SAMPLING VARIABILITY AND ESTIMATES OF DENSITY DEPENDENCE: A COMPOSITE-LIKELIHOOD APPROACH

Subhash R. Lele[1]

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta T6G 2G1 Canada

Abstract. It is well known that sampling variability, if not properly taken into account, affects various ecologically important analyses. Statistical inference for stochastic population dynamics models is difficult when, in addition to the process error, there is also sampling error. The standard maximum-likelihood approach suffers from large computational burden. In this paper, I discuss an application of the composite-likelihood method for estimation of the parameters of the Gompertz model in the presence of sampling variability. The main advantage of the method of composite likelihood is that it reduces the computational burden substantially with little loss of statistical efficiency. Missing observations are a common problem with many ecological time series. The method of composite likelihood can accommodate missing observations in a straightforward fashion. Environmental conditions also affect the parameters of stochastic population dynamics models. This method is shown to handle such nonstationary population dynamics processes as well. Many ecological time series are short, and statistical inferences based on such short time series tend to be less precise. However, spatial replications of short time series provide an opportunity to increase the effective sample size. Application of likelihood-based methods for spatial time-series data for population dynamics models is computationally prohibitive. The method of composite likelihood is shown to have significantly less computational burden, making it possible to analyze large spatial time-series data. After discussing the methodology in general terms, I illustrate its use by analyzing a time series of counts of American Redstart (*Setophaga ruticilla*) from the Breeding Bird Survey data, San Joaquin kit fox (*Vulpes macrotis mutica*) population abundance data, and spatial time series of Bull trout (*Salvelinus confluentus*) redds count data.

Key words: American Redstart; Breeding Bird Survey; bull trout; covariates in population dynamics; estimating functions; Gompertz model; hierarchical models; nonstationary population dynamics; San Joaquin kit fox; spatial time series; state-space models.

INTRODUCTION

Stochastic population dynamics models constitute an important component of applied and theoretical ecology. When analyzing population time-series data, DeValpine (2002) makes a strong case for using such mechanism-oriented models as compared to simple statistical regression models. These models are also used extensively in population viability analysis (Morris et al. 1999) with important implications in public-policy decision making. Statistical inference for population dynamics models has had a long history. Dennis et al. (1991) provide a review of this literature and an introduction to the method of maximum likelihood for estimation of parameters of population dynamics models. However, the approach used in Dennis et al. (1991) does not explicitly consider sampling variability. As argued by various researchers, not taking into account sampling variability can have significant impact on the statistical and scientific inferences (Shenk et al. 1998). Recently there have been several papers that attempt

to deal with the problem of parameter estimation of stochastic population dynamics models in the presence of sampling error (Holmes 2001, DeValpine 2002, DeValpine and Hastings 2002, Holmes and Fagan 2002, Staples et al. 2004; B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples, *unpublished manuscript*). These papers are based on the classical frequentist approach. On the other hand, Meyer and Millar (1999) suggest the use of a Bayesian approach to this problem. See also Carlin et al. (1992) for a Bayesian approach to statistical analysis of nonlinear, non-Gaussian state-space models. For recent ecological applications of the Markov-Chain Monte Carlo (MCMC)-based Bayesian approach to statistical inference for complex models in ecology, see Calder et al. (2003), Clark and Bjornstadt (2004), and references in both. However, following DeValpine and Hastings (2002) and B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples (*unpublished manuscript*), I here consider the frequentist approach to the problem. Since a thorough review of recent literature on the issue of incorporating sampling variability from a frequentist viewpoint is available in DeValpine (2002) and B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples (*unpublished manuscript*), instead of re-

[1] E-mail: slele@ualberta.ca

reviewing the literature I only discuss, in brief, the salient and distinguishing features of the problem and the approaches proposed in the literature so far. There are two components to this problem:

Component 1.—This consists of the stochastic population dynamics model that describes the relationship between population sizes from one time point to the next. Such models typically provide the one-step transition distributions. Given these transition distributions and the equilibrium distribution $f(N_1; \theta)$, one can write down the likelihood function, assuming we have T time points, in a straightforward fashion as

$$L(\theta \,|\, N) = f(N_1; \theta)f(N_2\,|\,N_1; \theta) \cdots f(N_T\,|\,N_{T-1}; \theta) \quad (1)$$

where θ indicates the parameters in the model.

Component 2.—In reality, one does not have the exact knowledge of N_t, the true population size. One may only have an *estimate* of the population size, denoted by $\hat{N}_t$. These estimates may be obtained through various sampling procedures such as capture–recapture methods or counts from pheromone traps (Seber 2002). Different sampling procedures lead to different sampling-error distributions. In general, these distributions might contain parameters of their own for taking into account observer effects and other covariates such as weather conditions that might affect the accuracy of the estimator. Let us assume that the form of the sampling distribution is known, based on the sampling procedure used. It is also reasonable to assume that conditionally on the true states, sampling errors from one time point to the other are independent of each other.

Now consider the likelihood function based on these *available* data, namely, the estimated population sizes $\hat{N}_t$. One cannot simply substitute $\hat{N}_t$ in place of N_t in the likelihood function in Eq. 1 because one has to explicitly account for the sampling error. The likelihood function that incorporates sampling error, however, can be written in a general form as

$$L(\theta, \phi \,|\, \hat{N})$$

$$= \int\int \cdots \int P(\hat{N}_1\,|\,N_1, \phi)P(\hat{N}_2\,|\,N_2, \phi) \cdots P(\hat{N}_T\,|\,N_T, \phi)$$

$$f(N_1, N_2, \ldots, N_T; \theta) \, dN_1 \cdots dN_T.$$

Here $P(\hat{N}_t\,|\,N_t, \phi)$ indicates the probability mass function or probability density function of the sampling-error distribution. Further, $f(N_1, N_2, \ldots, N_T; \theta)$ indicates the joint distribution of the original but unobserved time series of population sizes. This joint distribution is the same in form as the likelihood function given in Eq. 1.

Notice that the likelihood function that takes into account the sampling error involves T-dimensional integration. This is far more complicated than the likelihood function without sampling error that was presented in Eq. 1. This likelihood function becomes further complicated if each N_t is a vector-valued obser-

vation. Such is the case when one considers spatial time series models (Dennis et al. 1998, Lele et al. 1998) where an observation at a given time point consists of population estimates at S spatial locations or the community-structure models considered in Ives et al. (2003) where the observation consists of population estimates of S species in the community. In such cases, the likelihood function involves $S \times T$ dimensional integration where S corresponds to the number of spatial locations or number of species. Potentially this can be an extremely high-dimensional integral. Likelihood functions for the models that involve age or stage-structured populations (DeValpine 2004) or models that involve dispersal (Lele et al. 1998) are far too complicated to describe here.

As shown by Dennis et al. (1991), if the actual population sizes N_t are available and one can write down the likelihood function in Eq. 1, then it is relatively easy to obtain parameter estimates using the maximum-likelihood method of estimation. However, writing down the likelihood function and hence estimation by maximizing it is substantially difficult if only estimates of the population sizes are available. In some well-structured problems, the likelihood function in the presence of sampling error can be written down explicitly. B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples (*unpublished manuscript*), Holmes (2000), and Staples et al. (2004) exploit the structure of the Gompertz and exponential-growth models and the lognormal sampling-error distribution to explicitly write down the likelihood in the presence of sampling error. Even when such a likelihood function can be written down explicitly, as reported by B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples (*unpublished manuscript*), maximizing such a function can be a tricky task involving multiple modes and such.

When computing the likelihood function for less structured problems, one is confronted with the task of high-dimensional integration. Because of the high dimensionality, standard numerical-integration techniques are not practical. One way out of this difficulty is found in the form of Monte Carlo integration. But if the dimension of the integral is high, Monte Carlo integration is computer intensive due to the "curse of dimensionality" (Robert and Casella 1999:18). Thus brute-force integration, numerical or Monte Carlo, does not seem to be an effective answer to the problem. Kitagawa's algorithm (Kitagawa 1987) for nonlinear non-Gaussian time series, adapted by DeValpine and Hastings (2002) for ecological models, uses an ingenious combination of the structure of the problem, namely the first-order Markov property, and repeated use of one-dimensional numerical integration to reduce the computational burden. However, these structural advantages quickly disappear for vector time-series data such as those occurring in spatial-time series. For example, for spatial time-series data with S spatial locations and T time points, the application of Kitagawa's

algorithm involves repeated evaluations of several 2S-dimensional integrals and S-dimensional integrals. This is practically impossible if we have more than two spatial locations. In essence, one cannot use Kitagawa's algorithm for spatial time-series analyses. Simple introduction of covariates in the population dynamics models (Lele et al. 1998, Dennis and Otten 2000) makes the underlying population dynamics model nonstationary. These models are also computationally difficult to handle using the existing methods. Thus, there is still a need for a computationally simple statistical-inference method for stochastic population dynamics models in the presence of sampling error.

When the likelihood function is difficult to write or involves computationally difficult problems, the method of composite likelihood has been effectively used in the past (Besag 1975, Lindsay 1988, Lele 1997, Heagerty and Lele 1998, Lele and Taper 2002). The main attraction of the composite-likelihood method is its computational and pedagogical simplicity. Previous studies (Lele and Taper 2002, Henderson and Shimakura 2003, Zhao and Joe 2005) suggest that relative loss of efficiency as compared to full-fledged likelihood analysis is likely to be small. When the sample sizes are large or the model is complex, the statistically fully efficient, but computationally difficult, method of maximum-likelihood estimation may have to be abandoned in favor of a somewhat statistically less efficient but computationally feasible method of estimation such as the method of composite likelihood. This paper presents an application of the method of composite likelihood in the context of population dynamics models with sampling error.

The outline of the paper is as follows. I introduce the concept of composite likelihood; next I consider the stationary Gompertz model with Poisson errors in detail, and then I apply the methodology to reanalyze the American Redstart counts data reported in B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples (*unpublished manuscript*). The next section discusses application of the method to an important case of the nonstationary Gompertz model where the intrinsic growth rate is affected by environmental conditions. Currently available methods either are not applicable in this situation or have not been studied in this context. To illustrate the method, I reanalyze the San Joaquin kit fox population time-series data with amount of rainfall as an environmental factor affecting population growth. The next section addresses spatial time-series data. This extension allows analysis of a commonly occurring situation in ecology: availability of spatial replications of short time series (Dennis et al. 1998, Lele et al. 1998). I illustrate the method using the bull trout data reported in Dennis et al. (1998). Finally, I summarize the results and discuss further extensions.

A Brief Introduction to Composite-Likelihood Methods

A fairly elementary introduction to composite likelihood is available in Curriero and Lele (1999) and Lele and Taper (2002). Here I provide a brief description that will be sufficient for the application discussed in this paper. Let the joint distribution of the observations be given by $f(y_1, y_2, \ldots, y_n; \theta)$ where θ denotes the parameters. Then the full likelihood function for θ can be written as $L(\theta \mid y) = f(y_1, y_2, \ldots, y_n; \theta)$. However, suppose writing down the joint distribution of all the observations is difficult but one can write down the joint distribution of pairs of observations as $f(y_i, y_j; \theta)$ comparatively easily. Then, under regularity conditions it can be shown, using the theory of estimating functions (Godambe 1991), that the estimators obtained by maximizing $CL(\theta \mid y) = \Pi_{i=1}^{n-1} \Pi_{j=i+1}^{n} f(y_i, y_j; \theta)$ are consistent and asymptotically normal (Heagerty and Lele 1998). Thus, when computation of the joint distribution of all observations is difficult, but computation of the distributions of *pairs* of observations is relatively easy, then this provides a *feasible* method of estimation. This particular idea of considering pairs of observations has been used in various situations (Heagerty and Lele 1998, Lele and Taper 2002, Zhao and Joe 2005). In fact, one does not have to use only pairs of observations. If one-dimensional marginal distributions are easy to obtain, then one can consider the product of one-dimensional marginal distributions to construct a composite likelihood $CL_1(\theta \mid y) = \Pi_{i=1}^{n} f(y_i; \theta)$. In Markov random-field models, Besag (1975) notes that the joint distribution is difficult to write but that it is easier to specify the conditional distributions of the form $f(y_i \mid y_j, j \in N(i); \theta)$ where $N(i)$ denotes the neighborhood of the ith location. Using this fact, he considered $CL(\theta \mid y) = \Pi_{i=1}^{n} f(y_i \mid y_j, j \in N(i); \theta)$ and called it ''pseudo-likelihood.''

Notice the commonality in all of the above methods. Each one considers an *objective function* that is a product of functions such that each *component* function in the product is itself a proper likelihood (based only on the joint distribution of two observations or one observation or conditional distribution, etc.). Thus, Lindsay (1988) calls such an objective function ''composite likelihood.'' The method of composite likelihood is useful when writing the full likelihood is difficult but lower-dimensional distributions can be written somewhat easily. Simply taking the product of these lower-dimensional distributions provides an objective function that, when maximized, leads to reasonable estimators of the parameters.

The justification for the composite-likelihood method is not based on the quality of its approximation to the likelihood but through the estimating functions that it generates. The resultant estimating functions are zero unbiased and form a mixing sequence. Such estimating functions yield consistent and asymptotically normal

estimators. Thus, the main justification of the method of composite likelihood is that it is a computationally simple method that yields consistent, asymptotically normal estimators of the parameters. For this method, no general results about the loss of efficiency are available. However, past experience (Lele and Taper 2002, Henderson and Shimakura 2003) suggests that the loss of efficiency could be minimal. I suggest that this method be used only when the standard method of maximum likelihood is computationally or otherwise difficult to implement. This method is not a panacea. One has to be careful to first prove that the parameters are estimable using the marginal or conditional distributions. For example, consider the case of simple Gaussian auto-regressive time series of order 1. In this case, it is known that the one-dimensional marginal distribution is given by the following (Chatfield 1989):

$$f(y_i; \mu, \rho, \sigma^2) = \frac{1}{\sqrt{2\pi}[\sigma^2/(1 - \rho^2)]^{1/2}}$$
$$\times \exp\left\{-\frac{(1 - \rho^2)}{2\sigma^2}\left[y_i - \frac{\mu}{(1 - \rho)}\right]^2\right\}.$$

Suppose we consider the composite likelihood $CL_1(\theta \mid y) = \Pi_{i=1}^n f(y_i; \theta)$. It can be shown that in this case the only estimable parameters are $\mu/(1 - \rho)$ and $\sigma^2/(1 - \rho^2)$. Essentially, we can estimate the marginal mean and variance. However, that is not enough to tease out the autocorrelation or the three parameters (μ, σ^2, ρ) separately. Unfortunately such analysis of identifiability has to be conducted on a case-by-case basis and general results are difficult to specify. In the following section I consider the particular case of the Gompertz model and show that two-dimensional marginal distributions are sufficient to conduct the statistical analysis.

STATIONARY GOMPERTZ MODEL, SAMPLING VARIABILITY, AND COMPOSITE LIKELIHOOD

Let us now look at the stationary stochastic Gompertz model (Reddingius 1971, Dennis and Taper 1994, Dennis et al. 2004) in detail. As before, let N_t denote the population size and $\hat{N}_t$ denote the estimated population sizes at time t where $t = 1, 2, \ldots, T$. Basic structure of this model is as follows:

Actual population sizes are related to each other according to the relationship $N_{t+1} = N_t \exp(a + b \log N_t + \varepsilon_{t+1})$ where the parameter a may be considered per individual intrinsic growth rate and the parameter b indicates how the growth is affected by the density dependence. Logarithmic transformation of both sides linearizes the model, yielding the Eq. $\log N_{t+1} = a + (1 + b)\log N_t + \varepsilon_{t+1}$. We assume that $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ are independent random variables. Note that after log-transformation the Gompertz model corresponds to an autoregressive process of order 1 (Chatfield 1989). Hence, provided that $|(1 + b)| < 1$ and that the process has reached equilibrium, we can write down the joint distribution of the log-transformed population sizes log

N_t as multivariate normal with mean vector and the covariance matrix, respectively:

$$\mu = \begin{pmatrix} \theta_1 \\ \theta_1 \\ \vdots \\ \theta_1 \end{pmatrix}$$

$$\mathbf{V} = \theta_2 \begin{bmatrix} 1 & (1 + b) & (1 + b)^2 & \cdots \\ (1 + b) & 1 & (1 + b) & \cdots \\ (1 + b)^2 & (1 + b) & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

where $\theta_1 = -a/b$ and $\theta_2 = \sigma^2/(-b[2 + b])$. Given this result, one can write down the two-dimensional marginal distribution of a pair of observations (log N_i, log N_j) as a bivariate normal distribution with parameters

$$\mu = \begin{pmatrix} \theta_1 \\ \theta_1 \end{pmatrix}$$

$$\mathbf{V}_{ij} = \theta_2 \begin{bmatrix} 1 & (1 + b)^{|j-i|} \\ (1 + b)^{|j-i|} & 1 \end{bmatrix}.$$

Note that given the parameters of the bivariate distribution, namely (θ_1, θ_2, b), we can uniquely determine the values of the parameters (a, b, σ^2) by using the transformations $a = -\theta_1 b$ and $\sigma^2 = -\theta_2 b(2 + b)$. Hence, composite likelihood based on two-dimensional marginal distributions will be sufficient for estimation (at least when there is no sampling error). Now let us introduce sampling error into this model.

In biological field studies, several different methods are used to estimate population sizes (Seber 2002). Most of these sampling procedures provide not just estimated population sizes but also the standard errors or complete distributions associated with such estimated values. The procedure outlined below is applicable to any of these sampling procedures where the sampling-error distribution is completely known. If the sampling error is log-normal, then the model reduces to the standard linear Kalman-filter approach (Kalman 1960; B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples, *unpublished manuscript*). Statistical methods for the linear Kalman-filter situation are widely available. In this paper the goal is to deal with the case where the sampling distribution is not lognormal and hence is a particular case of a nonlinear Kalman filter. Kitagawa's solution (Kitagawa 1987), used by De Valpine and Hastings (2002), is a frequentist approach to the problem of nonlinear Kalman-filter models. In this paper, I apply the method of composite likelihood to alleviate the computational burden of Kitagawa's algorithm. For the sake of concreteness, let us assume that $\hat{N}_t \mid N_t \sim \text{Poisson}(N_t)$. In this model there are no additional parameters that may, for example, account for observer effects or other environmental

conditions that may affect sampling error. I discuss these general models later in the paper (see *Nonstationary Gompertz model with sampling variability: Negative binomial sampling error*, below).

Let us now proceed with the Poisson sampling distribution. We can write down the bivariate marginal distribution of $(\hat{N}_i, \hat{N}_j)$ as

$$f(\hat{N}_i, \hat{N}_j; \theta_1, \theta_2, b)$$

$$= \int\int \left[e^{-N_i} \frac{(N_i)^{\hat{N}_i}}{\hat{N}_i!} \right]\left[e^{-N_j} \frac{(N_j)^{\hat{N}_j}}{\hat{N}_j!} \right]$$

$$\times g(\log N_i, \log N_j \,\theta_1, \theta_2, b) \, d\log N_i \, d\log N_j. \quad (2)$$

In the above formula, $g(\log N_i, \log N_j)$ denotes the bivariate normal density with parameters given earlier. Conducting a two-dimensional numerical integration is computationally simple and hence it is feasible to obtain the bivariate marginal distributions of the pairs $(\hat{N}_i, \hat{N}_j)$. Now we apply the concept of composite likelihood to define an objective function $\mathrm{CL}_2(\theta_1, \theta_2, b \,|\, \hat{N})$ $= \Pi_{i=1}^{T-1} \Pi_{j=i+1}^{T} f(\hat{N}_i, \hat{N}_j; \theta_1, \theta_2, b)$. Notice that computation of this composite likelihood involves $T(T-1)/2$ two-dimensional integrals. Recall that as the temporal distance between two observations increases, these observations become almost independent. Thus little information is lost and computational effort can be greatly reduced by considering only those pairs that are within a certain finite neighborhood. For example, Heagerty and Lele (1998), when using composite likelihood for spatial binary data, only consider pairs that are in the immediate neighborhood. The properties of consistency and asymptotic normality are not affected by this reduction.

A straightforward, naïve maximization of the objective function $\mathrm{CL}_2(\theta_1, \theta_2, b \,|\, \hat{N}) = \Pi_{i=1}^{T} \Pi_{j=i+1}^{T} f(\hat{N}_i, \hat{N}_j; \theta_1, \theta_2, b)$ is feasible. One can use any of the standard maximization routines such as the Newton–Raphson or Nelder–Mead simplex algorithm (Press et al. 1992). However, maximization of a function over several variables, three parameters in the above case, requires a significant number of function evaluations. One can reduce the complexity of the maximization problems further.

Consider one-dimensional marginal distributions of $\hat{N}_i$. This is given by

$$f(\hat{N}_i; \theta_1, \theta_2) = \int e^{-N_i} \frac{(N_i)^{\hat{N}_i}}{\hat{N}_i!} g(\log N_i) \, d\log N_i. \quad (3)$$

As before $g(\log N_i)$ denotes the probability density function of a normal variable with mean θ_1 and variance θ_2. Notice that this involves only one-dimensional numerical integration. This is substantially simpler than two-dimensional integration. One can write down the composite likelihood based on these one-dimensional marginal distributions: $\mathrm{CL}_1(\theta_1, \theta_2 \,|\, \hat{N}) = \Pi_{i=1}^{T} f(\hat{N}_i; \theta_1, \theta_2)$. This composite likelihood is a function of (θ_1, θ_2) and using this we can estimate (θ_1, θ_2). We still need

to estimate the density-dependence parameter b. I suggest the following procedure to estimate (θ_1, θ_2, b):

Step 1.—Maximize the composite likelihood $\mathrm{CL}_1(\theta_1, \theta_2 \,|\, \hat{N}) = \Pi_{i=1}^{T} f(\hat{N}_i; \theta_1, \theta_2)$ with respect to (θ_1, θ_2).

Step 2.—Fix the value of (θ_1, θ_2) at the estimates obtained in Step 1 and then maximize the composite likelihood $\mathrm{CL}_2(\theta_1, \theta_2, b \,|\, \hat{N}) = \Pi_{i=1}^{T-1} \Pi_{j=i+1}^{T} f(\hat{N}_i, \hat{N}_j; \theta_1, \theta_2, b)$ with respect to the parameter b.

Confidence intervals, missing data, and non-Poisson sampling errors

Once the initial estimates are obtained, one can use a parametric bootstrap to obtain confidence intervals. The general procedure is explained elsewhere in detail (Dennis and Taper 1994; B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples, *unpublished manuscript*). Since the basic logic is identical for composite-likelihood-based estimators, I do not replicate the description here.

Missing data is another common problem in ecological data sets. If the data are missing completely at random, the bivariate marginal distribution of the pairs (that are formed only with the observed time points) is unaffected. Hence computation of the CL is unaffected by the presence of missing observations. On the other hand, to compute the likelihood, e.g., using Kitagawa's algorithm, requires integrating over the missing values. Thus the complexity of Kitagawa's algorithm increases in the presence of missing observations. It is the sequential nature of the computation of the likelihood function that makes handling missing data difficult using the maximum-likelihood estimators. On the other hand, the method of composite likelihood does not use the sequential nature and hence missing observations do not pose a problem for this method.

So far I have described the method of maximum-composite-likelihood (MCL) estimators for the Gompertz model with Poisson sampling errors. However, notice that *none* of the formulas depend on the Poisson error distribution. If one has different sampling error, one can simply replace the Poisson probability mass function in Eqs. 2 and 3 by the probability mass function for the appropriate sampling distribution. Later in the paper, we will consider sampling error induced by the capture–recapture method of sampling (see *Nonstationary Gompertz model with sampling error: Negative binomial sampling error*, below).

ILLUSTRATIVE DATA ANALYSIS AND STUDY OF STATISTICAL PROPERTIES OF MAXIMUM COMPOSITE-LIKELIHOOD (MCL) ESTIMATORS: STATIONARY CASE

In the following, I present a reanalysis of the American Redstart counts data set reported and analyzed in B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples (*unpublished manuscript*). This is strictly an illustrative analysis. I do not claim that all model assumptions are fully satisfied. For example, the

TABLE 1. Performance of the MCL (maximum composite-likelihood) estimators under the Gompertz Poisson model, based on 100 simulations.

Parameter	Performance measure	Estimates by time series, T		
		$T = 30$	$T = 50$	$T = 100$
$\theta_1 = -a/b = 1.9244$	mean	1.9239	1.9342	1.9212
	median	1.9245	1.9278	1.9254
	MAD	0.2285	0.1951	0.1485
$\sqrt{\theta_2} = \sqrt{\sigma^2/[-b(2 + b)]} = 0.4726$	mean	0.3879	0.4164	0.4504
	median	0.3886	0.3917	0.4541
	MAD	0.1296	0.0922	0.0853
$b = -0.24$	mean	-0.3469	-0.3249	-0.2760
	median	-0.3091	-0.2800	-0.2770
	MAD	0.2628	0.2329	0.1647

Notes: We use the estimated values obtained from the analysis of American Redstart count data (in the Breeding Bird Atlas Survey, 1966–1005) to conduct the Monte Carlo study; these estimates are shown in the first column. For all sample sizes, the estimator of the log (carrying capacity), θ_1, behaves quite well. The estimators of the variance, θ_2, and density-dependence parameter, b, are biased downward. This problem goes away as the sample size increases. Also notice that bias is very small compared to the mean absolute deviation (MAD).

assumption of Poisson sampling error may not be strictly correct. There are observer biases, environmental conditions affect the sampling error, and the parameters in the underlying Gompertz model might also vary with the environment. For these data, information about these complications is unavailable and hence, while acknowledging their existence, I am unable to consider them explicitly here. Some of these complications are addressed in other analyses presented in the paper.

Description of the American Redstart counts data set

The North American Breeding Bird Survey (BBS; Robbins et al. 1986, Peterjohn 1994) poses both challenges to and opportunities for ecological and statistical analyses. While the BBS data are problematic for various reasons, investigators remain optimistic that meaningful ''signals'' might be extracted (Link and Sauer 1997, 1998). The major advantages of the BBS data sets are their extensive spatial coverage and reasonable time-series lengths. The hope is that the extensiveness of the data might trump the imprecision. We consider the counts of American Redstart (*Setophaga ruticilla*) from the BBS record number 0214332808636 observed during 1966–1995. These data were reported and analyzed in B. Dennis, J. M. Ponciano, S. R. Lele, M. L. Taper, and D. F. Staples, *unpublished manuscript*) using the maximum-likelihood (ML) and restricted maximum-likelihood methods (REML) under lognormal sampling-error distribution.

Here we consider the Gompertz model, and a simple, yet reasonable, Poisson distribution as the model for sampling error. I use the algorithm described above to obtain maximum-composite-likelihood estimates. Following Heagerty and Lele (1998) I use only the nearest neighbors to form pairs. This decreases the computational burden substantially by reducing the number of pairs at the same time allowing for the estimation of

all the parameters. Table 1 provides the estimated values of the parameters obtained under the Poisson assumption. I further present a simulation study that studies various properties of the estimation procedure described above. First, we study the effect of not taking into account the sampling variability. These are the estimators (ML-NAÏVE) obtained by using the method described in Dennis et al. (1991) ignoring the fact the observations are not true population sizes but only estimated population sizes. Then we study the maximum-composite-likelihood estimators that take into account the fact that we have estimated population sizes (MCL-GP [Gompertz Poisson]). I provide evidence for reduced bias, consistency, and asymptotic normality of the maximum-composite-likelihood (MCL-GP) estimators as compared to the ML-NAÏVE estimators. Furthermore, I wish to provide some assurance that the loss in efficiency for MCL estimators is probably small. As discussed earlier, the computation of the maximum-likelihood estimators for the full model that takes into account sampling error is difficult. Hence we cannot do a direct comparison of the variance of the MCL estimator with the ML estimator based on the estimated population sizes. However, when there is no sampling error, it is possible to compute both the ML (ML-ORIG [original data]) and the MCL (MCL-ORIG) estimator. Hence we compare the efficiency of the MCL-ORIG estimator with that of the ML-ORIG estimator in the situation where there is no sampling error. The implicit argument is that if the loss of efficiency for MCL estimators is small in the no-sampling-error situation, then it is likely to be small in the case when there is sampling error.

Table 1 summarizes the distributional properties of the MCL-GP estimator for time series of lengths 30, 50, and 100. These simulations were conducted under the parameter values obtained from fitting the model to the American Redstart data. In Fig. 1 we provide
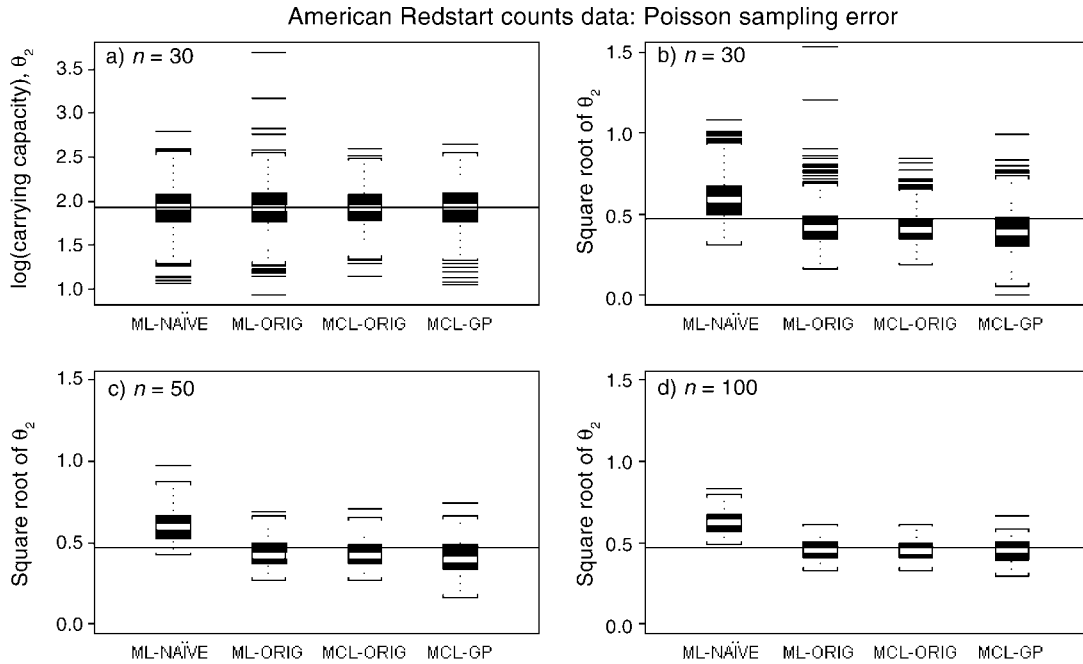
FIG. 1. Comparison of the performance of four estimators (ML-NAÏVE, ML-ORIG, MCL-ORIG, and MCL-GP) of (a) θ_1, the log(carrying capacity), and (b–d) the square root of θ_2, the variance. The horizontal line corresponds to the true value. All estimators behave quite well for estimating log(carrying capacity). The ML-NAÏVE estimator of the variance is biased, and this bias does not go away, even with larger sample sizes. The bias characteristics of the ML-ORIG, MCL-ORIG, and MCL-GP estimators of variance are quite comparable. Their bias goes to zero as sample size increases. Variability in MCL-GP is higher than in ML-ORIG and MCL-ORIG. This is to be expected, given the presence of sampling variability. However, notice that the variability of MCL-ORIG is very slightly larger than ML-ORIG, indicating very little loss of efficiency with the use of the maximum-composite-likelihood (MCL) over the maximum-likelihood (ML) method.

comparison of ML-ORIG, MCL-ORIG, MCL-GP and ML-NAÏVE estimators for the parameters θ_1 and θ_2. The exact algorithm for this simulation study is described in Appendix A. The results of this simulation study are summarized below.

a) Fig. 1a indicates that estimation of the log(carrying capacity) θ_1 works quite well using any of the estimators. Since this parameter is estimated well by all methods, its behavior is not presented for varying sample sizes.

b) ML-NAÏVE estimator of θ_2 has large positive bias for all sample sizes. Moreover, this bias is not reduced with increased sample size. Thus, ML-NAÏVE provides an inconsistent estimator of process variability. This estimator consolidates sampling variability with the process variability.

c) It appears that the MCL-GP estimator of θ_2 is also biased; however the bias is smaller than that for the ML-NAÏVE. Thus, by taking into account sampling variability, we have improved the analysis. Moreover, the variance, as indicated by the spread of the box plots for MCL-GP, also decreases as the sample size increases (Fig. 1b, c, and d). These observations together lend support to the claim of consistency of the MCL-GP estimators. Also notice that the distributions are symmetric around the median supporting asymptotic normality of the MCL-GP estimators.

d) The ML-ORIG and MCL-ORIG estimators of θ_2 are also biased. Thus, even if there is no sampling error, estimation of θ_2 is difficult. However, as the sample size increases from 30 to 50 to 100 (Fig. 1b, c, and d), this bias decreases. Also note that this suggests that the problem of bias in MCL-GP is probably not a fault of the composite-likelihood method but rather a problem intrinsic to the model itself.

e) The distributions of ML-ORIG and MCL-ORIG are quite similar to each other. This indicates a minor loss of efficiency by using the composite-likelihood estimator as compared to the maximum-likelihood estimator when there is no sampling error. This result provides indirect evidence for possibly only a small loss of efficiency using MCL in the presence of sampling error.

Fig. 2 shows the comparison of the estimators of the density-dependence parameter, b, for time series of lengths 30, 50, and 100. The first striking thing to notice is the negative bias in the ML-NAÏVE estimator. If sampling variability is not taken into account, we underestimate the density dependence substantially. The MCL-GP estimator of b is also biased but less so than ML-NAÏVE. The MCL-GP improves the analysis by accounting for sampling error. Moreover, as the sample size increases the distribution of the MCL-GP esti-
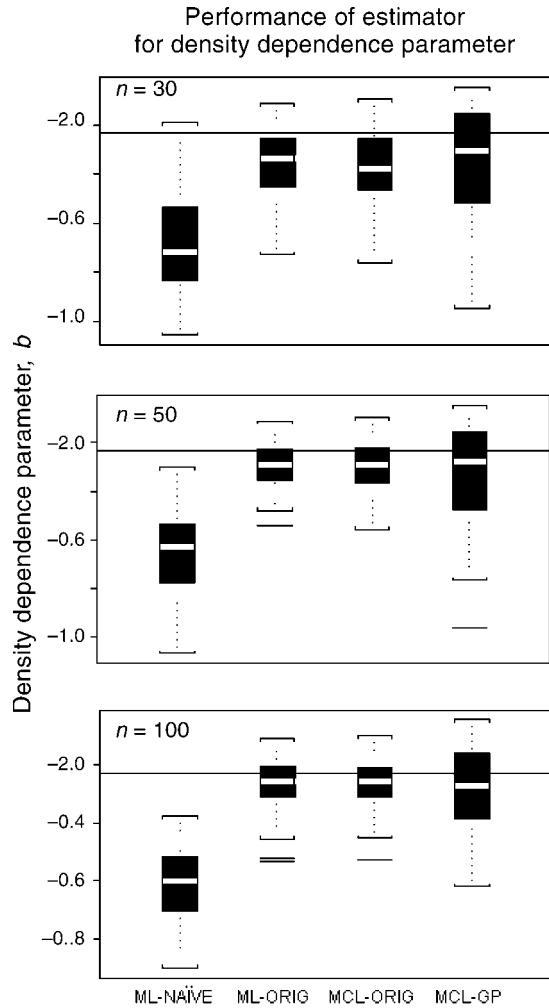
Performance of estimator
for density dependence parameter

FIG. 2. Comparison of the distribution of the ML-NAÏVE, ML-ORIG, MCL-ORIG, and MCL-GP estimators of the density-dependence parameter, b. The horizontal line indicates the true value. It appears that even the ML-ORIG and MCL-ORIG estimators are biased downward and that MCL-GP estimators are quite comparable with them in terms of bias. MCL-GP estimators are, as expected, more variable than ML-ORIG estimators. This is mostly due to the presence of sampling error. From comparing the distribution of ML-ORIG and MCL-ORIG, it appears that the loss of efficiency by using the MCL instead of ML method is quite small when the original data are available. One can implicitly conclude that the loss of efficiency in using MCL-GP estimators in place of ML-GP estimators must also be small.

mators becomes more concentrated around the true value. This result lends support to the claim that MCL-GP is a consistent estimator. Symmetry of the distribution also supports the claim of asymptotic normality. Furthermore, the ML-ORIG and MCL-ORIG estimators of b are also biased for small sample sizes. Thus, the bias in the MCL-GP is not aberrant. Moreover, also note that the distributions of ML-ORIG and MCL-ORIG estimators are quite similar. This indicates the loss in efficiency for MCL-ORIG is small, indirectly

suggesting that loss in efficiency for MCL-GP probably is also small.

NONSTATIONARY GOMPERTZ MODEL
WITH SAMPLING VARIABILITY

To illustrate the power and versatility of the method of composite likelihood, we now consider a case that has not been dealt with in previous papers. This case is extremely important for various applications of stochastic population dynamics models. In the stationary model considered above, it is assumed that the intrinsic growth parameter, a, and the density-dependence parameter, b, do not vary from time to time. Ecologically it is more sensible to model these parameters as dependent on the environmental conditions. For example, higher availability of food should lead to higher growth rate and vice versa. Environmental conditions vary from year to year; so should the parameters in the stochastic growth model.

Suppose there are p covariates summarizing environmental conditions that may affect the intrinsic growth rate. Let $\mathbf{Z}_t = (1, Z_{t1}, Z_{t2}, \ldots, Z_{tp})$ denote the covariate vector corresponding to time point t. I have included "1" to account for the intercept. Let the regression parameters be denoted by $\beta^T = (\beta_0, \beta_1, \ldots, \beta_p)$. The nonstationary Gompertz model is specified by $\log N_t = \alpha_t + (1 + b)\log N_{t-1} + \varepsilon_t$ where $\alpha_t = Z_t\beta$. Thus the intrinsic growth parameter is time varying and depends on the environmental conditions at that time. For this model, it can be shown that the joint distribution is multivariate normal with means, variances, and covariance given by

$$E(\log N_t) = (1 + b)^{t-1}\log N_1$$

$$+ \sum_{k=0}^{t-2} (1 + b)^k Z_{t-k}\beta$$

$$\text{var}(\log N_t) = \frac{\sigma^2}{1 - (1 + b)^2}[1 - (1 + b)^{2(t-1)}]$$

$$\text{cov}(\log N_t, \log N_s) = (1 + b)^{t-s}\text{var}(\log N_s) \quad \text{for } t > s$$

(see Appendix B for mathematical derivations).

Notice that the mean is a function of the initial population size and environmental conditions at all the intermediate time points. The variances and covariances, although independent of the covariates, are time dependent. Since marginal distributions are easy to compute, we can apply the method of composite likelihood to estimate the parameters. Somewhat counterintuitively, the nonstationary case is simpler than the stationary case. This is because in the nonstationary case, the one-dimensional distributions are sufficient to identify the parameters of interest, namely, $(\log N_1, \beta, b, \theta_2)$. We can use one-dimensional integrals instead of two-dimensional integrals. This makes the computational problem easier to deal with. Thus, to obtain the MCL (maximum composite-likelihood) estimators we maximize the following objective function:

TABLE 2. Analysis of San Joaquin kit fox time series under the nonstationary Gompertz model with Poisson and negative binomial (NB) sampling errors.

Parameter	Estimate	90% CL	Stationary process fit
Intercept β_0	0.5154 (NB: 0.5491)	$(-0.8594, 1.7311)$	2.6364
Rainfall β_1	0.0693 (NB: 0.0694)	(0.0534, 0.0931)	
Density dependence	-0.2800 (NB: -0.2873)	$(-0.5203, -0.0046)$	-0.5323
$\sqrt{\theta_2}$	0.2237241 (NB: 0.2194)	(0.00012, 0.4473)	0.4494

Notes: The estimated values and 90% confidence limits (CL) for the parameters of the nonstationary Gompertz model fit to the San Joaquin kit fox data. The environmental factor of rainfall is significant in determining the population dynamics. Also notice the reduction in the environmental variation in the stationary fit to the model that accounts for rainfall, indicating the importance of the covariate in the model. Estimated carrying capacity under the stationary model is 141.5773; under the nonstationary model at the average rainfall value it is 164.2728, and the mean of the observed population size is 156.2308.

$$CL_1(N_1, \beta, b, \theta_2 \mid \log N) = \prod_{t=1}^{T} f(\log \hat{N}_t; N_1, \beta, b, \theta_2)$$

where

$$f(\log \hat{N}_t; N_1, \beta, b, \theta_2)$$

$$= \int P(\hat{N}_t \mid N_t) g(\log N_t; N_1, \beta, b, \theta_2) \, d \log N_t.$$

As before, $g(\cdot)$ indicates the probability density function of the normal random variable. We want to emphasize again that we have to estimate the regression parameters as well as the initial population size. We consider Poisson sampling error in the following.

Algorithm for nonstationary Gompertz model with Poisson sampling error

Step 1.—Estimate the initial log-population size by $\log \tilde{N}_1 = \log(\hat{N}_1 + 0.5)$.

Step 2.—Fix $N_1 = \tilde{N}_1$ and maximize the objective function $CL_1(\tilde{N}_1, \beta, b, \theta_2 \mid \log N) = \prod_{t=2}^{T} f(\log \hat{N}_t; \tilde{N}_1, \beta, b, \theta_2)$ with respect to (β, b, θ_2).

Notice that this objective function only involves one-dimensional marginal densities and hence requires only one-dimensional numerical integration, reducing the computational burden substantially.

To illustrate this method and to study the statistical properties of these estimators under realistic parameter values, I reanalyze the population time series of San Joaquin kit fox (*Vulpes macrotis mutica*) from 1985 through 1994 reported in Dennis and Otten (2000). This population time series is accompanied by the environmental covariate of amount rainfall. As noted by Dennis and Otten (2000), the population sizes are estimated values using the mark–recapture method. They also provide standard errors associated with the estimated population size. I refer the reader to the original paper for more details on the data collection. To check the reasonableness of the Poisson sampling error, I plotted the estimated population sizes vs. the reported squared standard errors associated with these estimates. I observed that the plot (not presented here) was linear with slope close to 1. There was some indication that var-

iance might be slightly larger than the mean, but given the small sample size (12 years) and the illustrative nature of our analysis, *initially* I chose to ignore this minor deviation from the Poisson model.

Table 2 presents the estimated parameters along with the parametric bootstrap-based 90% confidence intervals. Given the small sample size, these confidence intervals are expectedly wide. However it is remarkable that even for such a small sample size, there is a clear indication that rainfall has a significant effect on the population growth (confidence interval excludes zero). The strength of the density-dependence parameter is also (although marginally) significant at the 10% level of significance. In any case, for such a small sample size, one cannot expect to obtain strong inferences. It is also worth noting that these scientific inferences are similar to the ones obtained by Dennis and Otten (2000). For comparison, in Table 2 the estimated parameter values under the stationary Gompertz model are presented. It is clear that the estimate of environmental variance is substantially higher in the stationary case. Much of the variation is explained by the rainfall effect on the population growth. Thus, in the nonstationary case, the estimate of the environmental variance is substantially smaller and is another indication that the covariates might a play significant role in modeling population dynamics.

The sample size for the original data is quite small. To study the behavior of these estimators for larger sample sizes, I conducted a Monte-Carlo study. I considered time-series lengths of 30, 50, and 100. The covariate values for longer time series were generated from a normal distribution with mean and variance equal to mean and variance of rainfall in the original time series. The results are shown in Fig. 3 and are summarized below.

1) Regression coefficients indicating how environmental variables affect the intrinsic growth rate are estimated quite well for all sample sizes.

2) The density dependence parameter, *b*, is also estimated quite well for all sample sizes.

3) The environmental variance parameter is estimated with a negative bias for smaller samples. How-

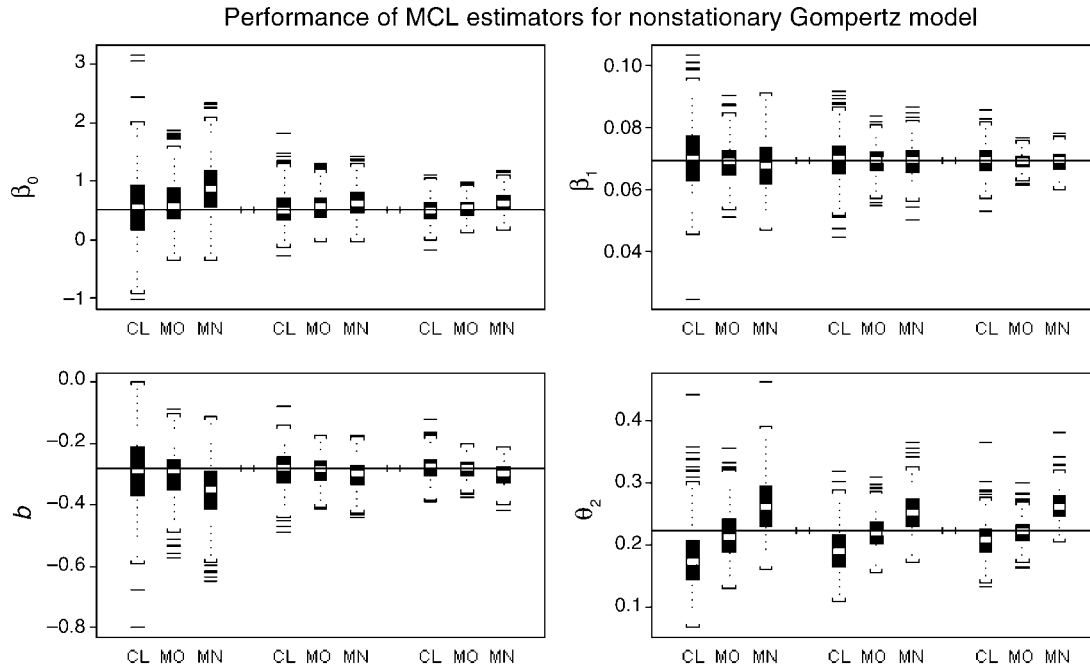## Performance of MCL estimators for nonstationary Gompertz model

FIG. 3. Monte Carlo study of the behavior of MCL estimators for a nonstationary Gompertz model for sample sizes 30, 50, and 100, based on 500 simulations. The horizontal line corresponds to the true parameter value. (In this figure we have used CL, MO, and MN in place of MCL-GP, ML-ORIG, and ML-NAÏVE, respectively, to conserve space. In each panel, the group of three box plots at the left has sample size $n = 30$; for the middle three, $n = 50$; and for the right-hand three, $n = 100$. Except for the variance parameter (θ_2), the environmental (β_0, β_1) and density-dependence (b) parameters are estimated quite effectively for all sample sizes. As the sample size increases, the variance parameter is well estimated also. The performance of MCL is comparable to that of ML-ORIG estimators, whereas the ML-NAÏVE estimator of variance is biased upward for all sample sizes.

ever this bias is reduced as the sample size increases indicating consistency of the estimator. Notice that the ML-NAÏVE estimator of variance has positive bias even for large sample sizes. As in the stationary case, the ML-NAÏVE estimator consolidates the sampling variability with the process error. Further the ML estimator of the environmental variance, even in the absence of sampling variability, is biased. This again shows that the behavior of the MCL estimator is not aberrant.

4) The distribution of the MCL estimator is symmetric, indicating reasonableness of the asymptotic normality of these estimators.

The above simulations are based on the composite likelihood that is based on only one-dimensional marginal distributions. To see if consideration of two-dimensional distributions reduces the bias in the estimator of environmental variance, I fixed the parameters (β_0, β_1, c) at the estimates obtained from one-dimensional composite likelihood and used the two-dimensional distributions, with nearest neighbors forming the pairs, to obtain the estimate of the environmental variance θ_2. The new estimated value was 0.0914, clearly larger than the estimated value obtained using the one-dimensional composite likelihood. The estimator based on one-dimensional distribution is negatively biased and hence it is reassuring that consideration of bivariate

distributions improves the performance of the MCL. I conducted a simulation study (500 replications) to study the behavior of the estimator of θ_2 based on two-dimensional distributions. The median of the bootstrap distribution of the MCL estimator with the one-dimensional distribution was 0.034, while with the two-dimensional distribution it was 0.056. Although it is not surprising, it is reassuring that consideration of two-dimensional distributions reduces bias in the MCL estimators substantially. However, this improvement in the performance is obtained at an increased computational cost of using two-dimensional integrals in place of one-dimensional integrals.

Negative binomial sampling error

As mentioned earlier, the plot of estimated values versus their variances (square of the standard errors) indicated that the relationship was not quite linear with slope equal to 1. Variance appeared larger than the mean, which indicates overdispersion. To account for overdispersion, I fitted the variance function corresponding to the negative binomial distribution, namely, var = mean $+(\text{mean})^2/\text{size}$. This achieved a better fit than the variance function corresponding to the Poisson distribution, var = mean. The estimated value of the parameter ''size'' was 625. The value of the size parameter is fairly large indicating the difference between

Poisson and negative binomial sampling model might be negligible. However, to illustrate the fact that the method of composite likelihood is not limited to using only the Poisson sampling error, I considered the negative binomial distribution as the sampling-error model. Based on the one-dimensional composite likelihood, the estimated values of the Gompertz parameters are reported in Table 2. Notice that these estimates are similar to the ones obtained under the Poisson distribution except that the estimated environmental variance is marginally smaller. The estimates do not change substantially because the size parameter is large enough that the difference between the Poisson and the negative binomial distribution is fairly small.

One can potentially consider a nonstationary Gompertz model where the density-dependence parameter b is also a function of the covariates. However, it seems more natural that the environmental conditions would affect the intrinsic growth rate rather than the density dependence parameter, which is an innate characteristic of the organism. Nonetheless, it will be an interesting study to analyze several data sets using the general model where both a and b depend on the environmental conditions and to check how often the submodel with only a depending on the environmental conditions provides an adequate fit.

Spatial Time Series and Sampling Variability

It should be quite apparent from the simulation results presented thus far that estimation of the parameters of a stochastic population dynamics model is a difficult task, with or without sampling variability. If the length of the time series is small, even with no sampling error, maximum-likelihood estimators are biased and highly variable. However, it is the ecological reality that many ecological time series are short. One reasonable way to boost the sample size is by using spatial replications (Lele et al. 1998). Likelihood-based statistical inference for spatial time series in the presence of sampling error is substantially more computer intensive and complicated than for a single time series. For example, to evaluate the likelihood function at a single parameter value using Kitagawa's (1987) sequential-integration method involves numerical evaluation of T, $2S$-dimensional integrals and T, S-dimensional integrals where S corresponds to the number of spatial locations. Further, to obtain the maximum-likelihood estimator, numerical maximization requires repeated evaluations of the likelihood function. This is prohibitive and practically impossible even for a small number of spatial locations.

However, one can use the composite likelihood for the analysis of spatial time-series data very easily. This extension to the spatial time-series data entails minimal increase in the computational complexity. The composite-likelihood method requires only two-dimensional integrals and not S-dimensional integrals as are required by Kitagawa's algorithm (DeValpine and Has-

tings 2002). This reduction in the dimension provides substantial computational advantage because the computational burden for numerical integration increases exponentially with the dimension of the integral. What is practically an impossible analysis for the maximum-likelihood method is a feasible analysis for the composite-likelihood approach.

Let us start with a description of a spatial Gompertz model. Let S denote the number of spatial locations. Then at each time point, an observation is an $S \times 1$ vector consisting of log(population sizes) at each of the locations. Let us denote this vector by $\log \mathbf{N_t}$. The spatial Gompertz model is given by $\log \mathbf{N_t} = \mathbf{a} + (\mathbf{I} + \mathbf{B})\log \mathbf{N_{t-1}} + \boldsymbol{\varepsilon_t}$. The vector $\mathbf{a} = (a_1, a_2, \ldots, a_s)$ corresponds to the intrinsic growth-rate parameters at the S spatial locations. The matrix $\mathbf{B} = \text{diag}(b_1, b_2, \ldots, b_s)$ is a diagonal matrix of dimension $S \times S$ and $\mathbf{I}$ is an $S \times S$ identity matrix. The process error is given by $\varepsilon_t \sim$ multivariate $\mathcal{N}(0, \Sigma)$. In this general model, the spatial aspect comes from the fact that process errors are spatially correlated. This completely general model has $S + S + S(S + 1)/2$ number of parameters. To reduce the number of parameters, one generally needs to assume a structure on the spatial correlations such as an exponential covariogram (Cressie 1993). One can model the spatial heterogeneity of the intrinsic growth-rate parameters as well as that of density-dependence parameters by considering location-specific environmental covariates. One can also consider spatial version of the random coefficient models such as those considered in Zeng et al. (1998). The kind of model chosen will depend on the availability of relevant information.

Let us consider the bull trout (*Salvelinus confluentus*) data representing counts of redds (breeding sites) recorded during 1980–1993 on four tributaries of the Middle Fork of the Flathead River, Montana, USA (Dennis et al. 1998). This data set provides no information on the environmental covariates, has only four spatial locations and 14 time points. Also, the spatial coordinates of the locations were not available to the author. Given this fairly limited information, we consider a model with spatially homogeneous intrinsic growth and density dependence parameters. We also assume spatially homogeneous process-error variance and equi-correlation between the locations. In summary, our model is based on $\mathbf{a} = (a, a, a, a)$, $\mathbf{B} = \text{diag}(b, b, b, b)$ and

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{pmatrix}.$$

To apply the method of composite likelihood, we need to compute one-dimensional and two-dimensional distributions. Using the theory of multivariate autoregressive processes (Chatfield 1989), these can be written as follows. Let $\theta_1 = -a/b$, $\theta_2 = \sigma^2/[-b(2 + b)]$.

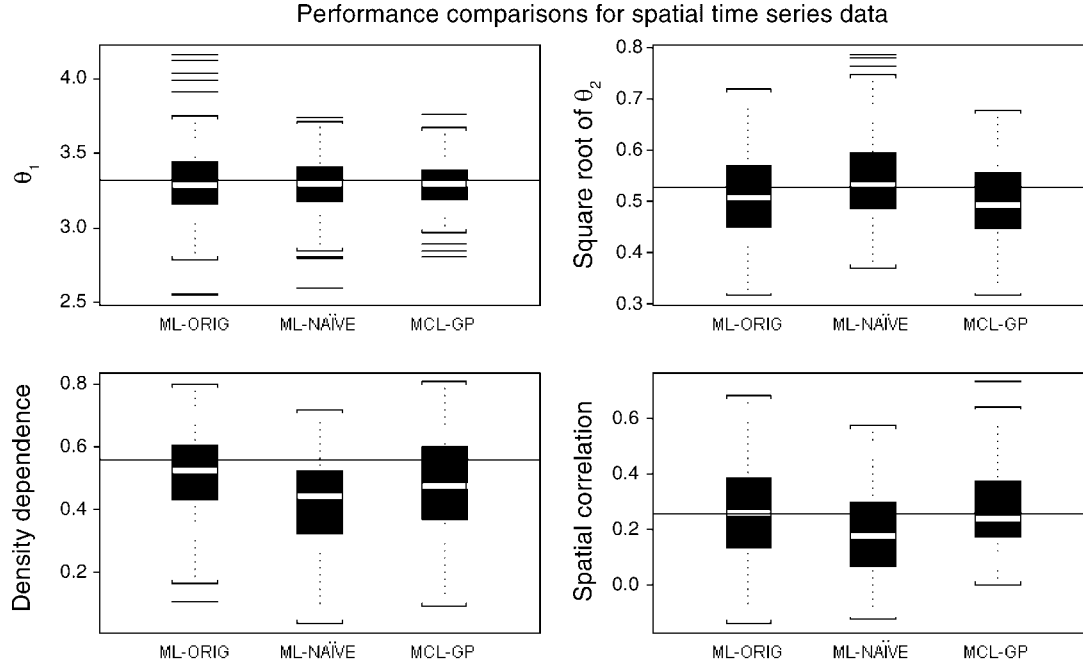### Performance comparisons for spatial time series data

FIG. 4. Simulation comparison of ML-ORIG, ML-NAÏVE, and MCL-GP estimators for spatial time-series data. As in the single time-series case, the MCL-GP estimators and ML-ORIG estimators are quite similar in their biases. ML-NAÏVE estimators are positively biased for variance and negatively biased for density-dependence and spatial-correlation parameters.

Then parameters for our spatial Gompertz model are $(\theta_1, \theta_2, b, \rho)$. In terms of these parameters, the one-dimensional marginal distributions of observations log $N_{t,s}$ are normal with mean θ_1 and variance θ_2 for $t = 1$, $2, \ldots, T$ and $s = 1, 2, \ldots, S$. Furthermore, the pairs that are between time points (i, j) but within the same spatial location s, namely, (log $N_{i,s}$, log $N_{j,s}$) are distributed as bivariate normal random variable with mean vector

$$\begin{pmatrix} \theta_1 \\ \theta_1 \end{pmatrix}$$

and covariance matrix

$$\theta_2 \begin{bmatrix} 1 & (1+b)^{|j-i|} \\ (1+b)^{|j-i|} & 1 \end{bmatrix}.$$

The pairs that are formed within the same time point but between spatial locations, namely, (log $N_{t,l}$, log $N_{t,m}$) are distributed as bivariate normal random variables with mean vector

$$\begin{pmatrix} \theta_1 \\ \theta_1 \end{pmatrix}$$

and covariance matrix

$$\theta_2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Using these results, we can write down the algorithm for computing composite-likelihood estimators for spatial Gompertz model with sampling error as follows.

Step 1.—Use the one-dimensional marginal distributions to construct $CL_1(\theta_1, \theta_2; \hat{N}_1, \hat{N}_2, \ldots, \hat{N}_S) = \Pi_{s=1}^{S} \Pi_{i=1}^{T} f(\hat{N}_{i,s}; \theta_1, \theta_2)$. Maximize this function with respect to the parameters (θ_1, θ_2).

Step 2.—Use the two-dimensional marginal distributions of the pairs formed within a spatial location but between time points to construct the objective function $CL_2(\theta_1, \theta_2, b \mid \hat{N}_1, \hat{N}_2) = \Pi_{s=1}^{S} \Pi_{i=1}^{T-1} \Pi_{j=i+1}^{T} f(\hat{N}_{i,s}, \hat{N}_{j,s}; \theta_1, \theta_2, b)$. Fix the (θ_1, θ_2) parameters at the estimates obtained in Step 1 and maximize with respect to b.

Step 3.—Use the pairs formed between spatial locations but within a time point to construct the objective function $CL_2(\theta_1, \theta_2, \rho \mid \hat{N}_1, \hat{N}_2, \ldots, \hat{N}_S) = \Pi_{s=1}^{S-1} \Pi_{s'=s+1}^{S} \Pi_{i=1}^{T} f(\hat{N}_{i,s}, \hat{N}_{i,s'}; \theta_1, \theta_2, \rho)$. Maximize this function with respect to the parameter ρ with (θ_1, θ_2) fixed at the estimated values from step 1.

In Fig. 4, we study the performance of MCL-GP estimators. Clearly pooling information across space has improved the performance of these estimators; they are more stable and less variable as compared to what was observed in the kit fox data where there were only 12 observations. The ML-NAÏVE estimator of the variance is larger than the MCL-GP estimators because it does not separate out the sampling variability. It also has negative bias in both density-dependence parameter and spatial correlation. These conclusions are similar to the ones that were noted for single time-series analysis. As before, MCL-GP estimators are somewhat biased but this behavior is not aberrant; the ML-ORIG estimators for data without sampling variability are bi-

ased as well. The most important thing to note here is that the method of composite likelihood facilitated the analysis of the spatial time-series data without much computational difficulty. This would have been a nearly impossible task for the method of maximum likelihood.

Summary and Conclusions

The problem of estimation of parameters of stochastic population dynamics models in the presence of sampling variability is a computationally difficult problem. The standard frequentist approach to statistical estimation is based on the likelihood function. The application of likelihood to statistical analysis has been generalized through the use of estimating functions (Godambe 1991). For the problem of inference for stochastic population dynamics models with sampling variability I noted that writing down the likelihood function is difficult. To overcome this problem, I proposed the method of composite likelihood. The main benefit of the MCL (maximum composite likelihood) is the reduction in the dimensionality of the integration to two, no matter what the time-series length or the number of spatial locations is. This method thus provided a practical, feasible approach to the problem of estimation in the presence of sampling variability. Although composite likelihood can be looked upon as an approximation to the full likelihood, the main justification of the method of composite likelihood is via the theory of estimating functions. It generates estimating functions that, under fairly general conditions, lead to consistent and asymptotically normal estimators. These estimators have somewhat larger asymptotic variance as compared to the maximum-likelihood estimators. But in the cases we study, maximum-likelihood estimators are computationally difficult. Hence, in using the method of composite likelihood we paid a price in statistical efficiency but gained in computational feasibility. However, previous studies (Lele and Taper 2002, Henderson and Shimakura 2003) as well as the simulation studies presented in this paper indicate that the loss in statistical efficiency is likely to be small.

Although I did not explicitly illustrate it, the method of composite likelihood is also applicable in the presence of missing observations. Computing the likelihood using Kitagawa's algorithm, which is sequential in nature, requires integration over the missing observations. This increases the computational complexity. On other hand, the definition of the composite likelihood is based on the marginal distributions. These are unaffected by the presence of missing data. One forms the composite likelihood based on the single or pairs of observations that are actually present in the data set. I also showed that the method of composite likelihood can entertain any sampling-error model. We started with an analysis of a single time-series data using the Gompertz growth model with Poisson sampling error. It was shown that the method of composite likelihood is easy to apply and that the estimators have reasonable

statistical properties. Then we studied the rarely considered nonstationary Gompertz model where the intrinsic growth parameter varies with time. The sampling-error model used was based on the mark–capture method of estimation of population sizes. The method of composite likelihood was easy to apply even in this nonstandard situation and the estimators had reasonable statistical properties. From the simulation studies, we learned that to obtain reliable estimates and short confidence intervals one needs large sample sizes. Most ecological time series tend to be short, but spatial replications are usually available. In such situations, the method of composite likelihood was shown to have even bigger computational advantage over the existing methods. Computational simplicity of the method of composite likelihood also facilitates computation of parametric bootstrap confidence intervals. One need not depend on asymptotic methods to obtain standard errors or confidence intervals. The method of composite likelihood involves computation of at most two-dimensional numerical integrals. These routines are readily available in many standard computer packages such as R, MATLAB, GAUSS, or MATHEMATICA. Programming the method of composite likelihood is straightforward and easier than the existing methods for the computation of the maximum-likelihood estimators.

As is true of any method of estimation, there are limitations to the application of the method of composite likelihood. In this paper, I used only one- or two-dimensional marginal distributions to construct the composite likelihood. This particular construction is applicable when such marginal distributions are easy to compute. Furthermore, to be successful, these marginal distributions should be informative about the parameters of interest. This may not hold true for all models of ecological interest. No general recipe can be provided for the construction of suitable estimating functions for models where small dimensional marginal distributions are not informative. On the other hand, I have shown that when such marginal distributions are easy to obtain, the method of composite likelihood is a reasonable alternative to more complicated methods for analyzing stochastic population dynamics models in the presence of sampling variability. Although the discussion in this paper was based on the Gompertz growth model, the method is equally applicable to other stochastic population dynamics models such as the logistic growth model or the delayed density-dependence models. However, some of the technical details are more involved than for the Gompertz model and, hence, these extensions will be discussed elsewhere.

LITERATURE CITED

Besag, J. E. 1975. Statistical analysis of non-lattice data. The Statistician **24**:179–195.

Calder, C., M. Lavine, P. Muller, and J. S. Clark. 2003. Incorporating multiple sources of stochasticity into dynamic population models. Ecology **84**:1395–1402.

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

Chatfield, C. 1989. The analysis of time series analysis: an introduction. Fourth edition. Chapman and Hall, New York, New York, USA.

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

Cressie, N. A. C. 1993. Statistics for spatial data. John Wiley and Sons, New York, New York, USA.

Curriero, F., and S. Lele. 1999. A composite likelihood approach to semivariogram estimation. Journal of Agricultural, Biological and Environmental Statistics **4**:9–28.

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

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

DeValpine, 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.

DeValpine, P. 2004. Monte Carlo state-space likelihoods by weighted posterior kernel density estimation. Journal of American Statistical Association **99**:523–535.

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

Godambe, V. P., editor. 1991. Estimating functions. Oxford University Press, Oxford, UK.

Heagerty, P., and S. R. Lele. 1998. A composite likelihood approach to binary data in space. Journal of American Statistical Association **93**:1099–1111.

Henderson, R., and S. Shimakura. 2003. A serially correlated gamma frailty model for longitudinal count data. Biometrika **90**:355–366.

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

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.

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

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

Lele, S. 1997. Estimating functions for semivariogram estimation. Pages 381–396 *in* I. V. Basawa, V. P. Godambe, and R. L. Taylor editors. Selected Proceedings of the Conference on Estimating Functions. Institute of Mathematical Statistics, Hayward, California, USA.

Lele, S. R., and M. L. Taper. 2002. A composite likelihood approach to (co)variance components estimation. Journal of Statistical Planning and Inference **103**:117–135.

Lele, S. R., M. L. Taper, and S. Gage. 1998. Statistical analysis of population dynamics in space and time using estimating functions. Ecology **79**:1489–1502.

Lindsay, B. G. 1988. Composite likelihood methods. Contemporary Mathematics **80**:221–239.

Link, W. A., and J. R. Sauer. 1997. Estimation of population trajectories from count data. Biometrics **53**:63–72.

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.

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

Morris, W., D. Doak, M. Groom, P. Kareiva, J. Fieberg, L. Gerber, P. Murphy, and D. Thomson. 1999. A practical handbook for population viability analysis. The Nature Conservancy, Arlington, Virginia, USA.

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

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical recipies 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.

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

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

Seber, G. 2002. The estimation of animal abundance and related parameters. Blackburn Press, Caldwell, New Jersey, 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.

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.

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

Zhao, Y., and H. Joe. 2005. Composite likelihood estimation in multivariate data analysis. Canadian Journal of Statistics **33**:335–356.

APPENDIX A

Computational details for the simulation study (*Ecological Archives* E087-008-A1).

APPENDIX B

Mathematical details of the computation of the means, variances, and covariance of the nonstationary Gompertz process (*Ecological Archives* E087-008-A2).