

Continuous structured population models for *Daphnia magna*

Erica M. Rutter*, H.T.Banks*, Gerald LeBlanc[@], and Kevin Flores*

*Center for Research in Scientific Computation

Department of Mathematics

North Carolina State University

Raleigh, NC 27695

and

[@]Department of Biological Sciences

North Carolina State University

Raleigh, NC 27695

December 21, 2016

Abstract

We continue our efforts in modeling *Daphnia magna*, a species of water flea, by proposing a continuously structured population model incorporating density-dependent and density-independent fecundity and mortality rates. Our model is fit to experimental data using the generalized least squares framework and we present confidence intervals on parameter estimates. We modify the model to incorporate more complexity into the density-dependent death rate, but discover that the simpler model outperforms the more complex model in terms of standard errors and is not inferior to the more complex model in terms of information content, using the Akaike Information Criteria.

Key words: continuous structured population models, inverse problems, generalized least squares, model selection, information content, residual plots.

1 Introduction

Structured population models (SPMs) track the density of a population of individuals over time with respect to a physiologically structured variable, such as age or size. SPMs have been used to describe a wide variety of ecological data, see [16, 17, 20, 29, 23, 24] and the references therein. SPMs are desirable because they describe the life history of the organism and allow for dependence of age or density on growth, survival, and fecundity rates. SPMs can be both discretely structured [32] or continuously structured [36].

A continuously structured population model can be preferable to a discrete model for several reasons. When using discrete structured population models (SPMs), parameter estimation may be computationally unstable when parameters are time or age-dependent [8, 40]. Previous work [6, 13] has indicated that the Sinko-Streifer model [36], a continuously structured population model, is more amenable to estimating age-dependent parameters than discretely-structured models. Further, we have previously [1] compared discrete and continuous SPMs for the density-independent *Daphnia magna* survival data and found that the Sinko-Streifer model generated a better fit to data.

Daphnia magna is a species of water flea widely used in ecotoxicology to assess the hazard of chemicals, such as pesticides, on ecosystems, and as a model organism in biomedical research [31, 34, 38, 39]. Ecological risk assessments that use daphnids are most commonly performed at the organismal response level. To enable the causal association of organismal responses to ecosystems adversity, mathematical models are needed to quantitatively connect and propagate organismal assessment information to the population level [4, 18, 28]. Recent daphnid structured population modeling efforts have lacked age-dependent demographics [25], an estimation of density-dependent parameter uncertainty, or have focused on qualitative model analysis instead of model validation [19, 21, 22, 26, 30]. Thus, current daphnia SPMs do not accurately capture the long-term dynamics of aggregate [11] population data.

We propose a continuously structured SPM for *Daphnia magna* that includes both density-independent and density-dependent growth, fecundity, and mortality. We fit our model to *Daphnia magna* data using a vector generalized least-squares framework and perform cross-validation. We determine local sensitivities on all estimated parameters to understand how the model is affected by parameter changes. Further, we calculate confidence intervals and standard errors for our parameters. We also propose a slight variation on our model, allowing for more flexibility in the relationship between density-dependent mortality and biomass. We estimate the additional parameters and show, using the Akaike Information Criteria score, that our simpler model does not perform worse. We also obtain realistic confidence intervals using the simpler model.

2 Mathematical Model

We employ the Sinko-Streifer equations [36] that describe the continuous-time dynamics of a population structured over a continuous variable, which in this case we take to be age (a). $u(t, a)$ represents the population at time t of age a .

$$\frac{\partial u(t, a)}{\partial t} + \frac{\partial u(t, a)}{\partial a} = -\mu_{ind}(a)\mu_{dep}(a, M(t))u(t, a) \quad (1)$$

The mortality rate is a product of a density-independent rate, $\mu_{ind}(a)$, and a density-dependent rate $\mu_{dep}(a, M(t))$. The density-dependent rate depends on age as well as total biomass at time t , given by $M(t)$. Our boundary condition represents the introduction of neonates into the population and is given by:

$$u(t, 0) = \int_0^{a_{max}} k_{ind}(s)k_{dep}(M(t - \tau))u(t, s)ds \quad (2)$$

The fecundity kernel in the recruitment term is similarly described by density-independent and density-dependent rates. Following the assumptions validated in previous results [2], the density-dependent rate is delayed and depends on the biomass τ days ago.

We use total population length as a surrogate measure of population biomass, which is given by:

$$M(t) = \int_0^{a_{max}} u(t, s) \frac{KM_0e^{rs}}{K + M_0(e^{rs} - 1)} ds. \quad (3)$$

Based on previous results [2], the length of each daphnid is assumed to follow a logistic growth curve.

We assume that the density-dependent rate of mortality is a non-decreasing function of population biomass, i.e., $\frac{\partial \mu_{dep}}{\partial M} \geq 0$, and that it is a non-increasing function of age, i.e., $\frac{\partial \mu_{dep}}{\partial a} < 0$. We use a hill function to describe the effect of age and a linear function to describe the effect of biomass:

$$\mu_{dep}(a, M(t)) = 1 + c_1 M(t) \frac{c_3^{h_2}}{c_3^{h_2} + a^{h_2}}. \quad (4)$$

We make the additional assumption that μ_{dep} only affects non-reproductive individuals (size class 1s) and has no affect on reproductive individuals (adults). To model this assumption, we take $c_3 = 8$ days old (the age just before the first offspring are produced) and the hill coefficient $h_2 = 10$, which describes a sharp cutoff in the age at which density has an affect on mortality.

We assume that the density-dependent rate of fecundity is a non-increasing function of population biomass, i.e., $\frac{\partial k_{dep}}{\partial M} \leq 0$. To model this behavior, we used the following hill function:

$$k_{dep}(M(t - \tau)) = \frac{q^{h_3}}{q^{h_3} + M(t - \tau)^{h_3}} \quad (5)$$

The density-independent rates of fecundity and mortality were estimated from individual level data. The daily data for density-independent fecundity was used to directly parameterize the density-independent fecundity, $k_{ind}(a)$, as a function of age. For $\mu_{ind}(a)$, we used an age-varying function that we previously estimated within a density-independent Sinko-Streifer framework using piecewise linear splines [7, 1].

The observables for the data set we collected are the total counts of daphnids within two size classes, which are described by an age cutoff a_1 that we previously determined by estimating the relationship between age and size [2]:

$$S_1(t) = \int_0^{a_1} u(t, s) ds \quad (6)$$

$$S_2(t) = \int_{a_1}^{a_{max}} u(t, s) ds \quad (7)$$

The total population size is the sum of $S_1(t)$ and $S_2(t)$, or

$$N(t) = \int_0^{a_{max}} u(t, s) ds \quad (8)$$

The integral in these population counts ends at a specified age, a_{max} , which we take to be 90 days. This limit is older than the last surviving daphnid we found in density-independent experiments, and allows for finitely defined age-mesh in the numerical PDE solver.

3 Methods

3.1 Data

The data used to fit our model are obtained as previously described [2], but we will briefly describe the data for completeness. A longitudinal study was performed, in duplicate, over 102 days. The daphnid media, reconstituted from deionized water as previously described [5], was seeded with five 6-day old female daphnids. Daphnids were counted every Monday, Wednesday and Friday for the first three weeks of the experiment and weekly thereafter. Daphnia were separated into two size classes with a fine mesh 1.62-mm pore size net.

3.2 Generalized Least Squares Parameter Estimation and Uncertainties

In order to fit the model to our available data, we use a vector generalized least squares (GLS) approach outlined in [11, 12]. We calculate the total number of daphnids in each size class: size class one is $S_1(t, \theta) = \int_0^{a_1} u(t, s) ds$ and size class two is $S_2(t, \theta) = \int_{a_1}^{a_{max}} u(t, s) ds$. Our forward solution observations for a set of parameters θ is given by the vector $\mathbf{f}(t, \theta) = [S_1(t, \theta), S_2(t, \theta)]^T$.

The statistical model we consider allows proportional errors and has the form (here N is the number of observations)

$$\mathbf{Y}_j = \mathbf{f}(t_j; \boldsymbol{\theta}) + \mathbf{f}^\gamma(t_j; \boldsymbol{\theta})\mathcal{E}_j, \quad j = 1, 2, \dots, N, \quad (9)$$

where \mathcal{E}_j are independent and identically distributed (*i.i.d.*) with zero mean and covariance $V_0 = \text{diag}(\sigma_{01}^2, \sigma_{02}^2)$. We estimate our parameters by seeking to minimize a weighted least squares

$$\sum_{j=1}^N \mathbf{w}_j [\mathbf{y}_j - \mathbf{f}(t_j; \boldsymbol{\theta})]^2 \quad (10)$$

where \mathbf{y}_j are the data and the weights \mathbf{w}_j depend on $\boldsymbol{\theta}$. This leads to the so-called generalized least squares (GLS) formulation defined by the solution to the *normal equations*

$$\sum_{j=1}^N [\mathbf{y}_j - \mathbf{f}(t_j; \boldsymbol{\theta})]^T V^{-1}(t_j; \boldsymbol{\theta}) \nabla_{\theta_k} \mathbf{f}(t_j; \boldsymbol{\theta}) = 0, \quad k = 1, \dots, \kappa_\theta, \quad (11)$$

where κ_θ is the number of parameters being estimated. We define

$$V(t_j; \boldsymbol{\theta}) = \text{diag} (f_1(t_j; \boldsymbol{\theta})^{2\gamma_1} \hat{\sigma}_{01}^2, f_2(t_j; \boldsymbol{\theta})^{2\gamma_2} \hat{\sigma}_{02}^2) \quad (12)$$

and

$$\hat{\sigma}_{0i}^2 = \frac{1}{N-p} \sum_{j=1}^N \left(\frac{y_j^i - f_i(t_j; \boldsymbol{\theta})}{f_i(t_j; \boldsymbol{\theta})^{\gamma_i}} \right)^2, \quad i = 1, 2, \quad (13)$$

-see Sections 3.2.5 and 3.2.6 of [11]. The iterative algorithm we use is further explained in [11, 12]. We note that in the case of $\gamma = 0$, this reduces to the vectorized weighted least squares.

Using asymptotic theory [12, 11], the vector generalized least-squares (GLS) estimator has a limiting distribution: $\boldsymbol{\theta}_{GLS} \sim \mathcal{N}(\boldsymbol{\theta}_0, \Sigma_0^N)$ using the true parameter values $\boldsymbol{\theta}_0$. Since these are unknown, we can approximate using our estimated parameter vector, $\hat{\boldsymbol{\theta}}$, and obtain $\boldsymbol{\theta}_{GLS} \sim \mathcal{N}(\boldsymbol{\theta}_0, \Sigma_0^N) \approx \mathcal{N}(\hat{\boldsymbol{\theta}}, \hat{\Sigma}^N)$ where

$$\hat{\Sigma}^N \approx \left(\sum_{j=1}^N D_j^T(\hat{\boldsymbol{\theta}}) V(t_j; \hat{\boldsymbol{\theta}}) D_j(\hat{\boldsymbol{\theta}}) \right)^{-1} \quad (14)$$

with

$$D_j = \begin{pmatrix} \frac{\partial f_1(t_j; \hat{\boldsymbol{\theta}})}{\partial \theta_1} & \dots & \frac{\partial f_1(t_j; \hat{\boldsymbol{\theta}})}{\partial \theta_{\kappa_\theta}} \\ \frac{\partial f_2(t_j; \hat{\boldsymbol{\theta}})}{\partial \theta_1} & \dots & \frac{\partial f_2(t_j; \hat{\boldsymbol{\theta}})}{\partial \theta_{\kappa_\theta}} \end{pmatrix}.$$

and $V(t_j; \boldsymbol{\theta})$ as defined above. We use our optimized estimates $\hat{\boldsymbol{\theta}}$ from equations (11). From this distribution, we can obtain standard errors and 95% confidence intervals to quantify

the uncertainty in the estimation of each parameter. Standard errors for each parameter i are given by $\text{SE}(\hat{\theta}_i) = \sqrt{\hat{\Sigma}_{ii}^N}$. These standard errors [12] are then used to create a 95% confidence interval around each parameter as $[\hat{\theta}_i - 1.96\text{SE}(\hat{\theta}_i), \hat{\theta}_i + 1.96\text{SE}(\hat{\theta}_i)]$.

In order to determine the correct statistical error model, in terms of which value of γ to use, we look towards finite differencing of the data as previously described in [9]. The advantage of this method is that we are able to examine pseudo-error residuals with respect to time and do not introduce bias associated with model error. We examine the second-order differencing of the data (the so-called pseudo errors discussed in [9])

$$\hat{\epsilon}_i = \frac{1}{\sqrt{6}} (y_{i+1} - 2y_i + y_{i-1}) \quad (15)$$

From this estimation of measurement errors, we then can define our modified pseudo-errors by:

$$\eta = \frac{\hat{\epsilon}_i}{|y_i - \hat{\epsilon}_i|^\gamma}. \quad (16)$$

Here $\gamma = 0$ corresponds to vector ordinary least squares and $\gamma \neq 0$ represents generalized least squares. We separate our replicate data into size classes 1 and 2 and calculated the modified residuals using Eq. (16) for various values of γ .

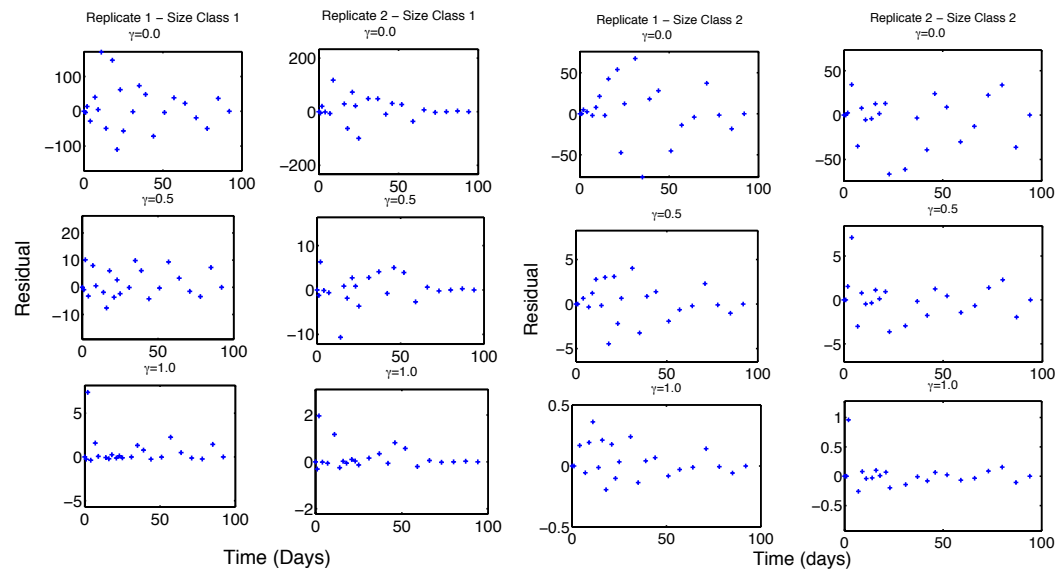


Figure 1: Modified residuals calculated by Eq. (16). Each column represents either size class 1 (left) or size class two (right). Each row increases values of γ .

Figure 1 shows the modified residuals for each replicate for both the size classes 1 and 2 populations. The best statistical model is chosen for plots where the distribution of points

appears random. We note that for the size class 1, the juveniles, a value of $\gamma = 0.5$ appears to be the most random. We reject $\gamma = 0$ for size class 1 because the magnitude of the residuals appear to be correlated with time. For size class 2, the adult population, however, $\gamma = 0$ is sufficiently random. We note that other works have incorporated varying γ values for certain classes of observables [14, 10].

We only optimize over a subset of our parameter space. Parameters K , r , and M_0 describe how individual daphnids contribute to the population biomass, i.e., total length $M(t)$, as described in equation (3). The information for these parameters is given in Table 1. Due to the small variance, we use the r , K , and M_0 parameters from Table 1 for all daphnids. The remaining fixed parameters for the model are given in Table 2. The parameters in the hill function describing the density-dependent death rate (equation (4)) for size class 1s include the age at which the density dependent falls to 0 for 8-day old daphnids (c_3) very steeply (h_2). For the parameters in the density-dependent fecundity rate (equation (5)), only h_3 , the power of the hill function, is fixed.

| Parameter | K | r | M_0 |
|-------------------------|-----------|-----------|-------------------------|
| Fixed effect mean value | 3.7346 | 0.0157 | .7333 |
| Random effect variance | 0.0010533 | 0.0048239 | 6.8978×10^{-7} |

Table 1: Mean parameter estimates and variances along with individual daphnid parameter estimates for the logistic equation using a nonlinear mixed effects model.

The parameters that are optimized via the above vector generalized least-squares framework are τ , the time delay for the effect of density on fecundity, q , the half-maximum for the effect of density on fecundity, and c_1 , the slope of the linear relationship between density-dependent mortality and total mass, $M(t)$.

The model equations are solved numerically in MATLAB using the hpde package [35]. In order to perform the optimization, we will need an initial guess for MATLAB’s LSQNONLIN. We use direct search [27], a non-gradient-based algorithm, in order to obtain a suitable initial guess. Direct search requires upper and lower limits for the parameters, which we choose to be 10^{-10} and 1000, respectively. The optimization is performed until parameter values change by less than 0.1%.

| Parameter | Value |
|-----------|--------|
| a_1 | 3 days |
| c_3 | 8 days |
| h_2 | 10 |
| h_3 | 2 |

Table 2: Fixed parameter values for the model

3.3 Cross validation

Previous work indicated that $\tau > 6$ provided the best fit to population data [2]. Here we perform a cross-validation over our two replicates to determine the best value of τ to use in our optimization. We allow τ to vary in integer values from 6 to 24. Integer values were chosen because of the prohibitive computational costs associated with a finer grid.

If we look over our entire data domain $D = \{d_1, d_2\}$, where each d_i is our replicate data, our best fit for a given value of τ is generated by the parameters $\hat{\theta} = (\hat{q}, \hat{c}_1)$ that minimize the residual sum of squares (RSS) and has a value R_{fit} :

$$R_{fit}(\tau) = \min_{\theta} [RSS(D, \theta, \tau)] \quad (17)$$

$$\hat{\theta}(\tau) = \operatorname{argmin}_{\theta} [RSS(D, \theta, \tau)] \quad (18)$$

Since we have two replicates of data, we can independently determine the parameter optimizations on each data set:

$$\hat{\theta}_i(\tau) = \operatorname{argmin}_{\theta} [RSS(d_i, \theta, \tau)] \quad (19)$$

With these new optimized parameters $\hat{\theta}_i$, we compute $RSS(D, \hat{\theta}_i, \tau)$ which represents the error generated over the full domain D using the parameter estimate $\hat{\theta}_i$ for a specific value of τ . We then generate:

$$R_i(\tau) = -\frac{N}{2} \ln(2\pi) + \ln(RSS(D, \hat{\theta}_i, \tau) - 1) \quad (20)$$

where N is the total number of observation points in D . Once these are computed for each replicate, our cross validation score for a value of τ is the average of these quantities:

$$R_{CrossVal}(\tau) = \frac{1}{2} \sum_{i=1}^2 R_i(\tau) \quad (21)$$

The value of τ the maximizes equation (21) will be used in our simulations.

4 Results

4.1 Cross-Validation

Figure 2 shows the result for the cross-validation scores (equation (21)) for integer values of τ , the density-dependent fecundity delay parameter, ranging from 6 days to 24 days. We can see that the maximum cross-validation score is generated for $\tau = 17$ days. The remaining figures and results use this value of τ .

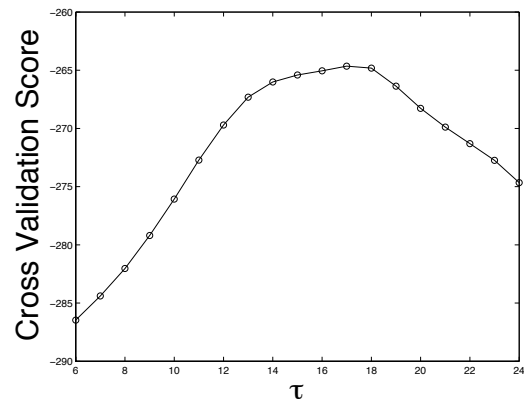


Figure 2: Cross validation scores using the two replicates for various integer values of τ , the fecundity delay parameter. $\tau = 17$ days provides the best cross-validation score.

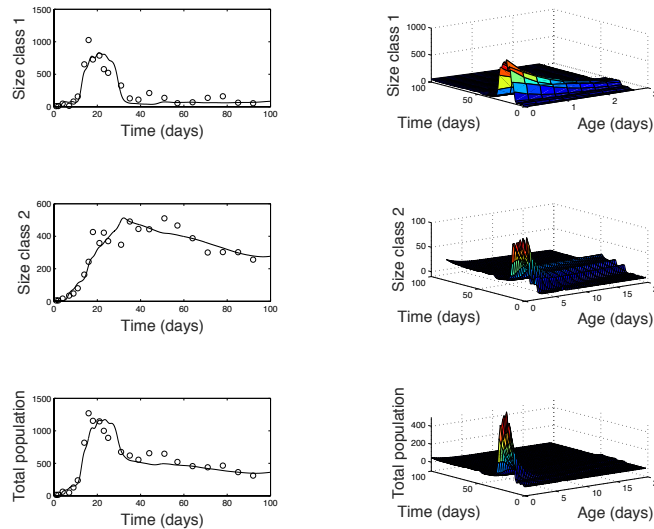


Figure 3: Density-dependent SPM fit to data for replicate 1. Daphnids separated into size class one (top), size class 2 (middle) and total population (bottom). On the left, the total number of daphnids in each size class per day, on the right a surface representing age, time and population number.

4.2 SPM Parameter Estimation and Uncertainty

Figure 3 displays the density-dependent SPM fit to the data for replicate 1 with optimized parameters. The fits are broken up into each size class as well as the total population. On the left side are the total population numbers at each day for each class and on the right is a surface representing dynamics of age, time, and population number. The data is shown with open circles and the model fit in a solid black line. The model solution for replicate one appears to fit each size class as well as the total population.

A similar plot for replicate 2 is shown in Figure 4. We can see that in replicate 2, the optimized fit does not appear as accurate, especially with size class one during the early portion of the simulation. During this time, we can see that there is a sharp increase in size class one, which is even higher than the similar increase in replicate 1. Size class two, however, does appear to have an accurate fit.

The resulting parameters, their confidence intervals, and standard errors are in Table 3, separated into the first replicate (top) and second replicate (bottom). We note that all parameters have non-negative confidence intervals, which we require for biological relevance.

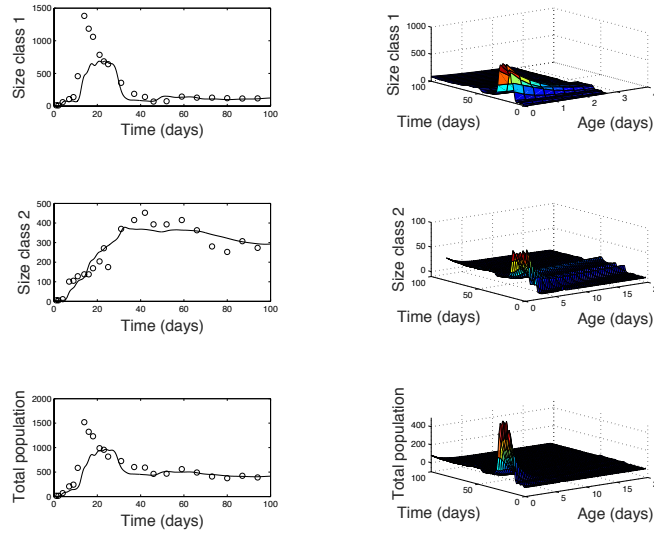


Figure 4: Density-dependent SPM fit to data for replicate 2. Daphnids separated into size class one (top), size class 2 (middle) and total population (bottom). On the left, the total number of daphnids in each size class per day, on the right a surface representing age, time and population number.

The standard errors are relatively small. Although the optimized parameter q for replicate 2 is outside the range of the 95% confidence interval generated in replicate 1, the optimized parameter q for replicate 1 is inside the range of the 95% confidence interval generated in replicate 2. The parameter c_1 for each replicate remains outside of the others' confidence interval. Additional replicates would be useful in determining which set of parameters is more descriptive of population dynamics.

We would also like to examine the residuals for each replicate with respect to time and model value. This will help to insure that the statistical model chosen, of $\gamma = 0.5$ for size class 1 and $\gamma = 0$ for size class 2, is correct. Figure 5 displays the residuals for both replicate one (left) and replicate two (right). Within each subfigure, the left panels show residuals against time and the right panels show residuals against model value. The top figures represent size class one and the bottom figures represent size class two. The residuals for all cases appear to be evenly distributed, indicating that the choice of statistical error model is correct.

| Parameter | Estimate (Rep1) | 95% CI (Rep1) | SE (Rep1) |
|-----------|-----------------|------------------------|-----------|
| q | 105.0066 | (78.0305 , 131.9826) | 13.8339 |
| c_1 | 0.0142 | (0.0128, 0.0156) | 7.0377e-4 |
| Parameter | Estimate (Rep2) | 95% CI (Rep2) | SE (Rep2) |
| q | 145.2262 | (92.1917 , 198.2607) | 27.1972 |
| c_1 | 0.0191 | (0.0170 , 0.0212) | 0.0011 |

Table 3: Optimal parameters, confidence intervals, and standard errors for replicates 1 and 2.

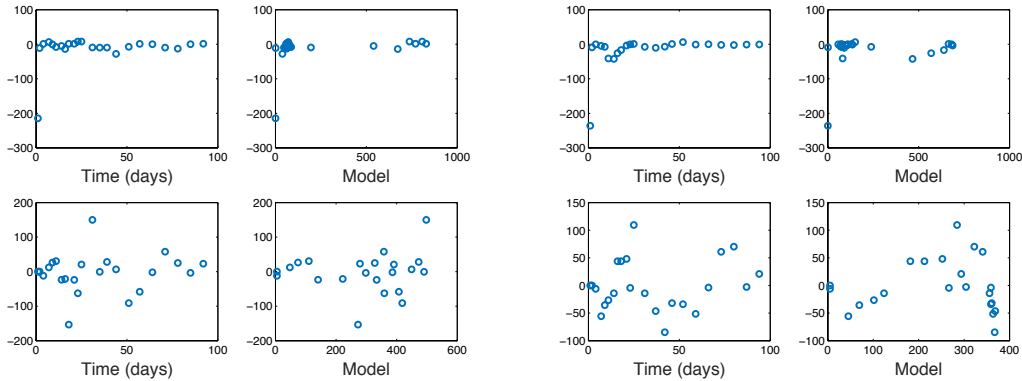


Figure 5: Residual plots for replicate one (right) and replicate two (left) for each class size and against time and model value. Residuals appear random, implying the statistical error model is sufficient.

4.3 Sensitivity Analysis

We perform a local sensitivity analysis for the parameters for each replicate and each size class as well as the full population size. Figure 6 displays the results of this sensitivity analysis. Replicate one is shown in black and replicate two is shown in red. The vertical dashed line is the time at which the daphnia population reaches its peak during the experiments and occurs at the same time for both replicates. Sensitivities were calculated using the complex-step method [33] and were corroborated with the finite-differencing.

The left panels of Figure 6 display the sensitivity of q , which is present in the density-dependent fecundity rate (equation (5)). Increasing q leads to higher fecundity rates and, thus, higher population levels for all size classes. Since q is heavily involved in the fecundity rate, it is sensible that there are slight oscillations in the sensitivities, since the daphnia give birth every third day, meaning that those days may be more sensitive to changes in parameter q . We varied our adjoint step size to ensure that the sensitivities converged and the oscillations were not due to numerical instability.

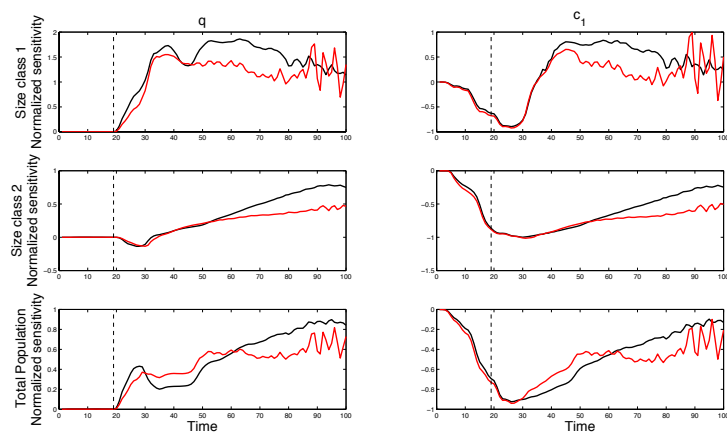


Figure 6: Sensitivity analysis for each parameter for each size class and total populations. Replicate one is shown in black and replicate 2 is shown in red. The vertical dashed line represents the time at which the population reaches its maximum in the experiments

The right panels in Figure 6 show the sensitivity of c_1 . The parameter c_1 is the linear relationship between density-dependent mortality and biomass, $M(t)$ (equation (4)). For the total population size, we see that increasing c_1 results in lower population levels, which is sensible, as it increases the density-dependent death rate. However, the dynamics are very different depending on whether we are looking at size class 1 or size class 2. For size class 1, increasing c_1 results in lower size class 1 populations for approximately the first 30 days of the experiment, after which increasing c_1 results in higher size class 1 populations. This is also somewhat intuitive since populations of size class 1s are very large in the first 30 days, meaning that the biomass is quite large, which increases the death rate. However, in later times, population levels are low, which means that increasing c_1 may not be enough to overcome the small values of $M(t)$.

In general, the sensitivities for replicate 1 and replicate 2 have similar overall shapes, although replicate 2 tends to show more oscillatory behavior.

4.4 Model Comparison

In addition to our current model, we also proposed a slight variation in the density-dependent death term, changing the linear relationship between biomass and density-dependent death rate to a hill function. This results in changing equation (4) to

$$\mu_{dep}(a, M(t)) = 1 + \frac{c_1 M(t)^{h_1}}{c_2^{h_1} + M(t)^{h_1}} \frac{c_3^{h_2}}{c_3^{h_2} + a^{h_2}}. \quad (22)$$

where c_2 and h_1 are optimized along with q and c_1 . This change allows further range of flexibility in how the density-dependent death rate depends on the total biomass, $M(t)$.

We perform the same optimization as outlined above including calculating the cross-validation score, optimizing parameters, and determining confidence intervals. We obtain a similar cross-validation, with $\tau = 16$ days being the optimal τ , maximizing equation (21). The resulting optimal parameter fits are displayed in Figures 7 and 8 for replicates 1 and 2, respectively.

By eye, the fits to data for the linear density-dependent death rate and the hill function density-dependent death rate appear similar. The optimized parameters for the hill function density-dependent death rate are given in Table 4. We see that our estimates for q are similar to the 2-parameter model. However, when we examine the confidence intervals and the standard errors, we can see a large decrease in performance from the 2-parameter model. Three of the four parameters for replicate one have confidence intervals which contain negative values, which is not biologically feasible. For replicate two, there are two parameters which also contain negative values.

It appears that by incorporating the hill function, we are over-parameterizing our model. This is corroborated by examining how the density-dependent death rate, equation (22), behaves with the optimized parameters in Table 4. For replicate one, the optimal parameters generate a linear relationship with respect to the biomass, $M(t)$, which explains why

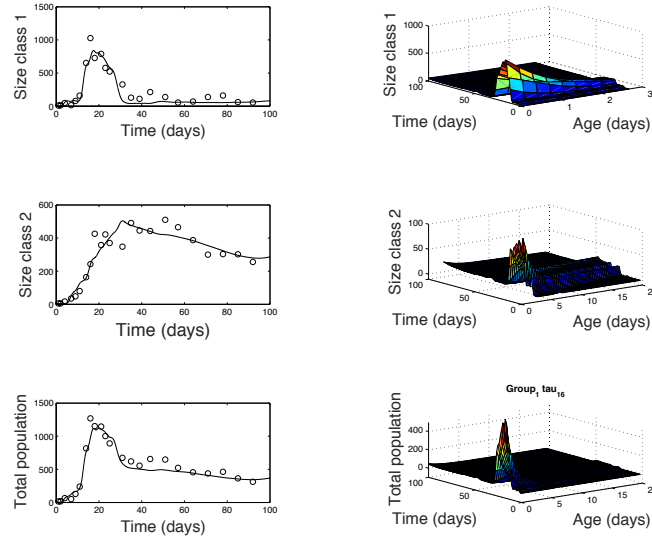


Figure 7: Density-dependent SPM fit to data for replicate 1 using density-dependent death rate given by equation (22). Daphnids separated into size class one (top), size class 2 (middle) and total population (bottom). On the left, the total number of daphnids in each size class per day, on the right a surface representing age, time and population number.

| Parameter | Estimate (Rep1) | 95% CI (Rep1) | SE (Rep1) |
|-----------|-----------------|--------------------------|-----------|
| q | 94.4182 | (16.9937 , 171.8428) | 39.7049 |
| c_1 | 42.5335 | (-382.2441 , 467.3111) | 217.8347 |
| c_2 | 1533.9 | (-14107 , 17175) | 8021.2 |
| h_1 | 1.4986 | (-1.4233 , 4.4206) | 1.4984 |
| Parameter | Estimate (Rep2) | 95% CI (Rep2) | SE (Rep2) |
| q | 154.2988 | (-62.2655 , 370.8632) | 111.0586 |
| c_1 | 16.5795 | (13.7050 , 19.4541) | 1.4741 |
| c_2 | 452.3498 | (356.5672 , 548.1324) | 49.1193 |
| h_1 | 18.6978 | (-33.2882 , 70.6837) | 26.6595 |

Table 4: Optimal parameters, confidence intervals, and standard errors for the 4-parameter model. In this case, many confidence intervals include negative values which is not biologically feasible.

none of the parameters associated with the density-dependent death term are able to be

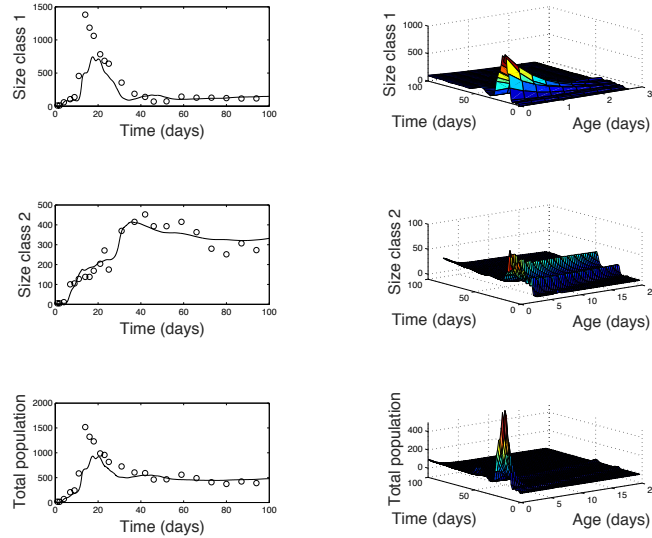


Figure 8: Density-dependent SPM fit to data for replicate 2 using density-dependent death rate given by equation (22). Daphnids separated into size class one (top), size class 2 (middle) and total population (bottom). On the left, the total number of daphnids in each size class per day, on the right a surface representing age, time and population number.

accurately estimated.

The small standard errors suggest that the 2-parameter model may be more advantageous to use, however, we want to ensure that by simplifying the model, we are not losing our ability to describe the dynamics. In order to compare the two models, we turn to the Akaike Information Criteria (AIC) score [3]. The AIC score is an unbiased measure of how well a model fits data. The AIC score is given by:

$$AIC = Nb \ln \left(\frac{RSS}{Nb} \right) + Nb(1 + \ln(2\pi)) + 2(p + 1) \quad (23)$$

where N represents the number of data observations, b represents the number of observables, in this case size class 1 (juveniles) and size class 2 (adults), p is the number of parameters being estimated, and RSS is the residual sum of squares. A lower AIC score implies higher accuracy.

The AIC scores need to be corrected when there are few data points as compared to parameters being estimated. Since $\frac{N}{p} < 40$, we will instead use the AIC_C which is given

| Replicate | 2-Parameter Model | 4-Parameter Model |
|-----------|-------------------|-------------------|
| 1 | 731.7744 | 716.8394 |
| 2 | 796.7948 | 799.6426 |

Table 5: AIC_C scores for the 2-parameter model and 4-parameter model for the two replicates.

by:

$$AIC_C = AIC + 2 \frac{\tilde{p}(b + p + 1)}{N - (b + p + 1)} \quad (24)$$

where \tilde{p} represents the total number of free parameters in the mathematical and statistical models. In our case $\tilde{p} = p + 2$ since we estimate σ_{0i}^2 in our statistical error model.

The resulting AIC_C scores for replicates 1 and 2 for the 2-parameter and 4-parameter model are given in Table 5. The AIC_C scores for replicate 1 show that there is slight improvement using the 4-parameter model. However, for replicate 2, the 2-parameter model has a slightly lower AIC_C score. We are interested in determining whether there really is improvement in one model versus the other. To do this, we can calculate the probability of the correct model [37] using Akaike weights [15].

In order to compute the Akaike weights, we need to determine the difference between the best AIC_C score:

$$\Delta_i(AIC_C) = AIC_{Ci} - \min AIC_C \quad (25)$$

The Akaike weights are computed using this measure of AIC_C differences as:

$$w_i(AIC_C) = \frac{\exp \left[-\frac{1}{2} \Delta_i(AIC_C) \right]}{\sum_{k=1}^K \exp \left[-\frac{1}{2} \Delta_k(AIC_C) \right]} \quad (26)$$

Note that $\sum_{k=1}^K w_k(AIC_C) = 1$, since the Akaike weights represent the probability that model i is the correct model. Therefore, the model with the higher Akaike weight is considered to be the better model [37]. We calculate the $w_i(AIC_C)$ for replicates 1 and two for the 2-parameter and 4-parameter models. For replicate 1, the Akaike weights are 0.0006 and 0.9994 for the 2-parameter and 4-parameter models, respectively. This suggests that the 4-parameter model outperforms the 2-parameter model. However, for replicate 2, the Akaike weights are 0.8059 and 0.1941 for the 2-parameter and 4-parameter models. This suggests that the 2-parameter model outperforms the 4-parameter model.

The Akaike weights are inconclusive as to whether the 2-parameter model or 4-parameter model performs better. Therefore, we argue that the benefits of the 2-parameter model, such as non-negative confidence intervals, outweigh the benefits of the 4-parameter model.

5 Discussion and Conclusions

We proposed a continuously structured population model that included both density-dependent and density-independent growth, fecundity, and mortality for *Daphnia magna* populations. The model that we proposed was used to fit longitudinal experimental data of population dynamics of *Daphnia magna*.

The model was fit to data using a generalized least-squares framework. We also generated parameter estimates and their associated confidence intervals. For the two parameters that were estimated, we performed local sensitivity analysis to further understand the effect of changing their values. Additionally, we proposed a slightly more complicated model with two more parameters and showed that our two-parameter model was not inferior at fitting the data, according to the Akaike Information Criteria score. Our two-parameter model had sensible confidence intervals around parameters (non-negative) while the four-parameter model included negative values in the confidence intervals. We learned that, for our data, the density-dependent death rates (equation (4) for the 2-parameter model and equation (22) for the 4-parameter model) should be linear with respect to total biomass, $M(t)$.

Data collection for this experiment was labor-intensive and we hypothesized whether we could have generated as good a fit with less frequent data collection. To this end, we ran some initial simulations reducing the total number of data points to examine which points were most important to determining parameter estimates. Initial results suggest that such frequent data collection is unnecessary but that consistent collection of data may be more important. We examine the standard errors (SEs) for weekday data collections (M-F every week), 3x a week collection (M/W/F every week), 2x a week collection (M/F or T/T every week) and once weekly collection. The SEs for all of these alternative schedules, with the exception of once weekly, were lower than the SE obtained with our collection schedule (M/W/F for first 3 weeks, thereafter weekly), even though, in some cases, the total number of data points collected were similar. The once-weekly collection SE was on par with the SE generated by our collection schedule for replicate 1 and slightly larger SE value for replicate 2.

The above results suggest that data collection timing is extremely important to the success of our modeling efforts. In order to shrink our standard errors, experimental design will emerge as a powerful tool to dictate experimental efforts. We plan to fully investigate optimal design of experiments in a subsequent paper.

Acknowledgements

This research was supported in part by the Air Force Office of Scientific Research under grant number AFOSR FA9550-15-1-0298, in part by the National Science Foundation under NSF grant number DMS-0946431, and in part by the EPA under US EPA STAR grant RD-835165.

| Collection Schedule (# time points) | Replicate 1 SE $q(105.0066)$ | Replicate 1 SE $c_1(0.0142)$ | Replicate 2 SE $q(145.2262)$ | Replicate 2 SE $c_1(0.0191)$ |
|--|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| M-F Weekly (67) | 6.6426 | 4.0209e-04 | 13.7059 | 6.0155e-04 |
| M/W/F Weekly (40) | 8.5031 | 51848e-04 | 17.7146 | 7.7890e-04 |
| Tu/Th Weekly (27) | 10.6405 | 6.3693e-04 | 21.6359 | 9.4698e-04 |
| M/F Weekly (26) | 10.6868 | 6.3771e-04 | 22.0443 | 9.5791e-04 |
| Weekly (14) | 14.4262 | 9.0486e-04 | 30.0295 | 0.0014 |
| Our Schedule (25) | 13.8339 | 7.0377e-4 | 27.1972 | 0.0011 |

Table 6: SE values for various data collection schedules for Replicate 1 and Replicate 2 for the optimized parameters.

References

- [1] Adoteye, K., Banks, H., Flores, K. B., and LeBlanc, G. A. (2015a). Estimation of time-varying mortality rates using continuous models for daphnia magna. *Applied Mathematics Letters*, 44:12–16.
- [2] Adoteye, K., Banks, H. T., Cross, K., Eytcheson, S., Flores, K. B., LeBlanc, G. A., Nguyen, T., Ross, C., Smith, E., Stemkovski, M., et al. (2015b). Statistical validation of structured population models for daphnia magna. *Mathematical Biosciences*, 266:73–84.
- [3] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723.
- [4] Ankley, G. T., Bennett, R. S., Erickson, R. J., Hoff, D. J., Hornung, M. W., Johnson, R. D., Mount, D. R., Nichols, J. W., Russom, C. L., Schmieder, P. K., et al. (2010). Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment. *Environmental Toxicology and Chemistry*, 29(3):730–741.
- [5] Baldwin, W. S. and Leblanc, G. A. (1994). Identification of multiple steroid hydroxylases in daphnia magna and their modulation by xenobiotics. *Environmental Toxicology and Chemistry*, 13(7):1013–1021.
- [6] Banks, H., Banks, J. E., Dick, L. K., and Stark, J. D. (2007). Estimation of dynamic rate parameters in insect populations undergoing sublethal exposure to pesticides. *Bulletin of Mathematical Biology*, 69(7):2139–2180.
- [7] Banks, H. and Davis, J. L. (2007). A comparison of approximation methods for the estimation of probability distributions on parameters. *Applied Numerical Mathematics*, 57(5):753–777.

- [8] Banks, H., Davis, J. L., Ernstberger, S. L., Hu, S., Artimovich, E., and Dhar, A. K. (2009). Experimental design and estimation of growth rate distributions in size-structured shrimp populations. *Inverse Problems*, 25(9):095003.
- [9] Banks, H. T., Catenacci, J., and Hu, S. (2015). Use of difference-based methods to explore statistical and mathematical model discrepancy in inverse problems. *Journal of Inverse and Ill-posed Problems*, 24(4):413–433.
- [10] Banks, H. T., Everett, R. A., Hu, N. M., and Tran, H. T. (2016) Mathematical and statistical model misspecifications in modeling immune response in renal transplant recipients. *Technical Report CRSC-TR16-14*, Center for Research in Scientific Computation, N C State University, Raleigh, NC, December. *Inverse Problems in Science and Engineering*, submitted.
- [11] Banks, H. T., Hu, S., and Thompson, W. C. (2014). *Modeling and Inverse Problems in the Presence of Uncertainty*. CRC Press, Boca Raton.
- [12] Banks, H. T. and Tran, H. (2009). *Mathematical and Experimental Modeling of Physical and Biological Processes*. CRC Press, Boca Raton.
- [13] Banks, J. E., Dick, L., Banks, H., and Stark, J. D. (2008). Time-varying vital rates in ecotoxicology: selective pesticides and aphid population dynamics. *Ecological Modelling*, 210(1):155–160.
- [14] Baraldi, R., Cross, K., McChesney, C., Poag, L., Thorpe, E., Flores, K. B., and Banks, H. (2014). Uncertainty quantification for a model of HIV-1 patient response to antiretroviral therapy interruptions. *Technical Report CRSC-TR13-13*, Center for Research in Scientific Computation, N C State University, Raleigh, NC, October, 2013. In *2014 American Control Conference*, pages 2753–2758. IEEE.
- [15] Burnham, K. P. and Anderson, D. R. (2002) *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer, New York.
- [16] Caswell, H. (1989). *Matrix Population Models: Construction, Analysis, and Interpretation* Sinauer Associates, Sunderland, MA.
- [17] Caswell, H. (2005). (ed.) *Food Webs: From Connectivity to Energetics* Advances in Ecological Research 36. Elsevier Academic Press, San Diego, California.
- [18] Council, N. R. (2013). *Assessing Risks to Endangered and Threatened Species from Pesticides*. The National Academies Press, Washington, DC.
- [19] Diekmann, O., Gyllenberg, M., Metz, J., Nakaoka, S., and de Roos, A. M. (2010). Daphnia revisited: local stability and bifurcation theory for physiologically structured population models explained by way of an example. *Journal of Mathematical Biology*, 61(2):277–318.

- [20] Diekmann, O., Gyllenberg, M., and Metz, J. (2007). *Physiologically Structured Population Models: Towards a General Mathematical Theory*. Springer-Verlag, Berlin Heidelberg.
- [21] El-Doma, M. (2011). Stability analysis of a size-structured population dynamics model of daphnia. *International Journal of Pure and Applied Mathematics*, 70(2):189–209.
- [22] El-Doma, M. (2012). A size-structured population dynamics model of daphnia. *Applied Mathematics Letters*, 25(7):1041–1044.
- [23] Ellner, S. P., Childs, D. Z., and Rees, M. (2016). *Data-driven Modelling of Structured Populations: A Practical Guide to the Integral Projection Model*. Springer-Verlag, Berlin.
- [24] Ellner, S. P. and Guckenheimer, J. (2011). *Dynamic Models in Biology*. Princeton University Press. Princeton, NJ.
- [25] Erickson, R. A., Cox, S. B., Oates, J. L., Anderson, T. A., Salice, C. J., and Long, K. R. (2014). A daphnia population model that considers pesticide exposure and demographic stochasticity. *Ecological Modelling*, 275:37–47.
- [26] Farkas, J. Z. and Hagen, T. (2007). Linear stability and positivity results for a generalized size-structured daphnia model with inflow. *Applicable Analysis*, 86(9):1087–1103.
- [27] Finkel, D. E. and Kelley, C. T. Convergence analysis of the direct algorithm. *Optimization Online*, 14(2):1–10.
- [28] Hanson, N. and Stark, J. D. (2011). A comparison of simple and complex population models to reduce uncertainty in ecological risk assessments of chemicals: example with three species of daphnia. *Ecotoxicology*, 20(6):1268–1276.
- [29] Keyfitz, N. and Caswell, H. (2005). *Applied Mathematical Demography, Third edition* Springer-Verlag, New York, NY
- [30] Kramer, V. J., Etterson, M. A., Hecker, M., Murphy, C. A., Roesijadi, G., Spade, D. J., Spromberg, J. A., Wang, M., and Ankley, G. T. (2011). Adverse outcome pathways and ecological risk assessment: Bridging to population-level effects. *Environmental Toxicology and Chemistry*, 30(1):64–76.
- [31] LeBlanc, G. A., Wang, Y. H., Holmes, C. N., Kwon, G., and Medlock, E. K. (2013). A transgenerational endocrine signaling pathway in crustacea. *PloS One*, 8(4):e61715.
- [32] Leslie, P. H. (1945). On the use of matrices in certain population mathematics. *Biometrika*, 33(3):183–212.
- [33] Martins, J. R. R. A., Kroo, I. M., and Alonso, J. J. An automated method for sensitivity analysis using complex variables. *38th Aerospace Sciences Meeting and Exhibit*.

- [34] Rider, C. V. and LeBlanc, G. A. (2005). An integrated addition and interaction model for assessing toxicity of chemical mixtures. *Toxicological Sciences*, 87(2):520–528.
- [35] Shampine, L. (2005). Solving hyperbolic pdes in matlab. *Applied Numerical Analysis & Computational Mathematics*, 2(3):346–358.
- [36] Sinko, J. W. and Streifer, W. (1967). A new model for age-size structure of a population. *Ecology*, 48(6):910–918.
- [37] Wagenmakers, E. and Farrell, S. (2004) AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1):192–196.
- [38] Wang, H. Y., Olmstead, A. W., Li, H., and LeBlanc, G. A. (2005). The screening of chemicals for juvenoid-related endocrine activity using the water flea daphnia magna. *Aquatic Toxicology*, 74(3):193–204.
- [39] Wang, Y. H., Kwon, G., Li, H., and LeBlanc, G. A. (2011). Tributyltin synergizes with 20-hydroxyecdysone to produce endocrine toxicity. *Toxicological Sciences*, 123(1):71–79.
- [40] Wood, S. (1994). Obtaining birth and mortality patterns from structured population trajectories. *Ecological Monographs*, 64(1):23–44.