $f_\varepsilon(\varepsilon_t)$, respectively.

A brief digression into probability theory will simplify the rest of this section. Suppose X is a random variable with pdf $f_X(x)$, and $y = h(x)$ is a (differentiable) function of x. How can we calculate the pdf of Y, $f_Y(y)$, from $f_X(x)$ and the function h? To answer this, we must remember that probability densities provide the probability of any *range* of random outcomes, not of specific, exact outcomes. By saying X has pdf $f_X(x)$, we mean that the probability that X falls in the range $[x, x + \delta x]$ is $f_X(x)\delta x$, for infinitesimal δx. Similarly, the probability that Y falls in the range $[y, y + \delta y]$ is $f_Y(y)\delta y$. In particular, we are interested in the range of values $[y, y + \delta y]$ that corresponds to the range $[x, x + \delta x]$, i.e. either (i) $y = h(x)$ and $y + \delta y = h(x + \delta x)$ or (ii) $y = h(x + \delta x)$ and $y + \delta y = h(x)$. The probability that X is in the range $[x, x + \delta x]$ must equal the probability that Y is in the corresponding range $[y, y + \delta y]$. Thus $f_Y(y)|\delta y/\delta x| = f_X(x)$. The absolute value is used to cover both cases (i) and (ii) just mentioned, i.e. to avoid negative probability. In the infinitesimal limits, $\delta y/\delta x$ is the derivative $dy/dx = dh(x)/dx$, which will be written $h'(x)$. The derivative is evaluated at the value of x that gives $y = h(x)$, which is the inverse of h, $x = h^{-1}(y)$. So, we have the general rule that

$$f_Y(y) = \frac{f_X[h^{-1}(y)]}{|h'[h^{-1}(y)]|}. \tag{20}$$

As a simple example, suppose X is uniformly distributed between 0 and 1, and $y = h(x) = 10x$. By intuition one can imagine that Y must be uniformly distributed between 0 and 10. The pdf of X is equal to 1 for x between 0 and 1, which integrates to 1, as it should to be a proper pdf. Then the pdf of Y should be constant between 0 and 10, and must also integrate to 1, so it must be equal to 0.1. Thus the pdf of Y at the value y is equal to the pdf of X at $x = h^{-1}(y)$ divided by the slope $h'(x) = 10$. The informal reasoning behind this justification is the same as the reasoning for a change of variables for integration in calculus. Readers interested in more detail are referred to introductory texts in probability theory or mathematical statistics, such as Rice (1988: 54–58).

Calculating pdf's from the population and observation models.— Pdf's for the observation probability.—We now return to Eq. 16, and we consider first $f_{Y_t|n_t}(y_t)$. This is the pdf of an observation given a state, and so will be derived from $y_t = G(n_t, \varepsilon_t)$. Here we view n_t as fixed, ε_t as a random variable with a known pdf (for a given $\boldsymbol{\Theta}$), and y_t as a function of ε_t. Then, by the Eq. 20 rule, the pdf of Y is

$$f_{Y_t|n_t}(y_t) = \frac{f_\varepsilon[G^{-1}(n_t, y_t)]}{|G'[n_t, G^{-1}(n_t, y_t)]|} \tag{21}$$

where G' is the derivative of G with respect to ε, and $G^{-1}(n_t, y_t)$ is the inverse of G for fixed state n_t: the value ε that would have produced the observation $y_t = G(n_t, \varepsilon)$. For the simple case that the expected value of the observation is the true value of the state and the errors due to ε are linear, we have

$$y_t = n_t + \varepsilon_t \qquad f_{Y_t|n_t}(y_t) = f_\varepsilon(y_t - n_t). \tag{22}$$

Pdf's for the state probability.—We now move to Eq. 17. We could derive $f_{N_t|n_{t-1}}(n_t)$ from $n_t = F(n_{t-1}, \nu_{t-1})$ by viewing n_{t-1} as fixed, so that n_t is a function of ν_{t-1}, for which we know the pdf. However, there is an alternative approach for the examples in this paper that turns out to be helpful for implementing these calculations numerically. In this paper we will use process models with additive noise:

$$n_t = F(n_{t-1}, \nu_{t-1}) = F_0(n_{t-1}) + \nu_{t-1} \tag{23}$$

where the subscript 0 on F indicates that this is the function of n_{t-1} obtained by $F(n_{t-1}, 0)$.

This form allows us to simplify Eq. 17. The interpretation of Eq. 17 is that the pdf of n_t requires consideration of all possible values of n_{t-1} and ν_{t-1} that give $n_t = F(n_{t-1}, \nu_{t-1})$. An equivalent view using Eq. 23 is to handle the "deterministic" function $n_d = F_0(n_{t-1})$ first and then the additive combination of two random variables, $n_t = n_d + \nu_{t-1}$, second to obtain the pdf of n_t.

The pdf of the "deterministic" part comes from applying the Eq. 20 rule to $n_d = F_0(n_{t-1})$:

$$f_{N_d|\mathbf{y}^{t-1}}(n_d) = \frac{f_{N_{t-1}|\mathbf{y}^{t-1}}[F_0^{-1}(n_d)]}{|F_0'[F_0^{-1}(n_d)]|} \tag{24}$$

where F_0' and F_0^{-1} are the derivative and inverse, respectively, of F_0.

The pdf of n_t is then

$$f_{N_t|\mathbf{y}^{t-1}}(n_t) = \int f_{N_d|\mathbf{y}^{t-1}}(n_t - \nu) f_\nu(\nu)\, d\nu. \tag{25}$$

This equation is known as a convolution integral, and we use the trick of Fourier transforms to calculate it numerically.

Relation to the Kalman filter

When F and G are linear functions in both n and ν or ε, respectively, and ν and ε are Gaussian, the above equations simplify enormously. Linear functions of Gaussian distributions as well as conditional distri-

butions of multivariate Gaussian distributions are themselves Gaussian and require only simple calculations for the mean and variance of each distribution; these are the Kalman-filter equations (Harvey 1989, 1993). For nonlinear functions and/or non-Gaussian noises, the above equations can be closely approximated by numerically discretizing the distributions (Kitigawa 1987), which is the approach we use here. An intermediate approach to handle nonlinearity is to approximate functions of distributions by a Taylor series, using derivatives of F and G and forcing the distributions to remain Gaussian by tracking only their mean and variance. We used the more precise full numerical calculations because ecological models can be substantially nonlinear and hence produce non-Gaussian distributions and because with few data points in most ecological time series and advances in computer speed, it is feasible to do the full, accurate statistical calculations with them.

Testing the Method

The models

A crucial step in evaluating a model-fitting method is to generate data from a known model and see how well the method estimates parameter values. This makes it possible to apply methods to real data guided by knowledge of their estimation properties. We studied the properties of the new method by comparing it to least-squares methods and testing agreement of its likelihood ratios with theoretical asymptotic distributions. We used two models, the stochastic Beverton-Holt model,

$$M_t = M_{t-1}\frac{\lambda}{1 + \gamma M_{t-1}}e^{\nu_{t-1}} \tag{26}$$

and the stochastic Ricker model,

$$M_t = M_{t-1}e^{r-bM_{t-1}}e^{\nu_{t-1}}. \tag{27}$$

For both models, M_t is the population size at time t, measured on an arbitrary but fixed scale such as 100s of individuals, and ν_t is the process noise variable, which is normally distributed with mean 0 and variance σ_ν^2.

For the deterministic part of the Beverton-Holt model, λ is the population growth rate when the population is small, $(\lambda - 1)/\gamma$ is the equilibrium population size, λ/γ is the limit of the maximum of M_t and occurs as $M_{t-1} \to \infty$, and when $\gamma > 0$ there is density dependence. For the deterministic part of the Ricker model, e^r is population growth rate when the population is small, r/b is the equilibrium population size, e^{r-1}/b is the maximum of M_t and occurs for $M_{t-1} = 1/b$, and when $b > 0$ there is density dependence. A fundamental difference between the two models involves the strength of density dependence. For the Ricker model, large M_{t-1} produces small M_t as a result of strong density dependence. For the Beverton-Holt model, large M_{t-1} produces M_t close to λ/γ, so large populations at one time do not produce population crashes at the next time.

The natural-logarithm form of these models is convenient for parameter estimation (e.g., Turchin and Taylor 1992, Dennis and Taper 1994). Define $n_t = \log(M_t)$, so

$$\begin{aligned} n_t &= F_{\mathrm{BH}}(n_{t-1}, \nu_{t-1}) \\ &= n_{t-1} + \log(\lambda) - \log(1 + \gamma e^{n_{t-1}}) + \nu_{t-1} \end{aligned} \tag{28}$$

for the Beverton-Holt model, and

$$n_t = F_{\mathrm{R}}(n_{t-1}, \nu_{t-1}) = n_{t-1} + r - be^{n_{t-1}} + \nu_{t-1} \tag{29}$$

for the Ricker model. For both models we consider observations of n_t that are unbiased but inaccurate counts:

$$y_t = G(n_t, \varepsilon_t) = n_t + \varepsilon_t \tag{30}$$

where ε_t is the observation error and is normally distributed with mean 0 and variance σ_ε^2.

Our choice of additive Gaussian process noises and observation errors simplifies comparison of the NISS (numerically integrated state-space) method with least-squares methods. The NISS method could accommodate a wide variety of non-additive and non-Gaussian noise and error models. However, least-squares methods assume additive Gaussian noise or error. The probability density functions (pdf's) involved in the NISS calculations are still highly non-Gaussian because of the nonlinear structure of the Ricker and Beverton-Holt models.

For the Ricker model, we considered three parameter choices with different deterministic dynamics: $r = 1.5$, 2.4, and 2.6. These parameters give a stable equilibrium, a stable two-cycle, and a stable four-cycle, respectively, when the model is purely deterministic (Fig. 1). The Ricker parameter b does not determine stability or cyclic behavior of the model, so we chose it to create a constant equilibrium value across different choices of r. Since the equilibrium value is $\hat{M} = r/b$, we used $b = r/100$. This is arbitrary since the units of M remain unspecified. The deterministic Beverton-Holt model produces only stable equilibrium dynamics, so we considered only one choice of parameters for it. We chose $\lambda = 4.48$ and $\gamma = (\lambda - 1)/100$ in order to have the same equilibrium and growth rate for small populations as the $r = 1.5$, $b = 0.015$ Ricker model.

Process-noise and observation-error combinations

For each choice of parameters, we used three combinations of process noise and observation errors: large standard deviation for both and large for one but small for the other. The large standard deviations were 0.2, and the small standard deviations were 0.05. In terms of the models before log transformation, about 67% of the time M_{t+1} will be within about $100\sigma_\nu$% of its deterministic value given M_t and about 95% of the time it will be within twice that range. These values are

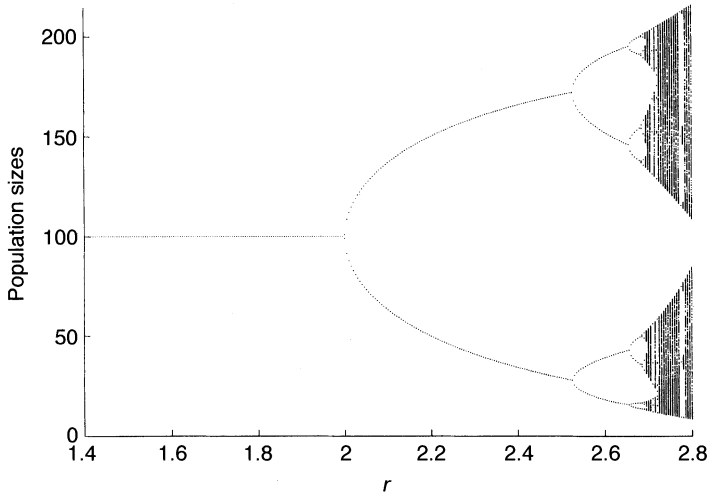


FIG. 1. Bifurcation diagram for the Ricker model. Parameter r varies along the x-axis, and parameter b is chosen by fixing equilibrium population size, r/b, at 100. For each r, the model is iterated for 1000 time steps, and the next 1000 time steps are plotted as population size (y-axis) values for that value of r. As r increases past 2.0, model dynamics bifurcate from a single stable equilibrium to a cycle alternating between two values (a two-cycle) and then four values (a four-cycle).

approximate because $e^x \approx 1 + x$ for small x, but they give a feel for the noises involved. These choices for "large" and "small" standard deviations are ad hoc; in some systems 0.20 might be small, and in others 0.05 might be large.

For each choice of parameters, process noise, and observation error, we generated 300 time series of length 20. The first observation of each time series was drawn essentially from a stationary distribution by iterating the model 1000–2000 times before taking observations. We fit these time series using the numerically integrated state-space method (NISS), least squares with only process noise (LSPN), and least squares with only observation error (LSOE). For the Ricker case of $r = 1.5$ and the corresponding Beverton-Holt case of $\lambda = 4.48$, we fit the time series using the correct model as well as the wrong model to examine the ability of the different fitting methods to distinguish correct model structure using information criteria.

For our examples, the NISS method generalizes the LSPN and LSOE methods in the sense that if one assumes $\sigma_\varepsilon^2 = 0$ or $\sigma_\nu^2 = 0$, then NISS is almost equivalent to LSPN or LSOE, respectively. In each case the difference has to do with treatment of the first observation. LSPN does not calculate a probability for the first observation, and LSOE attempts to estimate the first observation, while NISS calculates its probability from a stationary distribution. However, after the first observation, the comparisons are exact. Under NISS with no observation error, the conditional state distribution (Eq. 18) will describe the observation value as the only possible state of the system (i.e., the distribution will be like a Dirac delta function, which places all probability on a single value), which is the way LSPN treats observations. Under NISS with no process noise, the conditional state distribution (Eq. 18) will describe a deterministic state as the only possible value and will be unaffected by the observations. Thus the trajectory of conditional state distributions will be a deterministic trajectory, as in LSOE. This equivalence occurs because in the logarithmic form of the example models here (Eqs. 28, 29, and 30), the observation errors and process noises are normal and additive, which match the assumptions of LSPN and LSOE, respectively.

Implementation of NISS

Discretization of n and f(n).—We implemented the numerically integrated state-space method by discretizing the range of possible values of n and y. We chose minimum and maximum values, $n_{\min}$ and $n_{\max}$, so that all the data (for a particular time series), as well as the extreme tails of the range of possible state and observation values under the model (estimated from a stationary distribution as calculated below) with parameters Θ, fell between $n_{\min}$ and $n_{\max}$. We used 4096 values of n, which will be indexed with brackets, "[]":

$$n[1] = n_{\min}$$

$$n[i] = n_{\min} + (i - 1)\Delta n \qquad 1 \le i \le 4096$$

$$n[4096] = n_{\max}$$

$$\Delta n = \frac{n_{\max} - n_{\min}}{4095}.$$

Note that the $n[i]$ serve as a range of possible values for both n and y in the model.

A pdf for a random variable that takes values in the range $[n_{\min}, n_{\max}]$ is approximated by a set of values $f[i] = f(n[i])$. To approximate the value of $f(n)$ for an arbitrary value of n, we can find the index i so that $n[i] \le n < n[i + 1]$ and approximate $f(n)$ by linear interpolation as

$$f(n) \approx f[i] + (n - n[i])\left(\frac{f[i + 1] - f[i]}{\Delta n}\right). \quad (31)$$

Using these discretizations for n and any pdf's of n or y, we approximated the calculations of Eqs. 11–19.

Projection of state distributions.—The trickiest calculations to approximate are the state projections, given for general t by Eq. 17. We discuss the simplified version (Eqs. 24 and 25) for the case of additive process noise. We treated this as follows. For clarity, we will now index the discrete n values as $n[j]$ for n_t and as $n[i]$ for n_{t-1}. Consider the approximations $f_{N_t}[j]$ for $f_{N_t}(n_t)$ and $f_{N_{t-1}}[i]$ for $f_{N_{t-1}}(n_{t-1})$. The value $n[j]$ may be viewed as the center of a probability bin for n_t that ranges from $n[j] - 1/2\Delta n$ to $n[j] + 1/2\Delta n$. Then the probability that n_t is in this bin is approximately $f_{N_t}[j]\Delta n$. This must be the same as the probability that n_{t-1} is between $n_{\text{low}} = F_0^{-1}(n[j] - 1/2\Delta n)$ and $n_{\text{high}} = F_0^{-1}(n[j] + 1/2\Delta n)$. For simplicity in this explanation, we assume $n_{\text{low}} < n_{\text{high}}$ and $F_0(n)$ increases between them; the reverse situation follows similarly. To obtain the probability that $n_{\text{low}} < n_{t-1} < n_{\text{high}}$, we need to approximate

$$\int_{n_{\text{low}}}^{n_{\text{high}}} f_{N_{t-1}}(n_{t-1})\, dn_{t-1}.$$

To approximate this, we need to consider how n_{low} and n_{high} are located among the $n[i]$. One possibility is that there is an index i_s such that $n[i_s] < n_{\text{low}} < n_{\text{high}} < n[i_{s+1}]$, which means that n_{low} and n_{high} are between the same pair of neighboring $n[i]$'s. In this case,

$$\begin{aligned}\int_{n_{\text{low}}}^{n_{\text{high}}} & f_{N_{t-1}}(n_{t-1})\, dn_{t-1} \\ &\approx (n_{\text{high}} - n_{\text{low}})\left[f_{N_{t-1}}[i] + \left(\frac{n_{\text{high}} + n_{\text{low}}}{2} - n[i]\right)\right. \\ &\qquad \left.\times \left(\frac{f_{N_{t-1}}[i+1] - f_{N_{t-1}}[i]}{\Delta n}\right)\right]. \quad (32)\end{aligned}$$

Other possibilities are that n_{low} and n_{high} are located between different pairs of neighboring $n[i]$ values. Following similar lines as (31) and (32), one can obtain linear approximations of the integral of the pdf between any values of n_{low} and n_{high}. Note also that for the Ricker model, there are usually two values of the inverse, so that for each $n[j]$, there are two pairs of values of n_{low} and n_{high}, and the probability that $n[i]$ is between either pair must be calculated.

There is a special circumstance that arises when $F_0^{-1}(n[j] + 1/2\Delta n)$ does not exist. This occurs for the maximum $n[j]$ such that $n[j] - 1/2\Delta n$ is less than the maximum value of the Ricker or Beverton-Holt functions, i.e., near the "turning point" of the Ricker at which the Ricker switches from increasing to decreasing. In this case we used the smaller and larger of the two inverses $F_0^{-1}(n[j] - 1/2\Delta n)$ for n_{low} and n_{high}, respectively. Integrating between these points gives the probability that $n[i]$ is in the region that produces the highest possible values of $F_0(n[i])$.

One reason for this numerical approach is that the Ricker map of a pdf produces a singularity at the point where the Ricker has zero derivative, and the Beverton-Holt map behaves similarly for large n_{t-1}. Our approach allows numerical near conservation of probability at the cost of small inaccuracy in handling the singularity. An alternative approach might be to calculate each value of f_{N_t} "directly" from Eq. 24, that is, from

$$f_{N_t}[j] = \frac{f_{N_{t-1}}[F_0^{-1}(n[j])]}{|F_0'[F_0^{-1}(n[j])]|}$$

where $f_{N_{t-1}}(F_0^{-1}(n[j]))$ is approximated by (31). However, this approach would have the difficulty that values near the singularity are unstable to small differences in the grid arrangement: small grid changes can lead to large changes in $f_{N_t}[j]$ when $F_0'(F_0^{-1}(n[j]))$ is near zero. Our approach was stable to changes in grid arrangement and, moreover, produced only negligible losses in total probability: $\Sigma_i f_N[i]\Delta n$ stayed very close to 1. Nevertheless, to limit accumulation of numerical inaccuracies in total probability, which would affect likelihood values, we multiplied all $f_N[i]$ values by a correction factor (very close to 1) after each time step to achieve $\Sigma f_N[i]\Delta n = 1$. A second useful feature of our numerical approach is that, for a particular choice of Θ, the inverses $F_0^{-1}(n[j] \pm 1/2\Delta n)$ can be calculated once and then reused for each time step.

Adding process noise.—The final part of a state projection is to add process noise (Eq. 25). We accomplished this using a fast Fourier transform (FFT) as implemented by Frigo and Johnson (1997). The FFT may be thought of as a fast computational tool for calculating integrals like Eq. 25, known as convolution integrals. Without this trick Eq. 25 would be computationally intense because for each value n_t, one must sum over many values of n_{t-1} and v_{t-1}. The algorithm taking advantage of the FFT is to calculate the Fourier transform of $f_{N_d}(n_d)$ and $f_v(v)$ (implemented as the discrete Fourier transform, or FFT, of $f_{N_d}[i]$ and $f_v[i]$), multiply them at each frequency value, and calculate the inverse Fourier transform of the product, which gives the desired answer, $f_{N_t}(n_t)$ (implemented as $f_{N_t}[i]$). This method is especially efficient for discrete Fourier transforms using a number of points that is a power of 2, hence our choice of 4096 discretization points. Further explanation of the theory behind Fourier transforms is beyond the scope of this paper and is introduced by Press et al. (1992) and many applied mathematics texts, such as Strang (1986). In addition to the other considerations for choosing $n_{\max}$, we ensured that $n_{\max}$ was large enough that the FFT method of adding process noise did not incorrectly move probability from one end of f_N to the other, which can happen since Fourier transforms treat functions as being periodic (Press et al. 1992).

Likelihood of an observation and conditional state distribution.—To approximate the likelihood of a single observation, Eq. 16, with the additive error model (Eq. 22), using our numerical discretization, we used

$$f_{Y_t}(y_t) \approx \sum_i f_N[k - i] f_\varepsilon(i(\Delta n))\Delta n$$

where the index k is chosen so that $n[k] - 1/2\Delta n < y_t < n[k] + 1/2\Delta n$. This sum considers all values of n_t and ε_t that could produce y_t. This approach treats the y_t's as having resolution only as good as Δn, but since we used very small Δn, this was very accurate. It was also efficient because, for any value of σ_ε, we needed to calculate $f_\varepsilon(i(\Delta n))$ only once for each i.

Conditional distribution of states.—Following similar lines, we approximated the conditional distribution of states given an observation, (Eq. 18), by

$$f_{N_t|Y_t}[k - i] \approx \frac{f_N[k - i] f_\varepsilon(i(\Delta n))}{f_{Y_t}(y_t)}\Delta n.$$

Stationary distribution.—The stationary distribution of n used to initiate the fitting process was obtained by iterating Eq. 24, implemented numerically as described in *Projection of state distributions,* above, until the distribution converged to a repeating distribution, the stationary distribution. Convergence was assessed by requiring the sum of squared differences between f_n and $f_{n_{t-1}}$ over all points in the discretization to be smaller than a convergence criterion of 0.01.

Likelihood maximization

To maximize the likelihood, we minimized the negative log likelihood using a Nelder-Mead simplex algorithm adapted from Press et al. (1992). We also considered a conjugate gradient algorithm (Press et al. 1992), which can be more efficient when it works, but for the present problem we found it to be less reliable than the simplex algorithm. We initialized the simplex algorithm with a random set of parameters drawn from a large distribution around the true parameters. For some parameter ranges there were multiple optima. We handled this by restarting the algorithm with random parameters up to 20 times (the exact number varied based on results for different parameters) and used the best local optimum. We used diagnostic runs to confirm that this was a highly repeatable procedure that does not depend on the distribution of random starting point.

The least-squares fits

To calculate the sum of squares with only process noise, we used the standard approach of treating each observation as perfect and using one-step-ahead predictions from each observation to calculate each process noise for a given set of parameters (e.g., Polacheck et al. 1993, Dennis et al. 1995, Higgins et al. 1997, Hilborn and Mangel 1997). The sum of the squares of these process noises is then the measure of model fit to be minimized. To minimize the sum of squares, we again used the Nelder-Mead simplex algorithm. The least-squares surfaces were generally simple, so a single run of the simplex algorithm was usually sufficient; we used two runs to be safe. We could have calculated the LSPN (least squares with only process noise) maximum-likelihood parameters with standard regression formulae, but we used the numerical optimization methods to test them and treat the different estimation schemes as similarly as possible.

To calculate the sum of squares with observation error, we calculated an entire sequence of predictions from a single initial value and used the discrepancy between each prediction and observation as an observation error. Observation errors were then squared and summed to produce a sum of squares. With this method the true initial value was treated as an extra model parameter to be estimated (e.g., Polacheck et al. 1993, Hilborn and Mangel 1997). The least-squares surfaces in parameter space using this method were complicated. For the Beverton-Holt model, we solved this problem by using up to 100 random restarts. For the Ricker model, the surface was extremely complicated and we found that a simulated annealing extension of the Nelder-Mead simplex algorithm (Press et al. 1992) with 20 restarts worked well, i.e., was fairly repeatable. The complexity of these least-squares surfaces may be interpreted as reason for caution about this method.

To facilitate comparison among the methods, we express the sums of squares (SS) from the least-squares methods as likelihood values. Using standard statistical theory, the likelihood is

$$L(\text{SS}) = \frac{e^{-n/2}}{(2\pi)^{n/2}(\text{SS}/n)^{n/2}} \tag{33}$$

where n is the number of observations and SS is the sum of squared deviations between predictions and observations (e.g., Johnson and Wichern 1992). For least squares with process noise only, $n = 19$ because the first data point is not compared to a prediction. Although one might think of using the stationary distribution to estimate the probability of the first data point—based on ideas from the state-space method—we stick to the standard practice of using the first data point only as a starting point for predicting the second data point. For LSOE (least squares with observation error only), $n = 20$ because this method includes estimation of the initial population state.

Likelihood-ratio distributions

Since the numerically integrated state-space method maximizes likelihoods, its estimators should follow certain properties of maximum-likelihood estimators as the number of observations gets large (e.g., see Rice [1988], Tong [1990], Edwards [1992], and Hilborn and Mangel [1997] for general introductions). The most common asymptotic likelihood-ratio results apply for statistical models of repeated, independent sampling, which is different than time-series situations. However, similar results often extend to time series, and recently have been extended to a large class of state-space model settings by Bickel et al. (1998) for discrete models and Jensen and Petersen (1999) for continuous models.

TABLE 1. Summary of distributions of parameters estimated by three methods for fitting models to time series of population abundances: NISS (numerically integrated state-space), LSPN (least squares with only process noise), and LSOE (least squares with only observation error).

	Parameter estimation properties of each fitting method								
	Large PN, large OE			Large PN, small OE			Small PN, large OE		
Fitting method	Bias	Variance	MSE	Bias	Variance	MSE	Bias	Variance	MSE
Beverton-Holt, $\log(\lambda) = 1.5$									
NISS	0.37	0.69	0.82	0.16	0.62	0.64	0.65	0.62	1.05
LSPN	3.97	21.52	37.27	2.48	15.27	21.43	5.51	19.07	49.40
LSOE	1.65	14.65	17.36	1.09	11.25	12.45	2.18	14.50	19.26
Richer, $r = 1.5$									
NISS	−0.01	0.15	0.15	0.07	0.07	0.08	−0.25	0.20	0.27
LSPN	−0.24	0.06	0.12	−0.04	0.05	0.05	−0.44	0.05	0.24
LSOE	0.23	0.51	0.56	0.37	0.32	0.46	−0.15	0.72	0.74
Ricker $r = 2.4$									
NISS	−0.019	0.021	0.021	−0.009	0.012	0.012	0.004	0.002	0.002
LSPN	−0.214	0.044	0.090	−0.027	0.013	0.014	−0.178	0.013	0.044
LSOE	−0.026	0.101	0.102	0.171	0.106	0.135	0.005	0.004	0.004
Ricker $r = 2.6$									
NISS	−0.011	0.018	0.018	−0.010	0.008	0.009	−0.003	0.003	0.003
LSPN	−0.186	0.035	0.069	−0.023	0.009	0.009	−0.178	0.013	0.045
LSOE	0.241	0.033	0.090	0.239	0.032	0.089	0.066	0.007	0.012

Notes: Each model was fit to each of 300 time series simulated with different growth-rate parameters (λ for Beverton-Holt model and r for Ricker model) and different amounts of process noise (PN; large or small) and observation error (OE; large or small). Parameter estimation properties: Bias = the difference between the average estimate and the true value; Variance = the variance of estimates around their mean; and MSE = mean square error = variance plus the square of bias.

Strictly speaking, we have not shown whether our examples fit the setting of Jensen and Petersen (1999), so our evaluation of whether standard asymptotic results apply is mildly cavalier. However, since we consider relatively simple settings, and since the asymptotic theory is only just being developed for nonlinear, non-Gaussian state-space models, it is reasonable to compare our results to standard asymptotic results, which seem likely to apply.

Near convergence to asymptotic likelihood-ratio distributions would be extremely useful because it would facilitate hypothesis testing. In particular we expect that, for large T,

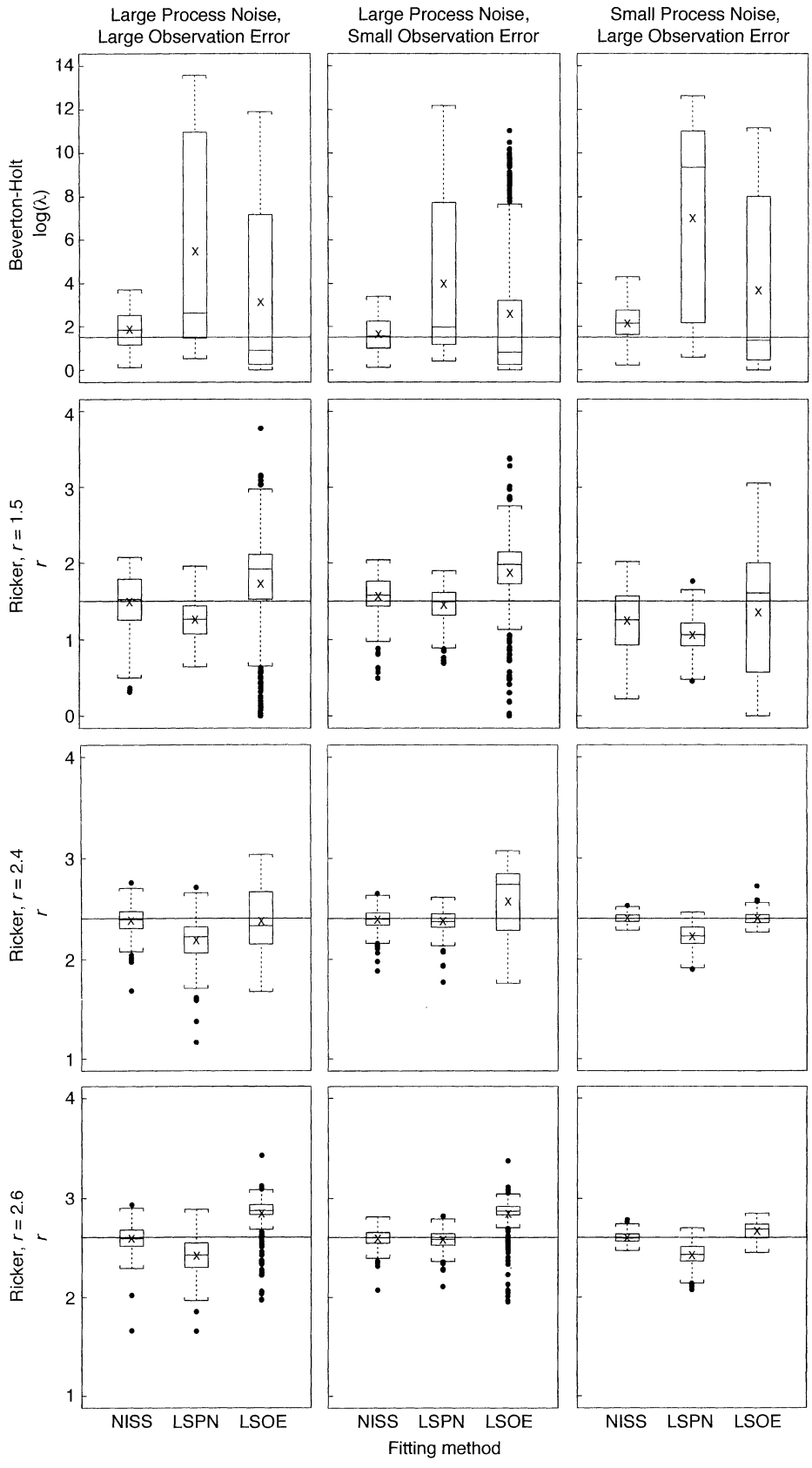
$$-2(\log[L(\boldsymbol{\Theta}_{\text{true}})] - \log[L(\boldsymbol{\Theta}_{\text{alt}}]) \sim \chi^2_p \qquad (34)$$

where $\boldsymbol{\Theta}_{\text{true}}$ are the true parameters (playing the role of a null or constrained hypothesis), $\boldsymbol{\Theta}_{\text{alt}}$ are the maximum-likelihood parameters (estimated under an "alternative" or unconstrained hypothesis), and p is the difference in the number of parameters estimated under the two hypotheses, which serves as the degrees of the freedom for the chi-squared distribution. We examined these distributions for our simulations and compared them to chi-squared distributions with four degrees of freedom using quantile–quantile plots. In this case p is 4 because four parameters are estimated under the unconstrained hypothesis (λ and γ under the Beverton-Holt, r and b under the Ricker, and σ_v^2 and σ_e^2 under either), and no parameters are estimated under the constrained hypothesis.

It may seem odd to use the true parameters to look at the likelihood-ratio distribution, when they will be unknown for data from natural systems. However, the basis of likelihood-ratio tests is to consider the distribution of likelihood ratios as if the parameters under a null hypothesis with no free parameters are the true parameters. In this study we are interested in how well the asymptotic approximate distribution really works, and for that purpose using the true parameters in the role of the null hypothesis is appropriate. For data from a natural system we could consider parameters other than the maximum-likelihood parameters as possible null hypotheses, and calculate approximate likelihood-ratio P values for them. This is the basis of constructing confidence intervals: include all parameters that, as a null hypothesis, would give $P > 0.05$. Thus, the idea of null and alternative hypotheses for statistical testing is implicit in our focus on maximum-likelihood parameter estimation and asymptotic likelihood-ratio distributions.

Information criteria

Finally, we evaluated the ability of the Akaike information criterion (see Burnham and Anderson [1998] for a general introduction) and other related information criteria to distinguish between the Ricker and Beverton-Holt models using each of the fitting methods.



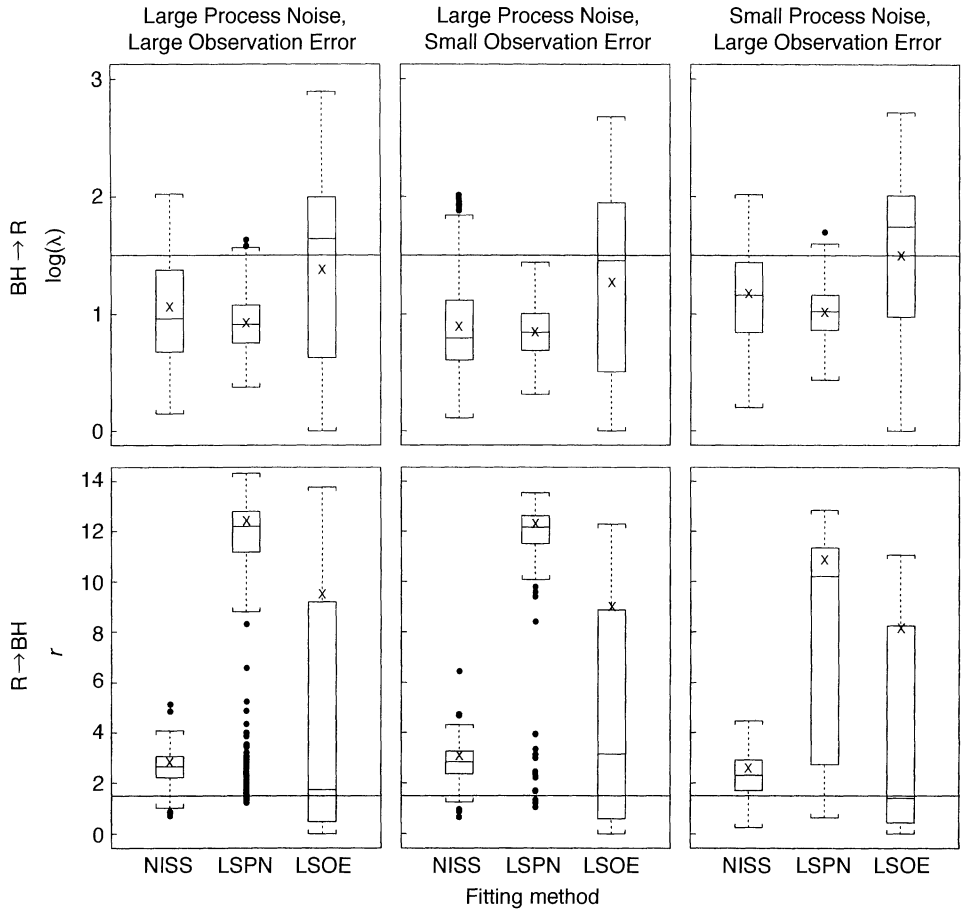


FIG. 3. Maximum-likelihood estimates of growth rate when data are fitted to the wrong model. "BH → R" (top row): Data were generated with a Beverton-Holt model with $\lambda = 4.48$ and fit to a Ricker model. "R → BH" (bottom row): Data were generated with a Ricker model with $r = 1.5$ and fit to a Beverton-Holt model. Horizontal lines show the true growth-rate parameters, $\log(\lambda) = r = 1.5$.

Since the number of parameters for each model is the same, the information-criterion methods, which differ only in the adjustment they make for number of parameters and data points, amount to picking the model with the highest maximum likelihood. Information criteria are becoming popular tools for model selection in ecology (Anderson and Burnham 1992, Burnham et al. 1994, Hilborn and Mangel 1997, Burnham and Anderson 1998, Dennis et al. 1998, Zeng et al. 1998).

RESULTS

Estimator properties

Our evaluation of parameter estimation properties of each fitting method focuses on estimates of the growth-rate parameters, λ for the Beverton-Holt and r for the Ricker. In all cases there was some correlation—often strong—between estimates of growth-rate parameters and density-dependence parameters (λ and γ for the

←

FIG. 2. Maximum-likelihood estimates of growth-rate parameters, λ for the Beverton-Holt model and r for the Ricker model. For the Beverton-Holt model (top row), the true value of λ is 4.48, shown by the horizontal line ($\log(4.48) = 1.5$) in each panel. For the Ricker model (next three rows), the true value of r is 1.5, $r = 2.4$ (two-cycle dynamics), or $r = 2.6$ (four-cycle dynamics). For each fitting method, a box-and-whisker plot summarizes the distribution of estimates. The box extends from the 25th to the 75th percentiles, the median is shown by a line in the middle of the box, the mean is marked with an "×," the dashed "whiskers" extend to the largest estimate less than 1.5 interquartile distances (i.e., length of the box) above the 75th and below the 25th percentiles, and all other estimates are individually plotted. The left, middle, and right columns of figures are grouped by the process noise (PN) and observation error (OE) variances used to generate data, as labeled above each column. Note the log transformation of λ estimates for the Beverton-Holt model, indicating that these estimates ranged over many orders of magnitude for the LSPN and LSOE methods. (See Table 1 for definitions of fitting-methods acronyms.)

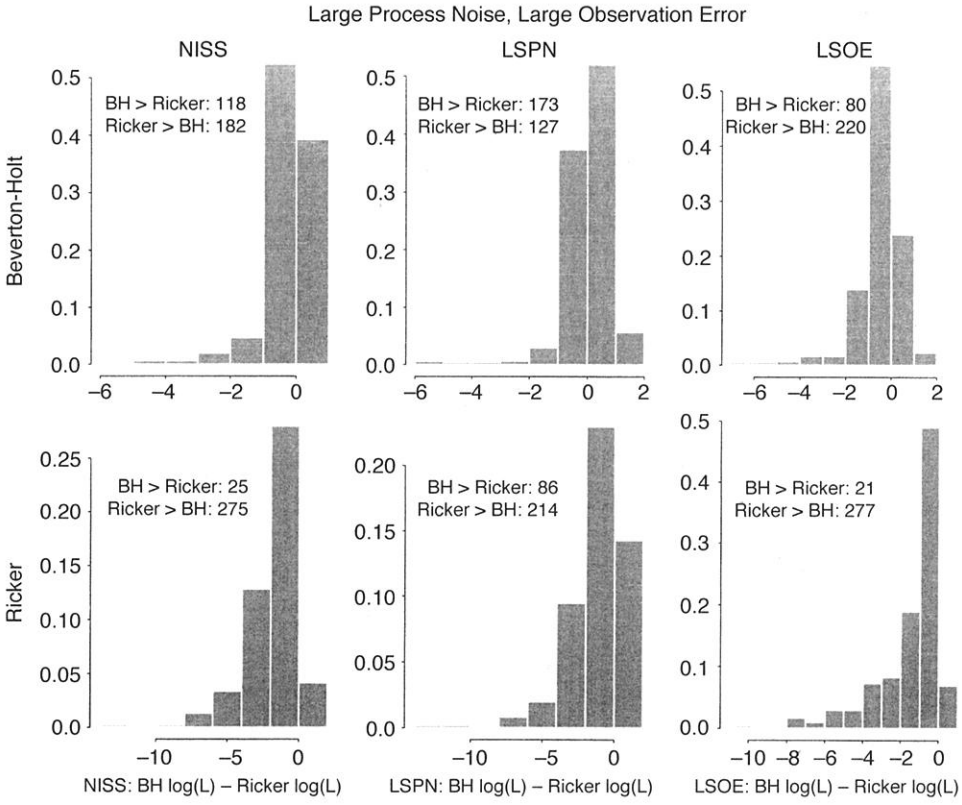


FIG. 4. Frequency density of log-likelihood [log (L)] differences for models with large process noise and large observation error. Each histogram gives the density of the difference between maximum log likelihood of the Beverton-Holt (BH) model and maximum log likelihood of the Ricker model. The proportion of differences that fall within the range of x-axis values for each bar is equal to the area of the bar. When data were generated with a BH model (top row), differences >0 correspond to correct model identification (larger maximum likelihood for the true model). When data were generated with the Ricker model (bottom row), differences <0 correspond to correct model identification. The left, middle, and right vertical pairs of histograms are grouped by fitting method, as indicated by the label above each pair (see Table 1 for acronym explanations). The notation "BH > Ricker: k" indicates that the BH model was chosen over the Ricker model k out of 300 times.

Beverton-Holt, r and b for the Ricker), but little or no correlation between estimates of growth-rate parameters and equilibrium population size ($(\lambda - 1)/\gamma$ for the Beverton-Holt, r/b for the Ricker). The distinctions between the fitting methods were clearer for the growth-rate parameters than for the equilibrium population sizes, hence our focus on the former. When comparing results from the Beverton-Holt and Ricker models, note that the order of magnitude of $\log(\lambda)$ is comparable to that of r.

We report results in terms of bias, variance, and mean squared error of estimates (Table 1), as well as the distributions of estimates (Fig. 2). Bias is defined as the difference between the average estimate and the true value. Variance is the variance of estimates around their mean. Mean squared error is variance plus the square of bias, so it combines these two components of estimation inaccuracy.

For the Beverton-Holt model, the NISS (numerically integrated state-space) method was less biased and had lower variance than both the least-squares methods for all three noise and error combinations (Fig. 2). Estimating growth rate, λ, is difficult for a Beverton-Holt model because M_t, the population size at time t, is nearly constant for large M_{t-1}. All of the fitting methods were biased towards putting a flat line between M_t and M_{t-1}, thus overestimating the growth-rate part of the curve, but only the NISS method obtained an average estimate close to the true value and a distribution of estimates within an order of magnitude of the true value. The two least-squares methods were not useful for the Beverton-Holt model, with estimates for λ ranging over many orders of magnitudes.

For the Ricker model with the dynamically simplest parameters, $r = 1.5$, each of the three methods was less biased than the others for one of the noise assumptions (Fig. 2). With small process noise, LSOE (least squares with only observation error) was least biased. Conversely, with small observation error, LSPN (least squares with only process noise) was least biased. With both noises large, NISS was least biased, and indeed nearly unbiased. In all cases, the LSPN estimator had the smallest variance. It should be remembered that the noise and error variances will not be

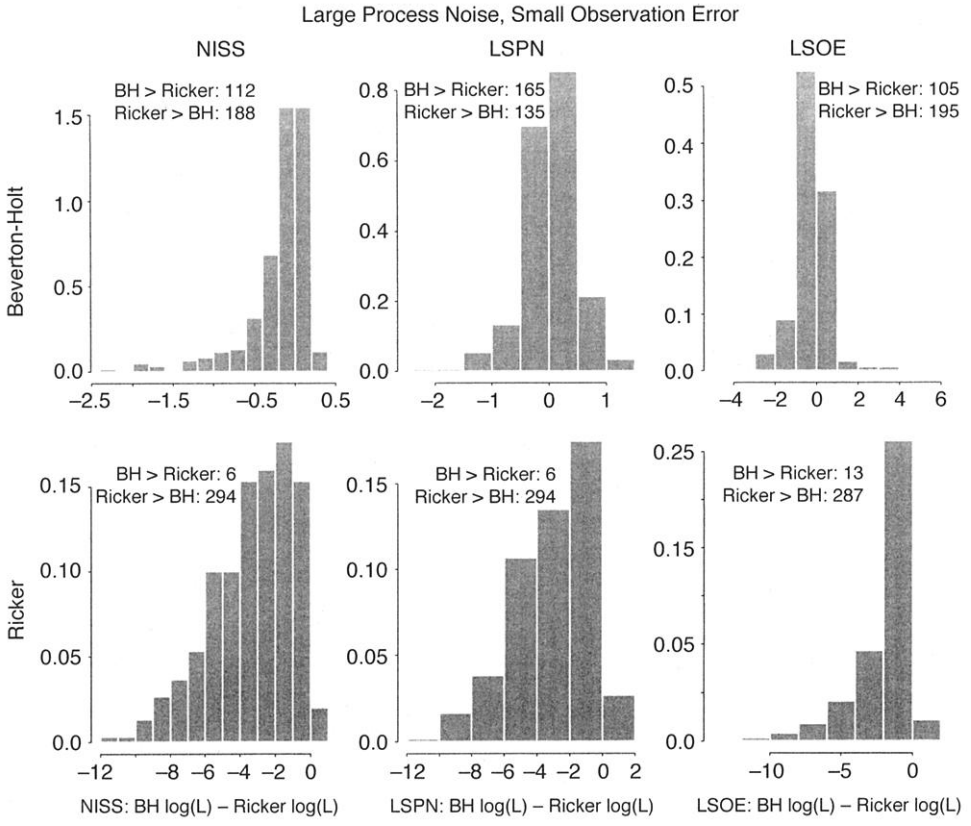


FIG. 5. Frequency density of information criterion differences for models with large process noise and small observation error. The format is as in Fig. 4.

known for real data, and the biases in the NISS estimator are smaller than those for the LSOE and LSPN estimators when the noise and error conditions are least suited to them. For example, with large process noise and small observation error, the assumptions of the LSOE method are least close to correct, and the corresponding bias in the LSOE estimates is large. The NISS method is more robust to lack of prior knowledge of noise and error variances. In theory if one knew the observation error variance, one could fit with LSPN (because it has the lowest estimator variance) and bootstrap a bias correction, but if observation noise is unknown this could be difficult because the bias will depend on the error variance and the bootstrap procedure would add variance to the overall estimation scheme.

With respect to estimating the correct dynamical regime, the NISS estimator virtually never estimates $r > 2.0$, the bifurcation boundary for this parameter (Figs. 1 and 2; a few replicates produced $2.0 < r < 2.05$). LSPN is also always <2.0, but that is related to its negative bias. LSOE frequently chooses parameters that would lead to dynamically wrong interpretations, such as spuriously estimating two-cycles.

For the Ricker model with two-cycle parameters, $r = 2.4$, and four-cycle parameters, $r = 2.6$, NISS was always the best parameter estimator (Fig. 2). For each of the noise choices, it produced virtually unbiased estimates with the lowest estimator variance of the three methods. The other methods had small bias and variance only under the noise conditions that approximated their assumptions, and even then the NISS method was superior. When the true noises did not fit their assumptions, the LSPN and LSOE methods were strongly biased. This suggests that in dynamically complicated regimes even small noises are sufficient to make NISS a better method, and the more complicated the regime (four-cycle vs. two-cycle), the stronger this conclusion. The NISS method also does a better job of estimating parameters with correct dynamical properties (compare to Fig. 1). For the two-cycle parameters, it strays occasionally into four-cycle parameter space. LSPN doesn't do this, but again it is biased low. LSOE strays far from dynamically similar parameters except when the true conditions are most appropriate for it (i.e., small process noise). A similar comparison of the three methods holds for the four-cycle parameters.

Model identification

Comparison of maximum likelihoods of data produced from either a Beverton-Holt or Ricker model and fit with both models shows a systematic bias toward larger maximum likelihoods with the Ricker model

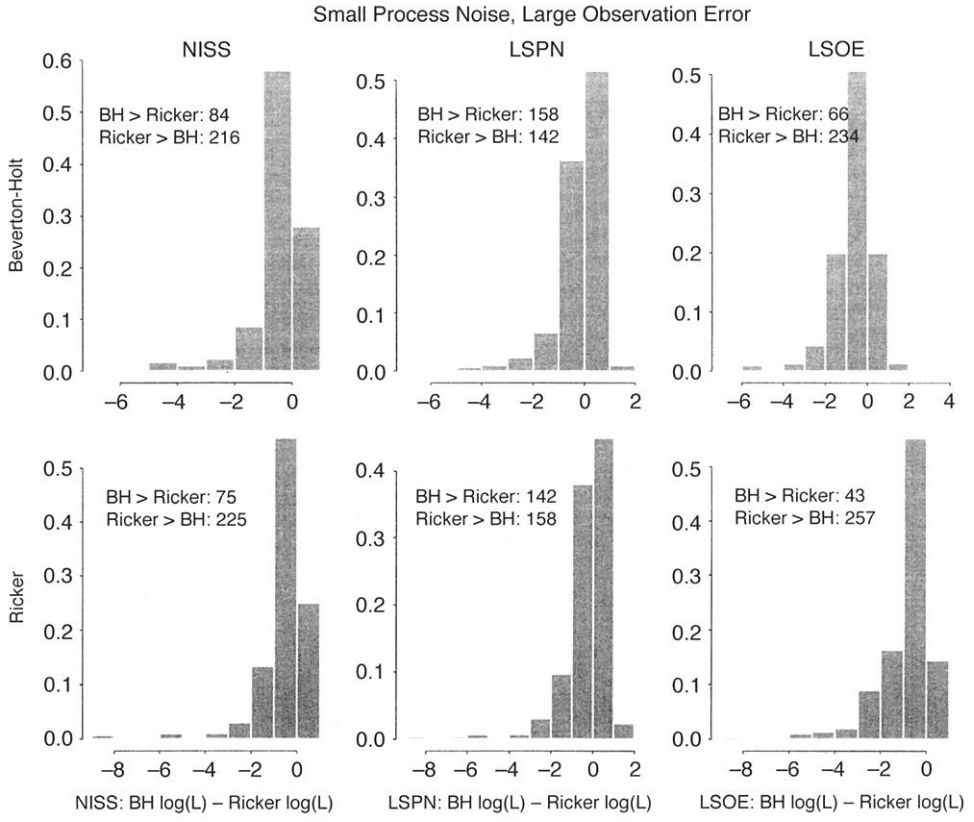


FIG. 6. Frequency density of information criterion differences for models with small process noise and large observation error. The format is as in Fig. 4.

(Figs. 3–6). Under all three noise assumptions, NISS (incorrectly) picked Ricker models for about two-thirds of the runs from a Beverton-Holt model and (correctly) picked Ricker models for most runs from a Ricker model. LSOE was similarly biased toward Ricker models. LSPN was less biased and was the only method with better than 50% success under all noise and error assumptions and both source models. It remains unclear whether LSPN is generally superior at model identification or whether its biases are just fortuitously useful for the Beverton-Holt vs. Ricker problem.

For all methods, many of the cases of misidentification involved small log-likelihood differences. The likelihood ratio itself can be interpreted as the ratio of the probability that under one model the data would be produced to the probability that under the other model the data would be produced (Edwards 1992, Royall 1997). Royall (1997) explores the interpretation of likelihood ratios and suggests that ratios of the magnitudes common in Figs. 4, 5, and 6 (expressed as differences of logs) should not be considered as strong evidence one way or the other between two models.

Asymptotic likelihood-ratio distributions

If the likelihoods calculated with NISS converge quickly (i.e., with few data points) to the asymptotic chi-squared likelihood-ratio distribution, hypothesis testing would be greatly simplified. Quantile–quantile plots show varying levels of convergence (Fig. 7). For the Beverton-Holt model, convergence is poor. For the Ricker model with $r = 1.5$, convergence is good for all noise assumptions. For the Ricker model with $r = 2.4$ and 2.6, large observation error seems to hinder convergence, but convergence is still better than under the Beverton-Holt model. For the Ricker $r = 1.5$ case, the good convergence can be used to construct confidence regions.

DISCUSSION

Our results indicate that the numerically integrated state-space (NISS) method can be an important tool for statistically relating population models to data. This method incorporates both process noise and observation error, can be applied to any model structure, and allows simple treatment of missing data points (one just projects the state distribution again without updating by an observation) and unobserved state variables. For the Beverton-Holt model we would virtually always recommend NISS over least-squares methods. For the Ricker model, we would recommend it over least-squares methods unless one has a priori knowledge that either process noise or observation error is

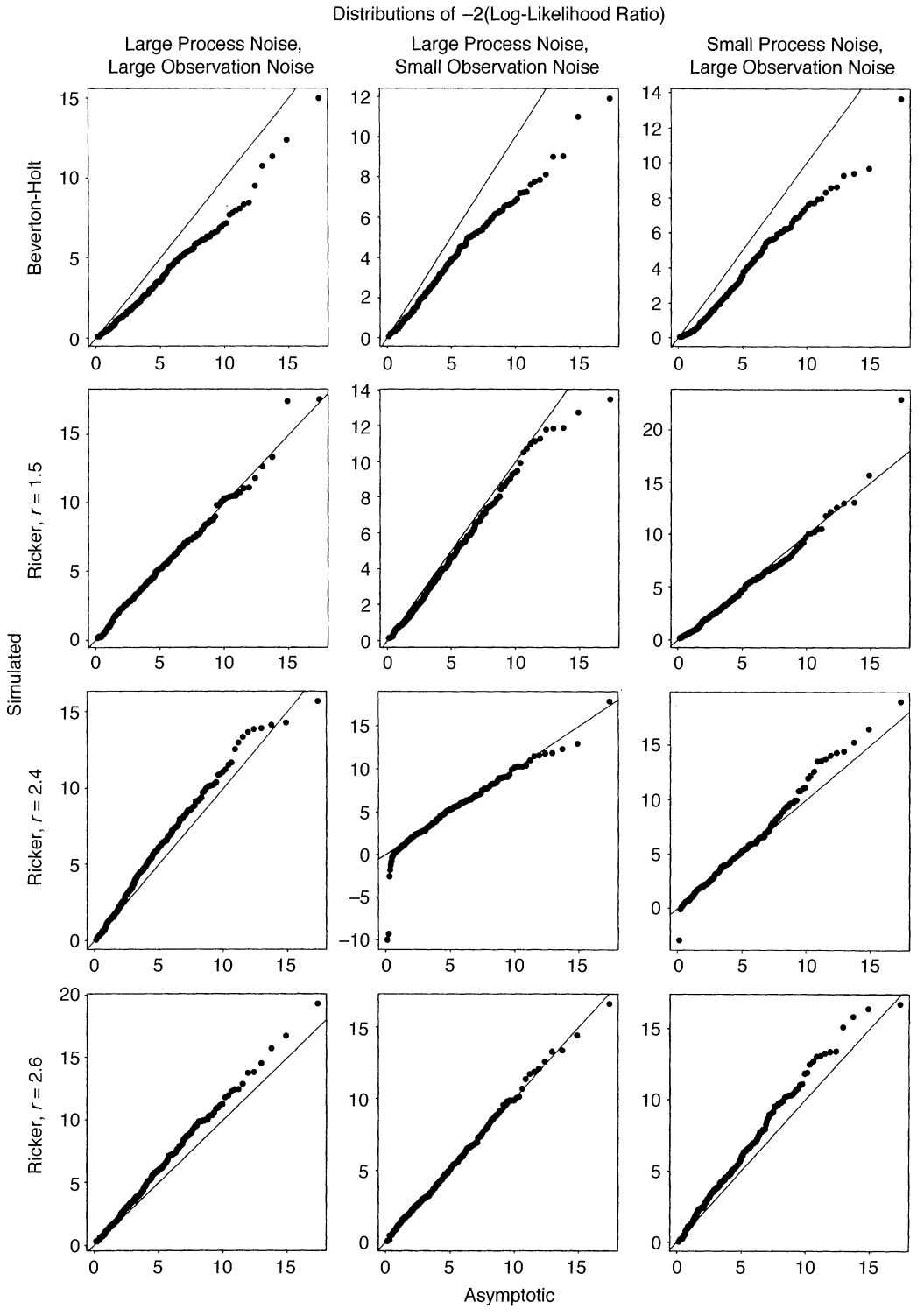


FIG. 7. Quantile–quantile plots comparing distributions of simulated likelihood ratios (likelihood with true parameters/maximum likelihood) to theoretical asymptotic likelihood-ratio distributions. Likelihood-ratio distributions are transformed to −2(log-likelihood ratio) and compared to a chi-squared distribution with 4 df (see Eq. 34). Points along the identity line indicate agreement with the theoretical asymptotic distribution. The top row shows Beverton-Holt model results; the next three rows show Ricker model results with $r = 1.5$, $r = 2.4$, and $r = 2.6$. Columns correspond to different combinations of process noise and observation error variances, as labeled.

small and the dynamics are simple. Others have found substantial biases in fitting biomass dynamics and catch-age models when process noise and observation error are not incorporated into the fitting process (Hilborn 1979, Uhler 1980, Ludwig and Walters 1981, 1989, Ludwig et al. 1988, Polacheck et al. 1993, Schnute and Richards 1995, Kimura et al. 1996).

Convergence to an asymptotic likelihood-ratio distribution was not generally sufficient with 20 data points to provide precise significance levels for hypothesis testing. However, for the Ricker model with simple dynamics ($r = 1.5$), convergence was good. For the other cases, the asymptotic distributions could be used to construct only approximate confidence intervals. An alternative approach for approximate confidence intervals would be to use a parametric bootstrapping procedure (e.g., Efron and Tibshirani 1993, Dennis and Taper 1994, Dennis et al. 1995). In all cases our results suggest that evaluation of estimation properties of different methods with different models can provide important baseline information before applying methods to real data.

Some of our results may be related to the amount of informative coverage of different population sizes provided by a given time series for each model. For a Ricker model, both large and small population sizes provide information about the shape of the function, while for the Beverton-Holt model, data from small population sizes are important for estimating the growth-rate part of the curve, while data from large population sizes are relatively uninformative about this. This raises the related issue that in a particular application the focus of interest might be a function of parameters such as maximum sustainable yield in fisheries, and it would be important to consider the distribution of estimates of such quantities explicitly. Similar issues were discussed by Hilborn (1979), Schaffer et al. (1986), and Ludwig and Walters (1989), among others.

Two important conclusions emerged from studying the model identification properties of the three fitting methods. First, it is possible to have systematic biases toward better fits with one model over another, in this case the Ricker over the Beverton-Holt. In this sense the Ricker is a more flexible model than the Beverton-Holt. It would be important to understand such biases before attempting model identification with data from natural systems. Second, using an information criterion alone—in which the better model is chosen regardless of how close the models are—can be misleading. It would be better to interpret the likelihood ratio and, if possible, to bootstrap a distribution of likelihood ratios (or differences in information criteria) and estimate a significance level of the actual ratio.

The problem of bootstrapping with the NISS method is not trivial because it is computationally intense, and with more state variables, i.e., higher dimensional distributions, this could be a limitation. However, in addition to faster computers, two approximations may be useful for higher-dimension problems. First, in many cases the more commonplace "extended" Kalman filter may be satisfactory. We did not use it here because when the state distribution spans zero or near-zero derivatives of the Ricker model or Beverton-Holt models, the projected distribution is highly non-normal. However, in simpler cases it could prove useful, and it is easier to find or program. Second, Markov chain Monte Carlo numerical integration methods can give a stochastic approximation to the probability-distribution calculations of state-space models (Carlin et al. 1992, Gilks et al. 1996). Although this is an approximation compared to our implementation, it works, can be quite accurate, and is easier to implement (Gilks et al. 1996). Meyer and Millar (1999) and Bjornstad et al. (1999*a*) used this approach to analyze fisheries data in a Bayesian framework. Although Markov chain Monte Carlo methods are often discussed in the context of Bayesian analyses, they can also be used for frequentist analyses (Geyer and Thompson 1992, Geyer 1996).

We have evaluated a model-fitting method little used in ecology and found that it has advantages over other methods. Relating population models to data statistically is important for testing hypotheses with both observational and experimental time series. Development and testing of model-fitting methods underlies analysis of real data. Although time-series analysis has most often been applied to observational data, many experimental time series from systems with short generation times can also be analyzed by fitting population models. The approach taken here offers one step toward providing better tools for such analyses and improving the connections between models and data and ecology.

Acknowledgments

We thank Brian Dennis for suggesting we look into state-space models and Kalman filters and R. Shumway for pointing out the Kitigawa (1987) paper. We thank Brian Inouye, Don Ludwig and Drew Tyre for helpful comments, and Allan Stewart-Oaten for extraordinarily helpful and detailed suggestions. P. de Valpine was supported by an NSF Graduate Research Fellowship.

Literature Cited

Anderson, D. R., and K. P. Burnham. 1992. Demographic analysis of Northern Spotted Owl populations. Appendix C in Recovery Plan for the Northern Spotted Owl. U.S. Fish and Wildlife Service, Portland, Oregon, USA.

Bickel, P. J., Y. Ritov, and T. Rydén. 1998. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. Annals of Statistics **26**:1614–1635.

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

Bjornstad, O. N., N. C. Stenseth, and T. Saitoh. 1999*b*. Synchrony and scaling in dynamics of voles and mice in northern Japan. Ecology **80**:622–637.

Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Burnham, K. P., D. R. Anderson, and G. C. White. 1994.

Estimation of vital rates of the Northern Spotted Owl. Appendix J in Final Supplemental Environmental Impact Statement on Management for Old-Growth Forest Related Species Within the Range of the Northern Spotted Owl. U.S. Department of Agriculture and U.S. Department of the Interior, Portland, Oregon, USA.

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.

Carpenter, S. R., K. L. Cottingham, and C. A. Stow. 1994. Fitting predator–prey models to time series with observation errors. Ecology **75**:1254–1264.

Collie, J. S., and M. P. Sissenwine. 1983. Estimating population size from relative abundance data measured with error. Canadian Journal of Fisheries and Aquatic Sciences **40**:1871–1879.

Costantino, R. F., R. A. Desharnais, J. M. Cushing, and B. Dennis. 1997. Chaotic dynamics in an insect population. Science **275**:389–391.

Dennis, B. 1996. Discussion: Should ecologists become Bayesians? Ecological Applications **6**:1095–1103.

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

Dennis, B., W. P. Kemp, and M. L. Taper. 1998. Joint density dependence. Ecology **79**:426–441.

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

Edwards, A. W. F. 1992. Likelihood. Expanded edition. Johns Hopkins University Press, Baltimore, Maryland, USA.

Efron, B., and R. J. Tibshirani. 1993. An introduction to the bootstrap. Chapman and Hall, New York, New York, USA.

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 **151**:425–440.

Ellner, S., and P. Turchin. 1995. Chaos in a noisy world: new methods and evidence from time-series analysis. American Naturalist **145**:343–375.

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.

Frigo, M., and S. G. Johnson. 1997. The fastest fourier transform in the west. Technical Report MIT-LCS-TR-728. Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

Geyer, C. J. 1996. Estimation and optimization of functions. Pages 241–258 *in* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. Markov chain Monte Carlo in practice. Chapman and Hall, New York, New York, USA.

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

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.

Grenfell, B. T., K. Wilson, B. F. Finkenstaedt, T. N. Coulson, S. Murray, S. D. Albon, J. M. Pemberton, T. H. Clutton-Brock, and M. J. Crawley. 1998. Noise and determinism in synchronized sheep dynamics. Nature **394**:674–677.

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

Hanski, I., I. Woiwood, and J. Perry. 1993. Density dependence, population persistence, and largely futile arguments. Oecologia **95**:595–598.

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

Harvey, A. C. 1993. Time series models. Second edition. The MIT Press, Cambridge, Massachusetts, USA.

Hastings, A., C. L. Hom, S. Ellner, P. Turchin, and H. C. J. Godfray. 1993. Chaos in ecology: Is mother nature a strange attractor? Annual Review of Ecology and Systematics **24**:1–33.

Higgins, K., A. Hastings, J. N. Sarvela, and L. W. Botsford. 1997. Stochastic dynamics and deterministic skeletons: population behavior of Dungeness crab. Science **276**:1431–1435.

Hilborn, R. 1979. Comparison of fisheries control systems that utilize catch and effort data. Canadian Journal of Fisheries and Aquatic Sciences **36**:1477–1489.

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.

Jensen, J. L., and N. V. Petersen. 1999. Asymptotic normality of the maximum likelihood estimator in state space models. Annals of Statistics **27**:514–535.

Johnson, R. A., and D. W. Wichern. 1992. Applied multivariate statistical analysis. Third edition. Prentice Hall, New Jersey, USA.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME Journal of Basic Engineering, **D 82**:35–45.

Kemp, W. P., and B. Dennis. 1993. Density dependence in rangeland grasshoppers (Orthoptera: Acrididae). Oecologia **96**:1–8.

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.

Kohn, R., and C. F. Ansley. 1987. Comment. Journal of the American Statistical Association **82**:1041–1044.

Ludwig, D., and C. J. Walters. 1981. Measurement errors and uncertainty in parameter estimates for stock and recruitment. Canadian Journal of Fisheries and Aquatic Sciences **38**:711–720.

Ludwig, D., and C. J. Walters. 1989. A robust method for parameter estimation from catch and effort data. Canadian Journal of Fisheries and Aquatic Sciences **46**:137–144.

Ludwig, D., C. J. Walters, and J. Cooke. 1988. Comparison of two models and two estimation methods for catch and effort data. Natural Resource Modeling **2**:457–498.

Meinhold, R. J., and N. D. Singpurwalla. 1983. Understanding the Kalman filter. American Statistician **37**:123–127.

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

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.

Newman, K. B. 1998. State-space modeling of animal movement and mortality with applications to salmon. Biometrics **54**:1290–1314.

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. M. Eggers, R. J. Marasco, C. Pautzke, and T. J. Quinn II, editors. Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations. Alaska Sea Grant College Program Report Number 93-02. University of Alaska, Fairbanks, Alaska, USA.

Polacheck, T., R. Hilborn, and A. E. Punt. 1993. Fitting surplus production models: comparing methods and measuring uncertainty. Canadian Journal of Fisheries and Aquatic Sciences **50**:2597–2607.

Pollard, E., K. H. Lakhani, and P. Rothery. 1987. The detection of density-dependence from a series of annual censuses. Ecology **68**:2046–2055.

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.

Reed, W. J., and C. M. Simmons. 1996. Analyzing catch-effort data by means of the Kalman filter. Canadian Journal of Fisheries and Aquatic Sciences **53**:2157–2166.

Rice, J. A. 1988. Mathematical statistics and data analysis. Wadsworth & Brooks/Cole, Pacific Grove, California, USA.

Royall, R. M. 1997. Statistical evidence: a likelihood paradigm. Chapman & Hall, New York, New York, USA.

Saitoh, T., O. N. Bjornstad, and N. C. Stenseth. 1999. Density dependence in voles and mice: a comparative study. Ecology **80**:638–650.

Schaffer, W. M., S. Ellner, and M. Kot. 1986. Effects of noise on some dynamical models in ecology. Journal of Mathematical Biology **24**:479–523.

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

Schnute, J. T., and L. J. Richards. 1995. The influence of error on population estimates from catch-age models. Canadian Journal of Fisheries and Aquatic Sciences **52**:2063–2077.

Shumway, R. H., and D. S. Stoffer. 2000. Time series analysis and its applications. Springer-Verlag, New York, New York, USA.

Stenseth, N. C., W. Falck, K.-S. Chan, O. N. Bjornstad, M. O'Donoghue, H. Tong, R. Boonstra, S. Boutin, C. J. Krebs, and N. G. Yoccoz. 1998. From patterns to processes: phase and density dependencies in the Canadian lynx cycle. Proceedings of the National Academy of Sciences, USA **95**: 15430–15435.

Strang, G. 1986. Introduction to applied mathematics. Wellesley-Cambridge Press, Wellesley, Massachusetts, USA.

Stuart, A., and J. K. Ord. 1991. Kendall's advanced theory of statistics. Volume 2. Classical inference and relationships. Fifth edition. Griffin, London, UK.

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

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

Turchin, P. 1990. Rarity of density-dependence or population regulation with time lags? Nature **344**:660–663.

Turchin, P. 1995. Population regulation: old arguments and a new synthesis. Pages 19–40 *in* N. Cappuccino and P. Price, editors. Population dynamics: new approaches and synthesis. Academic Press, San Diego, California, USA.

Turchin, P., and A. D. Taylor. 1992. Complex dynamics in ecological time series. Ecology **73**:289–305.

Uhler, R. S. 1980. Least squares regression estimates of the Schaefer production model: some Monte Carlo simulation results. Canadian Journal of Fisheries and Aquatic Sciences **37**:1284–1294.

Walters, C. J., and D. Ludwig. 1981. Effects of measurement errors on the assessment of stock–recruit relationships. Canadian Journal of Fisheries and Aquatic Sciences **38**:704–710.

Wolda, H., B. Dennis, and M. L. Taper. 1994. Density dependence tests, and largely futile comments: answers to Holyoak and Lawton (1993) and Hanski, Woiwood, and Perry (1993). Oecologia **98**:229–234.

Zeng, Z., R. M. Nowierski, M. L. Taper, B. Dennis, and W. P. Kemp. 1998. Complex population dynamics in the real world: modeling the influence of time-varying parameters and time lags. Ecology **79**:2193–2209.