ESTIMATING DENSITY DEPENDENCE, PROCESS NOISE, AND OBSERVATION ERROR

Brian Dennis,[1,5] José Miguel Ponciano,[2] Subhash R. Lele,[3] Mark L. Taper,[4] and David F. Staples[4]

[1]*Department of Fish and Wildlife Resources and Department of Statistics, University of Idaho, Moscow, Idaho 83844 USA*
[2]*Initiative for Bioinformatics and Evolutionary Studies (IBEST), Department of Mathematics, University of Idaho, Moscow, Idaho 83844 USA*
[3]*Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1 Canada*
[4]*Department of Ecology, Montana State University, Bozeman, Montana 59717 USA*

Abstract. We describe a discrete-time, stochastic population model with density dependence, environmental-type process noise, and lognormal observation or sampling error. The model, a stochastic version of the Gompertz model, can be transformed into a linear Gaussian state-space model (Kalman filter) for convenient fitting to time series data. The model has a multivariate normal likelihood function and is simple enough for a variety of uses ranging from theoretical study of parameter estimation issues to routine data analyses in population monitoring. A special case of the model is the discrete-time, stochastic exponential growth model (density independence) with environmental-type process error and lognormal observation error.

We describe two methods for estimating parameters in the Gompertz state-space model, and we compare the statistical qualities of the methods with computer simulations. The methods are maximum likelihood based on observations and restricted maximum likelihood based on first differences. Both offer adequate statistical properties. Because the likelihood function is identical to a repeated-measures analysis of variance model with a random time effect, parameter estimates can be calculated using PROC MIXED of SAS.

We use the model to analyze a data set from the Breeding Bird Survey. The fitted model suggests that over 70% of the noise in the population's growth rate is due to observation error. The model describes the autocovariance properties of the data especially well.

While observation error and process noise variance parameters can both be estimated from one time series, multimodal likelihood functions can and do occur. For data arising from the model, the statistically consistent parameter estimates do not necessarily correspond to the global maximum in the likelihood function. Maximization, simulation, and bootstrapping programs must accommodate the phenomenon of multimodal likelihood functions to produce statistically valid results.

Key words: Breeding Bird Survey; environmental noise; Gompertz growth model; Kalman filter; measurement error; multimodal likelihood; observation error; process noise; sampling error; state-space model; stationary distribution; stochastic population model.

INTRODUCTION

The use of mathematical population models as the basis for analysis of time series population abundances has been a productive and useful area of research in the past decade (Dennis et al. 1991, Dennis and Taper 1994, Berryman 1999, Turchin 2003). The basic approach converts the deterministic models of ecology textbooks into statistical analysis tools by incorporating into the models terms representing the ubiquitous stochastic forces affecting populations (Dennis et al. 1995, Hilborn and Mangel 1997, Ives et al. 2003, Turchin 2003). Ecologists now widely acknowledge that multiple stochastic forces, including demographic and environmental noise, are inherent in population fluctuations,

even in the laboratory. This "process noise" exists independently of the error inherent in the observation or sampling methods by which population abundances are estimated, which we term "observation error." While models with process noise alone or observation error alone are relatively easy to apply, analyzing populations affected by both process noise and observation error has remained a difficult computational and statistical challenge (Shenk et al. 1998).

The problem of estimating both observation error and process noise in population ecology is starting to yield to progress in statistical methods. "State-space" models for analyzing time series of population abundances offer possibilities of jointly estimating the amount of observation error along with the amount of process noise (De Valpine 2002, De Valpine and Hastings 2002, Clark and Bjørnstad 2004). A state-space model has two components: a stochastic model for an unobserved variable (or group of variables), and a stochastic model for an

observed variable (or group of variables), the latter model usually formulated as a function of the unobserved variable and measurement or sampling error. Two essential difficulties have hindered the wider use of these approaches.

First, the combination of process noise and observation error produces significant computational difficulties. With process noise or observation error alone, the statistical dependence structure of the observations can be described with relatively simple mathematical formulas that permit easy computing. However, the combination of process noise and observation error alters the dependence structure, so that each observation is statistically related to all the previous observations in the time series in ways that defy representation with simple formulas. Extensive numerical simulation and numerical integration is required for conducting statistical inferences; the task of just fitting a model can become a major research project.

Second, the information contained in the time series data for estimating the magnitudes of both the process noise and the observation error simultaneously can be elusive. Parameters for both types of variability can have estimates that are highly related to each other, as manifested through ridge-shaped likelihood functions and large estimation intervals.

Even so, state-space models have been proposed and applied in ecological work. De Valpine and Hastings (2002) developed density-dependent population models in which the underlying population process had either a stochastic Moran-Ricker or a stochastic Beverton-Holt form, and the observed time series was the underlying process plus normal observation error. They used Kitagawa's (1987) discretization algorithm for calculating the integrals in the likelihood function, and they tested maximum likelihood (ML) estimation with simulations. Numerous other state-space population models have been implemented, especially in the fisheries literature. Many of the applications have used linear difference equations with normal process noise for the population, sometimes with multiple state variables to accommodate age structure (for example, Sullivan 1992, Reed and Simons 1996). The observation error in these density-independent state-space models typically is also taken to be normally distributed. Conveniently, the likelihood function for fitting the linear models with normal error to data can be computed readily with a set of recursion equations known as the Kalman filter (Harvey 1993, Schnute 1994). For nonlinear density-dependent cases, Schnute (1994) noted the possibility of using the "extended Kalman filter" involving linear approximation methods, but Kitagawa's (1987) algorithm was partly motivated by the knowledge in statistics that the extended Kalman filter approach frequently does not work well. A useful review of the nonlinear approaches was compiled by De Valpine (2002). Staudenmayer and Buonaccorsi (2005) tackled the inherently nonlinear problem of heteroskedastic observation error in otherwise linear time state-space models; their paper provides a current entry to the statistics literature on state-space models.

The numerical challenges of nonlinear state-space population models have inspired Bayesian approaches. Millar and Meyer (2000*a*) use the Bayesian framework to implement state-space models for fisheries stocks based on the nonlinear Moran-Ricker or Beverton-Holt forms. In other papers, they develop linear age-structured models and explain use of the BUGS (Bayesian inference using Gibbs sampling) software (Meyer and Millar 1999*a, b*, Millar and Meyer 2000*b*). More recently, Buckland et al. (2004) constructed detailed state-space models with multiple state variables, and Newman et al. (2006) implemented such a model for fisheries data. Clark and Bjørnstad (2004) surveyed different types of population variability and observational errors that can be accommodated in the Bayesian state-space framework.

A special case of the linear state-space models is the stochastic model studied by Holmes (2001; see also Holmes and Fagan 2002, Holmes 2004). The model consists of discrete-time exponential growth with process noise and observation error and represents an extension of the diffusion process model used in conservation biology (Dennis et al. 1991). Holmes proposed a regression method for estimating model parameters; Lindley (2003) observed that ML estimates for the model could be computed with the Kalman filter. Staples et al. (2004) subsequently demonstrated that improved estimation for the model can be attained with restricted maximum likelihood (REML) estimates based on second differences of the time series data; the model for the first differences, it turns out, can be written as a standard analysis of variance model with mixed effects.

A promising route for combining the computational convenience of linear state-space models with density dependence is to formulate a linear autoregressive model on the logarithmic scale (Myers and Cadigan 1993*a, b*, Fromentin et al. 2001, Ives et al. 2003). When process noise is added on the logarithmic scale, a model of environmental variability is produced (Dennis et al. 1995). Myers and Cadigan (1993*a, b*) and Fromentin et al. (2001) used multivariate autoregressive models to study survival in fisheries populations, while Solow and Sherman (1997) and Ives et al. (2003) used such models as the basis for estimating interaction strengths and stability measures in time series data of interacting species. Significantly, a univariate version of these models is a Gompertz-type model of density-dependent population growth.

Our goal in this paper is to offer a statistical approach for jointly estimating density dependence, process error, and observation error that (1) is relatively easy for ecologists to use and (2) is applicable in many population systems. To that end, we study a univariate, discrete-time, stochastic model for analyzing time series abundances of a single population. The model contains

density dependence in the form of a Gompertz-type population growth function. The model represents environmental stochasticity as lognormal process noise. As well, the model contains lognormal observation error. While other state-space models can be formulated to contain more realism, the representations of density dependence, process noise, and observation error in the model serve as a minimal set for adequately describing time series abundances in a variety of ecological settings. Furthermore, a practical threshold of usefulness of a statistical method tends to be the availability of user-friendly software for conducting the calculations. Because the likelihood function for the model we present is identical to that of a repeated measures analysis of variance with a random time effect, parameter estimates can be calculated with software packages for mixed-effect models, such as PROC MIXED in SAS. We demonstrate with simulations that the magnitudes of process variability and sampling variability, along with the strength of density dependence, can be sorted out when fitting the model to a single time series of estimated population abundances.

We begin by describing the model, its variants and special cases, and its main statistical properties. The resulting likelihood functions for various types of time series observations, including differenced observations, are presented. We then develop ML and REML approaches to parameter estimation for the model. Simulations of the ML and REML estimation methods follow, and analysis of a data set from the Breeding Bird Survey illustrates an application of the model. We clarify the advantages and limitations of the model, in comparison to other approaches, in a discussion section. In Appendix A, the equations and statistical results presented in the paper are derived using concepts from elementary mathematical statistics. Appendix B contains an SAS computer program for fitting the model to data, along with annotated output.

THE MODEL AND ITS PROPERTIES

Model form and notation

The underlying model we use for population abundance is a discrete time, stochastic Gompertz model. Let N_t denote population abundance at time t. The abundance model (the process model) is

$$N_t = N_{t-1} \exp(a + b \ln N_{t-1} + E_t) \qquad (1)$$

where a and b are constants, and E_t (the "process noise") has a normal distribution with mean 0 and variance σ^2 (we write $E_t \sim \mathcal{N}(0, \sigma^2)$). The random variables E_1, E_2, ... are assumed to be uncorrelated. The model has a long history in density-dependence modeling (see Reddingius 1971, Pollard et al. 1987, Dennis and Taper 1994). A frequently seen alternative is a stochastic version of the Moran-Ricker model (Dennis and Taper 1994), which uses N_{t-1} instead of $\ln N_{t-1}$ in the exponential function; in comparative data analysis studies, the Gompertz model has performed as well as the

Moran-Ricker (e.g., Rotella et al. 1996). In a recent extensive analysis of 1780 time series of mammals, birds, fish, and insects using a flexible growth function, the most common shape of the per-unit abundance growth rate, as a function of population abundance, was strongly concave (up), similar to the shape of the Gompertz (Sibly et al. 2005). On the logarithmic scale, the Gompertz abundance model is a linear, autoregressive time series model of order 1 [AR(1) process]:

$$X_t = X_{t-1} + a + bX_{t-1} + E_t = a + cX_{t-1} + E_t \qquad (2)$$

where $X_t = \ln N_t$ and $c = b + 1$.

Statistical properties of the stochastic Gompertz are well known. The probability distribution of X_t is a normal distribution with mean and variance that change as functions of time. If $-1 < c < 1$, the probability distribution of X_t eventually approaches a time-independent stationary distribution that is a normal distribution with a mean of $a/(1 - c)$ and a variance of $\sigma^2/(1 - c^2)$. The stationary distribution is the stochastic version of an equilibrium in the deterministic model, and is an important statistical manifestation of density dependence in the population growth model.

Real population abundances are rarely known exactly. Therefore, we assume that the sampling procedure used in monitoring the population yields an estimate, Y_t, of the logarithmic population abundance X_t. The observation error in Y_t is assumed to be adequately described by a normal distribution with mean 0 and variance τ^2:

$$Y_t = X_t + F_t \qquad (3)$$

where $F_t \sim \mathcal{N}(0, \tau^2)$ and $F_1, F_2, \ldots$ are uncorrelated. The model describes sampling procedures that generate, on the original scale, multiplicative lognormal observation errors in the estimates of population abundances; the model is expected to describe a variety of sampling protocols fairly well (we expand this point further in *Discussion*).

The model as constituted by Eqs. 2 and 3 is in the form of a linear, normal "state-space model," for which there is a large statistical literature (e.g., Harvey 1993). We term Eqs. 2 and 3 the Gompertz state-space (GSS) model. Note that the underlying Gompertz population model (Eq. 1) is not linear, but is a sigmoid growth curve, and that the sampling portion of the model describes lognormal error that can be realistically expected in many ecological sampling protocols. The GSS model provides the likelihood function for connecting time series of estimated population abundances with the unknown parameters a, c, σ^2, and τ^2.

Likelihood function

The data are a time series of estimated log population abundances, denoted y_0, y_1, ..., y_q. These observations are assumed to constitute a realization of the stochastic process Y_t defined by Eqs. 2 and 3.

One question arises immediately: do the data contain enough information to estimate both the process variance (σ^2) and sampling variance (τ^2)? Some insight into the question can be gained by rewriting Eqs. 2 and 3 in terms of Y_t alone. Substituting Eq. 2 for X_t in Eq. 3, and noting that $X_{t-1} = Y_{t-1} - F_{t-1}$, we find that

$$Y_t = a + cY_{t-1} + (E_t + F_t - cF_{t-1}). \qquad (4)$$

The observation process Y_t is seen to consist of a deterministic portion, $a + cY_{t-1}$ (identical to the deterministic trend for X_t in Eq. 2), plus a noise portion in parentheses, $E_t + F_t - cF_{t-1}$. The noise portion in Eq. 4 is substantially different from the noise in Eq. 2, in that a component, F_{t-1}, of the noise for Y_{t-1} appears in the noise for Y_t. Consequently, the noise term in Eq. 4 is autocorrelated. The correlation of the noise for Y_t with that of Y_{t-1} can be shown to be $-c\tau^2/[\sigma^2 + (1 + c^2)\tau^2]$. The process Y_t is not a purely autoregressive process like X_t; instead, Y_t is a type of autoregressive moving average (ARMA) process. The process Y_t does not even have the Markov property: the distribution of Y_t depends not only on Y_{t-1}, but on the entire past history Y_{t-2}, $Y_{t-3}, \ldots$ of the process. The fact that observation error induces autocorrelation in the noise of an autoregressive process is well-known in statistics (Walker 1960). For that reason, checking for autocorrelation of residuals in process-error-only models has been stressed as a diagnostic tool for model evaluation (Dennis et al. 1991, Dennis and Taper 1994, Dennis and Otten 2000).

The point is, there is additional information in the data that potentially could be exploited to estimate the sampling variability parameter τ^2; the information is manifested in the autocorrelation of the noise terms for the observations. Whether the information is enough to render σ^2 and τ^2 separately estimable is a major focus of our study. We can anticipate that estimating τ^2 will be at least as difficult as estimating an autocorrelation coefficient in a time series, which is known to require many observations for good precision.

The likelihood function for the observations $y_0, y_1, \ldots, y_q$ forms the basis for many statistical inferences about the unknown parameters a, c, σ^2, and τ^2. The likelihood function is constructed from the following result: the random variable Y_t, given $Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \ldots, Y_0 = y_0$, has a normal distribution, with a mean m_t and variance v_t^2 that are computed recursively using the model parameters and history of observations. We write

$$Y_t | (Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \ldots, Y_0 = y_0) \sim \mathcal{N}(m_t, v_t^2). \qquad (5)$$

The recursion relationships for m_t and v_t^2 are

$$m_t = a + c\left[m_{t-1} + \frac{v_{t-1}^2 - \tau^2}{v_{t-1}^2}(y_{t-1} - m_{t-1})\right] \qquad (6)$$

$$v_t^2 = c^2 \frac{v_{t-1}^2 - \tau^2}{v_{t-1}^2}\tau^2 + \sigma^2 + \tau^2. \qquad (7)$$

If the initial population is assumed to arise from the stationary distribution of Y_t, the recursions are initiated at the mean and variance of that stationary distribution: $m_0 = a/(1 - c)$, $v_0^2 = [\sigma^2/(1 - c^2)] + \tau^2$. The stationary assumption would be applicable if the observations commenced after the population has been fluctuating around carrying capacity for awhile. If the initial population, however, is away from equilibrium, the recursions are initiated at $m_0 = x_0$, $v_0^2 = \tau^2$, with x_0 treated as an additional unknown parameter. This non-stationary assumption would apply, for instance, when a population commences at low abundance and grows to carrying capacity. In some non-stationary situations, such as an experimental population introduction, the initial population might be known and therefore not treated as an unknown parameter.

The results given by Eqs. 5–7 are well known in the statistics literature (for instance, Harvey 1993). The recursion equations (Eqs. 6 and 7) for the conditional mean and variance are subsumed in four matrix equations collectively termed the "Kalman filter" in statistics (with main applications in electrical engineering). We derived the simplified representation presented here in order to focus on details relevant to population modeling. The published Kalman filter proofs tend to be somewhat general, with dense notation and (ecologically) obscure terminology. The literature also tends to emphasize the estimation and prediction of the underlying signal X_t rather than estimation of the process parameters. We believe applications in ecology will benefit from an increased emphasis on the likelihood function, parameter estimation, model selection, and model evaluation. Toward that end, we provide simple derivation of Eqs. 5–7 in Appendix A, using just elementary mathematical statistics concepts and properties of the bivariate normal distribution. As will be seen, using this model requires numerical maximization and attention to detail. For fitting the model to data and interpreting the results, it is helpful to have some familiarity with the underlying statistical concepts.

The probability density function (pdf) for Y_t, given the history $Y_{t-1} = y_{t-1}$, $Y_{t-2} = y_{t-2}, \ldots, Y_0 = y_0$, is thus that of a normal distribution:

$$f(y_t | y_0, y_1, \ldots, y_{t-1}) = (v_t^2 2\pi)^{-1/2}\exp\left[-\frac{(y_t - m_t)^2}{2v_t^2}\right]. \qquad (8)$$

Because $Y_0 = X_0 + F_0$ (Eq. 3), the distribution of the initial observation Y_0 is also normal, with X_0 being either a normal random variable under the stationary assumption or a constant (x_0) under the non-stationary assumption. The pdf for Y_0 is

$$f(y_0) = (v_0^2 2\pi)^{-1/2}\exp\left[-\frac{(y_0 - m_0)^2}{2v_0^2}\right] \qquad (9)$$

with m_0 and v_0^2 initiated according to the stationary assumption or non-stationary assumption.

The likelihood function is a function of the unknown model parameters. It is the joint pdf for random

variables $Y_0, Y_1, \ldots, Y_q$, evaluated at the recorded values $y_0, y_1, \ldots, y_q$. The likelihood function is thus

$$L(a, c, \sigma^2, \tau^2)$$
$$= f(y_0)f(y_1|y_0)f(y_2|y_0, y_1) \cdots f(y_q|y_0, y_1, \ldots, y_{q-1})$$
$$= (2\pi)^{-(q+1)/2}(v_0^2 v_1^2 \cdots v_q^2)^{-1/2}\exp\left[-\frac{1}{2}\sum_{t=0}^{q}\frac{(y_t - m_t)^2}{v_t^2}\right]. \tag{10}$$

Under the non-stationary case, the likelihood function would include x_0 as an unknown parameter. Calculations for parameter estimation (see *Parameter Estimation Methods*) are typically performed with the log-likelihood function given by

$$\ln L(a, c, \sigma^2, \tau^2) = -\frac{(q+1)}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=0}^{q}\ln v_t^2$$
$$-\frac{1}{2}\sum_{t=0}^{q}\frac{(y_t - m_t)^2}{v_t^2}. \tag{11}$$

Two special cases of the likelihood are noteworthy. If the underlying process X_t is purely deterministic, that is, if $\sigma^2 = 0$ in Eqs. 6 and 7, then m_t becomes the trajectory of the deterministic "skeleton" equation $m_t = a + cm_{t-1}$, and $v_t^2 = \tau^2$ is constant through time. The likelihood function becomes the product of pdf's for independent normal distributions with constant variances

$$L(a, c, 0, \tau^2) = (\tau^2 2\pi)^{-(q+1)/2}\exp\left[-\frac{1}{2\tau^2}\sum_{t=0}^{q}(y_t - m_t)^2\right]. \tag{12}$$

Eq. 12 is the likelihood function implied by fitting the deterministic trajectory to the observations using non-linear regression. On the other hand, if there is no sampling variability ($\tau^2 = 0$), then $y_t = x_t$ with probability 1 for all values of t. The mean becomes $m_t = a + cy_{t-1}$, the variance becomes $v_t^2 = \sigma^2$, and under the nonstationary case the likelihood becomes

$$L(a, c, \sigma^2, 0) = (\sigma^2 2\pi)^{-(q+1)/2}$$
$$\times \exp\left[-\frac{1}{2\sigma^2}\sum_{t=0}^{q}(y_t - a - cy_{t-1})^2\right]. \tag{13}$$

Eq. 13, the likelihood for a process-error-only model starting at fixed initial value y_0, leads to conditional least squares estimates of parameters (one-step-ahead fitting; Dennis and Taper 1994). For the stationary case, the likelihood in Eq. 13 would be multiplied by the stationary density [Eq. 9 with $m_0 = a/(1 - c)$, $v_0^2 = \sigma^2/(1 - c^2)$].

The recursion equations (Eqs. 6 and 7) represent merely a convenient method of calculating the likelihood or log-likelihood function. The observations $Y_0, Y_1, \ldots, Y_q$ in fact have a joint multivariate normal distribution (derivation in Appendix A). When the initial observation arises from the stationary distribution, the expected value (E) and variance (V) of each observation are as follows:

$$E(Y_t) = \frac{a}{1 - c} \tag{14}$$

$$V(Y_t) = \frac{\sigma^2}{1 - c^2} + \tau^2 \tag{15}$$

while the covariance of Y_t and Y_{t+s} is

$$\text{Cov}(Y_t, Y_{t+s}) = \frac{\sigma^2}{1 - c^2}c^s \qquad s > 0. \tag{16}$$

Let $\mathbf{Y} = [Y_0, Y_1, \ldots, Y_q]$ denote the observed population process written as a $(q + 1) \times 1$ random vector. The multivariate normal distribution for $\mathbf{Y}$ has a mean vector, denoted $\mathbf{m}$, with each element equal to the stationary mean in Eq. 14, and a variance–covariance matrix $\mathbf{V}$ with variances (diagonal elements) given by Eq. 15 and covariances (off-diagonal elements) given by Eq. 16. The likelihood function for the vector of time series data, denoted $\mathbf{y} = [y_0, y_1, \ldots, y_q]$, is the pdf of a multivariate normal distribution

$$L(a, c, \sigma^2, \tau^2) = (2\pi)^{-(q+1)/2}|\mathbf{V}|^{-1/2}$$
$$\times \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{m})\mathbf{V}^{-1}(\mathbf{y} - \mathbf{m})\right]. \tag{17}$$

Calculation of this likelihood (or its logarithm) requires inversion of the $(q + 1) \times (q + 1)$ matrix $\mathbf{V}$. This likelihood (Eq. 17) and the Kalman-iterated form (Eq. 10) are equivalent. The Kalman-iterated form essentially represents a partitioning scheme for inverting the matrix $\mathbf{V}$ (Eubank and Wang 2002).

When the initial value of the underlying population process is fixed at x_0, the elements of the mean vector $\mathbf{m}$ and the variance-covariance matrix $\mathbf{V}$ depend on time t, as the process has not become stationary. The means, variances, and covariances for $Y_0, Y_1, \ldots, Y_q$ are derived in Appendix A and are given by the following expressions:

$$E(Y_t) = a\frac{1 - c^t}{1 - c} + x_0 c^t \tag{18}$$

$$V(Y_t) = \sigma^2\frac{1 - (c^2)^t}{1 - c^2} + \tau^2 \tag{19}$$

$$\text{Cov}(Y_t, Y_{t+s}) = \sigma^2\frac{1 - (c^2)^t}{1 - c^2}c^s \qquad s > 0. \tag{20}$$

The likelihood function for the nonstationary case is given by Eq. 17, except that the elements of $\mathbf{m}$ and $\mathbf{V}$ come from Eqs. 18–20, and the initial value x_0 must be included in the list of unknown parameters.

Density independence

An important special case of the GSS model is density independence, and for completeness, we present some of the relevant statistical results. Density independence is represented by $c = 1$ ($b = 0$) in Eq. 2:

$$X_t = X_{t-1} + a + E_t. \tag{21}$$

The resulting process model for population abundance is a form of stochastic, discrete-time, exponential growth

(or decay) model. The model serves as the null hypothesis in various statistical tests for density dependence (Pollard et al. 1987, Dennis and Taper 1994), and is the discrete-time version of the diffusion process model described by Dennis et al. (1991) for use in population viability analysis (PVA).

The density-independent state-space model (Eq. 21, in conjunction with Eq. 3), represents a stochastic exponential growth model with observation error. The likelihood and model properties of the density-independent state-space model merit separate consideration from those of the density-dependent case. Setting $c = 1$ does not work in all of the above density-dependent formulas, because that c value is the boundary of the parameter space for which the process is (or eventually becomes) stationary. The resulting statistically pathological properties are a part of what is known in statistics as the "unit root" problem in time series analysis (Dickey and Fuller 1981). The population under Eq. 21 does not have a statistical equilibrium; no stationary distribution exists. The initial population size x_0 therefore becomes an unknown model parameter. The recursion equations for the conditional mean and variance of Y_t (given $Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \ldots, Y_0 = y_0$) become

$$m_t = a + m_{t-1} + \frac{v_{t-1}^2 - \tau^2}{v_{t-1}^2}(y_{t-1} - m_{t-1}) \qquad (22)$$

$$v_t^2 = \frac{v_{t-1}^2 - \tau^2}{v_{t-1}^2}\tau^2 + \sigma^2 + \tau^2 \qquad (23)$$

with $m_0 = x_0$, $v_0^2 = \tau^2$. The log-likelihood function is Eq. 11, in which m_t and v_t^2 are calculated with Eqs. 22 and 23, and x_0 is treated as unknown.

As in the density-dependent case, the observations $Y_0, Y_1, \ldots, Y_q$ under the density-independent model have a joint multivariate normal distribution, which provides an alternative method of writing the likelihood function. The means, variances, and covariances are derived in Appendix A and are given by

$$E(Y_t) = x_0 + at \qquad (24)$$

$$V(Y_t) = \sigma^2 t + \tau^2 \qquad (25)$$

$$Cov(Y_t, Y_{t+s}) = \sigma^2 t \quad s > 0. \qquad (26)$$

The likelihood function for the density-independent model is the multivariate normal pdf (Eq. 17), with the elements of $\mathbf{m}$ and $\mathbf{V}$ coming from Eqs. 24–26. The Kalman-iterated form (Eq. 10 using Eqs. 22 and 23) is equivalent.

The density-independent state-space model was proposed by Holmes (2001) and Holmes and Fagan (2002) as an improved basis for conducting PVA in the presence of observation error. They adapted a linear regression method for parameter estimation. However, likelihood-based estimation methods (see *Parameter Estimation Methods*) have some theoretically optimal large-sample

(Knight 2000) as well as small-sample (Godambe 1960) statistical properties. Simulations of parameter estimation for the density-independent state-space model have supported likelihood-based methods (Staples et al. 2004), especially REML estimates based on differences.

Differences

In nonstationary situations, a standard technique in applied time series analysis is taking the differences of the observations. The differences lead to alternative parameter estimation methods, and it is worthwhile to investigate whether any benefits might result for estimation under the density-dependent or density-independent models. Let $W_t = Y_t - Y_{t-1}$ ($t = 1, 2, \ldots, q$) denote the first-order differences, and $U_t = W_{t+1} - W_t$ ($t = 1, 2, \ldots, q - 1$) denote the second-order differences.

In the density-dependent model (Eqs. 2 and 3), the process becomes stationary after sufficient time has elapsed. Upon attaining stationarity, the first differences $W_1, W_2, \ldots, W_q$ have a joint multivariate normal distribution with

$$E(W_t) = 0 \qquad (27)$$

$$V(W_t) = 2\tau^2 + 2\frac{\sigma^2}{1 - c^2}(1 - c) \qquad (28)$$

$$Cov(W_t, W_{t+1}) = -\frac{\sigma^2}{1 - c^2}(1 - c)^2 - \tau^2 \qquad (29)$$

$$Cov(W_t, W_{t+s}) = -\frac{\sigma^2}{1 - c^2}(1 - c)^2 c^{s-1} \quad s \geq 2. \qquad (30)$$

(see Appendix A for proof).

Note that differencing eliminates one parameter from the density-dependent model. The parameter a does not appear in the joint distribution of the first differences. One might expect estimation of the remaining parameters to be improved, calling to mind the removal of fixed-effects parameters through REML estimation in mixed effects models (Searle et al. 1992). In fact, the GSS model is a type of mixed-effects model (see *Parameter Estimation Methods*). However, the benefits if any of REML estimation in the GSS model appear to be slight (see *Results*). Also, if the process has not become stationary, the mean, variance, and covariance terms in the GSS model are then all functions of t, and differencing the observations does not eliminate any parameters.

The density-independent process (Eqs. 21 and 3) is highly nonstationary, with a mean increasing or decreasing exponentially on the original scale, and linearly on the logarithmic scale. The first differences, however, are stationary. In Appendix A it is shown that $W_1, W_2, \ldots, W_q$ have a joint multivariate normal distribution with

$$E(W_t) = a \qquad (31)$$

$$V(W_t) = \sigma^2 + 2\tau^2 \qquad (32)$$

$$\text{Cov}(W_t, W_{t+1}) = -\tau^2 \tag{33}$$

$$\text{Cov}(W_t, W_{t+s}) = 0 \qquad s \geq 2. \tag{34}$$

Note that one parameter, the initial condition x_0, has been eliminated from the distribution.

Another parameter elimination in the density-independence model occurs with the second differences U_1, $U_2, \ldots, U_{q-1}$. The second differences are stationary, and they have a joint multivariate normal distribution with

$$\text{E}(U_t) = 0 \tag{35}$$

$$\text{V}(U_t) = 2\sigma^2 + 6\tau^2 \tag{36}$$

$$\text{Cov}(U_t, U_{t+1}) = -\sigma^2 - 4\tau^2 \tag{37}$$

$$\text{Cov}(U_t, U_{t+2}) = \tau^2 \tag{38}$$

$$\text{Cov}(U_t, U_{t+s}) = 0 \qquad s \geq 3. \tag{39}$$

(see Appendix A for proof).

Staples et al. (2004) reported that estimates of the process and observation variances (σ^2 and τ^2) in the density-independent model based on the second differences (REML estimates) had excellent statistical properties in simulations.

PARAMETER ESTIMATION METHODS

Maximum likelihood estimation

ML parameter estimation is accomplished through maximization of the likelihood or log-likelihood function. The likelihood function can be calculated either in the Kalman-iterated form (Eq. 10) or in the multivariate normal form (Eq. 17). For the GSS model in the stationary case, the unknown model parameters are a, c, σ^2, and τ^2; in the non-stationary case, the unknown parameters are a, c, σ^2, τ^2, and x_0. For the density-independent model, the unknown parameters are a, σ^2, τ^2, and x_0. The values of the unknown parameters that jointly maximize the likelihood or log-likelihood are the ML parameter estimates.

Closed formulas are not available for ML (or REML) estimates in the GSS model; numerical maximization is necessary. The Nelder-Mead simplex algorithm for unconstrained minimization/maximization is easy to program (Press et al. 1992) and is remarkably robust. Various computer packages for mathematical calculation such as GAUSS (Aptech Systems, Maple Valley, Washington, USA), MATLAB (The Math Works, Natick, Massachusetts, USA), and MATHCAD (Mathsoft, Cambridge, Massachusetts, USA) include the Nelder-Mead and other algorithms as pre-programmed library routines.

More conveniently, the PROC MIXED procedure in SAS can be used to obtain the parameter estimates for the stationary case of the GSS model. The multivariate normal likelihood (Eq. 17), with the mean and covariance structure given by Eqs. 14–16, is identical to the likelihood of an analysis of variance model with repeated measures and mixed effects. The analysis of variance model has $q + 1$ repeated measures on a single subject, with (1) a fixed subject mean, (2) an AR(1) covariance structure for the repeated measures, and (3) a random time effect. In the resulting model parameterization, the fixed subject effect (intercept) is the stationary mean of X_t [given by $a/(1 - c)$], the variance–covariance matrix of the repeated measures has main diagonal elements equal to the stationary variance of X_t [$\sigma^2/(1 - c^2)$] and off-diagonal elements equal to the covariance of X_t and X_{t+s} ([$\sigma^2/(1 - c^2)$]c^s), and the variance component for the random time effect is the observation error variance (τ^2). An SAS program implementing the calculations (ML or REML) is provided in Appendix B, along with annotated program output.

The most important and difficult problem in ML (and REML) estimation for the GSS model is that the likelihood functions tend to have multiple modes, often three. Two of the local maxima, when they occur, are typically related to the special cases involving no process noise ($\sigma^2 = 0$) or no observation error ($\tau^2 = 0$). The estimates of parameters at the special-case maxima are close to the estimates arising from using Eqs. 12 and 13 as likelihoods (see *Results*). If the correct model is assumed to have both $\sigma^2 > 0$ and $\tau^2 > 0$, the proper ML estimates correspond to a local maximum in the likelihood function at which the estimates of σ^2 and τ^2 are interior points of the parameter space. The special-case estimates at the edge of parameter space are not statistically consistent when the data were indeed produced with appreciable amounts of both process noise and observation error ("statistically consistent" means that estimates converge to the true parameter values as time series length increases). The theory of likelihood estimation indicates that when there are multiple maxima the likelihood has at most one local maximum corresponding to consistent parameter estimates (under regularity conditions ostensibly met here [Huzurbazar 1948, Rohatgi 1976]). Users of the SAS program (Appendix B) should be cautioned that PROC MIXED uses only local maximization with starting parameter values based on moment estimates. The SAS procedure has an option to calculate the likelihood function for a grid of user-specified starting values; the option might help an investigator to locate additional likelihood peaks.

Interestingly, the globally highest peak in a GSS likelihood does not necessarily correspond to the statistically consistent estimates. In simulations of observations from the GSS model, we have found that it is not uncommon for the highest likelihood peak to be one of the special-case maxima. In one set of GSS simulations (time series length 30), under appreciable amounts of both process noise and observation error, 818 out of 2000 likelihoods were highest at one of the special case

maxima. The likelihood surface should be fully explored with different initial parameter values in the maximization calculations in order to find the interior solution. In particular, automating such explorations in computer programs for simulation and bootstrapping is a challenge. We have had success with a genetic algorithm (Eiben and Smith 2003), which introduces random variation ("mutations") each iteration into parameter values and selects new parameters based on the height of the likelihood ("fitness"). We suggest using exponential distributions for the random variation, with initial means taken as the values from fitting the process-error-only model (simple regression formulas), with the initial mean for τ^2 set equal to that of σ^2. Simulated annealing algorithms might be useful as well (see Press et al. 1992, De Valpine and Hastings 2002). When multiple peaks exist, the region of initial parameter values for which local hill-climbing algorithms lead to the proper interior peak is frequently small.

Profile likelihoods help reveal the information contained in the data about particular parameters. A profile likelihood is calculated by fixing the value of an unknown parameter, say σ^2, and maximizing the (log) likelihood using all the remaining unknown parameters. The process is repeated for many different fixed values of σ^2, ultimately producing a graph of the maximized log-likelihood as a function of the fixed values of σ^2. A steep, concave-down, approximately parabolic profile likelihood is a fair indication that the parameter is well estimated from the data. A profile likelihood surface can be calculated by using a grid of fixed values of two parameters (perhaps σ^2 and τ^2) and maximizing the log-likelihood at each grid point over the remaining parameters.

Restricted maximum likelihood estimation

In the density-dependent GSS model, we established that the first differences have a multivariate normal distribution with a zero mean vector (Eqs. 36–39), provided the process has attained stationarity. We denote the recorded differences by $w_1 = y_1 - y_0$, $w_2 = y_2 - y_1$, ..., $w_q = y_q - y_{q-1}$, and $u_1 = w_2 - w_1$, $u_2 = w_3 - w_2$, ..., $u_{q-1} = w_q - w_{q-1}$.

In the stationary case of the density-dependent model, the likelihood function for the first differences becomes

$$L(c, \sigma^2, \tau^2) = \left[\frac{1}{(2\pi)^{q/2}|\Phi|^{1/2}}\right] \exp\left(-\frac{1}{2}\mathbf{w}'\Phi^{-1}\mathbf{w}\right) \quad (40)$$

where $\mathbf{w} = [w_1, w_2, ..., w_q]$, and Φ is the variance–covariance matrix having terms in the form of Eq. 28 on the main diagonal, terms in the form of Eq. 29 on the upper and lower subdiagonals, and terms in the form of Eq. 30 elsewhere (see Appendix A). The log-likelihood is then

$$\ln L(c, \sigma^2, \tau^2) = -\frac{q}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Phi|) - \frac{1}{2}\mathbf{w}'\Phi^{-1}\mathbf{w}. \quad (41)$$

The restricted maximum likelihood (REML) estimates

of c, σ^2, and τ^2 for the stationary, density-dependent model are found by numerically maximizing Eq. 41. Even though the log-likelihood for the REML estimates contains one fewer parameters than for the ML estimates, multiple peaks can and do occur, and all the precautions mentioned for calculating ML estimates apply here. Once the REML estimates of c, σ^2, and τ^2 are obtained, the parameter a is estimated by maximizing the multivariate normal likelihood for the undifferenced observations (Eq. 17, with elements in $\mathbf{m}$ and $\mathbf{V}$ given by Eqs. 14, 15, and 16), holding c, σ^2, and τ^2 fixed at their REML values. In fact, the maximization for a constant mean vector in the multivariate normal has a closed form solution; the result for the REML estimate for a is

$$a = (1 - c)\frac{\mathbf{j}'\mathbf{V}^{-1}\mathbf{y}}{\mathbf{j}'\mathbf{V}^{-1}\mathbf{j}} \quad (42)$$

where $\mathbf{j}' = [1\ 1\ \cdots\ 1]$ is a vector of 1's.

The SAS program we provided (see Appendix B) can calculate REML estimates. In fact, REML estimation is the default method in recent versions of PROC MIXED in SAS, due to the growing evidence of the frequent superiority of REML estimates over ML estimates in mixed effects models (Littell et al. 1996). Our simulations (see *Results*) are more equivocal. In any case, the same cautions about multimodal likelihoods should be heeded for REML estimation.

For the density-independent model, the second differences are required for eliminating all the parameters in the mean, in order to concentrate the information in the data for estimating σ^2 and τ^2. The likelihood function for the second differences becomes

$$L(\sigma^2, \tau^2) = \left[\frac{1}{(2\pi)^{q/2}|\Psi_2|^{1/2}}\right] \exp\left(-\frac{1}{2}\mathbf{u}'\Psi_2^{-1}\mathbf{u}\right) \quad (43)$$

(see Appendix A), and the log-likelihood is

$$\ln L(\sigma^2, \tau^2) = -\frac{q}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Psi_2|) - \frac{1}{2}\mathbf{u}'\Psi_2^{-1}\mathbf{u} \quad (44)$$

where $\mathbf{u}$ is the $(q - 1) \times 1$ vector of second differences, and the elements of the variance–covariance matrix Ψ_2 are in the forms given in Eqs. 36–39. The REML estimates of σ^2 and τ^2 are found by maximizing Eq. 44. Then, the REML estimate of a is found by maximizing the multivariate normal likelihood of the first differences, with the values of σ^2 and τ^2 fixed at their REML values. The result can be written analytically as follows:

$$a = \frac{\mathbf{j}'\Psi_1^{-1}\mathbf{y}}{\mathbf{j}'\Psi_1^{-1}\mathbf{j}}. \quad (45)$$

Here, Ψ_1 is the variance–covariance matrix containing elements given by Eqs. 31–34. Finally, the estimate of x_0 is obtained from the multivariate normal distribution of the original undifferenced observations, with a, σ^2, and τ^2 held fixed at their REML values. We find that

$$x_0 = \frac{\mathbf{j}'\Psi^{-1}(\mathbf{y} - \mathbf{q}a)}{\mathbf{j}'\Psi^{-1}\mathbf{j}}. \qquad (46)$$

Here, Ψ is the variance-covariance matrix with elements given by Eqs. 25–26, and $\mathbf{q} = [0, 1, 2, \ldots, q]$. The SAS program provided by Staples et al. (2004) as an online supplement can be used for the REML calculations.

Functions of parameters

Estimating various functions of parameters might be of interest. Recall that the mean and variance of the stationary distribution of the observations (log scale) are, respectively, $E(Y_\infty) = a/(1 - c)$, $V(Y_\infty) = \sigma^2/(1 - c^2) + \tau^2$, and the mean and variance for the stationary distribution of the underlying process are $E(X_\infty) = a/(1 - c)$, $V(X_\infty) = \sigma^2/(1 - c^2)$. These can be estimated by substituting ML or REML parameter estimates. On the original scale, the mean and variance of the lognormal stationary distribution of the observations are given by

$$E[\exp(Y_\infty)] = \exp\left[E(Y_\infty) + \frac{1}{2}V(Y_\infty)\right] \qquad (47)$$

$$V[\exp(Y_\infty)] = \{\exp[E(Y_\infty)]\}^2\{\exp[V(Y_\infty)] - 1\} \qquad (48)$$

and the mean and variance of the lognormal stationary distribution of the underlying population abundance are given by

$$E(N_\infty) = \exp\left[E(X_\infty) + \frac{1}{2}V(X_\infty)\right] \qquad (49)$$

$$V(N_\infty) = \{\exp[E(X_\infty)]\}^2\{\exp[V(X_\infty)] - 1\}. \qquad (50)$$

These functions also are estimated by substituting the values of parameter estimates. The deterministic concept of "carrying capacity" is left behind in stochastic models (e.g., Dennis and Patil 1984, Dennis and Costantino 1988, Dennis and Taper 1994), but the stationary distribution mean, $E(N_\infty)$, represents a long term expected size around which the population fluctuates, with the magnitude of the fluctuations measured by $V(N_\infty)$.

Percentiles of the stationary distributions also might be of interest. The percentiles of the two stationary normal distributions (i.e., observation Y_∞ and process X_∞) are obtained using the stationary means and variances in the usual location-scale transform of standard normal distribution percentiles. The stationary normal distribution percentiles are then transformed exponentially to obtain the percentiles of the lognormal stationary distributions. For instance, the $100 \times p$th percentile, denoted η_p, of the lognormal stationary distribution for the observations is given by

$$\eta_p = \exp\left[E(Y_\infty) + z_p\sqrt{V(Y_\infty)}\right] \qquad (51)$$

where z_p is the $100 \times p$th percentile of the standard normal distribution (value of z for which the area under a standard normal curve between $-\infty$ and z is p). To estimate η_p, $E(Y_\infty)$ and $V(Y_\infty)$ in Eq. 51 are calculated with the parameter estimates. One potential use for stationary distribution percentiles is in a population monitoring program. A lower percentile could be designated as a warning level; a system's change for the worse might be indicated if a population's abundance dropped outside some ordinary range of variability.

Various proportional measures of variance components can be defined and estimated. The quantity $\sigma^2 + \tau^2$ is the one-step variability of the observation Y_t given the previous log population size X_{t-1}. One definition of the proportional component of variance due to process noise, denoted ϕ_1, uses that one-step variability in the denominator as a reference. The proportion of one-step variability in Y_t, so defined, due to the underlying process variability becomes a simple function of σ^2 and τ^2:

$$\phi_1 = \frac{\sigma^2}{\sigma^2 + \tau^2}. \qquad (52)$$

Another definition uses the variance of the stationary distribution of the observations, $V(Y_\infty)$, as the reference. The proportion of variability in the stationary distribution due to the underlying process variability is

$$\phi_2 = \frac{V(X_\infty)}{V(Y_\infty)} = \frac{\sigma^2}{\sigma^2 + \tau^2(1 - c^2)}. \qquad (53)$$

The density-dependence parameter c appears in this variance component measure because the stationary distribution variance reflects a blend of stochastic and deterministic forces. Ives et al. (2003) discuss further the concepts of stability and equilibrium in stochastic systems.

One can estimate the current value of the underlying population process X_t with the GSS model. An easy way to do this is to use the expected value of X_t given the history of the Y_t values up to and including Y_t. Embodied in the derivation of the Kalman recursions in Appendix A (especially Eq. A.16) is the relationship given by

$$E(X_t|Y_t = y_t, Y_{t-1} = y_{t-1}, \ldots, Y_0 = y_0)$$
$$= m_t + \frac{v_t^2 - \tau^2}{v_t^2}(y_t - m_t) \qquad (54)$$

where m_t and v_t are the recursively computed moments (Eqs. 6, 7). Point estimates of the X_t values are obtained by substituting the ML or REML parameter estimates into Eq. 54. In the state-space literature, a population estimate produced by Eq. 54 is the "filtered" value, that is, the expected value of X_t given the observations up to and including time t. An alternative is to calculate the "smoothed" value for X_t as the expected value of X_t given all the observations Y_0, Y_1, ..., Y_q (that is, including those subsequent to time t); for details see Harvey (1993). Whether the filtered or smoothed value is a more accurate predictor depends on various factors, for example, on whether the process prior to t was mainly transient or stationary.
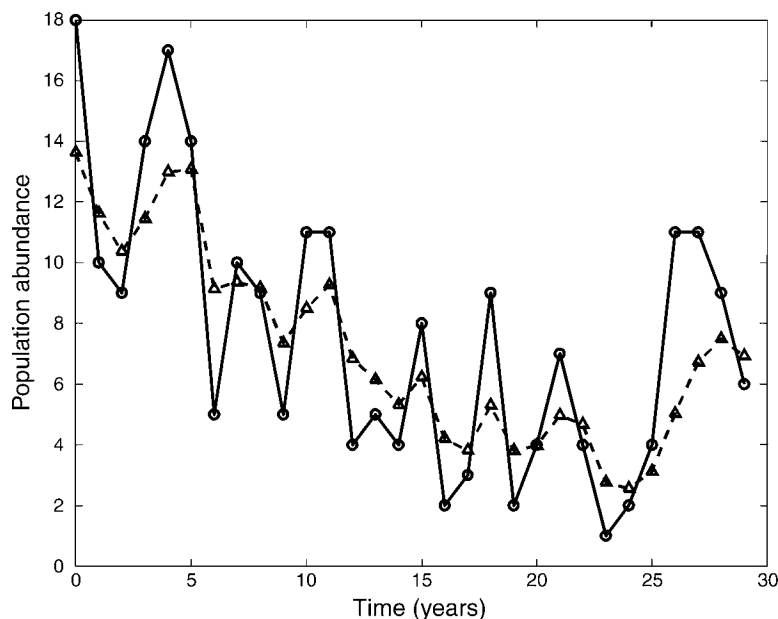
Fɪɢ. 1. Observed population counts of American Redstart, 1966–1995, from record number 0214332808636 of the North American Breeding Bird Survey (circles and solid line; see Peterjohn [1994] for description of data and sampling methods) and estimated population abundances from the fitted Gompertz state-space model (triangles and dotted line; see Table 1 legend).

Rᴇꜱᴜʟᴛꜱ

Example analyses using GSS model

The illustrative data set (Fig. 1) charts annual American Redstart (*Setophaga ruticilla*) counts at a survey location in the North American Breeding Bird Survey (BBS; Robbins et al. 1986, Peterjohn 1994). The data in the BBS are problematic in that the observations contain heterogeneous observation error. Survey routes are traversed by different observers with different training, abilities, hearing, eyesight, under different observing conditions. Nevertheless, the route locations are numerous and spatially extensive, and many of the time series exceed 30 yr in length. While the shortcomings of the BBS methodology are well documented, investigators remain confident that meaningful "signals" might be extracted which could prove valuable for monitoring and assessing trends in North American bird populations. Early statistical attempts to estimate trend in BBS data with linear regression (e.g., Bohning-Gaese et al. 1993) have been superceded by contemporary "overdispersed" models of count data that accommodate observer effects (Link and Sauer 1997, 1998).

We fitted the GSS model to the data with ML and REML estimation (Table 1, Fig. 1). We also calculated approximate 95% confidence intervals for the parameters under ML and REML estimation using parametric bootstrapping (Table 1). To obtain the bootstrap confidence intervals, we simulated 2000 data sets from the ML- and REML-estimated GSS models and recalculated estimates for each simulated data set. Our algorithm rejected solutions with σ^2 or τ^2 near zero and instead repeated and widened the search for an interior

local maximum using different initial values. The 2.5th and 97.5th empirical percentiles of the 2000 ML and REML values were taken as the confidence interval boundaries (Dennis and Taper 1994, Manly 1997). Thus, the confidence intervals reflect the sampling variability of the local interior solution and were not spread over separated intervals of parameter space. The ML and REML estimation methods yielded somewhat different values for the parameters. The ML estimate of a was considerably larger than the REML estimate, while the REML estimate of c was larger than the ML; both REML estimates were outside the ML-based confidence intervals (Table 1). The REML confidence intervals by contrast contained the ML estimates. The computer simulations presented below suggest that ML and REML estimates both have considerable sampling variability.

The fitted GSS model yielded estimates of the properties of the stationary distribution of population abundance (Table 2). The estimated stationary distribution of the log-scale observations under REML has larger variance than under ML, although the estimated means are similar. Translated to the original scale, the lognormal stationary distribution for the observed population abundance has 97.5th percentile ($\eta_{0.975}$) estimated to be almost twice as large under REML as under ML, but the 2.5th percentile ($\eta_{0.025}$) estimates are very close (Table 2). The ML estimates of the 2.5th and 97.5th percentiles for the lognormal stationary distribution of the underlying population process, N_t, are estimated to be (2.45, 18.28). Under the ergodic theorem for stochastic processes, the population abundance

TABLE 1. Maximum likelihood (ML) and restricted maximum likelihood (REML) estimates for the parameters a, c, σ^2, and τ^2 in the Gompertz state-space model.

Estimate	a	c	σ^2	τ^2
ML	0.3929	0.7934	0.09726	0.2315
	(0.15, 1.4)	(0.30, 0.92)	(0.020, 0.29)	(0.072, .38)
REML	0.1165	0.9427	0.06060	0.2622
	(0.015, 1.7)	(0.23, 0.99)	(0.0068, 0.26)	(0.076, 0.44)

Notes: The Gompertz state space model is given by $X_t = a + cX_{t-1} + E_t$, $Y_t = X_t + F_t$, where X_t is the natural logarithm of population abundance, Y_t is the observed or estimated value of X_t, E_t has a normal distribution with mean 0 and variance σ^2, F_t has a normal distribution with mean 0 and variance τ^2, a and c are constants, and t is time. Approximate 95% confidence intervals calculated with parametric bootstrapping appear in parentheses. The data were from record number 0214332808636 (American Redstart) of the North American Breeding Bird Survey, years 1966–1995. The observations are 18, 10, 9, 14, 17, 14, 5, 10, 9, 5, 11, 11, 4, 5, 4, 8, 2, 3, 9, 2, 4, 7, 4, 1, 2, 4, 11, 11, 9, 6.

would spend 95% of the time in the long run between those percentiles (discussion in Dennis and Costantino [1988]). The wide population fluctuations reinforce the idea that "carrying capacity" is better described by a stationary probability distribution than by a deterministic point equilibrium (Dennis and Patil 1984, Dennis and Costantino 1988).

The estimated variance components suggest that a substantial proportion of the observed population fluctuation is due to observation error (Table 2). Using the ML estimates, the proportional variance component due to the underlying process noise, under the definition given by Eq. 52, is estimated at $\hat{\phi}_1 = 0.2958$, and under the definition in Eq. 53, is estimated at $\hat{\phi}_2 = 0.5314$. Recall that these quantities measure different concepts and are not expected to have similar values; the value of

TABLE 2. Maximum likelihood (ML) and restricted maximum likelihood (REML) estimates for various functions of the parameters a, c, σ^2, and τ^2 in the Gompertz state-space model, using the data listed in Table 1.

Parameter	ML	REML
$E(X_\infty)$	1.902	2.033
$V(X_\infty)$	0.2625	0.5444
$V(Y_\infty)$	0.4940	0.8066
$E(N_\infty)$	7.637	10.03
$\eta_{0.025}$	1.689	1.314
$\eta_{0.975}$	26.56	44.41
ϕ_1	0.2958	0.1877
ϕ_2	0.5314	0.6749

Notes: Here $E(X_\infty) = a/(1 - c)$ [$= E(Y_\infty)$] is the mean of the stationary distribution of log population abundance X_t (and of the estimated log abundance Y_t), $V(X_\infty) = \sigma^2/(1 - c^2)$ is the variance of the stationary distribution of X_t, $V(Y_\infty) = [\sigma^2/(1 - c^2)] + \tau^2$ is the variance of the stationary distribution of Y_t, $E(N_\infty) = \exp[E(X_\infty) + V(X_\infty)/2]$ is the mean of the lognormal stationary distribution for population abundance N_t, $\eta_{0.025}$ and $\eta_{0.975}$ are, respectively, the 2.5th and 97.5th percentiles of the lognormal stationary distribution for the estimated population abundance $\exp(Y_t)$, and $\phi_1 = \sigma^2/(\sigma^2 + \tau^2)$ and $\phi_2 = \sigma^2/[\sigma^2 + (1 - c^2)\tau^2]$ are measures of the proportion of variability in the data due to processes error.
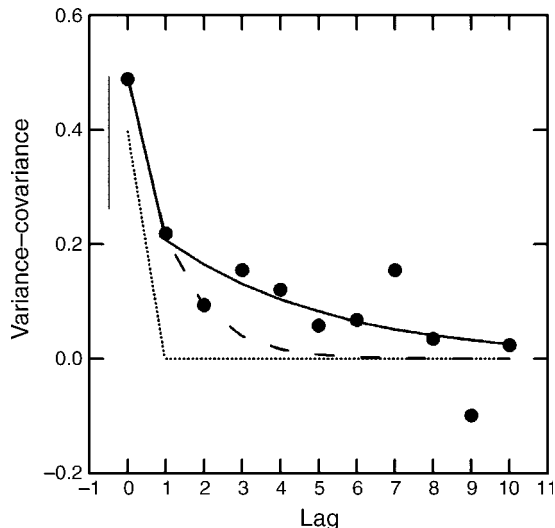

FIG. 2. Sample autocovariances (logarithmic scale; solid circles) at different lags, calculated from log-transformed American Redstart population data from Fig. 1. The solid line connects the theoretical covariances for the observations under the Gompertz state-space model (see Table 1 legend). The dashed line connects theoretical covariances for the underlying stochastic Gompertz model with process error only. The dotted line connects the theoretical covariances under the stochastic Gompertz model with observation error only. All models were fitted using maximum likelihood.

$\hat{\phi}_1$ suggests that more than 70% of the variation in the one-step fluctuations of the observations is observation error, while $\hat{\phi}_2$ suggests that observation error contributes almost half of the variability in the stationary distribution of the observations.

The GSS model fitted the data well. The estimates of the underlying local population abundances in the sampled area, calculated with Eq. 54 using the ML parameter values, tracked the observations, but with considerably less variability (Fig. 1). As reflected in the estimated variance components (above), the large volatility of the observations evident in Fig. 1 is estimated by the model to be due substantially to observation error. Recall that the information for such estimation is contained in the pattern of covariances displayed by the observations.

The GSS model described the covariance patterns of the data especially well, whereas the models with only process noise or observation error gave poor descriptions of the covariances. Fig. 2 shows the empirical covariances at various time lags, along with the theoretical covariances of the three models fitted to the data with ML estimation. For instance, the theoretical covariances of the GSS model (Eqs. 15, 16) were calculated using the ML parameter estimates for the GSS model (solid line, Fig. 2). Such excellent covariance description is important if a model is to be used for forecasting variability and first-passage properties. The extra sampling variance at the origin, termed the "nugget" in spatial statistics, is shown as a vertical bar
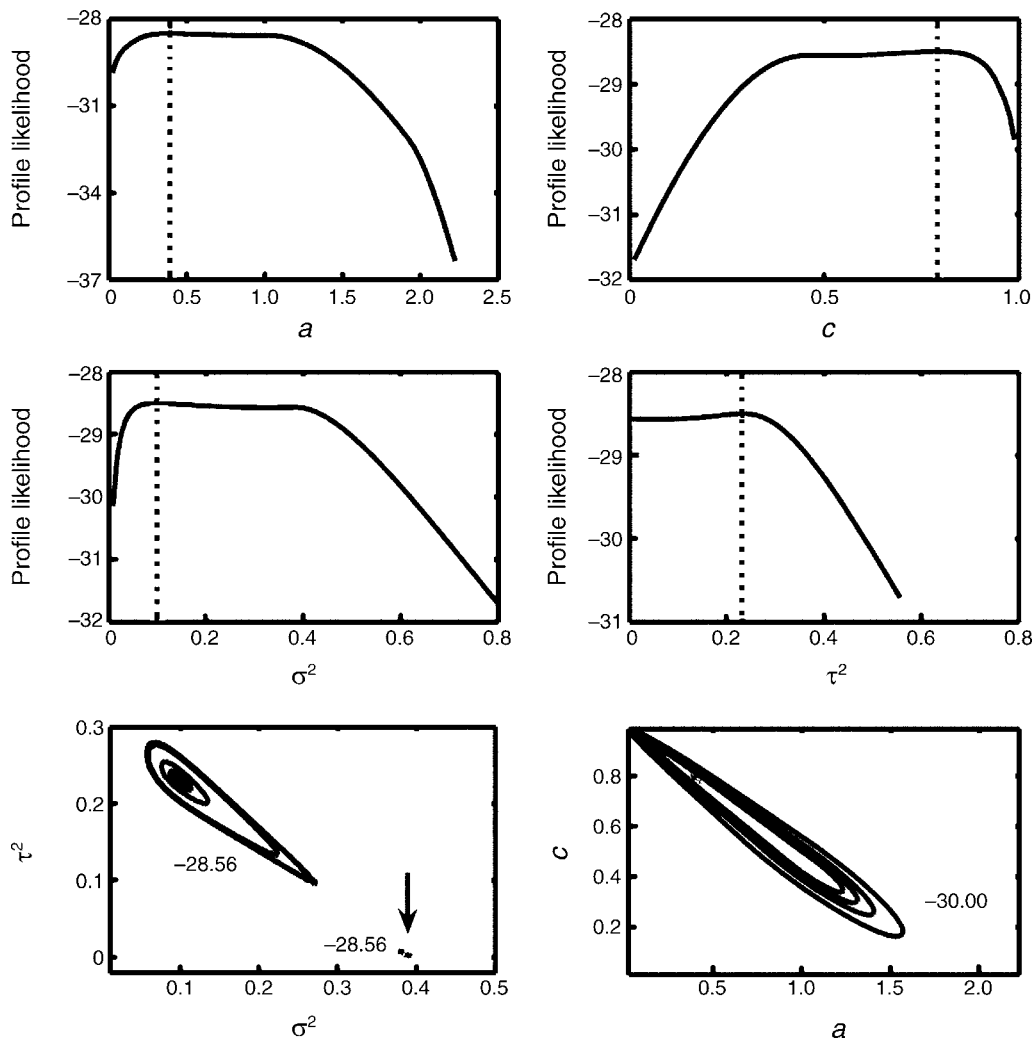
Fɪɢ. 3. In the top four graphs, solid curves are profile likelihoods for the parameters a, c, σ^2, and τ^2 in the Gompertz state-space model (see Table 1 legend), calculated with the American Redstart population data from Fig. 1. Dotted vertical lines indicate locations of maximum likelihood estimates. In the bottom two plots, solid curves are contours for joint profile likelihoods for the parameter pairs σ^2 and τ^2 (left) and a and c (right) calculated with the American Redstart population data from Fig. 1.

with length equal to the estimate of τ^2 near the vertical axis. The stochastic Gompertz model with process noise only has theoretical covariances identical to Eqs. 15 and 16 with τ^2 set to zero. Fitting the stochastic Gompertz model with only process noise yielded a model with theoretical covariances that failed to describe the shape of the empirical covariances (dashed line, Fig. 2). Because the process-error-only model lacks the extra parameter for observation error at a lag of 0, the resulting geometric series of theoretical covariances did not capture the break at lag 1 in the rate of decrease of the empirical covariances. An observation-error-only model, having theoretical covariances of zero, failed even worse to describe the empirical covariances (dotted line, Fig. 2).

The likelihood function for the GSS model with the BBS data set is multimodal, although the multimodality

is hard to illustrate in two dimensions. The upper two rows of Fig. 3 display the profile likelihoods individually for the four model parameters. Each profile was calculated by fixing the value of the parameter of interest, maximizing the log-likelihood over the remaining parameters, and repeating the process by fixing many values of the parameter of interest along the horizontal axis. The profiles appear ridge-shaped and even bimodal. A "shoulder" on the right of the σ^2 profile (Fig. 3, middle row, left) indicates the location of a local maximum and corresponds to the process-noise-only model. A local maximum at $\sigma^2 \approx 0$ that gives the observation-error-only model is not visible in the profile plots. The ML estimates are indicated by vertical lines in the graphs. Different sets of initial parameter values cause ordinary hill-climbing algorithms to converge to different local maxima. For some data sets we have

observed that the "basins of attraction," i.e., zones of initial parameter values for which an algorithm converges to a given local maximum, are banded. The bottom row of Fig. 3 shows joint profile likelihoods for two parameter pairs: σ^2 and τ^2 (left), and a and c (right). The profile likelihoods were computed by fixing the pair of parameter values at points over a grid, and maximizing the likelihood for the remaining two parameters; the panels display contours for the resulting landscapes. In the profile plot for σ^2 and τ^2, the arrow indicates the location of the local maximum for the process-noise-only model. The likelihood ridge in both joint profile plots is seen to be extremely narrow. It should be noted that greatly improved estimation of process variation can result when there is external information available about measurement error, as shown by Ferrari and Taper (2006) for the density-independent case.

We leave until a future paper the questions of whether alternative models provide better descriptions of the BBS data in Fig. 1, and of what statistical approaches should be used to conduct the model comparisons. For instance, the data give a visual impression of exponential decrease, although such wide fluctuations are consistent with, and well described by, the GSS model. The REML estimate of c near 1 (Table 1) suggests that the density-independent model (Eqs. 2 and 21) might be viable. As well, the data might reflect a density-dependent process decreasing to a new, lower, carrying capacity, in which case fitting the GSS model under the nonstationary assumption (Eqs. 17–20) might be necessary. The BBS data set we selected for illustrating the GSS model is not a poster child but rather exemplifies some of the practical problems in model selection and model fitting that can be encountered. For the GSS model family, statistical properties of parametric bootstrap hypothesis testing and of information-theoretic model selection indexes are not well understood and will require extensive simulation in order to gain confidence in these methods.

Simulation of parameter estimation

We simulated ML and REML estimation for the GSS model, under the stationary case (Fig. 4). Using the ML estimates for the BBS data in the role of the "true" parameter values (Table 1), we simulated 2000 data sets of length 30 (i.e., $q = 29$) and 2000 data sets of length 100 ($q = 99$), and calculated ML and REML estimates for each simulated data set. The longer length of 100 is of course unrealistic for ecological data, but it serves to reveal how slowly the estimates converge to the true parameter values. In each simulated data set, the initial population abundances were drawn from the stationary distribution. The box plots in Fig. 4 show the simulated parameter estimates divided by the true parameter values, so that the relative variability of the estimates can be compared across different parameters. For time series of length 30, ML estimates were variable but acceptable (Fig. 4, top row); however, skewness and small departures from centering can be discerned in the box plots. The REML

estimates were superior to the ML estimates, in that the REML estimates were better centered on the true parameter values (Fig. 4, second row).

For time series of length 100, the ML and REML estimates performed somewhat better than for length 30 in terms of bias (Fig. 4, bottom two rows). However, the reduction of variability was considerably slower than one generally expects in likelihood-based estimation. For instance, the simulated ML estimates were centered better on the true values, but the variability of the estimates was hardly diminished (compare graphs in the third row of Fig. 4 to corresponding graphs in first row). Similarly, the variability of the REML estimates hardly changed with time series length 100 (compare bottom row of Fig. 4 to second row). The pattern of improved centering but slow decrease in variability is common in estimation problems involving dependent observations.

Various functions of parameters were estimated well in the simulations under both ML and REML (Fig. 5). The 2000 parameter sets (time series length 30) were used to calculate estimates of stationary distribution quantities and variance components. The mean of the stationary distribution for Y_t (and X_t) was estimated especially well by both ML and REML (top row, Fig. 5). The estimated variances of the stationary distributions for both X_t and Y_t (second and third rows, respectively, Fig. 5) had small biases under ML but were more variable under REML. The two definitions of variance components (Eqs. 52, 53) had estimates under both ML and REML that displayed little bias; The estimates of ϕ_1 under ML and REML (fourth row, Fig. 5) were more variable than the estimates of ϕ_2 (bottom row, Fig. 5). The high quality of the estimates of the variance component ϕ_2, the proportion of long-term variability in the observations due to process error, suggests that the quantity deserves greater attention as a metric in ecological studies.

DISCUSSION

GSS contrasted with other models

Link and Sauer (1997, 1998) developed an overdispersed multinomial model for the BBS counts. Their approach takes into account the effects on sampling of different observers and changes through time in the abilities of observers. The underlying model for population abundance is not a population model per se, but rather is a flexible function of time for the purpose of estimating trend in the data. The approach is a comprehensive attempt to incorporate observer effects and other covariates into the model of the sampling process. The GSS model, by contrast, differs in taking all sampling effects as random lognormal noise, which certainly represents overdispersion, but not as explicitly. Also, the GSS model uses a process-oriented model for population abundance that contains the ecological processes of density dependence and environmental noise. The GSS model and its submodels have specific, fixed trends: no long-term trend for the full model,
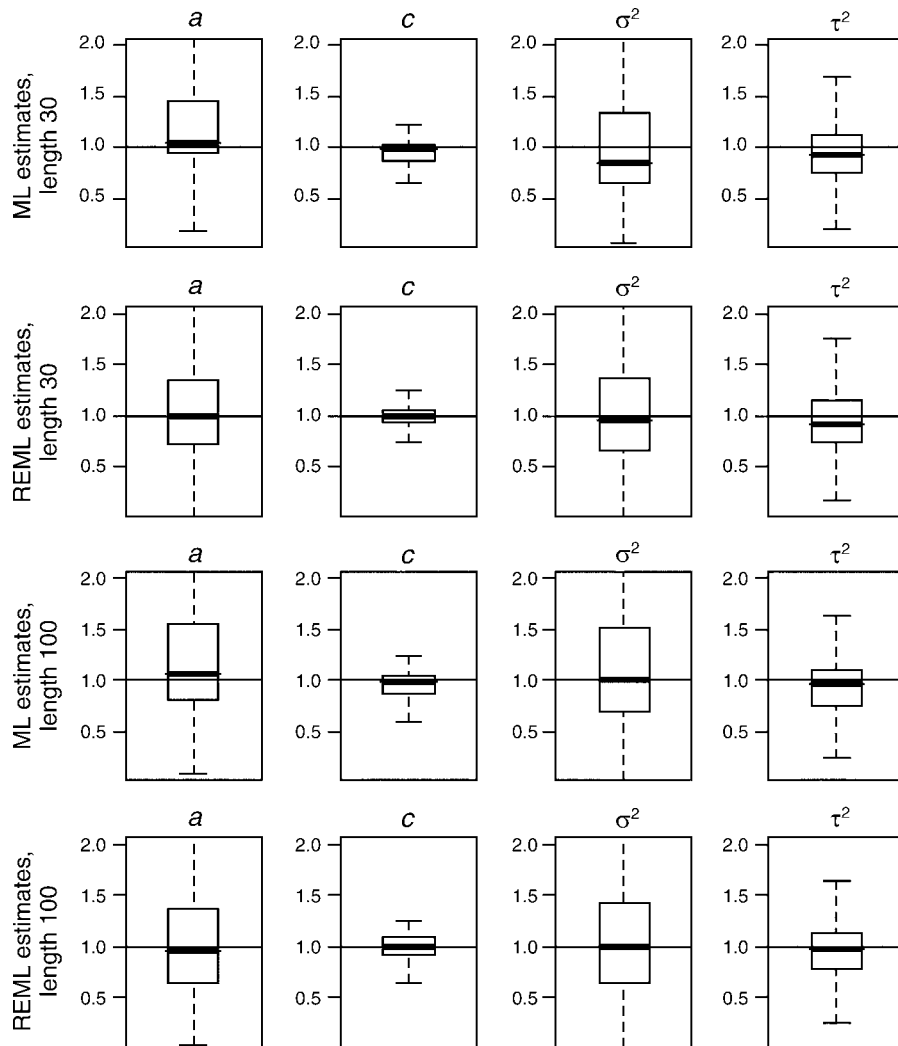
Fig. 4. Box plots (minimum, 25th percentile, median, 75th percentile, maximum) of maximum likelihood (ML) and restricted maximum likelihood (REML) parameter estimates divided by their true values, for the parameters a, c, σ^2, and τ^2 in the Gompertz state-space model, calculated for 2000 simulated time series data sets. Data sets were simulated from the Gompertz state-space model (see Table 1 legend) using the ML parameter estimates from Table 1 as the true values. Top row, 30 observations, ML; second row, 30 observations, REML; third row, 100 observations, ML; fourth row, 100 observations, REML.

exponential growth or decrease for the density-independent submodel. In using the GSS model, the investigator makes stricter assumptions about the form of population growth for purposes of evaluating those assumptions with data.

The state-space model of De Valpine and Hastings (2002) includes the GSS model as a special case. In their formulation, the underlying process model can be any single-population, stochastic, discrete-time growth model, and the observation process can have many distributional forms. In particular, De Valpine and Hastings study a stochastic Moran-Ricker model (Dennis and Taper 1994) for the underlying population process, combined with normal observation error. Their framework could accommodate discrete distributions, such as the Poisson or negative binomial, for more

realistic modeling of observation errors in count data, as might be more appropriate for BBS data. The computationally intensive algorithm of Kitagawa (1987) that they adopt for calculating the likelihood function basically mimics the Kalman filter derivation (see Appendix A) by simulation. Calculating ML estimates requires the additional task of numerical maximization, with the likelihood re-simulated at each maximization iteration. De Valpine and Hastings simulated ML estimation for their Ricker/normal state-space model, and they reported improved properties of the ML estimates over properties of estimates obtained with the misspecified model containing process error alone. De Valpine and Hastings noted the occurrence of multimodal likelihoods in their simulations, but do not seem to have adapted the simulations to the possibility
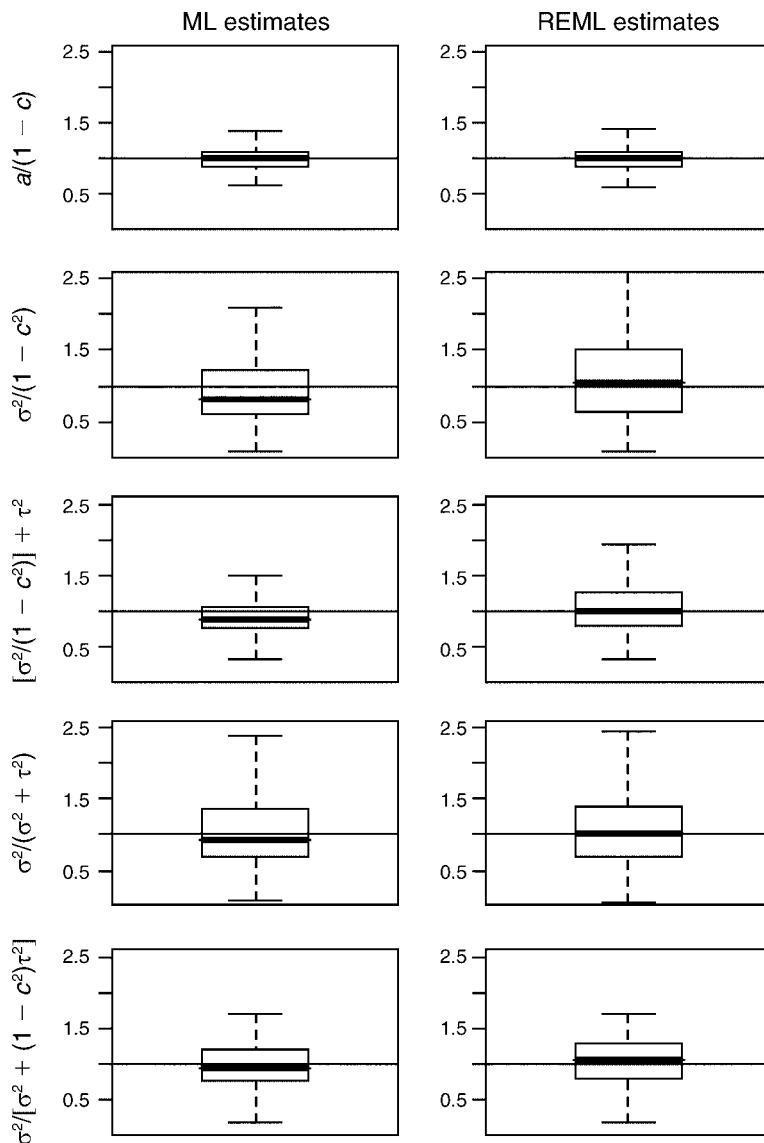
FIG. 5.   Box plots (minimum, 25th percentile, median, 75th percentile, maximum) of maximum likelihood (ML) and restricted maximum likelihood (REML) parameter estimates divided by their true values, for five functions of parameters a, c, σ^2, and τ^2 in the Gompertz state-space model (see Table 1 legend), calculated for 2000 simulated time series data sets with 30 observations each. Data sets were simulated from the Gompertz state-space model using the ML parameter estimates from Table 1 as the true values. First row: $a/(1 - c)$, the mean of the stationary distribution of log population abundance X_t (and of the estimated log abundance Y_t). Second row: $\sigma^2/(1 - c^2)$, the variance of the stationary distribution of X_t. Third row: $[\sigma^2/(1 - c^2)] + \tau^2$, the variance of the stationary distribution of Y_t. Fourth row: $\sigma^2/(\sigma^2 + \tau^2)$ ($= \phi_1$), a measure of the proportion of variability due to process error. Bottom row: $\sigma^2/[\sigma^2 + (1 - c^2)\tau^2]$ ($= \phi_2$), a second measure of the proportion of variability in the data due to processes error.

that the correct local maximum corresponding to the consistent likelihood root might not be the highest maximum. It is not known how often the multimodality occurs in the De Valpine and Hastings model, but the fact that multimodality occurs frequently in the GSS model, at an ecologically optimistic sample size of 30 observations, gives reason for pause. It is possible that the consistent ML estimates for the De Valpine and Hastings model perform considerably better than their simulations indicate.

The De Valpine and Hastings (2002) approach, while computationally intensive, is not Bayesian, but is rather a method of fitting a more realistic state-space model under a frequentist setting. By contrast, Bayesian approaches to state-space population models are being implemented more often (De Valpine 2002, Buckland et al. 2004, Clark and Bjørnstadt 2004). Ongoing development and improvement of software for the Bayesian calculations (for example, the BUGS software; see Meyer and Millar 1999*b*, Millar and Meyer 2000*b*) is helping to make Bayesian analysis more accessible.

Through the use of vague prior distributions, the peculiarities of the likelihood (such as in Fig. 3) are in essence smoothed over in the Bayesian setting. A note of caution, however, is appropriate: because convergence of the likelihood-based estimates to the true parameter values with increasing sample size is slow, the influence of even the vaguest priors on the analysis results can potentially remain large, say for sample sizes of 30 observations. Investigators should be aware that a Bayesian analysis is not just another statistical tool, but rather represents a substantially different approach to statistics and even to the philosophy of science (Lindley 1990, Dennis 1996, 2004, Mayo 1996).

The GSS model, in contrast to the above approaches, is minimalist. It uses just four parameters and has a relatively simple likelihood function, that of a multivariate normal distribution. Numerical maximization is the only computing required for obtaining point estimates. The simplicity of the model allowed insights into how statistical properties of population data with observation error differ from those of data with just process noise, how density dependence and observation error propagate into the covariance structure of the time series, and how joint estimation of process noise variance and observation error variance can be accomplished. Even with only four parameters, though, the model approaches the statistical limits of estimability; the ridge-shaped, multimodal likelihoods suggest that little additional information can be extracted from unreplicated single-species time series data.

Density independence

The statistical results included here for the density-independent case extend the work of Holmes (2001), Holmes and Fagan (2002), and Staples et al. (2004). Population viability analyses under density dependence or independence is potentially improved when observation error is accounted for and propagated into the extinction risk assessments (Holmes 2001). Many populations are expected on principle to be density dependent (Boyce 1992), but the density-independent case will remain an important model for situations in which the population is not limited by habitat, such as for new introductions or populations endangered by excessive harvesting. The simulations of Staples et al. (2004), as noted earlier, indicated that REML parameter estimates based on second differences had acceptable statistical properties for the density-independent state-space model. In fact, the REML estimates are better than they reported; the rare extreme over-estimates of the process noise variance (Fig. 1 in Staples et al. 2004) turn out to be instances where their simulation search algorithm settled on an incorrect likelihood maximum near the boundary of parameter space.

Observation error models

The observation error model given by Eq. 3 is likely to be a fairly robust description of many ecological sampling situations. The sampling model embodies many standard sampling methods, with added heterogeneity.

As an example, consider a plot- or area-based sampling method for estimating an unknown population abundance or density N_t. The distribution of counts in the plots, under ideal circumstances, is like a binomial or Poisson distribution. In such models, the variance of the estimate turns out to be proportional to N_t (Pielou 1977). Ordinary mark-recapture sampling models under ideal circumstances also feature estimator variance expressions that are proportional to the unknown population abundance (Otis et al. 1978). Because the mean of these estimators is equal to (unbiased or approximately unbiased estimate) or proportional to (in the case of abundance "indexes" such as light trap counts or spotlight drives) the abundance N_t, the coefficient of variation (standard deviation divided by mean) of the estimators changes with different N_t values; specifically, the coefficient of variation is proportional to $1/\sqrt{N_t}$.

Ecological sampling is not often ideal, however. Heterogeneity of sampling effort, observer abilities, observing or trapping conditions, trapabilities, and even local N_t values (as in aggregation) is rampant in ecological systems. One way to model heterogeneity is to assume that a sampling effort parameter varies from sampling occasion to occasion according to some probability distribution (Pielou 1977). For instance, if the observed population has a Poisson (λN_t) distribution, and λ varies according to a gamma distribution, then the unconditional distribution for the observed population is negative binomial (see Pielou 1977, Boswell et al. 1979). In such models, the variance of the observed population has a term proportional to N_t^2. The resulting coefficient of variation approaches a constant for large population sizes; that is, the coefficient of variation does not vary much with different N_t values.

The sampling model given by Eq. 3 takes the estimate of N_t to have a lognormal distribution. The lognormal sampling distribution is heavy-tailed, right-skewed, and has a constant coefficient of variation. Thus, it mimics key features of ecological sampling under heterogeneous conditions.

Further research

Overall, the REML estimates appeared to have slightly better statistical properties. However, the scope of our simulations was limited, and much more information is needed before blanket recommendations can be made. The ML and REML estimation methods should be evaluated for many different parameter values, and also for different functions of parameters (such as covariances and first passage quantities) which might be of specific interest in a population study. Many such problems left for future investigation could be studied with computer simulation.

We did not simulate ML estimation in the nonstationary case, a topic in which interesting problems are anticipated. On the one hand, starting the process far from equilibrium might be expected to improve the estimation of the density-dependence parameter c, because c measures the speed of recovery to equilibrium (see Dennis and Taper 1994, Ives et al. 2003). On the other hand, the nonstationary case contains an extra parameter (the initial log population abundance x_0), which consumes information in the data, unless the parameter value is known.

Estimation of first-passage properties for PVA remains a difficult statistical problem. Even with the simple exponential growth model with only process error, the confidence intervals for first-passage quantities are generally wide (Dennis et al. 1991). The GSS model represents a way of extending the Dennis et al. model to incorporate density dependence and observation errors into PVA. However, the degree of improvements in estimation that might thereby come from the improved description of the data are unknown, and should be studied by simulation. Simple formulas for first-passage quantities in the GSS model do not exist, so just obtaining the point estimates of first-passage quantities involves an additional layer of simulation.

Confidence intervals and hypothesis tests concerning various parameters (and functions of parameters) can be calculated with parametric bootstrapping. As well, parametric bootstrapping might help correct the biases in the estimates. Parametric bootstrapping is justified theoretically for any particular parameter estimation method, provided the parameter estimates are statistically consistent. In practice, parameter estimates can require large sample sizes to converge to the true parameter values. The bootstrapping procedures vary widely in quality, model by model and case by case. The statistical qualities of bootstrapping procedures, such as actual coverage probabilities for confidence intervals, have yet to be assessed for the GSS model.

Finally, problems related to model selection might usefully be studied with computer simulations. The GSS model contains various submodels as special cases, including stationary or not stationary, density independence, process noise only, observation error only, and various combinations of the deterministic and stochastic portions. The information-theoretic model selection indices such as AIC (Sakamoto 1986, Burnham and Anderson 2002) have been shown effective in simulations of a variety of model families (for instance, Hooten 1996). However, each model family is a different case, and the efficacy of using AIC and its relatives to choose from among the different special cases of the GSS model is at present unknown.

Concluding Remarks

Promising approaches exist for accomplishing the simultaneous estimation of observation error and process noise in density-dependent and density-inde-pendent populations. The approaches have demonstrable, proven usefulness in different ecological contexts. This paper has added one of the simpler approaches. The model we describe contains density dependence or independence, environmental process noise, and lognormal observation error, and the likelihood function for fitting the model to data can be written in closed form as the probability density function of a multivariate normal distribution. The combination of model characteristics likely represents a minimum level of complexity for adequately describing a variety of population time series data sets. The simplicity allows for (relatively) straightforward data analysis and could help investigators gain insights into the statistical properties of more complex population models. Nonetheless, we were surprised at the numerical and computational problems lurking inside the ordinary, single-variable Kalman filter model, problems which seem to have largely escaped attention in the voluminous statistical literature on state-space models.

We set out in this investigation to provide a statistical solution to the observation error problem that was accessible to busy empirical ecologists with standard training in contemporary statistical methods. A key to the widespread (maybe too widespread) use of earlier process-error-only models (Dennis et al. 1991, Dennis and Taper 1994) seemed to be that parameter estimation could be performed by ordinary linear regression. The GSS model, a model with the bare minimum of biological detail, requires attentive analysis, notwithstanding the fact that the necessary numerical maximizations can be performed with a SAS procedure. The multimodal likelihoods are a challenge for routine data analysis as well as for writing software to automate model fitting. Some knowledge of mathematical statistics, stochastic processes, and statistical simulation are exceedingly helpful for using the GSS model.

We cannot help but remark in closing that the various population models and contemporary statistical techniques are raising the bar on the level of quantitative preparation needed by ecologists, empirical and theoretical alike, who are seeking to understand population fluctuations. Judging from the recent literature, concepts such as stochasticity, covariance, likelihood, first passage distributions, stationary distributions, prior and posterior distributions, statistical simulation, and numerical optimization have joined nonlinear dynamics and stability as permanent parts of the landscape of ecological understanding (Turchin 2003, Cushing et al. 2003, Hilborn and Mangel 1997). Is training in only "applied" statistical methods, and not in statistical theory and stochastic modeling, adequate preparation for ecologists to have confidence in comprehending the very theories at the heart of our science (Dennis 2004)?

Literature Cited

Berryman, A. A. 1999. Principles of population dynamics and their application. Stanley Thomas, Cheltenham, UK.

Bohning-Gaese, K., M. L. Taper, and J. H. Brown. 1993. Are declines in North American insectivorous songbirds due to causes in the breeding range? Conservation Biology 7:76–86.

Boswell, M. T., J. K. Ord, and G. P. Patil. 1979. Chance mechanisms underlying univariate distributions. Pages 3–156 in J. K. Ord, G. P. Patil, and C. Taillie, editors. Statistical distributions in ecological work. International Co-operative Publishing House, Fairland, Maryland, USA.

Boyce, M. S. 1992. Population viability analysis. Annual Review of Ecology and Systematics 23:481–506.

Buckland, S. T., K. B. Newman, L. Thomas, and N. B. Koesters. 2004. State-space models for the dynamics of wild animal populations. Ecological Modelling 171:157–175.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and inference: a practical information-theoretic approach. Second edition. Springer-Verlag, New York, New York, USA.

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

Cushing, J. M., R. F. Costantino, B. Dennis, R. A. Desharnais, and S. M. Henson. 2003. Chaos in ecology: experimental nonlinear dynamics. Academic Press, San Diego, California, USA.

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

Dennis, B. 2004. Statistics and the scientific method in ecology (with commentary). Pages 327–378 in M. L. Taper and S. R. Lele, editors. The nature of scientific evidence: statistical, philosophical and empirical considerations. University of Chicago Press, Chicago, Illinois, USA.

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

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., P. L. Munholland, and J. M. Scott. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs 61:115–143.

Dennis, B., and M. R. M. Otten. 2000. Joint effects of density dependence and rainfall on abundance of San Joaquin kit fox. Journal of Wildlife Management 64:388–400.

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

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

De Valpine, P. 2002. Review of methods for fitting time series models with process and observation error and likelihood calculations for nonlinear, non-Gaussian state-space models. Bulletin of Marine Science 70:455–471.

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

Dickey, D. A., and W. A. Fuller. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. Econometrica 49:1057–1072.

Eiben, A. E., and J. E. Smith. 2003. Introduction to evolutionary computing. Springer-Verlag, New York, New York, USA.

Eubank, R. L., and S. Wang. 2002. The equivalence between the Cholesky decomposition and the Kalman filter. The American Statistician 56:39–43.

Ferrari, M. J., and M. L. Taper. 2006. Using auxiliary data on measurement error to improve estimates of process variation in ecological time series. Environmental and Ecological Statistics. In press.

Fromentin, J.-M., R. A. Myers, O. N. Bjørnstad, N. C. Stenseth, J. Gjøsæter, and H. Christie. 2001. Effects of density-dependent and stochastic processes on the regulation of cod populations. Ecology 82:567–579.

Godambe, V. P. 1960. An optimum property of regular maximum likelihood estimation. Annals of Mathematical Statistics 31:1208–1211.

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

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

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Sciences (USA) 98:5072–5077.

Holmes, E. E. 2004. Beyond theory to application and evaluation: diffusion approximations for population viability analysis. Ecological Applications 14:1272–1293.

Holmes, E. E., and W. F. Fagan. 2002. Validating population viability analysis for corrupted data sets. Ecology 83:2379–2386.

Hooten, M. M. 1996. Distinguishing forms of density dependence and independence in animal time series data using information criteria. Dissertation. Montana State University, Bozeman, Montana, USA.

Huzurbazar, V. S. 1948. The likelihood equation consistency, and maxima of the likelihood function. Annals of Eugenics 14:185–200.

Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. Ecological Monographs 73:301–330.

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

Knight, K. 2000. Mathematical statistics. Chapman & Hall/CRC, Boca Raton, Louisiana, USA.

Lindley, D. V. 1990. The 1988 Wald memorial lectures: the present position in Bayesian statistics (with comments). Statistical Science 5:44–89.

Lindley, S. T. 2003. Estimation of population growth and extinction parameters from noisy data. Ecological Applications 13:806–813.

Link, W. A., and J. R. Sauer. 1997. Estimation of population trajectories from count data. Biometrics 53:488–497.

Link, W. A., and J. R. Sauer. 1998. Estimating population change from count data: application to the North American Breeding Bird Survey. Ecological Applications 8:258–268.

Littell, R. C., G. A. Milliken, W. W. Stroup, and R. D. Wolfinger. 1996. SAS system for mixed models. SAS Publishing, Cary, North Carolina, USA.

Manly, B. F. J. 1997. Randomization, bootstrap and Monte Carlo methods in biology. Second edition. Chapman and Hall, London, UK.

Mayo, D. G. 1996. Error and the growth of experimental knowledge. University of Chicago Press, Chicago, Illinois, USA.

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

Meyer, R., and R. B. Millar. 1999b. BUGS in Bayesian stock assessments. Canadian Journal of Fisheries and Aquatic Sciences 56:1078–1087.

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

Millar, R. B., and R. Meyer. 2000b. Bayesian state-space modeling of age-structures data: fitting a model is just the beginning. Canadian Journal of Fisheries and Aquatic Sciences 57:43–50.

Myers, R. A., and N. G. Cadigan. 1993a. Density-dependent juvenile mortality in marine demersal fish. Canadian Journal of Fisheries and Aquatic Sciences 50:1576–1590.

Myers, R. A., and N. G. Cadigan. 1993b. Is juvenile natural mortality in marine demersal fish variable? Canadian Journal of Fisheries and Aquatic Sciences 50:1591–1598.

Newman, K. B., S. T. Buckland, S. T. Lindley, L. Thomas, and C. Fernández. 2006. Hidden process models for animal population dynamics. Ecological Applications 16:74–86.

Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62.

Peterjohn, B. G. 1994. The North American Breeding Bird Survey. Birding 26:386–398.

Pielou, E. C. 1977. Mathematical ecology. John Wiley, New York, New York, USA.

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 FORTRAN: the art of scientific computing. Cambridge University Press, Cambridge, UK.

Reddingius, J. 1971. Gambling for existence. A discussion of some theoretical problems in animal population ecology. Acta Biotheoretica 20 (Supplement):1–208.

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

Robbins, C. S., D. Bystrak, and P. H. Geissler. 1986. The Breeding Bird Survey: its first fifteen years, 1965–1979. Resource Publication 157. U.S. Fish and Wildlife Service, Washington, D.C., USA.

Rohatgi, V. K. 1976. An introduction to probability theory and mathematical statistics. John Wiley, New York, New York, USA.

Rotella, J. J., J. T. Ratti, K. P. Reese, M. L. Taper, and B. Dennis. 1996. Long-term population analysis of Gray Partridge in eastern Washington. Journal of Wildlife Management 60:817–825.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1986. Akaike information criteria statistics. D. Reidel, New York, New York, USA.

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

Searle, S. R., G. Casella, and C. E. McCulloch. 1992. Variance components. John Wiley and Sons, New York, New York, USA.

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

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

Solow, A. R., and K. Sherman. 1997. Testing for stability in a predator-prey system. Ecology 78:2624–2627.

Staples, D. F., M. L. Taper, and B. Dennis. 2004. Estimating population trend and process variation for PVA in the presence of sampling error. Ecology 85:923–929.

Staudenmayer, J., and J. P. Buonaccorsi. 2005. Measurement error in linear autoregressive models. Journal of the American Statistical Association 100:841–852.

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

Turchin, P. 2003. Complex population dynamics: a theoretical/empirical synthesis. Princeton University Press, Princeton, New Jersey, USA.

Walker, A. M. 1960. Some consequences of superimposed error in time series analysis. Biometrika 47:33–43.

APPENDIX A

Proofs of statistical properties of the density-dependent (Gompertz) and density-independent state-space models (*Ecological Archives* M076-012-A1).

APPENDIX B

An annotated SAS program for calculating parameter estimates for the Gompertz state-space model, along with annotated output (*Ecological Archives* M076-012-A2).

Brian Dennis, José Miguel Ponciano, Subhash R. Lele, Mark L. Taper, David F. Staples. 2006. Estimating density dependence, process noise, and observation error. *Ecological Monographs* **76:323-341.**

Appendix A. Proofs of statistical properties of the density dependent (Gompertz) and density independent state space models.

I. Kalman recursion relationships

The derivation of the recursion relationships for calculating the likelihood function is based on a well-known property of the bivariate normal distribution. Suppose the random variables X and Y are jointly distributed as bivariate normal, with means μ_X and μ_Y, variances σ_X^2 and σ_Y^2, and covariance σ_{XY}. We write

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \text{BVN}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} \right). \tag{A.1}$$

The property gives the conditional distribution of X, given $Y = y$, as normal with mean $\mu_X + \sigma_{XY}(\sigma_Y^2)^{-1}(y - \mu_Y)$ and variance $\sigma_X^2 - \sigma_{XY}(\sigma_Y^2)^{-1}\sigma_{XY}$. We write

$$X \,\big|\, Y = y \sim \text{N}\left(\mu_X + \sigma_{XY}\left(\sigma_Y^2\right)^{-1}(y - \mu_Y),\ \sigma_X^2 - \sigma_{XY}\left(\sigma_Y^2\right)^{-1}\sigma_{XY} \right). \tag{A.2}$$

The conditional mean is the familiar expression for the regression of X on Y for bivariate normal observations.

Case 1: initial population arises from stationary distribution.

The Gompertz state space model (Eqs. 2 and 3 in the paper) defines the (log-scale) population process X_t and the observation process Y_t. We first consider the case in which the initial value X_0 of the population process is assumed to arise from the normal

stationary distribution. Thus we have

$$X_0 \sim \mathrm{N}\left(\mu_0, \psi_0^2\right), \tag{A.3}$$

with mean μ_0 and variance ψ_0^2 given by

$$\mu_0 = \left(\frac{a}{1-c}\right)\left(= -\frac{a}{b}\right), \tag{A.4}$$

$$\psi_0^2 = \left(\frac{\sigma^2}{1-c^2}\right)\left(= -\frac{\sigma^2}{b(b+2)}\right). \tag{A.5}$$

The derivation occurs in five steps. First, we obtain the joint bivariate normal distribution of X_0 and Y_0. Second, we obtain the normal distribution of X_0, conditional on $Y_0 = y_0$. Third, the normal distribution of X_1, conditional on $Y_0 = y_0$, is obtained. Fourth, the joint bivariate normal distribution of Y_1 and X_1, given $Y_0 = y_0$, is established. Finally, the joint distribution from step four yields the marginal normal distribution of Y_1, given $Y_0 = y_0$, along with the recursion relationships for obtaining the conditional mean and variance of Y_1 from those of Y_0.

The first step of the derivation is to obtain the joint distribution of X_0 and Y_0. Observe that X_0 and Y_0 are a linear transformation of the bivariate normal (and independent) random variables X_0 and F_0: $X_0 = X_0$, and $Y_0 = X_0 + F_0$. Thus, X_0 and Y_0 jointly have a bivariate normal distribution. The mean and variance, denoted m_0 and v_0, of Y_0, are seen to be

$$\mathrm{E}(Y_0) \equiv m_0 = \mu_0, \tag{A.6}$$

$$\mathrm{V}(Y_0) \equiv v_0^2 = \psi_0^2 + \tau^2. \tag{A.7}$$

Writing $Y_0 - \mu_0 = X_0 - \mu_0 + F_0$ and substituting into the covariance expression $\mathrm{E}[(X_0 - \mu_0)(Y_0 - \mu_0)]$ helps establish that the covariance of X_0 and Y_0 is just the variance of X_0:

$$\mathrm{Cov}(X_0, Y_0) = \psi_0^2. \tag{A.8}$$

Thus:

$$\begin{bmatrix} X_0 \\ Y_0 \end{bmatrix} \sim \text{BVN}\left(\begin{bmatrix} \mu_0 \\ m_0 \end{bmatrix}, \begin{bmatrix} \psi_0^2 & \psi_0^2 \\ \psi_0^2 & v_0^2 \end{bmatrix} \right). \tag{A.9}$$

The second step is to establish the conditional distribution of X_0 given $Y_0 = y_0$. From the bivariate normal property (Eq. A.2), and from noting that $\psi_0^2 = v_0^2 - \tau^2$, we find that

$$X_0 \mid Y_0 = y_0 \sim \text{N}\left(m_0 + \frac{v_0^2 - \tau^2}{v_0^2}(y_0 - m_0), \frac{v_0^2 - \tau^2}{v_0^2}\tau^2 \right). \tag{A.10}$$

The third step is to obtain the form of the conditional distribution of X_1 given $Y_0 = y_0$. Noting from Eq. 2 that $X_1 = a + cX_0 + E_1$, and noting from Eq. A.10 the conditional mean and variance of X_0, we easily find that

$$\text{E}\left(X_1 \mid Y_0 = y_0 \right) \equiv \mu_1 = a + c\left[m_0 + \frac{v_0^2 - \tau^2}{v_0^2}(y_0 - m_0) \right], \tag{A.11}$$

$$\text{V}\left(X_1 \mid Y_0 = y_0 \right) \equiv \psi_1^2 = c^2 \frac{v_0^2 - \tau^2}{v_0^2}\tau^2 + \sigma^2. \tag{A.12}$$

It also follows, from the conditional normal distribution of X_0, that the distribution of X_1 given $Y_0 = y_0$ is normal with mean and variance given by Eqs. A.11 and A.12.

The fourth step is to establish the joint distribution of X_1 and Y_1, given $Y_0 = y_0$. Because X_1 given $Y_0 = y_0$ has a normal distribution, the joint distribution of X_1 and Y_1, given $Y_0 = y_0$, is easily obtained with the procedure used to obtain the joint distribution of X_0 and Y_0. Writing X_1 and Y_1 as a linear combination of X_1 and F_1 (i.e., $X_1 = X_1$ and $Y_1 = X_1 + F_1$), it follows that

$$\text{E}\left(Y_1 \mid Y_0 = y_0 \right) \equiv m_1 = \mu_1 = a + c\left[m_0 + \frac{v_0^2 - \tau^2}{v_0^2}(y_0 - m_0) \right], \tag{A.13}$$

$$V\left(Y_1 \,\middle|\, Y_0 = y_0\right) \equiv v_1^2 = \psi_1^2 + \tau^2 = c^2 \frac{v_0^2 - \tau^2}{v_0^2} \tau^2 + \sigma^2 + \tau^2, \quad \text{(A.14)}$$

$$\text{Cov}\left(X_1, Y_1 \,\middle|\, Y_0 = y_0\right) = \psi_1^2. \quad \text{(A.15)}$$

The joint distribution of X_1 and Y_1, given $Y_0 = y_0$, is bivariate normal, because X_1 and Y_1 are linear combinations of X_1 and F_1. We write

$$\begin{bmatrix} X_1 \\ Y_1 \end{bmatrix} \,\middle|\, Y_0 = y_0 \sim \text{BVN}\left(\begin{bmatrix} \mu_1 \\ m_1 \end{bmatrix}, \begin{bmatrix} \psi_1^2 & \psi_1^2 \\ \psi_1^2 & v_1^2 \end{bmatrix} \right). \quad \text{(A.16)}$$

Everything is now in place for step five and the main results. We see that Eq. A.16 gives the marginal distribution of Y_1, given $Y_0 = y_0$, as normal with mean m_1 and variance v_1^2:

$$Y_1 \,\middle|\, Y_0 = y_0 \sim \text{N}\left(m_1, v_1^2\right), \quad \text{(A.17)}$$

where m_1 and v_1^2 are related to m_0 and v_0^2 through the formulas A.13 and A.14. As well, Eq. A.16 provides, through the bivariate normal property (Eq. A.2), the distribution of X_1, conditional on $Y_1 = y_1$ *and* $Y_0 = y_0$. The conditioning on $Y_0 = y_0$ is carried along automatically because the joint distribution represented by Eq. A.16 is conditioned on $Y_0 = y_0$. One repeats steps two through four above, using X_1 and Y_1 in place of X_0 and Y_0, m_1 and v_1^2 in place of m_0 and v_0^2, etc. (basically, just incrementing all the subscripts upward by 1). Through this process, one arrives at the normal distribution of Y_2, given $Y_1 = y_1$ and $Y_0 = y_0$. The conditional mean, m_2, and conditional variance, v_2^2, of Y_2 turn out to have the same dependence on m_1 and v_1^2 that m_1 and v_1^2 have on m_0 and v_0^2 (Eqs. A.13 and A.14). Thus we find that

$$Y_t \,\middle|\, Y_{t-1} = y_{t-1}, Y_{t-2} = y_{t-2}, \dots, Y_0 = y_0 \sim \text{N}\left(m_t, v_t^2\right), \quad \text{(A.18)}$$

where the mean m_t and the variance v_t^2 are obtained through the recursion relationships given by

$$m_t = a + c\left[m_{t-1} + \frac{v_{t-1}^2 - \tau^2}{v_{t-1}^2}(y_{t-1} - m_{t-1})\right], \tag{A.19}$$

$$v_t^2 = c^2\frac{v_{t-1}^2 - \tau^2}{v_{t-1}^2}\tau^2 + \sigma^2 + \tau^2, \tag{A.20}$$

with the recursions initiated at $m_0 = a/(1 - c)$, $v_0^2 = [\sigma^2/(1 - c^2)] + \tau^2$.

Eqs. A.18-A.20 form the results (Eqs. 5-7 in the paper) used for building the likelihood function for the observations. Eqs. A.19 and A.20 are the Kalman recursions (corresponding to Eqs. 6 and 7 in the paper). Eqs. A.18-A.20 yield the normal pdf (paper, Eq. 8) for the observation Y_t conditioned on the observation history $Y_{t-1} = y_{t-1}$, $Y_{t-2} = y_{t-2}, \ldots, Y_0 = y_0$. The additional pdf (paper, Eq. 9) arising from the $N(m_0, v_0^2)$ distribution of Y_0 is incorporated in the likelihood function (paper, Eq. 10).

Case 2: initial population does not arise from stationary distribution.

If population monitoring commenced before the population reached stationarity, one can assume that the initial population is a fixed, but unknown, constant x_0. The derivation for the non-stationary case differs in the details about how the process is started up. For the non-stationary case, the derivation alters Eqs. A.3 to read:

$$\Pr[X_0 = x_0] = 1, \tag{A.21}$$

with the mean μ_0 and variance ψ_0^2 of X_0 given by

$$\mu_0 = x_0, \tag{A.22}$$

$$\psi_0^2 = 0, \tag{A.23}$$

instead of by Eqs. A.4 and A.5. Also, Y_0 and X_1 are found from x_0 through:

$$Y_0 = x_0 + F_0, \tag{A.24}$$

$$X_1 = a + cx_0 + E_1. \tag{A.25}$$

From Eq. A.24, we see that Y_0 has a N(m_0, v_0^2) distribution with $m_0 = x_0$ and $v_0^2 = \tau^2$.

In particular, Y_0 and X_1 are independent (because F_0 and E_1 are independent), and the conditional distribution of X_1 given $Y_0 = y_0$ is just the unconditional normal distribution of X_1, found from Eq. A.25:

$$X_1 \mid Y_0 = y_0 ~ \sim ~ N(\mu_1, \psi_1^2), \qquad (A.26)$$

with $\mu_1 = a + cx_0$ and $\psi_1^2 = \sigma^2$. Note that the relationships given by Eqs. A.11 and A.12 for the dependence of μ_1 and ψ_1^2 on μ_0 and ψ_0^2 are valid for this nonstationary case. The rest of the derivation, establishing the joint distribution of Y_1 and X_1 given $Y_0 = y_0$, then the distribution of Y_1 given $Y_0 = y_0$, and so on, follows exactly as in the stationary case above.

<center>II. Multivariate normal likelihood function</center>

Here we derive the full multivariate normal likelihood function for the univariate time series of observations Y_0, Y_1, ..., Y_q.

Density dependent model, stationary case

Under the AR(1) model (paper, Eq. 2), it is well-known that the true population abundances X_0, X_1, ..., X_q have a joint multivariate normal probability distribution (that is, X_t is a Gaussian process). Let $\mathbf{X} = [X_0, X_1, ..., X_q]'$, the population abundances collected into a $(q+1) \times 1$ column vector. The mean vector for $\mathbf{X}$ contains identical elements equal to the mean, $a/(1-c)$, of the stationary distribution when the initial population size X_0 arises out of the stationary distribution. The variance-covariance matrix $\mathbf{\Sigma}$ has diagonal elements equal to the stationary variance $\sigma^2/(1-c^2)$, and the off-diagonal element in row i and column j is $[\sigma^2/(1-c^2)]c^{|j-i|}$. These results are readily demonstrated by writing $\mathbf{X}$ as a linear transformation of independent, normally

distributed random variables in the vector $\mathbf{E} = [X_0, E_1, E_2, ..., E_q]$. Iteration of Eq. 2

(paper) establishes that

$$\mathbf{X} = \mathbf{g}a + \mathbf{CE}, \qquad (A.27)$$

where

$$\mathbf{g} = \begin{bmatrix} 0 & 1 & \dfrac{1-c^2}{1-c} & \dfrac{1-c^3}{1-c} & \cdots & \dfrac{1-c^q}{1-c} \end{bmatrix}', \qquad (A.28)$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ c & 1 & 0 & 0 & \cdots & 0 \\ c^2 & c & 1 & 0 & \cdots & 0 \\ c^3 & c^2 & c & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c^q & c^{q-1} & c^{q-2} & c^{q-3} & \cdots & 1 \end{bmatrix}. \qquad (A.29)$$

Because

$$\mathbf{E} \sim \mathrm{MVN}(\mathbf{h}, \boldsymbol{\Psi}), \qquad (A.30)$$

with

$$\mathbf{h} = \begin{bmatrix} \dfrac{a}{1-c} & 0 & 0 & \cdots & 0 \end{bmatrix}', \qquad (A.31)$$

$$\Psi = \sigma^2 \begin{bmatrix} \frac{1}{1-c^2} & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}, \qquad (A.32)$$

the vector $\mathbf{X}$ has a multivariate normal distribution with mean vector $\mathbf{g}a + \mathbf{Ch}$ and variance-covariance matrix $\mathbf{C}\boldsymbol{\Psi}\mathbf{C}'$. The matrix multiplications simplify when carried out algebraically. Thus

$$\mathbf{X} \sim \mathrm{MVN}\left(\frac{a}{1-c}\mathbf{j}, \boldsymbol{\Sigma}\right), \qquad (A.33)$$

where

$$\Sigma = \frac{\sigma^2}{1-c^2} \begin{bmatrix} 1 & c & c^2 & \cdots & c^q \\ c & 1 & c & \cdots & c^{q-1} \\ c^2 & c & 1 & \cdots & c^{q-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c^q & c^{q-1} & c^{q-2} & \cdots & 1 \end{bmatrix}, \tag{A.34}$$

and $\mathbf{j}$ is a $(q+1) \times 1$ vector of ones.

Now let $\mathbf{Y} = [Y_0, Y_1, ..., Y_q]'$. Eq. 3 is equivalent to writing

$$\mathbf{Y} = \mathbf{X} + \mathbf{F}, \tag{A.35}$$

where $\mathbf{F} \sim \text{MVN}(\mathbf{0}, \tau^2\mathbf{I})$, with $\mathbf{0}$ being a column vector of zeros and $\mathbf{I}$ being the identity matrix. Therefore,

$$\mathbf{Y} \sim \text{MVN}(\mathbf{m}, \mathbf{V}), \tag{A.36}$$

where $\mathbf{m} = [a/(1-c)]\mathbf{j}$ and $\mathbf{V} = \Sigma + \tau^2\mathbf{I}$. With $\mathbf{y}$ denoting the column vector $[y_0, y_1, ..., y_q]'$ of recorded data, the likelihood function is written as the pdf of a multivariate normal distribution:

$$L(a, c, \sigma^2, \tau^2) = (2\pi)^{-(q+1)/2} |\mathbf{V}|^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}-\mathbf{m})'\mathbf{V}^{-1}(\mathbf{y}-\mathbf{m})\right]. \tag{A.37}$$

Eqs. A.36 and A.37 embody the multivariate normal likelihood for the observations (paper, Eqs. 14-17).

Density dependent model, nonstationary case (X_0 fixed at x_0)

The derivation of the multivariate normal for the case of fixed initial condition follows the structure of the stationary case, with some redefinitions of notation. Define $\mathbf{X}_1$ as the $q \times 1$ vector of population abundances without the initial abundance: $\mathbf{X}_1 = [X_1, ..., X_q]'$. Define as well the vector $\mathbf{E}_1$ without the initial population abundance: $\mathbf{E}_1 = [E_1, E_2, ..., E_q]$. Then by iterating Eq. 2, one finds that $\mathbf{X}_1$ is a linear transformation

of independent, normally distributed random variables:

$$\mathbf{X}_1 = \mathbf{g}_1 a + \mathbf{h}_1 x_0 + \mathbf{C}_1 \mathbf{E}_1 , \tag{A.38}$$

where the vectors $\mathbf{g}_1$, $\mathbf{h}_1$, and the matrix $\mathbf{C}_1$ are given by

$$\mathbf{g}_1 = \left[1 \quad \frac{1 - c^2}{1 - c} \quad \frac{1 - c^3}{1 - c} \quad \cdots \quad \frac{1 - c^q}{1 - c} \right]' , \tag{A.39}$$

$$\mathbf{h}_1 = \left[c \ \ c^2 \ \ c^3 \ \cdots \ c^q \right]' , \tag{A.40}$$

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ c & 1 & 0 & 0 & \cdots & 0 \\ c^2 & c & 1 & 0 & \cdots & 0 \\ c^3 & c^2 & c & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c^{q-1} & c^{q-2} & c^{q-3} & c^{q-4} & \cdots & 1 \end{bmatrix} . \tag{A.41}$$

Because

$$\mathbf{E}_1 \sim \mathrm{MVN}\big(\mathbf{0}, \sigma^2 \mathbf{I} \big) , \tag{A.42}$$

the vector $\mathbf{X}_1$ has a multivariate normal distribution with mean vector $\mathbf{g}_1 a + \mathbf{h}_1 x_0$ and variance-covariance matrix $\sigma^2 \mathbf{C}_1 \mathbf{C}_1'$.

Now let $\mathbf{Y}_1 = [Y_1, ..., Y_q]'$, the vector of observations with the initial observation deleted. The model for the observations (paper, Eq. 3) can be rewritten in vector form as

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{F}_1 , \tag{A.43}$$

where $\mathbf{F}_1 \sim \mathrm{MVN}(\mathbf{0}, \tau^2 \mathbf{I})$ is $q \times 1$. Therefore,

$$\mathbf{Y}_1 \sim \mathrm{MVN}\big(\mathbf{g}_1 a + \mathbf{h}_1 x_0, \ \sigma^2 \mathbf{C}_1 \mathbf{C}_1' + \tau^2 \mathbf{I} \big) . \tag{A.44}$$

Because $Y_0 \sim \mathrm{normal}(x_0, \tau^2)$, and is independent of $Y_1, ..., Y_q$, the vector $\mathbf{Y} = [Y_0, Y_1, ..., Y_q]'$ has a multivariate normal distribution:

$$\mathbf{Y} \sim \mathrm{MVN}(\mathbf{m}, \mathbf{V}) , \tag{A.45}$$

where now $\mathbf{m} = \mathbf{g}a + \mathbf{h}x_0$, $\mathbf{V} = \mathbf{\Sigma} + \tau^2\mathbf{I}$, $\mathbf{I}$ is $(q+1) \times (q+1)$, with

$$\mathbf{g} = \begin{bmatrix} 0 & 1 & \dfrac{1-c^2}{1-c} & \dfrac{1-c^3}{1-c} & \cdots & \dfrac{1-c^q}{1-c} \end{bmatrix}', \tag{A.46}$$

$$\mathbf{h} = \begin{bmatrix} 1 & c & c^2 & c^3 & \cdots & c^q \end{bmatrix}', \tag{A.47}$$

$$\mathbf{\Sigma} = \left[\begin{array}{c|c} 0 & \mathbf{0}' \\ \hline \mathbf{0} & \sigma^2\mathbf{C}_1\mathbf{C}_1' \end{array}\right], \tag{A.48}$$

and with $\mathbf{0}$ being a column vector of zeros. Carrying out the laborious matrix operations produces the expressions for the elements of $\mathbf{m}$ and $\mathbf{V}$ (Eqs. 18, 19, and 20 in the paper).

Density independent model

With the density dependence parameter c appearing throughout the variance-covariance matrix of the density dependent model, it is not surprising that the model of density independence (Eqs. 21 and 3 in the paper) has an entirely different covariance structure. Because the density independent model does not have a stationary state, the only case is that of the initial condition fixed and unknown.

Iterating the density independent process model (paper, Eq. 21) easily produces

$$X_t = x_0 + at + E_1 + E_2 + \cdots + E_t. \tag{A.49}$$

Define $\mathbf{X}_1$ as the $q \times 1$ vector of population abundances without the initial abundance: $\mathbf{X}_1 = [X_1, ..., X_q]'$. Define as well the vector $\mathbf{E}_1$ as $\mathbf{E}_1 = [E_1, E_2, ..., E_q]$. Then one can write Eq. A.49 in matrix form as

$$\mathbf{X}_1 = \mathbf{j}x_0 + \mathbf{C}_1\mathbf{j}a + \mathbf{C}_1\mathbf{E}_1, \tag{A.50}$$

where the vector $\mathbf{j}$ is a $q \times 1$ vector of 1's, and the $q \times q$ matrix $\mathbf{C}_1$ is given by

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{A.51}$$

Because

$$\mathbf{E}_1 \sim \text{MVN}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right), \tag{A.52}$$

the vector $\mathbf{X}_1$ has a multivariate normal distribution with mean vector $\mathbf{j}x_0 + \mathbf{C}_1\mathbf{j}a$ and variance-covariance matrix $\sigma^2 \mathbf{C}_1 \mathbf{C}_1'$.

Now let $\mathbf{Y}_1 = [Y_1, ..., Y_q]'$. The observation model (paper, Eq. 3) is equivalent to writing

$$\mathbf{Y}_1 = \mathbf{X}_1 + \mathbf{F}_1, \tag{A.53}$$

where $\mathbf{F}_1 \sim \text{MVN}(\mathbf{0}, \tau^2 \mathbf{I})$ is $q \times 1$. Therefore,

$$\mathbf{Y}_1 \sim \text{MVN}\left(\mathbf{j}x_0 + \mathbf{C}_1\mathbf{j}a, \ \sigma^2 \mathbf{C}_1 \mathbf{C}_1' + \tau^2 \mathbf{I}\right). \tag{A.54}$$

Because $Y_0 \sim \text{normal}(x_0, \tau^2)$, and is independent of $Y_1, ..., Y_q$, the vector $\mathbf{Y} = [Y_0, Y_1, ..., Y_q]'$ has a multivariate normal distribution:

$$\mathbf{Y} \sim \text{MVN}(\mathbf{m}, \mathbf{V}), \tag{A.55}$$

where now $\mathbf{m} = \mathbf{j}x_0 + \mathbf{C}_2\mathbf{j}a$, $\mathbf{V} = \Sigma + \tau^2\mathbf{I}$, $\mathbf{I}$ is $(q+1) \times (q+1)$, $\mathbf{j}$ is $(q+1) \times 1$, and

$$\mathbf{C}_2 = \begin{bmatrix} \mathbf{0}' \\ \cdots \\ \mathbf{C}_1 \end{bmatrix}, \tag{A.56}$$

$$\Sigma = \begin{bmatrix} 0 & \vdots & \mathbf{0}' \\ ----- & \vdots & ---------- \\ \mathbf{0} & \vdots & \sigma^2 \mathbf{C}_1 \mathbf{C}_1' \end{bmatrix}, \tag{A.57}$$

with $\mathbf{0}$ being a column vector of zeros. Carrying out the matrix operations produces Eqs. 24, 25, and 26 for the elements of $\mathbf{m}$ and $\mathbf{V}$.

III. Differences

The forms of the multivariate normal likelihoods for the differenced observations, under the density dependent and density independent models, are obtained here.

Density dependent model, stationary case

The differences, $W_t = Y_t - Y_{t-1}$ ($t = 1, 2, ..., q$), are a simple linear transformation of the observations, and so the multivariate normal distribution of the differences is readily obtained. Let $\mathbf{W} = [W_1, W_2, ..., W_q]'$. It is evident that

$$\mathbf{W} = \mathbf{DY}, \tag{A.58}$$

where $\mathbf{Y} = [Y_0, Y_1, ..., Y_q]'$, and

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \tag{A.59}$$

is a $q \times (q + 1)$ matrix. The multivariate normal distribution of the observations is given in Eq. A.36; the linear transformation in Eq. A.58 produces

$$\mathbf{W} \sim \text{MVN}(\mathbf{0}, \mathbf{\Phi}), \tag{A.60}$$

where $\mathbf{\Phi} = \mathbf{D}[\mathbf{\Sigma} + \tau^2 \mathbf{I}]\mathbf{D}'$, with $\mathbf{\Sigma}$ given by Eq. A.34. The matrix multiplications, written out, yield Eqs. 27-30 in the paper. The likelihood function for the observations $w_1 = y_1 - y_0$, $w_2 = y_2 - y_1$, ..., $w_q = y_q - y_{q-1}$ is the pdf of a multivariate normal distribution with zero mean vector:

$$L\left(c, \sigma^2, \tau^2\right) = \left(\frac{1}{(2\pi)^{q/2}|\mathbf{\Phi}|^{1/2}}\right) \exp\left(-\frac{1}{2}\mathbf{w}'\mathbf{\Phi}^{-1}\mathbf{w}\right), \tag{A.61}$$

where $\mathbf{w} = [w_1, w_2, ..., w_q]'$.

Density independent model

The derivation for the joint distribution of the differences is similar to the density dependent case. The multivariate normal distribution of $\mathbf{Y} = [Y_0, Y_1, ..., Y_q]'$ for the density independent model is listed in Eqs. A.55, A.56, and A.57, with the definitions of $\mathbf{m}$ and $\mathbf{V}$ given in the accompanying text. The transformation to first differences is $\mathbf{W} = \mathbf{DY}$, where $\mathbf{D}$ is the matrix in Eq. A.59. The mean vector becomes $\mathbf{Dm} = \mathbf{j}a$. The distribution for $\mathbf{W}$ becomes

$$\mathbf{W} \sim \mathrm{MVN}(\mathbf{j}a, \mathbf{DVD'}). \tag{A.62}$$

The expressions in the paper (Eqs. 31-34) for the means, variances, and covariances of the differenced observations are obtained by writing out the matrix multiplications in Eq. A.62.

The second differences, $U_t = W_{t+1} - W_t$ ($t = 1, 2, ..., q - 1$), are in turn a linear transformation of the first differences. Letting $\mathbf{U} = [U_1, U_2, ..., U_{q-1}]'$, it is evident that $\mathbf{U} = \mathbf{DW}$. The mean vector of $\mathbf{U}$ is $\mathbf{Dj}a = \mathbf{0}$, and so the parameter a is eliminated, producing Eq. 35. The distribution for $\mathbf{U}$ is

$$\mathbf{U} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{DDVD'D'}). \tag{A.63}$$

The expressions for the variances and covariances of the second differences (paper, Eqs. 36-39) result from the matrix multiplications.

Brian Dennis, José Miguel Ponciano, Subhash R. Lele, Mark L. Taper, David F. Staples. 2006. Estimating density dependence, process noise, and observation error. *Ecological Monographs* **76:323-341.**

Appendix B: An annotated SAS program for calculating parameter estimates for the Gompertz state-space model, along with annotated output.

```
/*----------------------------------------------------------------------------*/
/*        PARAMETER ESTIMATES FOR THE GOMPERTZ STATE SPACE MODEL           */
/*   SAS program to calculate parameter estimates for the Gompertz state-  */
/*   space model, using time series population abundance estimates.  The   */
/*   GSS model is given by                                                 */
/*         X(t) = a + cX(t-1) + E(t)                                       */
/*         Y(t) = X(t) + F(t)                                              */
/*   where X(t) is the natural logarithm of population abundance N(t)      */
/*   (assumed unknown), Y(t) is the observed value of X(t), E(t) has a     */
/*   normal distribution with mean 0 and variance sigmasquared, F(t) has   */
/*   a normal distribution with mean 0 and variance tausquared (with no    */
/*   auto- or cross-correlations in E(t) and F(t)), and t is time.  Unknown */
/*   model parameters are a, c, sigmasquared, tausquared.  Data to be      */
/*   input into the program consist of observed or estimated population    */
/*   abundances O(0), O(1), O(2), .., O(q) (estimates of N(0), N(1), etc.), */
/*   along with the values of t.  The program currently does not accomodate */
/*   missing observations.                                                 */
/*                                                                         */
/*   Program transforms data to logarithmic scale:  Y(t) = ln[O(t)].  The  */
/*   program recasts the model as a linear mixed model with:  (1)  repeated */
/*   measures on one subject having an AR(1) covariance structure, and (2) */
/*   a random effect due to time (considered as a categorical variable).   */
/*   The random effect represents the extra variance component due to ob-  */
/*   servation error and produces a "nugget" (augmented main diagonal) in  */
/*   the var-cov matrix for the observations.                              */
/*                                                                         */
/*   The example data are from the North American Breeding Bird Survey     */
/*   (record # 0214332808636, American Redstart), and correspond to Table 1 */
/*   and Figure 1 of Dennis et al. (200X).                                 */
/*----------------------------------------------------------------------------*/

options nocenter;
data in;
input observed time;
y = log(observed);
cards;
 18 0
 10 1
  9 2
 14 3
 17 4
 14 5
  5 6
 10 7
  9 8
  5 9
11 10
11 11
  4 12
  5 13
  4 14
  8 15
  2 16
  3 17
  9 18
  2 19
  4 20
  7 21
  4 22
  1 23
```

```
 2 24
 4 25
11 26
11 27
 9 28
 6 29
;
proc mixed  method=ml alpha=.05 noitprint noinfo data = in;

/*  Restricted maximum likelihood (REML) is the default estimation method  */
/*  in PROC MIXED (SAS System for Windows Version 9.1).  Delete "method=    */
/*  ml" (or substitute "method=reml") in list of options in the above      */
/*  "proc mixed" statement for REML estimation if desired.  Also, the      */
/*  value of alpha, for asymptotic 100(1-alpha)% confidence intervals for  */
/*  parameters, be changed in the option list.                            */


class time;
model y= ;
random time;
repeated / type=ar(1) subject=intercept;
estimate 'intercept' intercept 1;

run;
quit;
/*---------------------------------------------------------------------------*/



/*---------------------------------------------------------------------------*/
/*            ANNOTATED OUTPUT OF THE GSS ESTIMATION PROGRAM              */
/*                                                                       */
/*  The following output was generated using SAS/STAT software, Version  */
/*  9.1 of the SAS System for Windows. Copyright (c) 2002-2003 SAS       */
/*  Institute Inc.  SAS and all other SAS Institute Inc. product or      */
/*  service names are registered trademarks or trademarks of SAS Institute */
/*  Inc., Cary, NC, USA.                                                  */

The SAS System

The Mixed Procedure

            Class Level Information

Class    Levels    Values

time        30      0 1 2 3 4 5 6 7 8 9 10 11 12
                    13 14 15 16 17 18 19 20 21 22
                    23 24 25 26 27 28 29


                  Covariance Parameter Estimates

Cov Parm     Subject      Estimate     Alpha      Lower       Upper

time                      0.2315       0.05       0.08439     1.7944
AR(1)        Intercept    0.7934       0.05       0.1859      1.4010
Residual                 0.2625       0.05       0.08314     3.9119

/*  In the "Estimate" column, the value listed for "time" is the estimate  */
/*  of tausquared, for "AR(1)" is c, and for "residual" is                 */
/*  sigmasquared/(1 - c*c) (the stationary variance of X(t)).  "Lower"    */
/*  and "Upper" columns give boundaries of asymptotic 95% confidence      */
/*  intervals for the parameters, based on inversion of the information   */
/*  matrix (Hessian of the log-likelihood).  The CIs have unknown coverage */
/*  properties for small- and moderate-lengthed time series.  The CI for  */
/*  c, along with the large value of the stationary variance upper bound,  */
/*  might suggest that the density independent model (c=1) is a viable    */
/*  model for the data.  A SAS program to fit the density independent     */
/*  state space model was provided as a supplement to Staples et al.      */
```

```
/*  (2004).                                                                  */


          Fit Statistics

-2 Log Likelihood              57.0
AIC (smaller is better)        65.0
AICC (smaller is better)       66.6
BIC (smaller is better)        70.6

/*  These "fit statistics" can be used for model selection, in comparison  */
/*  to other models fitted to the data.                                    */

                        Estimates

                  Standard
Label       Estimate      Error       DF     t Value     Pr > |t|

intercept     1.9021     0.2645       29       7.19       <.0001

/*  The estimate listed for "intercept" is the estimate of a/(1-c), the    */
/*  stationary mean of X(t).  The "Standard Error" is an asymptotic        */
/*  estimate based on the information matrix.  The t-test for the null     */
/*  hypothesis that a/(1-c)=0 is nonsensical in the context of the model.  */
/*                                                                         */
/*  Thus for the example BBS data, the ML parameter estimates are:         */
/*                                                                         */
/*           tausquared = 0.2315                                           */
/*           c = 0.7934                                                     */
/*           sigmasquared = 0.2625*(1 - c*c) = 0.09726                     */
/*           a = 1.9021*(1-c) = 0.3930                                      */
/*                                                                         */
/*  Compare with ML estimates, Table 1, Dennis et al. (200X).  Small       */
/*  numerical differences are due to roundoff error in SAS.                */
/*-----------------------------------------------------------------------*/


/*-----------------------------------------------------------------------*/
/*                         REFERENCES                                      */
/*                                                                         */
/*  Dennis, B., J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F.         */
/*     Staples.  2006.  Estimating density dependence, process noise, and  */
/*     observation error.  Ecological Monographs 76:323-341.               */
/*                                                                         */
/*                                                                         */
/*  Staples, D. F., M. L. Taper, and B. Dennis.  2004.  Estimating         */
/*     population trend and process variation for PVA in the presence of   */
/*     sampling error.  Ecology 85:923-929, with supplement in Ecological  */
/*     Archives E085-025-S1.                                               */
/*                                                                         */
/*-----------------------------------------------------------------------*/
```