



Published in final edited form as:

Ecology. 2014 August ; 95(8): 2069–2076.

Density dependent state space model for population abundance data with unequal time intervals

Brian Dennis^{1,3} and José Miguel Ponciano²

¹Department of Fish and Wildlife Sciences and Department of Statistical Science, University of Idaho, Moscow ID 83844-1136, USA

²Department of Biology, University of Florida, Gainesville, FL, 32611-8525, USA

Abstract

The Gompertz state-space (GSS) model is a stochastic model for analyzing time series observations of population abundances. The GSS model combines density dependence, environmental process noise, and observation error toward estimating quantities of interest in biological monitoring and population viability analysis. However, existing methods for estimating the model parameters apply only to population data with equal time intervals between observations. In the present paper, we extend the GSS model to data with unequal time intervals, by embedding it within a state-space version of the Ornstein-Uhlenbeck process, a continuous-time model of an equilibrating stochastic system. Maximum likelihood and restricted maximum likelihood calculations for the Ornstein-Uhlenbeck state-space model involve only numerical maximization of an explicit multivariate normal likelihood, and so the extension allows for easy bootstrapping, yielding confidence intervals for model parameters, statistical hypothesis testing of density dependence, and selection among sub-models using information criteria. Ecologists and managers previously drawn to models lacking density dependence or observation error because such models accommodated unequal time intervals (for example, due to missing data) now have an alternative analysis framework incorporating density dependence, process noise and observation error.

Keywords

density dependence; diffusion process; Gompertz model; lognormal distribution; mean-reverting process; Ornstein-Uhlenbeck process; stationary distribution; stochastic differential equation; stochastic population model

INTRODUCTION

Gaps in time series observations of population abundances pose challenges for analysis. Various statistical models for ecological time series data, especially models incorporating realistic population dynamics, require observations spaced at equal time intervals. However, ecological sampling involves constraints of time, personnel, and budgets that do not always

³brian@uidaho.edu.

live up to the designs and requirements of statistical models. As well, various ecological systems such as aquatic or microbial systems (Kirchman 2012) have intrinsically continuous-time dynamics and are sometimes sampled at unequal time intervals. Ordinarily equal time intervals will be stretched and shrunk into unequal intervals in systems in which time is measured in degree days (Metcalf and Luckmann 1994). The data that do exist in studies with missing data or unequal time intervals are potentially informative, and precluding such data from analysis could affect conclusions regarding the biological resources in question. The monitoring and management of biological populations would benefit from having better models for analyzing time series data with unequal time intervals.

Dennis et al. (2006) described a “state-space” population model for use in ecological time series analysis. The model, termed the Gompertz state space (GSS) model, is one of the simplest possible formulations containing density dependence, stochastic process variability, and stochastic observation or measurement error. The simplicity of the model allows for an explicit likelihood function and for parameter estimation through ordinary numerical maximization. Although the GSS model can accommodate missing observations from otherwise equally spaced time series, the GSS model does not allow observations collected at unequal, non-integer time intervals and hence is not suitable for ecological situations in which time is considered to be real-valued.

Various state space population models of greater complexity can handle unequally spaced data (de Valpine and Hastings 2002, de Valpine 2002, 2004, Clark and Bjørnstad 2004, Ionides et al. 2006, Lele et al. 2007, Ponciano et al. 2009). Such models require custom programming of simulation-intensive computer algorithms for fitting (parameter estimation) and other statistical inferences. For example, Lele et al. (2007) and Ponciano et al. (2009) used data cloning (a Monte Carlo Markov Chain method) to calculate maximum likelihood estimates of parameters in state space models for analyzing the well-known laboratory *Paramecium* (sp.) data of Gause (1934) which have missing observations on day 1. Adapting such models to other data sets with other configurations of missing observations would require reprogramming the calculations for each new case.

A special case of the GSS model is a density independent state space model. The exponential growth state space (EGSS) model was introduced by Holmes (2001), and parameter estimation was studied by Lindley (2003) and Staples et al. (2004). By contrast to the GSS model, the EGSS model has been generalized to apply to unequal, real-valued time intervals (Staudenmayer and Buonaccorsi 2006, Humbert et al. 2009).

This paper extends the full, density dependent GSS model to unequal, real-valued time intervals. The method used employs a stochastic version of the continuous-time, deterministic Gompertz growth model as the underlying model of population growth (process model). The stochastic version is a continuous-time diffusion process representing Gompertz-type density dependent growth perturbed by environmental noise. Transforming the process model to the logarithmic scale produces the Ornstein-Uhlenbeck process, a well-studied diffusion process. Finally, adding an observation error component produces a state-space version of the Ornstein-Uhlenbeck model. The state space model has discrete-time statistical properties identical to those of a GSS model, but has a likelihood function that can

be written for time series observations recorded at arbitrary real-valued time intervals. The likelihood makes accessible a variety of statistical inferences based on parametric bootstrapping. Four example data sets serve to illustrate statistical inferences with the model. Procedures for calculating confidence intervals for model parameters and for conducting bootstrap tests of density dependence are described in the online Appendix to this paper, and R programs for all statistical inferences are provided (Supplementary Material online).

POPULATION MODEL

Deterministic Gompertz model

The Gompertz curve (after Gompertz 1825) originally was an actuarial model of mortality, but since the 1920s biologists have been using the curve as a deterministic model of biological growth, variously to describe the growth of tumors, individuals, or populations (Winsor 1932). The model is frequently presented in the form of an ordinary differential equation:

$$\frac{dn(t)}{dt} = \theta n(t) [\log \kappa - \log n(t)]. \quad (1)$$

Here $n(t)$ is the abundance of the growing entity at time t , θ and κ are positive constants, with κ being the equilibrium abundance and θ being a measure of the speed of equilibration. The similar logistic growth model has the density dependence term proportional to $n(t)$ instead of $\log n(t)$. The solution trajectory for Eq. 1 is given by

$$n(t) = \kappa \exp \left[e^{-\theta t} \log (n_0 / \kappa) \right], \quad (2)$$

with $n_0 = n(0)$ denoting the initial population abundance. The Gompertz trajectory has an inflection point at κ/e if the initial population abundance is below that level (inflection for the logistic model is at $k/2$).

Stochastic Gompertz model

A stochastic version of the Gompertz model is represented by the following stochastic differential equation (SDE):

$$dN(t) = \theta N(t) [\log \kappa - \log N(t)] dt + \beta N(t) dW(t). \quad (3)$$

Here an infinitesimal increment of the Gompertz growth model (Eq. 1 written in differential form) is perturbed by a random noise term in which $dW(t)$ has a normal distribution with a mean of 0 and a variance of dt (where the correlation between $dW(t_i)$ and $dW(t_j)$ is assumed equal to 0 if $t_i \neq t_j$), and in which the intensity of the noise is scaled by the term $\beta N(t)$, with $\beta > 0$. The scaling term proportional to population abundance is a common model of environmental stochasticity (Tier and Hanson 1981, Dennis and Patil 1984). Such stochasticity can produce substantial population variability at high as well as low abundances. The assumption of zero-correlated process noise describes the unpredictability of growth increment fluctuations in response to environmental buffeting has theoretical and

empirical utility (Allen 2010, Dennis et al. 1991). Temporal correlation in process noise, if present, is difficult to estimate in the presence of sampling error (Staples et al. 2008).

The SDE as written in the form of Eq. 3 can be understood as an easy recipe for simulating a trajectory of population abundance for specified values of θ , κ , and β (for instance, Higham 2001, Allen 2010): (1) Over a tiny time interval dt , calculate an increment $dN(t)$ of abundance from a normal distribution with a mean of $\theta N(t)[\log \kappa - \log N(t)]dt$ and a variance of $[\beta N(t)]^2 dt$, with $N(t)$ being the current population abundance and t being the current time in the simulation. (2) Update population abundance as $N(t) + dN(t)$ and update time as $t + dt$. (3) Return to step (1) and repeat the process for a new time interval, generating a new normal random growth increment using the updated population abundance. The accuracy of the simulation (in terms of the statistical properties of $N(t)$) is degraded if the time interval dt is not small. While various alternative simulation algorithms can improve the numerical accuracy of SDEs in general (Allen 2010) and the OU process in particular (Gillespie 1996), the preceding “Euler-Maruyama” method for SDEs serves as an accessible pedagogical entry point to a highly mathematized subject.

Ornstein-Uhlenbeck model

Written as an SDE in the form of Eq. 3, population abundance is a type of stochastic process known as a diffusion process with infinitesimal mean function given by $m(n) = \theta n(\log \kappa - \log n)$ and infinitesimal variance function $v(n) = \beta^2 n^2$ (Karlin and Taylor 1981). A useful property of diffusion processes is that a smooth invertible transformation of a diffusion process is also a diffusion process (Karlin and Taylor 1981). Let $X(t) = g(N(t))$ be such a transformation (perhaps $\log N(t)$ or $\sqrt{N(t)}$, etc.). The infinitesimal mean and variance of

$X(t)$ are given by the Itô transformation formulas: $m_x(x) = m(n)g'(n) + \frac{1}{2}v(n)g''(n)$, $v_x(x) = v(n)[g'(n)]^2$, with $n = g^{-1}(x)$. The transformation property opens the possibility that a given diffusion process might be transformed to a scale for which results helpful for analysis and inference can be derived. Such a transformation exists for the stochastic Gompertz model.

The logarithmic transformation, $X(t) = \log N(t)$, converts population abundance $N(t)$ under the stochastic Gompertz model into a well-studied diffusion process known as an Ornstein-Uhlenbeck (OU) process (after Uhlenbeck and Ornstein 1930; see Allen 2010). Under the Itô transformation formulas, the infinitesimal mean and variance functions for $X(t)$ become

$$m_x(x) = \theta(\mu - x), \quad (4)$$

$$v_x(x) = \beta^2, \quad (5)$$

where $\mu = \log \kappa - \beta^2/(2\theta)$. The SDE for $X(t)$ takes the form

$$dX(t) = \theta[\mu - X(t)]dt + \beta dW(t), \quad (6)$$

which is a growth increment $\theta[\mu - X(t)]dt$ from a linear deterministic model perturbed by normally-distributed noise with constant variance $\beta^2 dt$. The SDE given by Eq. 6 defines an OU process $X(t)$. Among the well-known results (see Allen 2010) are that $X(t)$ has a normal transition probability distribution with a mean function given by

$$E[X(t) | X(0) = x_0] = \mu - (\mu - x_0)e^{-\theta t}, \quad (7)$$

where $x_0 = \log n_0$, and a variance function given by

$$V[X(t) | X(0) = x_0] = \frac{\beta^2}{2\theta} (1 - e^{-2\theta t}). \quad (8)$$

The covariance of the process at two different times is given by

$$Cov[X(t), X(t+s)] = \frac{\beta^2}{2\theta} (1 - e^{-2\theta t}) e^{-\theta s}, \quad t, s > 0, \quad (9)$$

and values of the process at multiple different times have a joint multivariate normal distribution. As time t becomes large, the transition distribution converges to a normal stationary distribution with mean

$$E[X(\infty)] = \mu, \quad (10)$$

and variance

$$V[X(\infty)] = \frac{\beta^2}{2\theta}. \quad (11)$$

The stationary distribution on the original abundance scale is a lognormal distribution and represents the stochastic counterpart to the point equilibrium or “carrying capacity” in the deterministic model (Dennis and Patil 1984).

Ornstein-Uhlenbeck state space model

We define an Ornstein-Uhlenbeck state space (OUSS) model by adding an observation error component. The observations can be taken at arbitrary times which need not be equally spaced. The observed population log-abundance at time t_i is modeled as

$$Y(t_i) = X(t_i) + F_i, \quad (12)$$

where F_i has a normal distribution with mean 0 and variance τ^2 , with F_i, F_j uncorrelated ($i \neq j$). Observation error on the original abundance scale thereby has a lognormal distribution, which can be regarded as a model of ecological sampling under heterogeneous sampling conditions (Dennis et al. 2006). Eqs. 6 and 12 together constitute the OUSS model. The underlying actual population log-abundance $X(t_i)$ is unobserved.

The observations $Y(0), Y(t_1), Y(t_2), \dots, Y(t_q)$ (estimates or indexes of log-population abundance) recorded at times $t_0 = 0, t_1, \dots, t_q$ need not be equally spaced. The observations under the OUSS model have a joint multivariate normal distribution, but the distribution

takes two different forms, non-stationary and stationary, depending on the situation being modeled. If the population abundances have commenced far from equilibrium, the observations have the non-stationary distribution in which the mean of each $Y(t_i)$ is

$$E[Y(t_i)] = \mu - (\mu - x_0) e^{-\theta t_i}, \quad (13)$$

the variance is

$$V[Y(t_i)] = \frac{\beta^2}{2\theta} (1 - e^{-2\theta t_i}) + \tau^2, \quad (14)$$

and the covariance of the process at two different times t_i and t_j is given by

$$Cov[Y(t_i), Y(t_j)] = \frac{\beta^2}{2\theta} (1 - e^{-2\theta \min(t_i, t_j)}) e^{-\theta |t_i - t_j|}. \quad (15)$$

If however the population abundances have become stationary (fluctuating stochastically around an equilibrium) before sampling commenced, the observations have the stationary distribution in which the mean of each $Y(t_i)$ is

$$E[Y(t_i)] = \mu, \quad (16)$$

the variance is

$$V[Y(t_i)] = \frac{\beta^2}{2\theta} + \tau^2, \quad (17)$$

and the covariance of the process at two different times t_i and t_j is given by

$$Cov[Y(t_i), Y(t_j)] = \frac{\beta^2}{2\theta} e^{-\theta |t_i - t_j|}. \quad (18)$$

In either the stationary or non-stationary case the joint distribution of the observations is similar to that of the underlying OU process, except that the quantity τ^2 is added to the variance of each $Y(t_i)$. Two different methods for fitting the model to data (Statistical Methods, below) arise from whether or not the population has become stationary.

The OUSS model can be regarded as a continuous time version of the GSS model. The GSS model is defined (in the notation of Dennis et al. 2006) by $X_t = a + cX_{t-1} + E_t$, $Y_t = X_t + F_t$, where X_t and Y_t are respectively the underlying and observed population log-abundances at times $t = 0, 1, 2, \dots, q$, and E_t and F_t have independent normal distributions with means of 0 and respective variances of σ^2 and τ^2 . The parameters a , σ^2 , and τ^2 are positive, and $-1 < c < +1$. If $0 < c < 1$, the transition probability distributions of the OUSS and GSS models evaluated at integer times coincide as identical normal distributions means and variances given by Eqs. 13 and 14, with the relationships between parameters given by the following transformations: $a = \mu(1 - e^{-\theta})$, $c = e^{-\theta}$, $\sigma^2 = (1 - e^{-2\theta})\beta^2/(2\theta)$; $\mu = a/(1 - c)$, $\theta = -\ln(c)$, $\gamma^2 = -[2\sigma^2 \ln(c)]/(1 - c^2)$, with τ^2 having the same value in both models. If $-1 < c < 0$ the, GSS model has a deterministic component with damped oscillatory behavior for which there is no

straightforward continuous time version. However, oscillatory behavior in population data, damped or otherwise, suggests the presence of nonlinear dynamic forces (May 1994), and one should explore alternative model structures with additional state variables and/or nonlinear interactions.

STATISTICAL METHODS

Maximum likelihood estimation for the OUSS stationary case

Under stationarity, the initial log-population abundance, $X(0)$, is assumed to arise from the stationary normal distribution of the population abundances. Such an assumption might be plausible if the monitoring program commenced after the population had been existing for a time long enough for stochastic equilibration in the particular environment. The model then has four unknown parameters: θ , μ , β^2 , and τ^2 .

Fitting the model to data means estimating the unknown model parameters. Using the multivariate normal distribution, the model can be fitted to data with maximum likelihood (ML) estimation. For ML estimation with the OUSS model, the likelihood function is the joint multivariate normal probability density for the observations evaluated at their realized data values. The ML estimates then are the parameter values that jointly maximize the likelihood or log-likelihood function. The multivariate normal log-likelihood for the stationary OUSS model is given by

$$\log L(\mu, \theta, \beta^2, \tau^2) = -\frac{(q+1)}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{V}|) - \frac{1}{2} (\mathbf{y} - \mathbf{m})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{m}), \quad (19)$$

where the elements of the column vector \mathbf{y} are the data values $y_0, y_1, y_2, \dots, y_q$ (realized values of the random variables $Y(0), Y(t_1), Y(t_2), \dots, Y(t_q)$) recorded at times $0, t_1, t_2, \dots, t_q$. Also, the column vector \mathbf{m} has all $q+1$ elements identically equal to μ (Eq. 16), the variances on the main diagonal of the variance-covariance matrix \mathbf{V} are all equal to $\tau^2 + [\beta^2/(2\theta)]$ (Eq. 17), and the (i, j) th and (j, i) th covariance elements in \mathbf{V} are both equal to $[\beta^2/(2\theta)]e^{-\theta S_{ij}}$, where $S_{ij} = |t_i - t_j|$ (Eq. 18).

The ML estimates must be calculated with numerical optimization (for instance, with the “optim” function in R; R Core Development Team 2006). An R program for such calculation is provided in the Supplementary Material online.

Restricted ML estimation for the stationary case

Restricted maximum likelihood (REML) estimation for the multivariate normal distribution uses a linear combination of the observations that eliminates the parameters in the mean vector, leaving only parameters in the variance-covariance matrix. The statistical properties of the resulting estimates for the GSS and EGSS models are frequently better than those of ML estimates (Dennis et al. 2006, Humbert et al. 2009). For the OUSS model under the stationary case, a linear combination that produces a zero-mean vector is the simple differencing of the observations. Let $W_i = Y(t_i) - Y(t_{i-1})$, $i = 1, 2, \dots, q$. The W_i 's have a joint multivariate normal distribution with a mean vector of all zeros and a variance-covariance matrix given by

$$\Phi = \mathbf{D}\mathbf{V}\mathbf{D}', \quad (20)$$

where \mathbf{D} is the $q \times (q+1)$ differencing matrix given by

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & 0 \cdots \\ 0 & -1 & 1 & 0 \cdots \\ 0 & 0 & -1 & 1 \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (21)$$

For REML estimation, the multivariate normal log-likelihood is a function of three of the parameters in the OUSS model:

$$\log L(\theta, \beta^2, \tau^2) = -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log(|\Phi|) - \frac{1}{2} \mathbf{w}' \Phi^{-1} \mathbf{w}. \quad (22)$$

Here the column vector of differenced observations is $\mathbf{w} = \mathbf{D}\mathbf{y}$. After numerical maximization of the REML log-likelihood, the mean parameter μ is estimated by

$$\hat{\mu} = \mathbf{j}' \hat{\mathbf{V}}^{-1} \mathbf{y} / (\mathbf{j}' \hat{\mathbf{V}}^{-1} \mathbf{j}), \quad (23)$$

where $\hat{\mathbf{V}}$ is the variance-covariance matrix \mathbf{V} (for the undifferenced observations, Eq. 19) evaluated with the REML estimates of θ , β^2 , and τ^2 , and \mathbf{j} is a $(q+1) \times 1$ column vector of ones. The R program for parameter estimation accompanying this paper (Supplementary Material online) calculates the REML estimates in addition to the ML estimates.

ML estimation for the nonstationary case

The nonstationary OUSS model should be considered for use when an appreciable proportion of the observations are recorded during a transient growth phase, as for instance when a population displays a sigmoid trajectory instead of a trajectory solely fluctuating around a steady state. Examples of such situations might occur when monitoring the translocation of a species by release of a few individuals into the wild, the colonization of a new area by an invading species, the increase of an agricultural pest through a growing season starting from low density, the recovery of a species after a catastrophic decline, or the growth of microbial populations in experimental laboratory cultures (Ponciano et al. 2005). The initial population log-abundance χ_0 usually becomes an additional unknown parameter in the nonstationary model, unless its value is known without sampling error (such as might be the case for a species translocation or a laboratory inoculation).

ML estimation is available for the nonstationary case, but REML estimation is not available. ML estimation uses the multivariate normal log-likelihood (Eq. 19) for the data values $y_0, y_1, y_2, \dots, y_q$, but with a different mean vector and variance-covariance matrix. The element in the mean vector \mathbf{m} corresponding to the data value y_i is given by Eq. 13, the variances on the main diagonal of \mathbf{V} are computed with Eq. 14, and the (i, j) th and (j, i) th covariance elements in \mathbf{V} are obtained with Eq. 15. The log-likelihood must be numerically maximized jointly for the five unknown parameters $\chi_0, \theta, \mu, \beta^2$, and τ^2 .

If the initial log-abundance χ_0 is known, the log-likelihood has just the four unknown θ , μ , β^2 , and τ^2 , with χ_0 fixed at its known value. Also, when χ_0 is known the initial observation y_0 is equal to χ_0 and is omitted from the likelihood (that is, omitted from the data vector \mathbf{y}). ML estimation uses the multivariate normal log-likelihood (Eq. 19) modified to apply to just the q observations y_1, y_2, \dots, y_q recorded at times t_1, t_2, \dots, t_q . The modifications consist of substituting a $q \times 1$ vector \mathbf{m} containing means defined by Eq. 13, a $q \times q$ variance-covariance matrix \mathbf{V} containing elements defined by Eqs. 14 and 15, and a leading term $-\frac{q}{2} \log(2\pi)$ instead of $-\frac{q+1}{2} \log(2\pi)$. The R program for parameter estimation accompanying this paper also optionally calculates estimates for the non-stationary case (Supplementary Material online).

Additional statistical inferences for the OUSS model include confidence intervals for parameters, hypothesis tests for density dependence, and model selection among submodels. The statistical justifications and calculations involved in the additional inferences are described in the Appendix.

EXAMPLES

Bobcat (*Lynx rufus*) abundances in Idaho seemingly held steady, albeit with substantial fluctuations, over a 25 yr period beginning in 1956 (Figure 1, upper left). Bobcat abundances in Maine displayed fluctuations of even greater magnitude (Figure 1, upper right). The bobcat data are furbearer harvest records listed in the Global Population Dynamics Database (NERC Centre for Population Biology 2010), data sets 212 (Idaho) and 216 (Maine). As an index of population abundance, fur harvest records can be assumed to arise from a variable and heterogeneous sampling process, and any models of such time series should incorporate an appropriate sampling component. In Idaho, an unexplained hiatus in the data occurs during the years 1966-69, while in the Maine data the years 1938-41, 1967 and 1969 are missing. Mark-recapture estimates of the elk (*Cervus elaphus*) population in the central valley of Grand Teton National Park, Wyoming, are missing the year 1983 (Figure 1, lower left; data reported by Boyce 1989 as used by Dennis and Taper 1994). Yearly sampling of adult rangeland grasshopper (Orthoptera: Acrididae, mixed species) abundances in Montana began in 1948, with the years 1949-50, 1977, and 1982 missing (Figure 1, lower right; data reported by Kemp and Dennis 1993). The gaps are the main reason for the selection of the particular data sets here to illustrate the OUSS model. All the data sets in Figure 1 are among those included in the R programs accompanying this paper (Supplementary Material online).

The parameter μ tends to be estimated with precision, according to the REML bootstrap confidence intervals (Figure 1). The parameters θ , β^2 and τ^2 are estimated well for these data, as suggested by the wide confidence intervals (Figure 1). The GSS model, and by implication the OUSS model, sometimes has ridge-like likelihoods, or likelihoods with multiple local maxima corresponding to sub-models with all observation error or all process noise (Dennis et al. 2006, 2011, Knappe 2008), especially for short time series. Replicating the sampling process one or more times at selected sampling times can substantially improve parameter estimation (Dennis et al. 2011), as can substantial amounts of transience in the

data (nonstationarity) resulting from initiating the time series far from equilibrium (Dennis and Taper 1994).

DISCUSSION

If unequal sampling intervals are due simply to data missing from otherwise equally spaced observations (such as the data in Figure 1), then the ordinary GSS model can in fact be used with modification. The GSS model, like the OUSS model, has observations with a multivariate normal log-likelihood (Eq. 19). To calculate ML estimates for the GSS model, the missing observations in question are deleted from the vector \mathbf{y} , the entries corresponding to the missing observations are deleted from \mathbf{m} , and the rows and columns corresponding to the missing observations are deleted from \mathbf{V} . The resulting expression is the proper log-likelihood function for the multivariate normal distribution of the remaining observations, under a standard property of the multivariate normal distribution (Seber 1984). While deleting observations from the GSS is easy in principal, programming software to automate the deletions in order to analyze many data sets would not be straightforward.

An additional problem with missing observations occurs for restricted maximum likelihood (REML) estimation. For the GSS model, REML estimates are calculated from the likelihood function for the first differences (Dennis et al. 2006). For the EGSS model, REML estimates are calculated from the likelihood function for the second differences (Staples et al. 2004). The problem is that each missing observation removes two observations from REML estimation in the GSS model and three observations from REML estimation in the EGSS model. REML estimation with missing data thus becomes costly, in terms of information about parameters, for the low sample sizes common in ecological time series.

Also, the GSS model cannot be used for systems sampled naturally at unequal time intervals. Plankton sampling schedules in aquatic systems, for instance, might not adhere to rigidly equal intervals. In insect ecology, degree days are routinely used as a time scale, and the application of a degree-day transformation to sampling times with equal intervals invariably produces unequal intervals on the new scale. In such cases, the unequal intervals cannot be handled merely by dropping observations from the multivariate normal log-likelihood function. A solution instead is to use the OUSS model.

Alternative ways for accommodating unequal intervals in state space models are computationally challenging. The basic approach treats the underlying population abundance as an unobserved “latent variable” in a hierarchical statistical model. Parameters are estimated with intensive MCMC simulation, coupled with either a Bayesian (Clark and Bjornstadt 2004) or frequentist (Lele et al. 2007) inferential method. While the approach allows the use of more complex ecological models of population growth and/or sampling, the problems of parameter estimability are not made any easier. Recent developments in estimation diagnostics (Lele et al. 2010, Ponciano et al. 2012) at least are helping investigators recognize when problems exist. As well, the MCMC calculations require considerable expertise in programming and statistical theory. By contrast, the OUSS model presented here brings density dependence, process noise, and observation error within reach of ordinary numerical maximization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

BD was supported by US Department of Defense SERDP project SI-1477 and National Marine Fisheries Service contract AB133F-06-SE-5682. JMP was supported by the National Institute of General Medical Sciences of the National Institutes of Health, under grant number 1R01GM103604-01.

LITERATURE CITED

- Allen, LJS. An introduction to stochastic processes with applications to biology. second edition. CRC Press; Boca Raton, Florida, USA: 2010.
- Boyce, MS. The Jackson elk herd: intensive management in North America. Cambridge University Press; Cambridge, UK: 1989.
- Clark JS, Bjørnstad ON. Population time series: process variability, observation errors, missing values, lags, and hidden states. *Ecology*. 2004; 85:3140–3150.
- Dennis B, Munholland PL, Scott JM. Estimating growth and extinction parameters for endangered species. *Ecological Monographs*. 1991; 61:115–143.
- Dennis B, Patil GP. The gamma distribution and weighted multimodal gamma distributions as models of population abundance. *Mathematical Biosciences*. 1984; 68:187–212.
- Dennis B, Ponciano JM, Lele SR, Taper ML, Staples DF. Estimating density dependence, process noise, and observation error. *Ecological Monographs*. 2006; 76:323–341.
- Dennis B, Ponciano JM, Taper ML. Replicated sampling increases efficiency in monitoring biological populations. *Ecology*. 2011; 91:610–620. [PubMed: 20392025]
- Dennis B, Taper ML. Density dependence in time series observations of natural populations: estimation and testing. *Ecological Monographs*. 1994; 64:205–224.
- de Valpine P. 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*. 2002; 70:455–471.
- de Valpine P. Monte Carlo state-space likelihoods by weighted posterior kernel density estimation. *Journal of the American Statistical Association*. 2004; 99:523–536.
- de Valpine P, Hastings A. Fitting population models with process noise and observation error. *Ecological Monographs*. 2002; 72:57–76.
- Gause, GF. *The struggle for existence*. Williams & Wilkins; Baltimore, Maryland, USA: 1934.
- Gillespie DT. Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Physical Review E*. 1996; 54:2084–2091.
- Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*. 1825; 115:513–583.
- Higham DJ. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*. 2001; 43:525–546.
- Holmes EE. Estimating risks in declining populations with poor data. *Proceedings of the National Academy of Sciences (USA)*. 2001; 98:5072–5077.
- Humbert J-Y, Mills LS, Horne JS, Dennis B. A better way to estimate population trend. *Oikos*. 2009; 118:1940–1946.
- Ionides EL, Breto C, King AA. Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences (USA)*. 2006; 103:18438–18443.
- Karlin, S.; Taylor, HM. *A second course in stochastic processes*. Academic Press; San Diego, California, USA: 1981.
- Kemp WP, Dennis B. Density dependence in rangeland grasshoppers (Orthoptera: Acrididae). *Oecologia*. 1993; 96:1–8.

- Kirchman, DL. Processes in microbial ecology. Oxford University Press; Oxford, UK: 2012.
- Knape J. Estimability of density dependence in models of time series data. *Ecology*. 2008; 89:2994–3000.
- Lele SR, Dennis B, Lutscher F. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*. 2007; 10:551–563. [PubMed: 17542934]
- Lele SR, Nadeem K, Schmuland B. Estimability and likelihood for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*. 2010; 105:1617–1625.
- Lindley ST. Estimation of population growth and extinction parameters from noisy data. *Ecological Applications*. 2003; 13:806–813.
- May, RM. Stability and complexity in model ecosystems. 2nd edition. Princeton University Press; Princeton, New Jersey, USA: 1974.
- NERC Centre for Population Biology, Imperial College. The Global Population Dynamics Database Version 2. 2010. <http://www3.imperial.ac.uk/cpb/databases/gpdd>
- Metcalf, RL.; Luckmann, WH., editors. Introduction to insect pest management. third edition. John Wiley; New York, New York, USA: 1994.
- Ponciano JM, Vandecasteele F, Hess T, Forney LJ, Crawford RL, Joyce P. A stochastic approach to model the effect of environmental factors on bacterial growth. *Applied and Environmental Microbiology*. 2005; 71:2355–2364. [PubMed: 15870322]
- Ponciano JM, Taper ML, Dennis B, Lele SR. Hierarchical models in ecology: confidence intervals, hypothesis testing and model selection using data cloning. *Ecology*. 2009; 90:356–362. [PubMed: 19323219]
- Ponciano JM, Burleigh G, Braun E, Taper ML. Assessing parameter identifiability in phylogenetic models using Data Cloning. *Systematic Biology*. 2012; 61:955–972. [PubMed: 22649181]
- R Core Development Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2006.
- Seber, GAF. Multivariate observations. John Wiley; New York, New York, USA: 1984.
- Staples DF, Taper ML, Dennis B. Estimating population trend and process variation for PVA in the presence of sampling error. *Ecology*. 2004; 85:923–929.
- Staples DF, Taper ML, Dennis B, Boik RJ. Effects of sampling error and temporal correlations in population growth on process variance estimators. *Environmental and Ecological Statistics* (online). 2008 DOI 10.1007/s10651-008-0097-5.
- Staudenmayer J, Buonaccorsi JP. Measurement error in a random walk model with applications to population dynamics. *Biometrics*. 2006; 62:1178–1189. [PubMed: 17156293]
- Tier C, Hanson FB. Persistence in density dependent stochastic populations. *Mathematical Biosciences*. 1981; 53:89–117.
- Uhlenbeck GE, Ornstein LS. On the theory of Brownian Motion. *Physical Review*. 1930; 36:823–841.
- Winsor CP. The Gompertz curve as a growth curve. *Proceedings of the National Academy of Sciences (USA)*. 1932; 18:1–8.

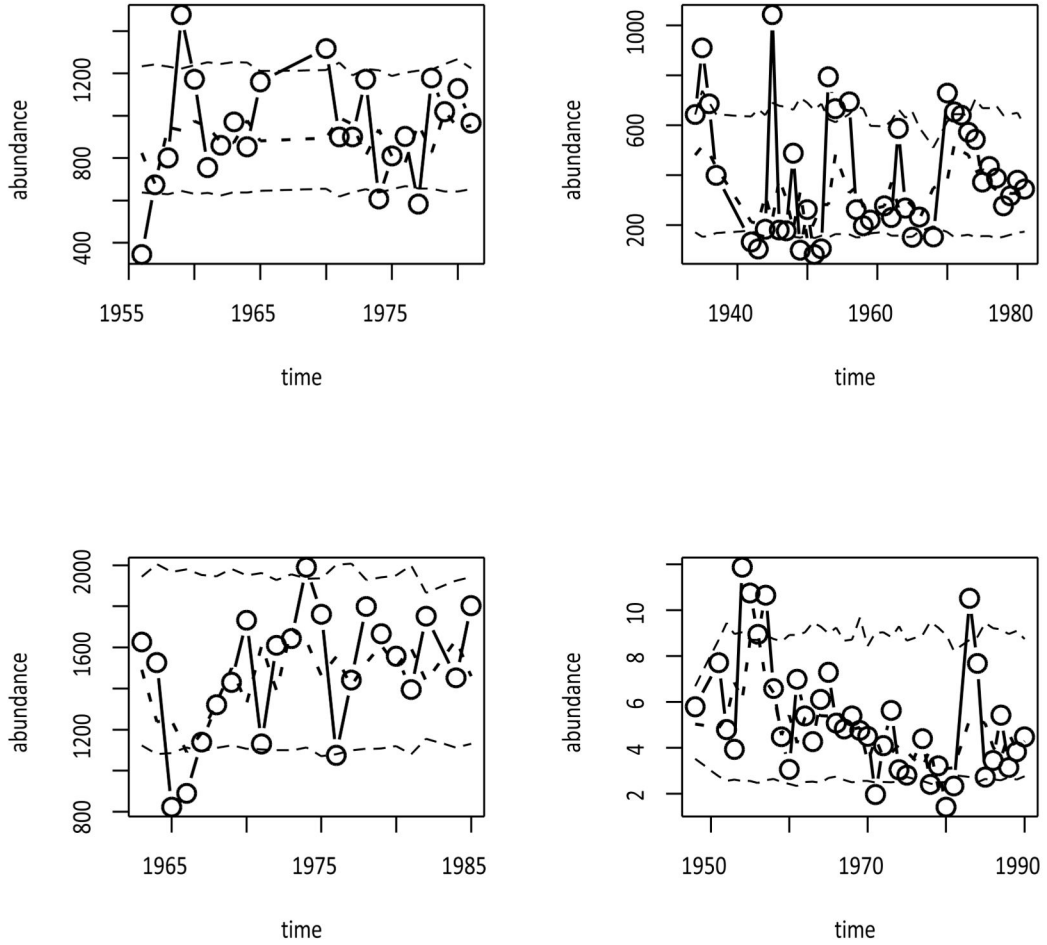


Figure 1.

Four population abundance time series (circles), with expected values of true abundances given the other observations (short dashed lines), and bootstrapped 95% confidence intervals for true abundances (long dashed lines), estimated with restricted maximum likelihood under the Ornstein-Uhlenbeck state space model. Upper left: bobcat furbearer harvest records from Idaho, USA (parameter estimates with bootstrapped 95% confidence intervals:

$$\hat{\mu}=6.79 (6.61, 6.97), \hat{\theta}=1.26 (3.91 \times 10^{-7}, 20.2), \hat{\beta}^2=0.272 (7.43 \times 10^{-3}, 3.94),$$

$$\hat{\tau}^2=7.48 \times 10^{-4} (9.14 \times 10^{-10}, 0.0847)).$$

Upper right: bobcat furbearer harvest records from Maine, USA ($\hat{\mu}=5.78 (5.47, 6.08)$, $\hat{\theta}=0.877 (0.0226, 10.3)$, $\hat{\beta}^2=0.735 (0.0202, 2.40)$,

$$\hat{\tau}^2=0.00475 (1.79 \times 10^{-8}, 0.334)).$$

Lower left: elk population estimates from Grand Teton National Park, Wyoming, USA ($\hat{\mu}=7.29 (7.14, 7.44)$, $\hat{\theta}=0.868 (0.229, 19.3)$,

$$\hat{\beta}^2=0.0990 (0.0296, 1.45), \hat{\tau}^2=9.80 \times 10^{-9} (2.93 \times 10^{-11}, 1, 22 \times 10^{-3})).$$

Lower right: grasshopper density estimates from the western mountainous region of Montana, USA

$$(\hat{\mu}=1.56 (1.31, 1.82), \hat{\theta}=0.722 (0.272, 2.01), \hat{\beta}^2=0.347 (0.160, 0.752), \\ \hat{\tau}^2=2.27 \times 10^{-7} (3.81 \times 10^{-8}, 2.27 \times 10^{-5})).$$