

Article

# Validation of a Mathematical Model for Green Algae (*Raphidocelis Subcapitata*) Growth and Implications for a Coupled Dynamical System with *Daphnia Magna*

Michael Stemkovski <sup>1</sup>, Robert Baraldi <sup>1</sup>, Kevin B. Flores <sup>1,2,3,\*</sup> and H.T. Banks <sup>1,3</sup>

<sup>1</sup> Center for Research in Scientific Computation, North Carolina State University, Raleigh, NC 27695-8212, USA; mstemko@ncsu.edu (M.S.); rjbarald@ncsu.edu (R.B.); htbanks@ncsu.edu (H.T.B.)

<sup>2</sup> Center for Quantitative Sciences in Biomedicine, North Carolina State University, Raleigh, NC 27695-8212, USA

<sup>3</sup> Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8212, USA

\* Correspondence: kbflores@ncsu.edu; Tel.: +1-919-513-0821

Academic Editor: Yang Kuang

Received: 31 March 2016; Accepted: 5 May 2016; Published: 18 May 2016

**Abstract:** Toxicity testing in populations probes for responses in demographic variables to anthropogenic or natural chemical changes in the environment. Importantly, these tests are primarily performed on species in isolation of adjacent trophic levels in their ecosystem. The development and validation of coupled species models may aid in predicting adverse outcomes at the ecosystems level. Here, we aim to validate a model for the population dynamics of the green algae *Raphidocelis subcapitata*, a planktonic species that is often used as a primary food source in toxicity experiments for the fresh water crustacean *Daphnia magna*. We collected longitudinal data from three replicate population experiments of *R. subcapitata*. We used this data with statistical model comparison tests and uncertainty quantification techniques to compare the performance of four models: the Logistic model, the Bernoulli model, the Gompertz model, and a discretization of the Logistic model. Overall, our results suggest that the logistic model is the most accurate continuous model for *R. subcapitata* population growth. We then implement the numerical discretization showing how the continuous logistic model for algae can be coupled to a previously validated discrete-time population model for *D. magna*.

**Keywords:** algae growth models; uncertainty quantification; asymptotic theory; bootstrapping; model comparison tests; *Raphidocelis subcapitata*; *Daphnia magna*

## 1. Introduction

Studies of the population dynamics of phytoplankton and their zooplankton predators in lentic habitats have found a variety of patterns. Plankton communities have been observed to either oscillate in low or high amplitude cycles or to remain relatively stable throughout the year [1]. The same lake may exhibit stability on a given year but switch to oscillation during the following year, and vice versa. A variety of explanations have been proposed for this behavior in the field, including predator-prey interactions, temperature fluctuations, and external influences on nutrient content [2–4]. Fewer studies have attempted to answer the question of what drives these population dynamics in the laboratory setting. Of the predator-prey models that have been proposed for plankton communities, most do not consider certain elements of zooplankton biology such as density-dependent mortality or age-specific fecundity. These traits are crucial for describing the

population growth of zooplankton such as *Daphnia magna* [5] and may be an important factor when modeling population dynamics that are observed in lakes and other ecosystems.

The aim of the present study is to validate and parametrize continuous models for the growth of green algae (*Raphidocelis subcapitata*) in the absence of predation by zooplankton. The broader goal is to couple a validated green algae model with a validated discrete-time population model for *Daphnia magna* [5], and ultimately with an analogous continuous-time model, in order to investigate the possibility of oscillations in the laboratory setting similar to those found in lentic environments. In our previous study of *D. magna* population dynamics [5], we carried out laboratory experiments in which green algae were fed to *D. magna* populations on a daily basis, and populations were reared in media optimized for daphnia survival, but not necessarily ideal for algae growth. In particular, it was not known whether green algae could proliferate in daphnia media to an extent that would affect daphnia population dynamics. Thus, it is of central importance to quantitate the rate at which green algae grow in daphnia media and whether this growth has the potential to significantly alter the fecundity and survival of daphnia. In theory, such changes would thereby affect the quantification of population level risk assessments involving experimental exposure of daphnia populations to environmental perturbations, e.g., toxins or temperature change.

We tested several commonly used growth models for organisms with a limiting nutrient: the Gompertz, the Logistic (continuous and discretized), and the Bernoulli population models. We note that the Bernoulli model is a generalization of the continuous Logistic model, which allows for nested model comparison. Each model has been used to describe populations across many scenarios associated with saturated growth processes in biology. We collected experimental data from three replicates of green algae grown in isolation of predation. We describe goodness of fit of several mathematical models for green algae growth in the context of asymptotic theory and bootstrapping techniques. We provide estimated parameter values and computed confidence intervals for the model predictions. Finally, we implement a numerical scheme that can be used to approximate the concentration of green algae on a daily time scale in order to combine the continuous green algae model that we had the most confidence in with a discrete time population model for *Daphnia magna*. We performed simulations of an unvalidated coupled green algae and daphnia model in order to explore the possible effects of green algae growth on daphnia population dynamics.

## 2. Data and Methods

### 2.1. Data

To observe the growth of *Raphidocelis subcapitata* populations (previously known as *Pseudokirchneriella subcapitata* and *Selenastrum capricornutum*), we seeded three beakers containing 1 L of media reconstituted from deionized water for *Daphnia magna* culturing (previously described in [6]) and recorded the population density for eight days. Each population was kept in an incubator (Thermo Fisher Scientific, Waltham, MA, USA) at 20 °C on a 16/8-h light/dark cycle (6 AM–10 PM light, 10 PM–6 AM dark). Media lost to evaporation was replaced daily with deionized water in order to retain a 1 L volume and avoid replenishing nutrients. The 1 L algae cultures were uncovered and inspected for contamination during measurements. We selected a seeding concentration of  $7 \times 10^7$  cells based on previous studies of algal growth in order to observe both the early (growth) and late (saturation) stage dynamics of the population [7–9]. We measured the density of each population replicate twice using a hemocytometer (Hausser Scientific, Horsham, PA, USA) at 9 AM, 3 PM, and 9 PM daily in order to obtain sufficient data points for parameter estimation and uncertainty quantification. The two measurements of density at each time point for each replicate were averaged to minimize human measurement error. This yielded a total of 23 data points for each of the three replicates.

### 2.2. Asymptotic Theory

The goal of this paper is to determine the most accurate model for algae growth in the absence of consumption. The uncertainty in parameters for each model can be quantified via asymptotic theory. In this section we provide the theory behind our asymptotic theory methodology. The estimation of parameters using asymptotic theory requires mathematical models of the form

$$\begin{aligned} \frac{dy}{dt} &= g(t, y(t), q), \\ y(t_0) &= y_0, \end{aligned} \tag{1}$$

and the corresponding observation process

$$f(t, \theta) = Cy(t, \theta), \tag{2}$$

where  $\theta = (q, \tilde{y}_0) \in \mathbb{R}^{p+\tilde{p}}$  is the vector of unknown parameters,  $q$  is a vector of  $p$  model parameters,  $\tilde{y}_0$  is the number  $\tilde{p}$  of initial conditions that is unknown, and  $C$  maps the model solution  $y(t, \theta)$  in  $\mathbb{R}^l$  to the observed states  $f(t, \theta)$ . We consider the initial condition to be unknown because of measurement error. In this investigation, the observation operator will always produce a scalar, and thus  $C$  maps  $\mathbb{R}^l$  to  $\mathbb{R}$ . In fact, in all our considerations we have  $\tilde{p} = l = 1$ , i.e., the models are scalar and  $C = I$ .

Due to the discrete nature of our experimental data, the observations for our statistical error model occur at  $n = 23$  discrete times  $t_i$ . Thus, the observations will be

$$f(t_i, \theta) = Cy(t_i, \theta), i = 1, \dots, n. \tag{3}$$

To account for measurement error, we use the statistical model

$$Y_i = f(t_i, \theta_0) + \mathcal{E}_i, i = 1, \dots, n \tag{4}$$

for our observations, where  $\mathcal{E}_i$  is a zero mean random variable representing identically, independently distributed (*i.i.d.*) noise that causes our observed data to deviate from our model solution, and  $\theta_0$  is the hypothesized “true” or “nominal” parameter vector that generates the observations  $\{Y_i\}_{i=1}^n$ . The existence of this “true” parameter vector  $\theta_0$  is a standard assumption in frequentist statistical formulations. The *i.i.d.* nature of the error in our model implies that  $\mathbb{E}(\mathcal{E}_i) = 0$  for each  $i$ , and that  $\mathcal{E}_i = 1, \dots, n$ , are identically distributed with variance  $\sigma_0^2$ .

Since  $\mathcal{E}_i$  is a random variable,  $Y_i$  is a random variable with corresponding realizations  $y_i$ . Asymptotic theory seeks to estimate  $\theta_0$  by creating a random variable  $\Theta$  whose realizations for a given data set  $y_i$  will be estimates  $\hat{\theta}$  of  $\theta_0$ . These estimates  $\hat{\theta}$  will approximate the true parameters  $\theta_0$ , and are obtained by minimizing the ordinary least squares (OLS) cost functional [10,11]

$$S(Y; \theta) = \sum_{i=1}^n [Y_i - f(t_i, \theta)]^2, \tag{5}$$

where  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$ . Thus, with  $\Omega$  being the space of admissible parameters and  $y_i$  being the realizations of the random variable  $Y_i$ ,

$$\theta_0 \approx \hat{\theta}_{OLS}^n = \operatorname{argmin}_{\theta \in \Omega} \sum_{i=1}^n [y_i - f(t_i, \theta)]^2 \tag{6}$$

provides an estimate for  $\theta_0$ . The process of estimating parameters from data is known as an inverse problem, and all inverse problems in this experiment are computed using `fmincon` in MATLAB (Mathworks, 2015b, Natick, MA, USA, 2015) with function and step tolerances of  $10^{-20}$  and 1000 iterations.

Once we have an estimate  $\hat{\theta}_{OLS}^n$ , we wish to ascertain the statistical properties of the estimator  $\Theta$ . Although we do not know the distribution of the estimator  $\Theta_{OLS}^n$ , we can approximate it under asymptotic theory (as  $n \rightarrow \infty$ ) by the multivariate Gaussian distribution [10–12]

$$\Theta_{OLS}^n \sim N(\theta_0, \Sigma_0^n) \tag{7}$$

where, based on previous assumptions, the covariance matrix  $\Sigma_0^n$  is approximated by

$$\Sigma_0^n \approx \hat{\Sigma}^n = \hat{\sigma}_{OLS}^2 [\chi^{nT}(\hat{\theta}) \chi^n(\hat{\theta})]^{-1}. \tag{8}$$

Here  $\chi^n$  is the sensitivity matrix

$$\chi_{jk}^n(\theta) = \frac{\partial f(t_i, \theta)}{\partial \theta_k}, i = 1, \dots, n; k = 1, \dots, p, \tag{9}$$

where  $\theta_k$  is the  $k^{th}$  component of the vector  $\theta \in \mathbb{R}^{1 \times p}$ . The unbiased estimator for  $\sigma_0^2$  is

$$\hat{\sigma}_{OLS}^2 = \frac{1}{n - p} \sum_{i=1}^n [y_i - f(t_i, \hat{\theta}_{OLS}^n)]^2 \tag{10}$$

where, for our own examples,  $n = 23$  and  $p = 3$  or  $4$  is the number of model parameters. Both  $\hat{\theta}_{OLS}^n$  and  $\hat{\sigma}_{OLS}^2$  are then used in Equation (8) (i.e.,  $\hat{\theta} = \hat{\theta}_{OLS}$  and  $\hat{\sigma}^2 = \hat{\sigma}_{OLS}^2$ ).

In our calculations, the sensitivity equations are calculated analytically by solving the differential equation at  $\hat{\theta}$

$$\frac{d}{dt} \left( \frac{\partial y}{\partial \theta} \right) = \frac{\partial g}{\partial y} \frac{\partial y}{\partial \theta} + \frac{\partial g}{\partial \theta}. \tag{11}$$

Note that  $\frac{dy}{dt} = g(t, y(t), \hat{\theta})$  is the differential equation for the green algae model and  $f(t_j, \hat{\theta}) = y(t_j, \hat{\theta})$  is the forward solution of each model. Because we know analytical formulas that provide solutions for  $\frac{dy}{dt} = g(t, y(t), \hat{\theta})$ , we can solve Equation (11) by setting up a differential equation in terms of the sensitivity [11].

The  $\chi^n$  matrix provides a measure for how sensitive the mathematical model is to each of its parameters. This can be used to estimate the  $p \times p$  covariance matrix,  $\Sigma_0^n$ ,

$$\Sigma_0^n \approx \hat{\Sigma}^n = \hat{\sigma}^2 [\chi^{nT}(\hat{\theta}) \chi^n(\hat{\theta})]^{-1}. \tag{12}$$

In order to determine the confidence we have in the parameter estimates, we also compute the asymptotic theory based standard error  $SE(\hat{\theta}_k) = \sqrt{\hat{\Sigma}_{kk}^n}$  for the  $k^{th}$  parameter.

### 2.3. Bootstrapping

We implemented bootstrapping techniques to complement our asymptotic theory approach with regards to estimating parameter uncertainty. We again assume that we have experimental data for a dynamical system from an underlying observation process in Equation (4) where  $\tilde{\epsilon}_i$  are *i.i.d.* with mean zero and constant variance  $\sigma_0^2$  and  $\theta_0$  is the “true value” hypothesized to exist in statistical treatments of data [10]. The random variable also has realizations

$$y_i = f(t_i, \theta_0) + \tilde{\epsilon}_i. \tag{13}$$

We use the following algorithm [10] to compute the bootstrapping estimate  $\hat{\theta}_{BOOT}$  of  $\theta_0$  and its empirical distribution.

1. First estimate  $\hat{\theta}^0$  from the entire sample  $\{y_i\}_{i=1}^n$  using OLS.
2. Using this estimate, define the standardized residuals

$$\bar{r}_i = \sqrt{\frac{n}{n-p}}(y_i - f(t_i, \hat{\theta}^0)) \tag{14}$$

for  $i = 1, \dots, n$ , where  $n$  is the number of data points, and  $p$  are the number of model parameters. Set  $m = 0$ , which will represent the total number of artificial samples we will create.

3. Create a bootstrapping sample of size  $n$  using random sampling with replacement from the data (realizations)  $\{\bar{r}_1, \dots, \bar{r}_n\}$  to form a bootstrapping sample  $\{\bar{r}_1^m, \dots, \bar{r}_n^m\}$ .
4. Create bootstrap sample points

$$y_i^m = f(t_i, \hat{\theta}^0) + r_i^m \tag{15}$$

for  $i = 1, \dots, n$ .

5. Obtain a new estimate  $\hat{\theta}^{m+1}$  from the bootstrapping sample  $\{y_i^m\}$  using OLS.
6. Set  $m = m + 1$  and repeat steps 3–5 until  $m \geq 1000$  (this can be any large value, but for these experiments we used  $M = 1000$ ).

We then calculate mean, standard error, and confidence intervals using the formulas:

$$\hat{\theta}_{BOOT} = \frac{1}{M} \sum_{m=1}^M \hat{\theta}^m, \tag{16}$$

$$Var(\theta_{BOOT}) = \frac{1}{M-1} \sum_{m=1}^M (\hat{\theta}^m - \hat{\theta}_{BOOT})^T (\hat{\theta}^m - \hat{\theta}_{BOOT}), \tag{17}$$

$$SE_k(\hat{\theta}_{BOOT}) = \sqrt{Var(\theta_{BOOT})_{kk}}, \tag{18}$$

where  $\theta_{BOOT}$  denotes the bootstrapping estimator. We present the results of these techniques as standard errors about the mean of the parameter estimates, as well as the parameter distributions created. This procedure is performed for each replicate in our experiments.

#### 2.4. Model Comparison: Nested Restraint Tests

We used nested model comparison tests to determine the loss in accuracy by constraining certain models, *i.e.*, holding some parameters constant. In general, we assume that we have an inverse problem for the model observations  $f(t, \theta)$  and are given  $n$  observations with the cost function described above in Equation (5). We are interested in using data to question whether the “true” parameter  $\theta_0$  can be found in a subset  $\Omega_H \subset \Omega$ , for which we make the same assumptions as Banks, Hu, and Thompson [10]. Thus, we want to test the null hypothesis  $H_0: \theta_0 \in \Omega_H$ , or that the constrained model provides an adequate fit to the data. We then define

$$\Theta_H^n(Y) = \operatorname{argmin}_{\theta \in \Omega_H} S^n(Y; \theta) \tag{19}$$

and

$$\hat{\theta}_H^n(y) = \operatorname{argmin}_{\theta \in \Omega_H} S^n(y; \theta), \tag{20}$$

where  $y$  is a realization of  $Y$ . It is important to note that  $S^n(y; \hat{\theta}_H^n) \geq S^n(y; \hat{\theta}^n)$ . We define the nonnegative test statistics and their realizations, respectively, by

$$T_n(Y) = S^n(Y; \hat{\theta}_H^n) - S^n(Y; \hat{\theta}^n) \tag{21}$$

and

$$\hat{T}_n = T_n(y) = S^n(y; \hat{\theta}_H^n) - S^n(y; \hat{\theta}^n). \tag{22}$$

We refer to [10] for a description of asymptotic convergence as  $n \rightarrow \infty$ , which yields the model comparison result

$$U_n(Y) = \frac{nT_n(Y)}{S^n(Y; \theta^n)} \tag{23}$$

with the corresponding realizations

$$\hat{u}_n = U_n(y) \tag{24}$$

which can be compared to a  $\chi^2$  distribution with  $r$  degrees of freedom. In this project we use a  $\chi^2(1)$  table when comparing the results from the Logistic model to those from the Bernoulli model.

### 2.5. Akaike Information Criterion

In some cases (such as comparison between the Logistic and the Gompertz), the models are not nested (although they are related through a limiting process—see below) and hence we cannot use the model comparison tests outlined above. However, we can use an alternative model evaluation framework and implement the Akaike Information Criterion ( $AIC_c$ ) with a small size sample correction [10] in the context of an ordinary least squares framework

$$AIC_c = n[1 + \ln(2\pi)] + n \ln \left( \frac{\sum_{i=1}^n (y_i - f(t_i, \hat{\theta}_{OLS}^n))^2}{n} \right) + 2(p + 1) + \frac{2p(p + 1)}{n - p - 1} \tag{25}$$

where  $n$  is the sample size and  $p$  is the number of unknowns (parameters). This will allow us to suggest which model provides a better fit for the data (models with smaller  $AIC_c$  values provide better fits). While other goodness of fit tests may be useful for selecting models, we chose to use  $AIC_c$ , since it is a widely adopted measure of model accuracy (see Sections 4.2, 4.3 of [10]).

## 3. Models

### 3.1. Logistic Model

The first model we consider is the widely used logistic model for bounded growth of a population  $P(t)$ , given by the differential equation

$$\frac{dP}{dt} = RP(t) \left( 1 - \frac{P(t)}{K} \right) \tag{26}$$

where  $R$  is the intrinsic growth rate, and  $K$  is the carrying capacity for the population under consideration.

### 3.2. Bernoulli Model

We also analyze the data within the context of a Bernoulli model due to Richards [13], given by the differential equation

$$\frac{dP}{dt} = RP(t) \left( 1 - \left( \frac{P(t)}{K} \right)^\beta \right). \tag{27}$$

The Bernoulli model has three model parameters,  $R$ ,  $K$ , and  $\beta$ . The parameter  $\beta$  extends the logistic model to allow flexibility in the growth dynamics by allowing the inflection point to change while keeping the carrying capacity approximately the same [13]. Setting  $\beta = 1$  yields the logistic model in Equation (26); hence, the logistic model is a special case of the Bernoulli model, thereby enabling us to use nested model comparison tests described above.

### 3.3. Gompertz Growth Model

The next model we consider is the Gompertz growth model, which is widely used for biological and economic phenomena where population growth is not symmetric about the point of inflection, *i.e.*, growth rates are time dependent. The differential equation form of this model is

$$\frac{dP}{dt} = \kappa P(t)(\log(K) - \log(P(t))) = \kappa P(t) \log\left(\frac{K}{P(t)}\right), \tag{28}$$

where  $K$  is the carrying capacity and  $\kappa$  scales the time. For both the Logistic and Gompertz models, we let  $X_0$  represent the initial condition, *i.e.*,  $P(t_0) = X_0$ .

The Logistic and Gompertz models, while not nested, are related through a limiting process. Since

$$\lim_{\nu \rightarrow \infty} \nu \left(1 - \left(\frac{P(t)}{K}\right)^{\frac{1}{\nu}}\right) = -\log\left(\frac{P(t)}{K}\right), \tag{29}$$

we find that the Gompertz model is the limit as  $\nu \rightarrow \infty$  of the *generalized* logistic model for  $\nu > 0$

$$\frac{dP}{dt} = \nu \kappa P(t) \left(1 - \left(\frac{P(t)}{K}\right)^{\frac{1}{\nu}}\right). \tag{30}$$

### 3.4. Logistic Model: Numerical Discretization

Another model we consider is a discrete numerical approximation of the Logistic model. We note that the continuous models described above were simulated using the `ode45` algorithm in Matlab. In order to ensure that the logistic model can be coupled to a discrete time model for a *D. magna* population in which the population size is updated once per day [5], we investigated the logistic model using a forward Euler scheme that was discretized on an hour basis. In this paper, we refer to this discrete Euler-method logistic model as the DEL model. This numerically discretized logistic model is given by the difference equation

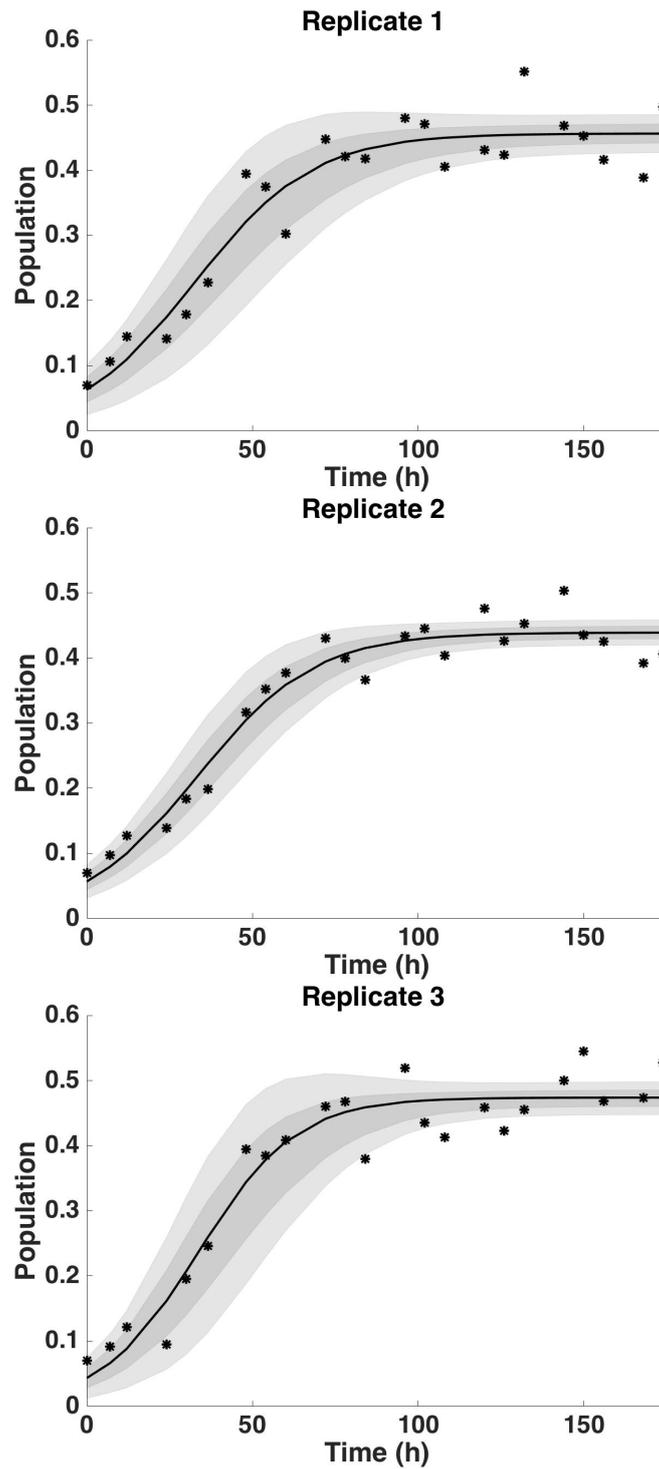
$$P_{t+1} = P_t + RP_t \left(1 - \frac{P_t}{K}\right) \tag{31}$$

where  $P_t$  is the population at time  $t$  hours and  $P_{t+1}$  is the population at the next time step. The parameters  $R$  and  $K$  are analogous to those of the continuous Logistic model and can be interpreted as the intrinsic population growth rate and carrying capacity, respectively. We refer to the initial population,  $P_{t=0}$ , as  $X_0$  in our results and data fitting procedure.

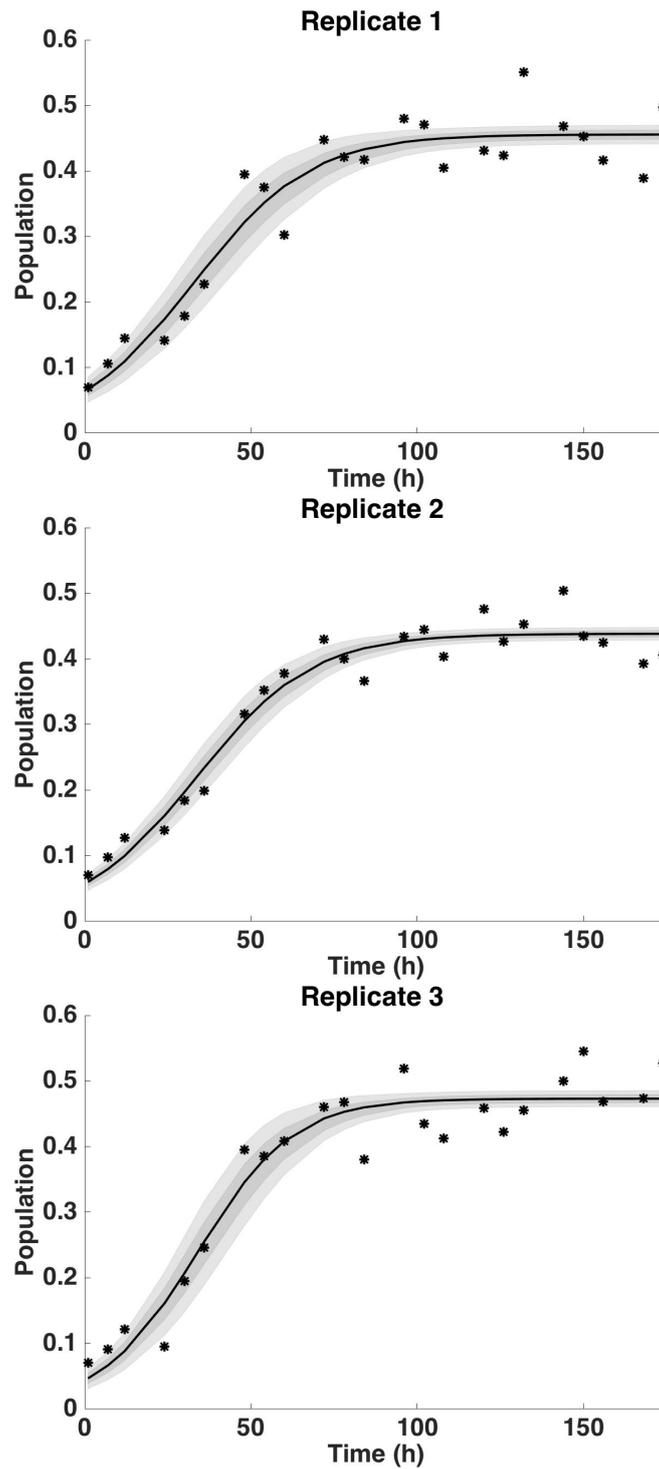
## 4. Results

### 4.1. Data fitting and Model Comparisons

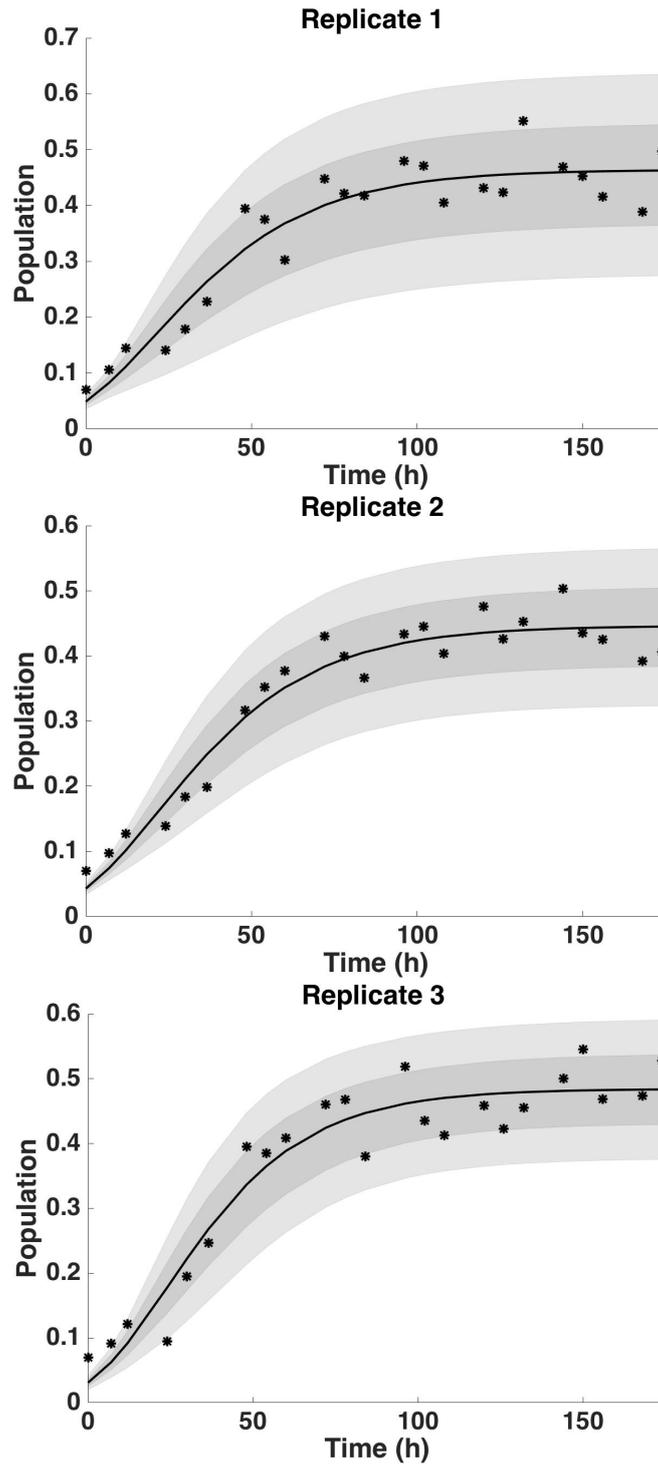
Overall, we found that all models provide a reasonable fit to the data. Figures 1–4 show results of the least squares estimation for the three different replicates of the Logistic, Bernoulli, DEL, and Gompertz models. These figures contain 68% and 95% confidence bands around the fits to data. These were constructed by generating 1000 random parameter sets from a normal distribution described by the mean and standard error obtained by the asymptotic theory results, computing the model for each of these parameter sets, and then calculating the respective confidence intervals from model generated points  $f(t_i, \theta^k)$ , where  $k = 1, \dots, 1000$  [5]. One primary difference between the fits to data that we found was that each model tends to underestimate the initial data and the Bernoulli model provided the closest fit (Figure 4). We note that we chose to estimate the initial condition due to measurement error associated with a low cell density as well as how much influence these discrepancies in error would affect the outcome of the model.



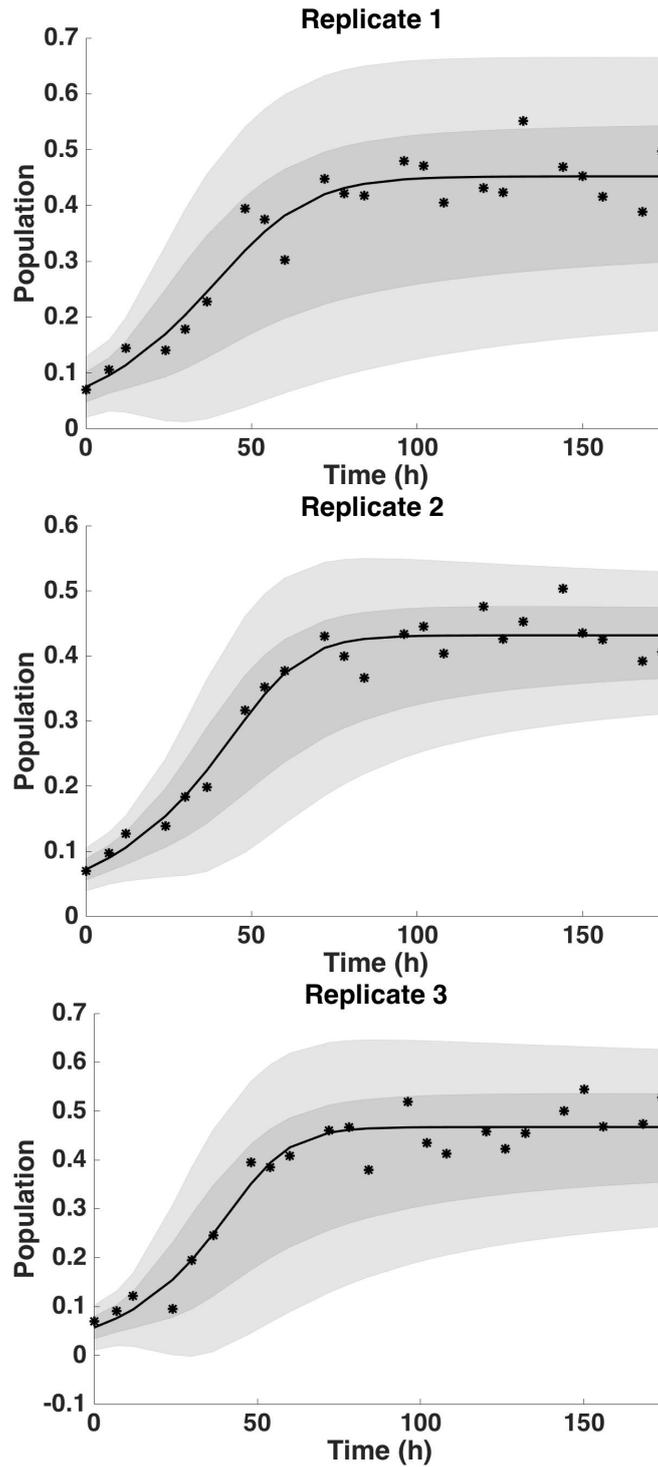
**Figure 1.** Plots of forward solutions for the Logistic curve for the three replicates of the data. Replicate one is on top and three is on the bottom. The lighter and darker shades of grey represent the 95% and 68% confidence bars on the model solution, respectively. The algae population is represented as  $cells \times 10^7/L$ . Data points are shown as “\*”.



**Figure 2.** Plots of forward solutions for the discrete Euler-method logistic (DEL) please confirm. model for the three replicates of the data from left to right. Replicate one is on top and three is on the bottom. The lighter and darker shades of grey represent the 95% and 68% confidence bars on the model solution, respectively. The algae population is represented as  $cells \times 10^7/L$ . Data points are shown as “\*”.



**Figure 3.** Plots of forward solutions for the Gompertz curve for the three replicates of the data. Replicate one is on top and three is on the bottom. The lighter and darker shades of grey represent the 95% and 68% confidence bars on the model solution, respectively. The algae population is represented as  $cells \times 10^7/L$ . Data points are shown as “\*”.



**Figure 4.** Plots of forward solutions for the Bernoulli curve for the three replicates of the data. Replicate one is on top and three is on the bottom. The lighter and darker shades of grey represent the 95% and 68% confidence bars on the model solution, respectively. The algae population is represented as  $cells \times 10^7/L$ . Data points are shown as “\*”.

The results in Table 1 show the small sample corrected Akaike Information Criterion ( $AIC_c$ ) scores based on Equation (25) for each replicate and each model. These results suggest that the discrete and continuous logistic population models are better able to describe the green algae

growth data than either the Gompertz or Bernoulli models, although the differences in  $AIC_c$  values were small between all models. The results of the model comparison test performed between the continuous Logistic model and the Bernoulli model are given in Table 2. In addition to fixing  $\beta$  to 1 to reduce the Bernoulli model to the Logistic model, we also fixed the initial condition  $X_0$  at the seed value to enunciate the model improvement with an unalterable initial condition. Of all comparisons, only the Logistic model with  $X_0$  fixed on replicates 2 and 3 yielded a significant result at the  $\alpha = 0.9$  confidence level (or  $\chi^2(1)_{.9} = 2.706$ ). This indicates that, in general, the OLS cost associated with the Bernoulli model was significantly improved by fixing  $\beta$  and reducing it to the Logistic model. However, restricting the Logistic model further by fixing  $X_0$  does not significantly affect the cost of the Bernoulli model. The Logistic model also has benefits with regards to identifiability, which will be seen in subsequent passages of this document.

**Table 1.** Corrected Akaike Information Criterion scores for each model and replicate.

Replicate	Gompertz	Logistic	Bernoulli	Discrete Euler-Method	Logistic (DEL)
1	−69.4203	−71.5919	−69.2189		−72.6155
2	−84.2435	−89.0016	−89.3905		−90.4114
3	−71.3972	−74.2560	−72.4414		−75.3515

**Table 2.** Model comparison  $\hat{u}_n$  scores for the continuous Logistic model compared to Bernoulli model for each replicate. We also chose to fix the initial condition  $X_0$  at the seed population value to enunciate model improvement if  $X_0$  was unalterable. Note that values less than 2.706 indicate the restricted model is better.

Replicate	Bernoulli Restricted to:	Bernoulli Restricted to:
	Logistic	Logistic with $X_0$ fixed
1	0.5935	0.7233
2	2.4718	3.6216
3	1.1733	3.4118

#### 4.2. Uncertainty Analysis

We compared results of parameter estimation and multiple uncertainty quantification techniques (asymptotic theory and bootstrapping) for the Logistic, Gompertz, and Bernoulli growth models, as well as the numerically discretized version of the Logistic model (DEL model), since each of them provided reasonable fits to the data. We first note that the usual assumption of *i.i.d.* residuals required for uncertainty analysis held for all models investigated (Supplementary Figures S1–S4). Although methods involving autocorrelation on residuals may be used to investigate the *i.i.d.* assumption, we investigated this assumption by visually inspecting residual plots, since there were not enough data to perform autocorrelation tests. Visual inspection of residual plots is a commonly used procedure when not enough longitudinal data are available (see [10], Section 3.6). We also note that the normality assumption for the parameter distributions in asymptotic theory was confirmed by our bootstrapping results in all but the Bernoulli model (Supplementary Figures S5–S17). We divide our analysis of the results from uncertainty quantification among the sets of parameters with similar interpretations for each model. For example, each model has an initial condition, a growth rate, and a parameter governing the saturation of growth due to population density.

##### 4.2.1. Uncertainty Analysis: Initial Condition

The bootstrapping distribution results of  $X_0$  estimation are presented in Supplementary Figures S5–S8. These appear to be approximately normally distributed, with some exceptions

occurring where the estimates are close to the zero boundary. Bootstrapping estimates of uncertainty for  $X_0$  are compared to asymptotic theory in Supplementary Tables S1–S4. We observed that the parameter estimates for  $X_0$  for all replicates were lowest in the Gompertz model and highest in the Bernoulli model. The standard errors also varied between the asymptotic and bootstrapping techniques depending on the model. For example, the order of magnitude for the standard errors from bootstrapping was greater than asymptotic theory for the Gompertz model and the DEL model.

#### 4.2.2. Uncertainty Analysis: Growth Rate

Each model that we investigated has a parameter that describes the population growth rate (Logistic and DEL:  $R$ , Gompertz  $\kappa$ , Bernoulli  $R$ ). Because numerical estimates for the growth rate parameters will not be equal across models, we analyzed their consistency and uncertainty across replicate data sets within the same model. The bootstrapping distributions for the growth rates for each model were normally distributed except for the Bernoulli model, which was skewed to the right (Supplementary Figures S9–S12). We postulate that one reason for this skewness may be that the Bernoulli model is over-parameterized. Similarly, the standard errors computed for the growth rate within each model were of reasonable size and of the same order of magnitude except for the Bernoulli model (Supplementary Tables S5–S8). The growth rate estimates for the logistic model differed between the continuous version and its numerical discretization using the euler method (DEL model). Specifically, the growth rate estimates for the DEL model were consistently higher. In addition, the asymptotic standard errors for the DEL model were lower than for the continuous logistic model.

#### 4.2.3. Uncertainty Analysis: Saturation Parameter

The saturation parameter  $K$  has the same interpretation for all models we investigated, it is the carrying capacity of the green algae population. The estimates for  $K$  were remarkably similar across all models (Supplementary Tables S9–S12). The standard errors were inconsistent between asymptotic theory and bootstrapping for the Bernoulli and Gompertz models. For example, the asymptotic standard error for the estimate of  $K$  in replicate 1 for the Bernoulli model was 0.1138, whereas the bootstrapping error was 0.0158 (Supplementary Table S12). These results are important, because the asymptotic standard error would result in a much wider confidence band around the model fit to the data, which is indeed the case for replicate 1 of the Bernoulli model (Figure 4, Top). Since the bootstrapping distributions for  $K$  for all of the models are normally distributed, this indicates that the bootstrapping standard errors are accurate (Supplementary Figures S13–S16).

#### 4.2.4. Uncertainty Analysis: Bernoulli Model Parameter $\beta$

The parameter  $\beta$  is unique to the Bernoulli model, and scales the rate at which the green algae population reaches carrying capacity. In particular, when  $\beta = 1$ , the Bernoulli model reduces to the logistic model. We found that the estimates for  $\beta$  with asymptotic theory and bootstrapping were  $>1$ . We can not confidently say that these estimates are accurate, since the corresponding standard errors are unreasonably high for both uncertainty techniques and for all three replicates (Table 3). Moreover, the bootstrapping distributions for  $\beta$  were not normally distributed and heavily skewed to the right for all three replicates (Supplementary Figure S17), indicating the possible presence of correlations with other model parameters. We speculate that the parameters  $\beta$  and  $K$  may be correlated for the Bernoulli model.

**Table 3.**  $\beta$  estimate and standard error for the Bernoulli model.

Asymptotic Results: $\beta$	Replicate	Estimate	SE
	1	2.1646	2.5440
	2	3.4574	2.8118
	3	2.8188	2.8758
Bootstrapping Results: $\beta$	Replicate	Estimate	SE
	1	38.31	113.72
	2	29.78	92.89
	3	43.27	113.81

### 4.3. Coupling to the Discrete-Time *Daphnia magna* Population Model

Our uncertainty analysis results indicate that the Logistic model and its numerically discretized counterpart, the DEL model, are the most accurate models among those we investigated. We forgo a summary of the evidence for this conclusion until the discussion section. Our ultimate aim of validating a model for green algae growth was to couple it to a model for *D. magna* population dynamics. In this section, we couple two validated models for algae and *Daphnia* population growth to create a theoretical, unvalidated model to explore potentially complex predator-prey dynamics.

The *D. magna* model we use is one that we recently validated with experimental population data [5]. The validated *D. magna* model is a specification of the following discrete-time discrete-age structured population model:

$$\begin{bmatrix} p(t+1,1) \\ p(t+1,2) \\ p(t+1,3) \\ \vdots \\ p(t+1,i_{max}) \end{bmatrix} = \begin{bmatrix} a(t,1) & a(t,2) & a(t,3) & \dots & a(t,i_{max}) \\ b(t,1) & 0 & 0 & \dots & 0 \\ 0 & b(t,2) & 0 & \dots & 0 \\ \vdots & & \ddots & \dots & \vdots \\ 0 & 0 & 0 & \dots & b(t,i_{max}-1) \end{bmatrix} \begin{bmatrix} p(t,1) \\ p(t,2) \\ p(t,3) \\ \vdots \\ p(t,i_{max}) \end{bmatrix}. \tag{32}$$

The population is divided into one-day age classes up to some maximum age  $i_{max}$  and the number of daphnids of age  $i$  at a time  $t$  is  $p(t,i)$ . The average fecundity of each age class  $i$  is given by  $a(t,i)$  and the survival rate for daphnids of age  $i$  is given by  $b(t,i)$ . The validated functional forms are  $a(t,i) = \alpha(i)(1 - q)^{M(t-\tau)}$  and  $b(t,i)$  is defined piecewise as  $\mu(1 - c)^{M(t)}$  if  $i \leq 4$  and  $\mu$  if  $i \geq 5$ . A summary of the parameters and variables in the model are listed in Table 4 (see [5] for further details).

We coupled the *D. magna* population model to the DEL green algae model by assuming that *D. magna* consumes green algae and that the density-dependent survival and fecundity rates of *D. magna* are influenced directly by the algae concentration. We modeled the algae population with predation as

$$P_{t+1} = P_t + RP_t \left( 1 - \frac{P_t}{K} \right) - \delta M_t P_t \tag{33}$$

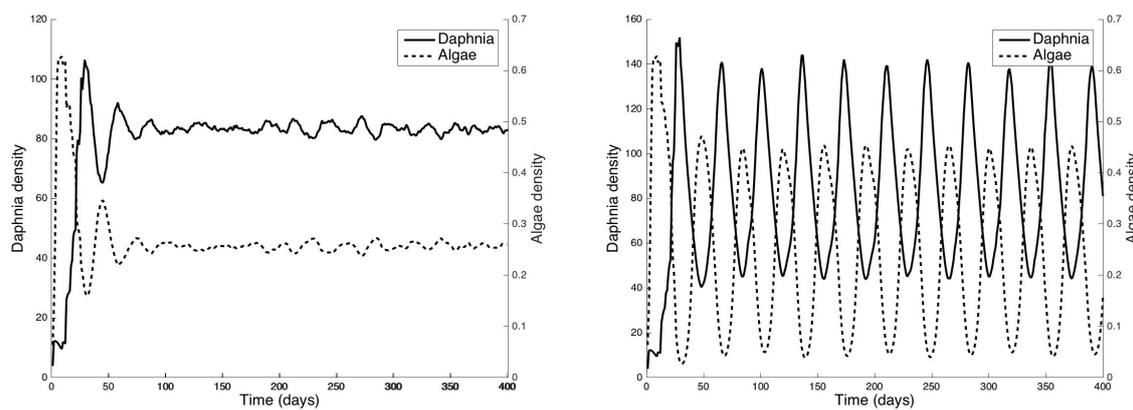
where  $\delta$  is a predation coefficient, and  $M_t$  is the total *Daphnia* biomass at time  $t$ . We chose this functional form based on the assumption that *Daphnia* consume algae at a rate proportional to the density of algae and the total biomass of the *Daphnia* population, as adult daphnids consume food at a higher rate than younger ones. We used a 24 h time discretization to model algae growth for our simulation study and transformed parameters accordingly, setting  $\delta = 0.001$ ,  $K = 0.4559$ ,  $R = 1.34$ , and the initial algae population  $P_0 = 0.0633$ . We modeled the algal influence on *Daphnia* fecundity and survivorship by setting  $a(t,i) = \alpha(i)(1 - q)^{1/P_t - \tau}$  and  $b(t,i) = \mu(1 - c)^{1/P_t}$  if  $i \leq 4$ . In this model, we changed the functional form of the *Daphnia* model based on the assumption that the negative density-dependent fecundity and survivorship effects that daphnids experience are driven by a lack

of food in the form of algae, represented as  $\frac{1}{P_i}$  and its time-delayed analogue  $\frac{1}{P_{t-\tau}}$ . The *Daphnia* matrix model is otherwise unchanged.

**Table 4.** Summary of *Daphnia magna* and algae model parameters and variables.

Parameter/Variable	Description	Units
$p(t, i)$	Number of daphnids of age $i$	# of daphnids
$N(t)$	Total population size at time $t := \sum_{i=1}^{i_{max}} p(t, i)$	# of daphnids
$q$	Density-dependent fecundity constant	dimensionless
$\alpha(i)$	Density-independent fecundity rates	# neonates · daphnid <sup>-1</sup> · day <sup>-1</sup>
$\mu$	Density-independent survival rate	day <sup>-1</sup>
$\tau$	Delay for density-dependent fecundity	days
$c$	Density-dependent survival constant	dimensionless
$M(t)$	Total biomass at time $t := \sum_{i=1}^{i_{max}} p(t, i) \frac{kZ_0 e^{ri}}{k + Z_0(e^{ri} - 1)}$	mm
$k$	Average maximum daphnid size (major axis)	mm
$r$	Average daphnid growth rate	mm/hour
$Z_0$	Average neonate size (major axis)	mm
$R$	Intrinsic growth rate of algae	cells · 10 <sup>7</sup> · L <sup>-1</sup> · day <sup>-1</sup>
$K$	Algal population carrying capacity	cells · 10 <sup>7</sup> · L <sup>-1</sup>
$\delta$	Density dependent predation constant	mm <sup>-1</sup> · cells · 10 <sup>-7</sup>

We found that a coupled model could result in both steady state dynamics as well as oscillatory behavior for different choices of parameter values in the *Daphnia* model (Figure 5). The deterministic simulations in Figure 5 left show steady state behavior with small deviations relative to the population size. The seemingly random perturbations are due to the age-dependency of the density-independent fecundity rate  $\alpha(i)$ . We found that lowering the density-dependent survival competition parameter  $c$  yielded sustained oscillations, and increasing it led to both populations reaching a steady state. Other parameter combinations may also yield similar dynamics, but detailing those values is not the aim of the present study.



**Figure 5.** Simulations of the coupled daphnia and green algae dynamics model resulting in steady state behavior (Left,  $c = 0.01$ ) and sustained oscillations (Right,  $c = 0.04$ ).

### 5. Discussion

Our results highlight the importance of performing uncertainty quantification in validated biological models, even in the simple case of saturating growth dynamics encountered for green algae. For example, the ordinary least squares regression seemed to indicate that each of the models we investigated provide reasonable fits to the algae growth data. In addition, parameter estimates were consistent between replicate data sets for each model. From this information alone,

one might conclude that the Bernoulli model was the best performing model, since it best fit the initial data  $X_0$  (Figure 4). However, a deeper investigation with uncertainty analysis allowed us to generate confidence bands around the fits to the data, showing that the Bernoulli model was the model for which we could have the least confidence. This result was corroborated by a thorough examination of the standard errors for each of the parameters with similar interpretations across all models. For example, the growth rates for the Bernoulli model had unreasonably high standard errors, while the uncertainty in the corresponding growth rates for the other models were relatively low. Our results also indicated that the Gompertz model had inconsistent standard errors between asymptotic theory and bootstrapping for the initial condition and saturation parameter, emphasizing the importance of using multiple uncertainty quantification techniques to ascertain the best validated model. We observed that the logistic and DEL models have different confidence regions for the model solution in Figures 1 and 2. We attribute the change in computed confidence regions to the differences in numerical discretization, the time step for the DEL model was equal to one hour while the logistic model had a coarser mesh.

We collected replicate data and our results had strong agreement across the three replicates. Although methodology exists to fit all three data sets simultaneously to the same model, we chose to fit them separately to test whether the model validation results were consistent across several repeated experiments. We noticed a slight trend in the residuals resulting from fits of the model to data for each replicate: the fit sometimes either over- or under-estimates the data in groups of threes (Supplementary Figures S1–S4). We surmise that this phenomenon may be explained by human measurement error; some people tend to over or under count the algae when using a hemocytometer. Since all models were confounded with this possibility for human error, we can assume that human error did not influence the analysis by favoring one model over any other. In future work, more accurate and consistent cell counts may be performed with a flow cytometer. Alternatively, a spectrophotometer may be used to approximate algae concentrations.

Overall, we suggest that the population growth of *Raphidocelis subcapitata* is most accurately modeled using the Logistic equation among the simple growth models we investigated. It is important to note that our findings are limited to controlled laboratory settings with unchanging temperature, constant photoperiod, and no change in nutrient availability. For example, we did not consider the possible influence of photoperiod (light/dark) conditions on algae growth. We also did not consider the affect of limiting nutrients such as carbon, nitrogen, phosphorous, or sulfur. The models we investigated here represent our first approximation of algae growth and seemed to fit the data well even without considering how light affects algae growth and that the affect of limiting nutrients could be described by saturating algae growth. The simplifying assumption we made that growth parameters are independent of light conditions may be investigated in future work to yield a closer fit to the data. Our work here serves the purpose of coupling our green algae model with one of zooplankton (*D. magna*) population growth in a laboratory setting, e.g., for toxicity testing, but should not be directly extrapolated to populations in lakes. In order to develop an accurate model of community fluctuations in the field, we will need to consider predation by various zooplankton and microbes, competition with other algae, nutrient fluctuation, abiotic drivers, and habitat heterogeneity. This study is, however, a useful step toward developing a more comprehensive model. In particular, our results showing that a coupling a validated green algae model with a validated daphnia model are important because it exemplifies the possibility of using a mathematical model to recapitulate the oscillatory dynamics seen in nature. In contrast, the previously validated daphnia model that did not include algae dynamics was not able to produce oscillations and only resulted in steady state behavior [5], a result that did not account for the broader range of plankton dynamics seen in natural systems [1]. We note that our ultimate goal is to validate a coupled continuous time daphnia/algae model since continuous time models, such as the Sinko-Streifer model, are described by partial differential equations (PDEs) with a continuously structured variable and can be more amenable to the estimation of age-dependent parameters than a discretely structured

model [14–17]. In this work, we investigated the dynamics of a coupled algae/daphnia discrete-time model as a coarse approximation to understand the qualitative impact of including a dynamic food source on daphnia populations. Our finding that increasing the density-dependent survival constant ( $c$ ) in a coupled predator-prey model yields oscillatory dynamics compliments previous work that has predicted limit cycles based on increased mortality [18]. In an ecological setting, changes in the parameter  $c$  could reflect differing nutrient (algae) requirements on the density-dependent survival of daphnia due to differences in species size (e.g., *Daphnia pulex* vs. *Daphnia magna*) or other increased sources of density-dependent mortality such as predation on daphnia. Changes in  $c$  may also be induced toxicologically. For example, endocrine toxins are known to alter the molt cycle of adult daphnids through incomplete ecdysis [19], which may have an indirect affect on density-dependent survival by lowering competition for algae. Together, these results suggest that a structural change in the validated daphnia model, *i.e.*, including predation, and not just a change in parameter values is required to reproduce population oscillations observed in laboratory and natural settings. This finding is important in the context of our previous and current ongoing efforts, since oscillations were not observed in our previous daphnia population experiments [5]. Although the experiments performed in this work did not involve toxins, the species investigated are commonly used in toxicity assessments. Thus, our results can be used to provide a baseline to compare effects in a toxicity setting in future work.

**Supplementary Materials:** The following are available online at [www.mdpi.com/2076-3417/6/5/155/s1](http://www.mdpi.com/2076-3417/6/5/155/s1).

**Acknowledgments:** This research was supported in part by the National Science Foundation under NSF Undergraduate Biomathematics grant number DBI-1129214, NSF grant number DMS-0946431, NSF grant number DMS-1514929 and in part by the Air Force Office of Scientific Research under grant numbers AFOSR FA9550-12-1-0188 and AFOSR FA9550-15-1-0298. The authors are grateful to Amanda Laubmeier and Kaska Adoteye for their assistance in collection of the data used in our efforts.

**Author Contributions:** Michael Stemkovski conceived and designed the experiments and performed simulations; Michael Stemkovski and Robert Baraldi performed the experiments, wrote the code, and performed the uncertainty analysis; Michael Stemkovski, Robert Baraldi, H. T. Banks, and Kevin B. Flores analyzed the results and wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. McCauley, E.; Murdoch, W.W. Cyclic and stable populations: Plankton as paradigm. *Am. Nat.* **1987**, *129*, 97–121.
2. Benincà, E.; Dakos, V.; van Nes, E.H.; Huisman, J.; Scheffer, M. Resonance of plankton communities with temperature fluctuations. *Am. Nat.* **2011**, *178*, E85–E95.
3. McCauley, E.; Nisbet, R.M.; Murdoch, W.W.; de Roos, A.M.; Gurney, W.S. Large-amplitude cycles of *Daphnia* and its algal prey in enriched environments. *Nature* **1999**, *402*, 653–656.
4. Scheffer, M.; Rinaldi, S.; Kuznetsov, Y.A.; van Nes, E.H. Seasonal dynamics of *Daphnia* and algae explained as a periodically forced predator-prey system. *Oikos* **1997**, *80*, 519–532.
5. Adoteye, K.; Banks, H.T.; Cross, K.; Eytcheson, S.; Flores, K.B.; LeBlanc, G.A.; Nguyen, T.; Ross, C.; Smith, E.; Stemkovski, M. Statistical validation of structured population models for *Daphnia magna*. *Math. Biosci.* **2015**, *266*, 73–84.
6. Wang, Y.H.; Wang, G.; LeBlanc, G.A. Cloning and characterization of the retinoid X receptor from a primitive crustacean *Daphnia magna*. *Gen. Comp. Endocrinol.* **2007**, *150*, 309–318.
7. Caperon, J. Population growth in micro-organisms limited by food supply. *Ecology* **1967**, *48*, 715–722.
8. Schanz, F.; Zahler, U. Prediction of algal growth in batch cultures. *Schw. Zeit. Hydr.* **1981**, *43*, 103–113.
9. Thompson, P.-A.; Couture, P. Short-and long-term changes in growth and biochemical composition of *Selenastrum capricornutum* populations exposed to cadmium. *Aqua. Toxic.* **1991**, *21*, 135–143.
10. Banks, H.T.; Thompson, W.C.; Hu, S. *Modeling and Inverse Problems in the Presence of Uncertainty*; CRC Press: Boca Raton, FL, USA, 2014.

11. Banks, H.T.; Tran, H.T. *Mathematical and Experimental Modeling of Physical and Biological Processes*; CRC Press: Boca Raton, FL, USA, 2009.
12. Seber, G.A.; Wild, C.J. *Nonlinear Regression*; Wiley: Hoboken, NJ, USA, 2003.
13. Richards, F.J. A flexible growth function for empirical use. *J. Exp. Bot.* **1959**, *10*, 290–300.
14. Banks, H.T.; Banks, J.E.; Dick, L.K.; Stark, J.D. Estimation of dynamic rate parameters in insect populations undergoing sublethal exposure to pesticides. *Bull. Math. Biol.* **2007**, *69*, 2139–2180.
15. Banks, H.T.; Davis, J.L.; Ernstberger, S.L.; Hu, S.; Artimovich, E.; Dhar, A.K. Experimental design and estimation of growth rate distributions in size-structured shrimp populations. *Inv. Probl.* **2009**, *25*, 095003, doi:10.1088/0266-5611/25/9/095003.
16. Banks, J.E.; Dick, L.K.; Banks, H.T.; Stark, J.D. Time-varying vital rates in ecotoxicology: Selective pesticides and aphid population dynamics. *Ecol. Model.* **2008**, *210*, 155–160.
17. Wood, S.N. Obtaining birth and mortality patterns from structured population trajectories. *Ecol. Monogr.* **1994**, *64*, 23–44.
18. May, R.M.; Conway, G.R.; Hassell, M.P.; Southwood, T.R.E. Time delays, density-dependence and single-species oscillations. *J. Anim. Ecol.* **1974**, *43*, 747–770.
19. Wang, Y.H.; Kwon, G.; Li, H.; LeBlanc, G.A. Tributyltin synergizes with 20-Hydroxyecdysone to produce endocrine toxicity. *Toxicol. Sci.* **2011**, *123*, 71–79.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).