

Approximations to Continuous Dynamical Processes in Hierarchical Models

Amanda R. Cangelosi and Mevin B. Hooten
 Department of Mathematics and Statistics
 Utah State University

Abstract

Models for natural non-linear processes, such as population dynamics, have been given much attention in applied mathematics. For example, species competition has been extensively modeled by differential equations. It is of both scientific and mathematical interest to implement such models in a statistical framework to quantify uncertainty. This study offers an alternative to common ecological modeling practices by using a bias-corrected truncated normal distribution to model the observations and latent process, both having bounded support. Parameters of an underlying continuous process are characterized in a Bayesian hierarchical context, utilizing a fourth-order Runge-Kutta approximation.

KEY WORDS: Population Dynamics, Truncated Normal, Differential Equations, Continuous Models, Numerical Approximations

1. Introduction

Population dynamics have been extensively studied from both deterministic and stochastic perspectives in mathematical ecology. This study addresses the need to quantify uncertainty associated with such mathematical models, highlighting the advantage of a statistical approach. While there are situations that naturally warrant the use of discrete models (*e.g.*, matrix models [Caswell, 2001]), this study focuses on the implementation of continuous models to describe underlying processes, even when the corresponding observations may be discrete. We show that within a hierarchical framework [Berliner, 1996], a bias-corrected truncated normal data model can be utilized to accommodate observations having bounded support. In addition, a similar bias-corrected truncated normal physical process model is introduced to describe both single-species population growth (Velhurst logistic growth) and interspecies competition (Lotka-Volterra equations) [Edelstein-Keshet, 1988]. We also utilize a fourth-order Runge-Kutta (RK4) approximation to the system of differential equations was utilized in conjunction with Markov chain Monte Carlo (MCMC) methods to implement the formal statistical model. While high-order approximation, such as RK4, is not new to statistics (*e.g.*, Shoji and Ozaki [1998]; Rumelin [1982]), its imple-

mentation within a hierarchical framework is underutilized.

To illustrate our methods, we use a historical data set [Gause, 1934] containing population measurements of *Paramecium aurelia* and *Paramecium caudatum* grown in nutrient medium both separately and together. While this particular data set has been used in numerous studies [Pascual and Kareiva, 1996; Leslie, 1957; Edelstein-Keshet, 1988; Lele et al., 2007], often serving as the poster child for biological logistic growth models, many previous studies utilize low-order discretizations, wherein population size is exclusively modeled at times coinciding with those of the observed data. In contrast, we augment the process so that it can be modeled at arbitrarily small time increments.

Regarding random variables having positive support, common practices include log-transformations for use in Gaussian settings. Specifically, in cases where the data include zeros, a common but *ad hoc* approach is to add a small number so that a log-transformation can be taken. While these practices are widespread, they often do not represent the dynamics well; this study offers an alternative through the incorporation of a bias-corrected truncated normal model.

1.1 The Truncated Normal Distribution

The truncated normal distribution can accommodate random variables with non-negative support, but a more appropriate model can be specified by implementing a bias correction in the truncated normal with a target expectation. To illustrate the suitability of such a bias correction, first suppose $X \sim \text{T.N.}(\mu, \sigma^2)$, where we desire $E(X) = \mu$. The actual expectation is not μ however, but rather

$$E(X) = \mu + \sigma \left(\frac{\phi\left(\frac{b_1 - \mu}{\sigma}\right) - \phi\left(\frac{b_2 - \mu}{\sigma}\right)}{\Phi\left(\frac{b_2 - \mu}{\sigma}\right) - \Phi\left(\frac{b_1 - \mu}{\sigma}\right)} \right), \quad (1)$$

where b_1 and b_2 are the lower and upper bounds of truncation, and ϕ and Φ are the standard normal pdf and cdf, respectively [Johnson et al., 1994; Horrace, 2005]. That is, the expectation is not equivalent to the parameter μ , as desired. However, if we specify $X \sim \text{T.N.}(\mu - \gamma, \sigma^2)$, where $E(X) = \mu$, then we need only find the bias correcting function $\gamma(\mu, \sigma^2, b_1, b_2)$ (see Figure 1). This implementation of a bias correction also decreases process uncertainty, particularly near truncation bounds, as often appears in data sets and is depicted in Figure 2. Specifically,

$$E(X) = \mu - \gamma + \sigma \left(\frac{\phi\left(\frac{b_1 - (\mu - \gamma)}{\sigma}\right) - \phi\left(\frac{b_2 - (\mu - \gamma)}{\sigma}\right)}{\Phi\left(\frac{b_2 - (\mu - \gamma)}{\sigma}\right) - \Phi\left(\frac{b_1 - (\mu - \gamma)}{\sigma}\right)} \right). \quad (2)$$

The expectation in (2) is a complicated nested integral equation that is analytically intractable, but it can be approximated numerically.

2. Methods

2.1 Constructing the Process Model

Let $\frac{d\mathbf{x}}{dt} = h(t, \mathbf{x})$ be a differential equation that models a continuous dynamical process \mathbf{x} in time t . Such a process can be well-approximated by the classical fourth-order Runge-Kutta method (RK4):

$$\mathbf{x}_{t+\Delta t} = f(\mathbf{x}_t, \theta) \quad (3)$$

$$= \mathbf{x}_t + \frac{\Delta t}{6}(a_1 + 2a_2 + 2a_3 + a_4), \quad (4)$$

where $a_1 = h(t, \mathbf{x}_t)$, $a_2 = h(t + \frac{\Delta t}{2}, \mathbf{x}_t + \frac{\Delta t}{2}a_1)$, $a_3 = h(t + \frac{\Delta t}{2}, \mathbf{x}_t + \frac{\Delta t}{2}a_2)$, $a_4 = h(t + \Delta t, \mathbf{x}_t + a_3\Delta t)$, and θ is a vector of parameters. The RK4 approximation can then be embedded within a hierarchical probability model at the process level. Note that, depending on the value of Δt , the process can be approximated to a desired level of precision. If the process is allowed to be stochastic at each discretization time t , while Δt is specified to be small, then a state-space model with Markov dependence and an augmented process component results; *i.e.*, $\mathbf{x}_t \sim [\mathbf{x}_t | f(\mathbf{x}_{t-\Delta t}, \theta)]$.

Here, $f(\mathbf{x}_{t-\Delta t}, \theta)$ is the RK4 approximation to the Lotka-Volterra model for species competition:

$$\frac{dx_1}{dt} = r_1 x_1 \frac{k_1 - x_1 - \beta_1 x_2}{k_1}, \quad (5)$$

$$\frac{dx_2}{dt} = r_2 x_2 \frac{k_2 - x_2 - \beta_2 x_1}{k_2}, \quad (6)$$

where the parameter vector θ contains $r_1, k_1, \beta_1, r_2, k_2$, and β_2 , the parameters controlling the dynamics of the system. Specifically, for the i^{th} species, \mathbf{x}_i is the underlying continuous process of population growth, \mathbf{r}_i is the growth rate, \mathbf{k}_i is the carrying capacity, and β_i is the competition parameter. Regarding competition parameters, β_1 denotes the deleterious effect of species 2 on species 1, and vice versa. Note that in the absence of one species, this particular process model defaults to Velhurst logistic growth of the other species.

2.2 The Hierarchical Representation

To introduce the hierarchical model, we begin with the model for observations (*i.e.*, likelihood). Let observations for the process under study at time t be denoted by \mathbf{y}_t . Then a probability model can be written for these measurements:

$$\mathbf{y}_t \sim \text{T.N.}(\mathbf{x}_t - \gamma_{\mathbf{y}_t}, \Sigma_{\mathbf{y}})_{b_1}^{b_2}, \quad \Sigma_{\mathbf{y}} = \sigma_{\mathbf{y}}^2 \mathbf{I}, \quad (7)$$

where $\gamma_{\mathbf{y}_t}$ is a bias correction such that $E(\mathbf{y}_t) = \mathbf{x}_t$, and b_1, b_2 are the lower and upper truncation boundaries, respectively. Note that the truncated distribution accommodates measurements with bounded support, such as

the case with all measurements of mass (*e.g.*, population size).

The data model in (7) is conditional on the second level in the hierarchy, *i.e.*, the process, \mathbf{x}_t . Let us adopt the formulation described above as a model for the process.

$$\mathbf{x}_t \sim \text{T.N.}(f(\mathbf{x}_{t-\Delta t}, \theta) - \gamma_{\mathbf{x}_t}, \Sigma_{\mathbf{x}})_{b_1}^{b_2}, \quad (8)$$

$$\Sigma_{\mathbf{x}} = \begin{pmatrix} \sigma_{x_1}^2 & 0 \\ 0 & \sigma_{x_2}^2 \end{pmatrix},$$

where $\gamma_{\mathbf{x}_t}$ is a bias correction such that $E(\mathbf{x}_t) = f(\mathbf{x}_{t-\Delta t}, \theta)$, and $f(\mathbf{x}_{t-\Delta t}, \theta)$ is the RK4 approximation to the Lotka-Volterra species competition model at time t .

The latent variables \mathbf{x}_t can be thought of as state vectors and may occur at a finer temporal resolution than the data, \mathbf{y}_t . In other words, if $\tau_{\mathbf{x}}, \tau_{\mathbf{y}}$ are finite sets of times, $|\{\mathbf{x}_t, \forall t \in \tau_{\mathbf{x}}\}| > |\{\mathbf{y}_t, \forall t \in \tau_{\mathbf{y}}\}|$, where “ $|\cdot|$ ” represents the size of the set. This point is critical for obtaining a precise approximation to the motivating system of differential equations, because in practice we could choose the set $\tau_{\mathbf{x}}$ to be large enough to attain any desired precision in the approximation.

The parameter model makes up the third and final level of the hierarchy, controlling the dynamics of the system (whence, θ) as well as additional stochasticity (whence, $\sigma_{\mathbf{x}}^2, \sigma_{\mathbf{y}}^2$).

$$\theta \sim \text{T.N.}(\mu_{\theta}, \Sigma_{\theta})_{d_1}^{d_2}$$

$$\mathbf{x}_0 \sim \text{T.N.}(\mu_0, \Sigma_0)_{b_1}^{b_2}$$

$$\sigma_{\mathbf{y}}^2 \sim \text{Inverse Gamma}(r_{\mathbf{y}}, q_{\mathbf{y}})$$

$$\sigma_{\mathbf{x}}^2 \sim \text{Inverse Gamma}(r_{\mathbf{x}}, q_{\mathbf{x}})$$

2.3 Model Implementation

The hierarchical model described above can be implemented in an MCMC setting in the usual manner by analytically identifying full-conditional distributions for parameters and latent state vectors, and then sampling from them sequentially. In this case, the truncated distributions imply nonconjugacy in all parameters and a Metropolis approach must be taken in the MCMC algorithm. Additionally, although the bias corrections are functions of \mathbf{x}_t , they are analytically intractable and thus must be approximated numerically by defining an objective function (see below) and optimizing over γ , using the numerical optimization routine of choice. The robust Nelder-Mead (1965) method [Nelder and Mead, 1965] was used for this particular model.

To illustrate the numerical method for the univariate case, suppose we desire a truncated normal random variable, y , to have expectation μ . Recall that, without a bias correction, the expectation would be as specified in (1). Thus, we construct a function γ such that $E(y)$ appears

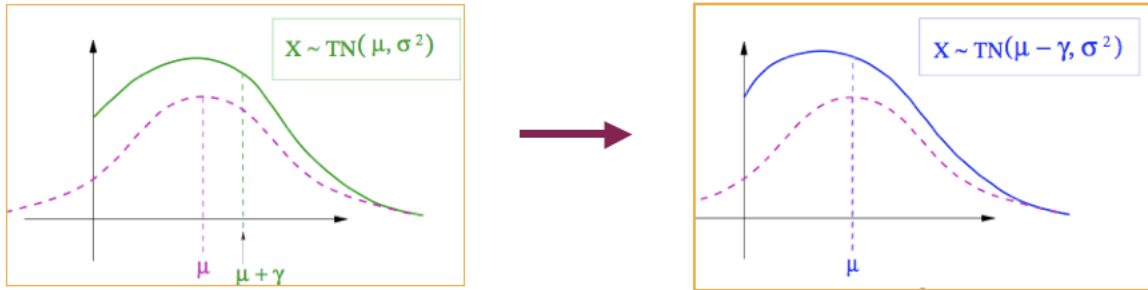


Figure 1: *Effect of bias correction when using truncated normal distribution in hierarchical model.* While the expectation of the normal curve (pink) is equivalent to the parameter μ , that of the left-truncated normal curve (green) is affected by $\gamma(\mu, \sigma^2, b_1, b_2)$. The blue curve is a truncated normal distribution with a bias correction that allows its expectation to be μ .

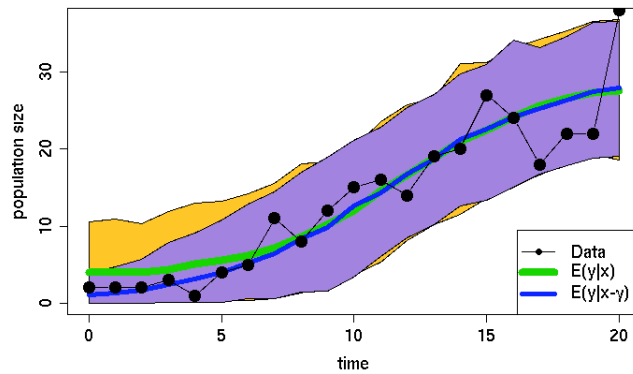


Figure 2: *Effect of bias correction on uncertainty.* Assume a truncated normal data model, left-truncated at zero. The purple area depicts a probability envelope for a data model containing a bias correction; the gold area is that for a model without a bias correction. Note the shift in the mean and reduction of uncertainty in the purple envelope near the truncation boundary.

as in (2) and define an objective function $Q = (E(y) - \mu)^2$ to be minimized with respect to γ . As needed in the MCMC algorithm, we can numerically find γ conditional on μ, σ^2, b_1, b_2 . This method is easily generalized for the multivariate case we are dealing with here.

3. Results

In fitting the model described in the previous section to data, we actually fit three separate models: One univariate logistic growth model to each of the single-species data, followed by a competition model using the two-species data. In this case, since populations were observed both in isolation and in competition, the advantage is that posterior distributions from the single-species models can be used to inform prior distributions for germane parameters (*i.e.*, r_1, r_2, k_1, k_2) in the two-species model. This focuses all available power on the estimation of the additional competition parameters (*i.e.*, β_1, β_2). We set $\Delta t = 1/4$, yielding an augmented process with four times the temporal resolution of the data and a very precise approximation to the differential equations. For the two-species case, the MCMC algorithm was run for 20,000 iterations with a burn-in of 2,000.

To assess the model's capability in situations similar to the Gause data, the model was evaluated using simulated data first; the results of this are summarized in Table 1.

Figures 3-6 depict results of the model, applied to Gause's *Paramecium* data, scaled by 1/10. Figure 3 shows the data for single-species logistic growth, overlaid with the augmented process 95% credible intervals. With regard to the model fit using the Gause data, Figure 4 shows posterior distributions for single-species logistic growth rates and carrying capacities. Posterior mean single-species growth rates (r_1 and r_2) were 0.685 and 0.886 with standard deviations 0.1786 and 0.4147, for *P. aurelia* and *P. caudatum*, respectively. Here, the prior distribution $[r_1] = [r_2] = \text{TN}(0.5, 1)_{d_{r_1}}^{d_{r_2}}$, having truncation bounds $d_{r_1} = 0$ and $d_{r_2} = \infty$. Posterior mean single-species carrying capacities (k_1 and k_2) were 56.1 and 20.8, with standard deviations 3.223 and 5.121, for *P. aurelia* and *P. caudatum*, respectively. Here, the prior distributions $[k_1] = [k_2] = \text{TN}(60, 10000)_{d_{k_1}}^{d_{k_2}}$, having truncation bounds $d_{k_1} = 0$, and $d_{k_2} = \infty$.

Figure 5 shows the data for two-species competition, overlaid with the augmented process credible intervals, after using the single-species growth rate and carrying capacity parameters to inform the two-species priors. Figure 6 shows posterior distributions for competition parameters (β_1 and β_2). Posterior means of the competition parameters were 2.58 and 0.554 with standard deviations 0.8258 and 0.3134, for *P. aurelia* and *P. caudatum*, respectively. Here, $[\beta_1] = [\beta_2] = \text{TN}(1, 10)_{d_{\beta_1}}^{d_{\beta_2}}$, having truncation bounds $d_{\beta_1} = 0$ and $d_{\beta_2} = \infty$.

4. Discussion

Regarding the simulated data summarized in Table 1, note that the posterior variance is reduced for most parameters in the competition model, as compared to the single-species cases (the exception being k_2). This suggests that the competition data hold additional information about the single-species growth parameters. Further, note that each of the parameters used for simulation is indeed captured by the model, suggesting that the model describes the simulated process accurately. Regarding the Gause data, the overlap in posterior growth rates for single-species (Figure 4a) provides little evidence of a significant difference, although the variability differs. The fact that the posterior variability of r_2 is greater than that of r_1 may seem surprising since the *P. caudatum* observations appear to have less spread than the *P. aurelia* observations. Perhaps the difference in variability is due to the fact that *P. caudatum* initially experiences a delayed growth relative to *P. aurelia*, yet *P. caudatum* appears to meet its carrying capacity earlier than *P. aurelia*. Figure 4b suggests that *P. aurelia* is naturally capable of attaining a larger population size than *P. caudatum* in the absence of the other species. Figure 8 evinces a substantial difference among competition parameters; given that *P. aurelia* achieved a larger population in competition, it is perhaps counterintuitive that *P. aurelia* is affected more negatively by the presence of *P. caudatum* than vice versa. One important aspect with regards to the augmented process is that, similar to prediction uncertainty in geostatistics, the credible intervals of the augmented process indicate increased uncertainty at time points farther away from data.

Regarding the methodology of our model, the proposed bias-corrected truncated normal models alleviate the need for a log-transformation and allow us to model a non-negative process. It should be noted that both the process and data models have bounded support, whereas other studies (*e.g.*, Stein, [1992]) have modeled bounded data which were assumed to arise from a measurement model conditioned on a process with real support, rendering no need for a bias correction. The implementation of a bias correction allows us to specify an appropriate physical process model with positive support. Furthermore, utilizing $\Delta t < 1$ improves stability of the dynamical system and avoids drawbacks pertaining to the representation of dynamics that accompany many analytically discretized models (*e.g.*, Ricker growth [Turchin, 2003]). The process augmentation, resulting from $\Delta t < 1$, also allows our approximation to be faithful to the motivating continuous dynamical model; that is, as $\Delta t \rightarrow 0$, our model preserves the dynamical properties of the differential equations. Additionally, in the presence of stochasticity, as $\Delta t \rightarrow 0$, our model converges to an underlying stochastic differential process. The incorporation of RK4 for continuous model approximation is useful for emphasizing parameter estimation and ensuring that process uncertainty is focused on model choice rather than model

Table 1: Results from simulated data

Parameter	Truth	Posterior Mean	Posterior S.D.	Prior Mean	Prior S.D.
Single-Species 1					
r_1	0.6	0.5889	0.1625	0.5	1
k_1	60	60.90	3.713	60	100
Single-Species 2					
r_2	0.8	0.8293	0.2484	0.5	1
k_2	40	41.78	1.956	60	100
Species Competition					
r_1	0.6	0.7290	0.2038	0.5889	0.1625
k_1	60	60.30	3.991	60.90	3.713
β_1	2.5	0.8226	0.8272	1	3.162
r_2	0.8	0.7422	0.2173	0.8293	0.2484
k_2	40	42.10	1.872	41.78	1.956
β_2	0.5	0.2500	0.2759	1	3.162

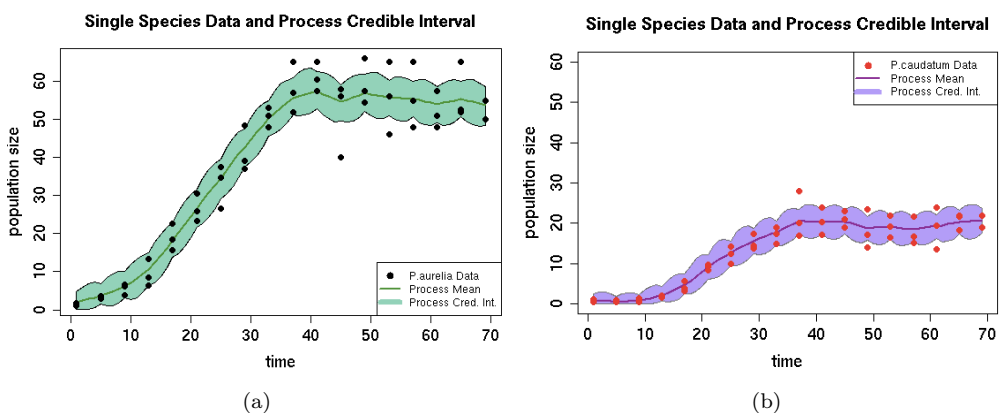


Figure 3: *Paramecium aurelia* (a) and *Paramecium caudatum* (b) observations with augmented process 95% credible interval overlaid.

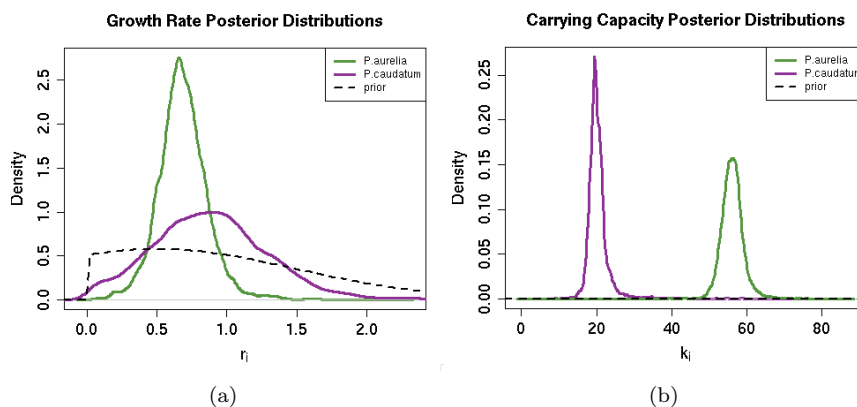


Figure 4: Posterior distributions for single-species logistic growth rates (a) and carrying capacities (b), with prior distribution overlaid.

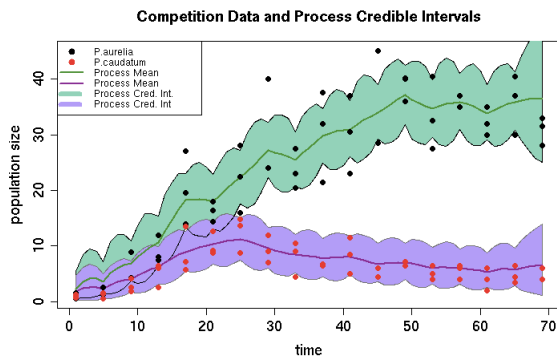


Figure 5: Mixed *P. aurelia* and *P. caudatum* observations with augmented process 95%credible intervals overlaid.

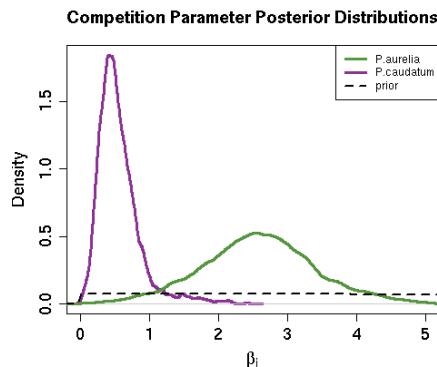


Figure 6: Posterior distributions for competition parameters, with prior distribution overlaid.

approximation.

It should be noted that, as Gause’s *Paramecium* data are counts, a Poisson data model would also be reasonable. We have found that this yields similar results in terms of the posterior process and parameters. Although the Poisson specification addresses the discrete nature of the data in this case, it imposes a distinct mean-variance relationship. The truncated normal specification introduced here also implies a mean-variance relationship, yet it is more flexible than the Poisson due to its two-parameter model formulation. Furthermore, the model we use here is also more robust in that it allows for linear transformations of the data (*e.g.*, scaling) as well as various types of data; for example, the proposed model is applicable to data having bounded continuous support (*e.g.*, rainfall data; percent quadrat cover; plant basal area). Application to such examples is the subject of ongoing research.

REFERENCES

- Berliner, L.M. (1996), “Hierarchical Bayesian Time-Series Models,” In *Maximum Entropy and Bayesian Methods*, K. Hanson and R. Silver, editors, Kluwer Academic Publishers, 15-22.
- Buckland, S.T., K.B. Newman, C. Fernandez, L. Thomas, and J. Harwood (2007), “Embedding Population Dynamics Models in Inference,” *Statistical Science*, **22**, 44-58.
- Caswell, H. (2001), *Matrix Population Models: Construction, Analysis, and Interpretation*, Sinauer Associates.
- Dueker, M. (2005), “Kalman Filtering with Truncated Normal State Variables for Bayesian Estimation of Macroeconomic Models,” Federal Reserve Bank of St. Louis, Working Paper Series.
- Edelstein-Keshet, L. (1988), *Mathematical Models in Biology*, McGraw-Hill, Inc.
- Gause, G.F. (1934), *The Struggle for Existence*, The Williams and Wilkins Company.
- Horrace, W.C. (2005), “Some Results on the Multivariate Truncated Normal Distribution,” Department of Economics, Syracuse University Working Paper Series.
- Johnson, D.S., J.M. London, M. Lea, and J.W. Durban (2007), “Continuous-Time Correlated Random Walk Models for Animal Movement Data,” *Ecology*, In Press.
- Johnson, N.L., S. Kotz, and N. Balakrishnan (1994), *Continuous Univariate Distributions*, John Wiley and Sons.
- Kotecha, J.H. and P.M. Djuric (1999), “Gibbs Sampling Approach for Generation of Truncated Multivariate Gaussian Random Variables”, In *Proceedings of the Acoustics, Speech, and Signal Processing, IEEE International Conference*, **03**, 1757-1760.
- Lele, S.R., B. Dennis, and F. Lutscher. (2007), “Data Cloning: Easy Maximum Likelihood Estimation for Complex Ecological Models Using Bayesian Markov Chain Monte Carlo Methods,” *Ecology Letters*, **10**, 551-563.
- Leslie, P.H. (1957), “An Analysis of the Data for Some Experiments Carried out by Gause with Populations of the Protozoa, *Paramecium aurelia* and *Paramecium caudatum*,” *Biometrika*, **44**, 314-327.
- Nelder, J.A. and R. Mead (1965), “A Simplex Method for Function Minimization,” *Computer Journal*, **7**, 308-313.
- Pascual, M.A. and P. Kareiva (1996), “Predicting the Outcome of Competition Using Experimental Data: Maximum Likelihood

- and Bayesian Approaches," *Ecology*, **77**, 337-349.
- Rumelin, W. (1982), "Numerical Treatment of Stochastic Differential Equations," *SIAM Journal on Numerical Analysis*, **19**, 604-613.
- Shoji, I. and T. Ozaki (1998), "A Statistical Method of Estimation and Simulation for Systems of Stochastic Differential Equations," *Biometrika*, **85**, 240-243.
- Stein, M.L. (1992), "Prediction and Inference for Truncated Spatial Data," *American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America*, **1**, 91-110.
- Turchin, P. (2003), *Complex Population Dynamics: A Theoretical/Empirical Synthesis*, Princeton University Press.