QUARTERLY OF APPLIED MATHEMATICS
VOLUME LXXVII, NUMBER 1
MARCH 2019, PAGES 1–18
https://doi.org/10.1090/qam/1514
Article electronically published on August 28, 2018

PARAMETER IDENTIFICATION AND SENSITIVITY ANALYSIS FOR A PHYTOPLANKTON COMPETITION MODEL

By

THOMAS STOJSAVLJEVIC (*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O.Box 413, Milwaukee, WI 53201-0413*),

GABRIELLA PINTER (*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O.Box 413, Milwaukee, WI 53201-0413*),

ISTVAN LAUKO (*Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O.Box 413, Milwaukee, WI 53201-0413*),

AND

NICHOLAS MYERS (*Department of Mathematics, NC State University, P.O.Box 8205, Raleigh, NC 27695*)

Abstract. Phytoplankton live in a complex environment with two essential resources, light and nutrients, forming various gradients. Light supplied from above is never homogeneously distributed in a body of water due to refraction and absorption from biomass present in the ecosystem and other sources. Nutrients in turn are typically supplied from below mixed-up by diffusion from the benthic region. Here we present a model of two phytoplankton species competing in a deep freshwater lake for light and two nutrients, one of which is assumed to be preferred. The model is comprised of a system of nonlinear, non-local partial differential equations with appropriate boundary conditions. The parameter space of the model is analyzed for parameter identifiability - the ability for a parameter's true value to be recovered through optimization, and for global sensitivity - the influence a parameter has on model response. The results of these analyses are interpreted within their biological context.

1. Introduction. The vertical distribution of phytoplankton in a water column has been the subject of numerous investigations [3, 5, 11, 13, 16, 21, 26]. Prominent vertical patterns observed include deep chlorophyll maxima (DCMs), benthic layers, and surface

Received September 20, 2017.
2010 *Mathematics Subject Classification.* Primary 35Q92, 35R30, 49K40.
This research was supported in part by the National Science Foundation, UBM-Institutional grant: DUE-1129056.
Email address: tgs@uwm.edu
Email address: gapinter@uwm.edu
Email address: iglauko@uwm.edu
Email address: nmyers2@ncsu.edu

scums [21, 26]. Deep chlorophyll maxima is the layering phenomenon where biomass accumulation happens beneath the surface of the water column. Surface scums is the layering pattern where up to 90% of the biomass concentration is near the surface resulting in heavy shading of the water. Benthic layers occur in stratified bodies of water. Stratification occurs due to energy influx induced temperature and density gradients as well as advection within the body of water. The thickness of these layers is controlled by the degree of mixing, and can vary from several centimeters to tens of meters [21]. These distributions affect primary production [24] and foodweb structure in aquatic ecosystems, and are affected by changing environmental conditions and global warming [12].

From laboratory experiments, it is known that phytoplankton require inorganic nitrogen, sulfur, and phosphorous compounds along with trace elements and vitamins. For freshwater environments, phosphorus and nitrogen are often nutrients present in low amounts that limit biomass growth [4, 27]. In particular, free phosphorus is available mostly from geochemical sources within aquatic ecosystems [17] and from recycling mediated by bacteria. Nitrogen, in comparison to phosphorus, is typically more abundant in aquatic ecosystems [11]. In the environment, nitrogen is usually available in the form of ammonium, nitrate, or dissolved organic nitrogen. The relationship between ammonium and nitrate uptake and its impact on growth has long been studied [6, 11]. Classically, it is understood that ammonium is the preferred form of nitrogen compared to nitrate due to the extra step of nitrate reduction to ammonium. In this view, ammonium is taken up as long as it is available, and only when it is depleted nitrate uptake begins.

Various mathematical models have been developed to describe the dynamics of phytoplankton populations [5, 13, 21, 26, 29]. While certain models attribute the layering phenomenon to physical forces such as wind shear, advection within the body of water and mixing processes, other models suggest that layer formation is due to phytoplankton competing in a stratified environment with non-homogeneous light and nutrient distributions in the vertical dimension. Light is never homogeneously distributed in aquatic environments since it forms a gradient over biomass and other light-absorbing substances [15, 16]. Further, continuous residence in the illuminated layers is neither necessary nor optimal for growth [32].

In this paper we build on a model described in [21, 26] and incorporate preferential nutrient uptake for two phytoplankton species. Model simulations will be compared to layering profiles from Lake Michigan. While the qualitative agreement between our simulation results and the data is evident, many of the parameters used in our model are approximate at best. Some of the parameters are difficult to measure while other parameters have not been specifically calibrated to the environment in Lake Michigan. With enough data available, it might be reasonable to identify these parameters computationally. To this end, we first studied the inverse problem to see which parameters can be computationally identified. We followed ideas in [2] and generated synthetic data using the model itself. Next, we solved the inverse problem under increasing noise levels added to the synthetic data to examine the impact using noisy data on the identifiability of the parameters. Additionally, we performed a global sensitivity analysis to study which parameters are most influential in phytoplankton dynamics. The identification of the

most important parameters guides laboratory efforts to determine them, and is a crucial step towards achieving better quantitative agreement between our model and the data.

2. Model development. The proposed model studies the coexistence of two phytoplankton species, b_1 and b_2, in a water column where there are two limiting nutrients, R_1 and R_2, present with the assumption that one nutrient, R_1, is preferentially taken up. The full model consists of a non-local, non-linear system of integro-partial differential equations for the depth distributions of biomass densities $b_1(z,t)$ and $b_2(z,t)$, the limiting nutrient concentrations $R_1(z,t)$ and $R_2(z,t)$, and light $I(z,t)$. For simplifying purposes we consider a one dimensional water column indexed by z, where $0 \leq z \leq z_b$ and $z = 0$ represents the surface and $z = z_b$ represents the depth of the water column.

Biomass densities are assumed to be limited by the availability of light and nutrients. The functions $f_{I,1}(I(z,t))$ and $f_{I,2}(I(z,t))$ will represent the phytoplankton growth rate when light is limiting for species 1 and species 2, respectively, while the functions $f_{R,1}(R_1(z,t), R_2(z,t))$ and $f_{R,2}(R_1(z,t), R_2(z,t))$ represent the phytoplankton growth rate when nutrients are limiting for species 1 and 2, respectively. The gross phytoplankton growth rate of each species follows the Liebig law of the minimum so that the per-capita growth rate will be given by the equations $g_k(z,t) = \min[f_{I,k}(I(z,t)), f_{R,k}(R_1(z,t), R_2(z,t))] - m_k$, where m_k is the loss rate, and the index $k = 1, 2$ represents the respective species.

Each phytoplankton species is assumed to move within the water column. Their movement is affected by the processes of diffusion and active movement. Passive movement is modeled by eddy diffusion with diffusion coefficients D_{b_1} and D_{b_2} which are assumed to be uniform throughout the water column. This assumption is not necessary and, in general, depth dependent diffusion is permissible [14,26]. To model the active movement, we introduce a velocity function ν which is dependent on the gradient of the growth rate, $\frac{\partial g_k}{\partial z}$, i.e., $\nu = \nu_k\left(\frac{\partial g_k}{\partial z}\right)$ for $k = 1, 2$ [28]. Given how phytoplankton can regulate their position, to an extent, depending on whether light or nutrients are limiting, the biological assumption is introduced that phytoplankton will move up if growth conditions are better above than they are below, and phytoplankton will move down if the conditions are better below than they are above, and phytoplankton will not move if the conditions are worse above and below [21,26]. Positive velocity is oriented upward and associated to the negative z direction. Specifically we assume $\nu(\cdot)$ is an odd, decreasing function which approaches a value $\nu_{k_{max}}$ as $\frac{\partial g_k}{\partial z}$ approaches negative infinity and approaches $-\nu_{k_{max}}$ as $\frac{\partial g_k}{\partial z}$ approaches positive infinity for $k = 1, 2$ [19, 21, 26, 28].

The resulting partial differential equations for biomass densities are given by

$$\begin{aligned}\frac{\partial b_1}{\partial t} &= \min(f_{I,1}(I), f_{R,1}(R_1, R_2))b_1 - m_1 b_1 + D_{b_1}\frac{\partial^2 b_1}{\partial z^2} + \frac{\partial}{\partial z}\Big[\nu_1\Big(\frac{\partial g_1}{\partial z}\Big)b_1\Big] \\ &= [\text{Growth}] - [\text{Loss}] + [\text{Passive movement}] + [\text{Active movement}] \end{aligned} \tag{2.1}$$

and

$$\begin{aligned}\frac{\partial b_2}{\partial t} &= \min(f_{I,2}(I), f_{R,2}(R_1, R_2))b_2 - m_2 b_2 + D_{b_2}\frac{\partial^2 b_2}{\partial z^2} + \frac{\partial}{\partial z}\Big[\nu_2\Big(\frac{\partial g_2}{\partial z}\Big)b_2\Big] \\ &= [\text{Growth}] - [\text{Loss}] + [\text{Passive movement}] + [\text{Active movement}]. \end{aligned} \tag{2.2}$$

These partial differential equations are given no flux boundary conditions

$$\left[D_{b_1}\frac{\partial b_1}{\partial z} + \nu_1\left(\frac{\partial g_1}{\partial z}\right)b_1\right]\Bigg|_{z=0} = \left[D_{b_1}\frac{\partial b_1}{\partial z} + \nu_1\left(\frac{\partial g_1}{\partial z}\right)b_1\right]\Bigg|_{z=z_b} = 0 \tag{2.3}$$

and

$$\left[D_{b_2}\frac{\partial b_2}{\partial z} + \nu_2\left(\frac{\partial g_2}{\partial z}\right)b_2\right]\Bigg|_{z=0} = \left[D_{b_2}\frac{\partial b_2}{\partial z} + \nu_2\left(\frac{\partial g_2}{\partial z}\right)b_2\right]\Bigg|_{z=z_b} = 0. \tag{2.4}$$

The functions $f_{R,1}(R_1, R_2)$ and $f_{R,2}(R_1, R_2)$ take the form of a modified Michaelis-Menten relation given by

$$f_{R,1}(R_1(z,t), R_2(z,t)) = r_1\left(\frac{R_1(z,t)}{K_{R_1,1} + R_1(z,t)} + \frac{R_2(z,t)}{K_{R_2,1} + R_2(z,t)}\,\frac{\lambda^n}{\lambda^n + R_1(z,t)}\right)$$

and

$$f_{R,2}(R_1(z,t), R_2(z,t)) = r_2\left(\frac{R_1(z,t)}{K_{R_1,2} + R_1(z,t)} + \frac{R_2(z,t)}{K_{R_2,2} + R_2(z,t)}\,\frac{\lambda^n}{\lambda^n + R_1(z,t)}\right),$$

where r_1 and r_2 are the maximal growth rate of species 1 and species 2, respectively, and the parameters $K_{R_1,1}$ and $K_{R_2,1}$ represent the saturation constants for species 1 with the corresponding nutrient, the parameters $K_{R_1,2}$ and $K_{R_2,2}$ represent the saturation constants for species 2 with the corresponding nutrient, $K_{I,1}$ and $K_{I,2}$ represent the light saturation constants for species 1 and 2, respectively, and λ is an inhibition coefficient. Following [18] the functions $f_{I,k}$ for $k = 1, 2$ are hyperbolic tangent functions of the form

$$f_{I,1}(I(z,t)) = r_1 \tanh\left(\frac{I(z,t)}{K_{I,1}}\right)$$

and

$$f_{I,2}(I(z,t)) = r_2 \tanh\left(\frac{I(z,t)}{K_{I,2}}\right).$$

The use of hyperbolic tangent functions over the classic Michaelis-Menten relation is due to better results from regression analysis on experimental data [18].

As in [26], the active movement functions ν_1 and ν_2 take the form

$$\nu_1\Big(\frac{\partial g_1}{\partial z}\Big) = -\nu_{1_{max}}\,\frac{\frac{\partial g_1}{\partial z}}{|\frac{\partial g_1}{\partial z}| + K_{swim}}$$

and

$$\nu_2\Big(\frac{\partial g_2}{\partial z}\Big) = -\nu_{2_{max}}\,\frac{\frac{\partial g_2}{\partial z}}{|\frac{\partial g_2}{\partial z}| + K_{swim}}.$$

Nutrients in the water column are impacted by diffusion processes and by phytoplankton through consumption and recycling. Let D_{R_1} and D_{R_2} represent the diffusion coefficients of the two nutrients, respectively, and let $\varepsilon_{1,1}$, $\varepsilon_{2,1}$, $\varepsilon_{1,2}$, and $\varepsilon_{2,2}$ represent the proportion of nutrients that are available in the water column from dead phytoplankton. Since both species use both nutrients to grow, it is necessary to determine the proportion of each nutrient lost to uptake processes. Define the functions γ and ρ by

$$\gamma_k(R_1) = r_k \, \frac{R_1}{K_{R_1,k} + R_1}$$

and

$$\rho_k(R_1, R_2) = r_k \, \frac{R_2}{K_{R_2,k} + R_2} \, \frac{\lambda^n}{\lambda^n + R_1}$$

for $k = 1, 2$. These functions will be used to give the appropriate proportion of each nutrient uptaken by each species. Using these modifications and the physical and biological assumptions the partial differential equations for the nutrients are given by

$$\begin{aligned}
\frac{\partial R_1}{\partial t} =& - \frac{b_1}{Y_1} \min(f_{I,1}(I), f_{R,1}(R_1, R_2)) \, \frac{\gamma_1(R_1)}{f_{R,1}(R_1, R_2)} \, + \varepsilon_{1,1} m_1 \, \frac{b_1}{Y_1} \\
& - \frac{b_2}{Y_2} \min(f_{I,2}(I), f_{R,2}(R_1, R_2)) \, \frac{\gamma_2(R_1)}{f_{R,2}(R_1, R_2)} \, + \varepsilon_{2,1} m_2 \, \frac{b_2}{Y_2} \\
& + D_{R_1} \frac{\partial^2 R_1}{\partial z^2} \\
=& -[\text{Species 1 Uptake}] + [\text{Species 1 Recycling}] - [\text{Species 2 Uptake}] \\
& + [\text{Species 2 Recycling}] + [\text{Mixing}]
\end{aligned} \tag{2.5}$$

and

$$\begin{aligned}
\frac{\partial R_2}{\partial t} =& - \frac{b_1}{Y_1} \min(f_{I,1}(I), f_{R,1}(R_1, R_2)) \, \frac{\rho_1(R_1, R_2)}{f_{R,1}(R_1, R_2)} \, + \varepsilon_{1,2} m_1 \, \frac{b_1}{Y_1} \\
& - \frac{b_2}{Y_2} \min(f_{I,2}(I), f_{R,2}(R_1, R_2)) \, \frac{\rho_2(R_1, R_2)}{f_{R,2}(R_1, R_2)} \, + \varepsilon_{2,2} m_2 \, \frac{b_2}{Y_2} \\
& + D_{R_2} \frac{\partial^2 R_2}{\partial z^2} \\
=& -[\text{Species 1 Uptake}] + [\text{Species 1 Recycling}] - [\text{Species 2 Uptake}] \\
& + [\text{Species 2 Recycling}] + [\text{Mixing}].
\end{aligned} \tag{2.6}$$

For boundary conditions on the nutrient equations, assume that nutrients cannot enter or leave the water column from the surface, and that both nutrients are supplied from the bottom. Assume nutrients in the sediment have a constant concentration. Let the concentration of nutrient 1 in the sediment be denoted by R_{in_1} and let the concentration of nutrient 2 in the sediment be denoted by R_{in_2}. Then the boundary conditions for (2.5) and (2.6) are given, respectively, by

$$\left.\frac{\partial R_1}{\partial z}\right|_{z=0} = 0, \quad \left.\frac{\partial R_1}{\partial z}\right|_{z=z_b} = h(R_{in_1} - R_1(z_b)) \tag{2.7}$$

and

$$\left.\frac{\partial R_2}{\partial z}\right|_{z=0} = 0, \quad \left.\frac{\partial R_2}{\partial z}\right|_{z=z_b} = h(R_{in_2} - R_2(z_b)), \tag{2.8}$$

where h represents the permeability of the sediment-water interface.

Finally, to describe the change in light at depth z, we modify the Lambert-Beer law to accommodate the presence of multiple phytoplankton species. The Lambert-Beer law becomes

$$I(z,t) = I_{in} \exp\left[- \int_0^z (a_1 b_1(w,t) + a_2 b_2(w,t) + a_{bg}) dw \right], \tag{2.9}$$

where the parameters I_{in}, a_1, a_2 and a_{bg} represent the incident light, species specific light attenuation coefficients, and background attenuation coefficient, respectively [20]. The incident light is considered to be constant in our simulations, although this assumption can be relaxed to admit a time dependent function $I_{in}(t)$. This model is based on [21] with the addition of another phytoplankton species, the competition for two nutrient resources, one of which is assumed to be preferred, and light dependence on growth functions consistent with experimental results.

3. Model simulation. In order to simulate the model, we use a finite volume method as described for a related system in [14]. Under this approach the spatial differential operators are replaced by discrete approximations, and the integral term for the light attenuation in (2.9) is approximated by Simpson's rule. For the convective term we used a third order upwinding scheme, while the diffusive term is approximated by centered difference [14]. This results in a large non-linear system of ordinary differential equations in time of the form

$$\frac{d\,\mathbf{W}(t)}{dt} = \mathbf{F}(\mathbf{W}(t)),\ t \geq 0, \tag{3.1}$$

where the vector $\mathbf{W} \in \mathbb{R}^{4N}$ contains the components $b_i(z_j,t)$ and $R_i(z_j,t)$, for $i = 1, 2$, $j = 1, \ldots N$, resulting from the uniform spatial discretization of the depth, $0 \leq z \leq z_b$. System (3.1) is solved by numerical integration. Since the ODE model (3.1) is generally stiff, to solve the system numerically, an implicit integration method is needed. In this treatment, we implemented the model in MATLAB, and integrated using MATLAB's ODE solver ode15s [1].

Parameter interpretation and values used for the simulations can be found in Table 1. The constants for light and nutrients used in the growth rate affect the proportion of light and nutrients needed by a species. For this reason, these parameters are chosen such that $K_{I,2} < K_{I,1}$, $K_{R_1,2} > K_{R_1,1}$, and $K_{R_2,2} > K_{R_2,1}$. This way the competing species will have growth functions defined in such a way that the first species has a growth rate with proportionally larger requirement on light, while the second species has a growth rate with larger nutrient requirements. With these assumptions, the phytoplankton species have potential natural niches located at different depths that they will want to occupy. Finally, if $m_1 = m_2$, then the principle of competitive exclusion would hold, and the model would give the extinction of one of the phytoplankton species. Thus we choose $m_2 < m_1$.

TABLE 1. Parameters used in the simulation of the model given by equations (2.1)-(2.9).

Parameter	Explanation	Value	Source
N	Spatial discretization level	200	
z_b	Water column depth (m)	100	[21]
R_{in_1}	Sediment concentration Nutrient 1 (μg L^{-1})	150	
R_{in_2}	Sediment concentration Nutrient 2 (μg L^{-1})	100	[21]
h	Sediment-water column permeability (m^{-1})	10^{-2}	[21]
I_{in}	Incoming light (μmol photons m^{-2} s^{-1})	1,400	[21]
a_{bg}	Background attenuation coefficient (m^{-1})	0.35	[21], [22]
a_1	Species 1 algal attenuation coefficient (m^{-1} [cells ml^{-1}]$^{-1}$)	10^{-5}	[21], [22]
a_2	Species 2 algal attenuation coefficient (m^{-1} [cells ml^{-1}]$^{-1}$)	10^{-5}	[21], [22]
D_{b_1}	Species 1 biomass diffusion coefficients (m^2 d^{-1})	10	[21]
D_{b_2}	Species 2 biomass diffusion coefficients (m^2 d^{-1})	10	[21]
D_{R_1}	Nutrient 1 diffusion coefficient (m^2 d^{-1})	10	[21]
D_{R_2}	Nutrient 2 diffusion coefficient (m^2 d^{-1})	10	[21]
$\nu_{1_{max}}$	Species 1 swimming speed (m d^{-1})	10	[21]
$\nu_{2_{max}}$	Species 2 swimming speed (m d^{-1})	10	[21]
r_1	Species 1 maximum growth rates (d^{-1})	0.4	[21]
r_2	Species 2 maximum growth rates (d^{-1})	0.4	[21]
m_1	Species 1 loss rate (d^{-1})	0.2	[21]
m_2	Species 2 loss rate (d^{-1})	0.1	
$K_{R_1,1}$	Nutrient 1 saturation constant for Species 1 (μg L^{-1})	1	[21]
$K_{R_1,2}$	Nutrient 1 saturation constant for Species 2 (μg L^{-1})	1	[21]
$K_{R_2,1}$	Nutrient 2 saturation constant for Species 1 (μg L^{-1})	10	
$K_{R_2,2}$	Nutrient 2 saturation constant for Species 2 (μg L^{-1})	10	
$K_{I,1}$	Light saturation constant for Species 1 (μmol photons m^{-2} s^{-1})	50	[21]
$K_{I,2}$	Light saturation constant for Species 2 (μmol photons m^{-2} s^{-1})	5	
Y_1	Species 1 yield coefficient (cells ml^{-1} [μg L^{-1}]$^{-1}$)	10^3	[21]
Y_2	Species 2 yield coefficient (cells ml^{-1} [μg L^{-1}]$^{-1}$)	10^3	[21]
$\varepsilon_{i,1}$	Species i Nutrient 1 recycling coefficient (dimensionless)	0.9	[26]
$\varepsilon_{i,2}$	Species i Nutrient 2 recycling coefficient (dimensionless)	0.9	[26]
K_{swim}	Swimming constant (m^{-1} d^{-1})	0.001	[26]
λ	Nutrient uptake inhibition factor (dimensionless)	0.8	
n	Shaping parameter	4	

For initial conditions assume $b_1(z,0) = b_2(z,0) = 10^4$ cells ml^{-1}, $R_1(z,0) = 2.25$ μg L^{-1}, and $R_2(z,0) = 2.1$ μg L^{-1} holds for all z, $0 \leq z \leq z_b$. To investigate the phytoplankton layering patterns, model simulations are run for a time period long enough for numerical changes in phytoplankton and nutrient distributions to approximately stabilize. For simplifying purposes we consider situations when $\nu_{max_1} = \nu_{max_2}$ and $\varepsilon_{1,1} = \varepsilon_{1,2} = \varepsilon_{2,2} = \varepsilon_{2,2}$, although these assumptions can be relaxed.

As hypothesized, the model under this parametrization gives a vertically heterogeneous water column. Studying panel A in Figure 1 we can see that the two species form layers whose thickness and location within the water column vary. While this result supports the hypothesis of MacArthur and Levins [25] to explain why various phytoplankton species coexist, it is not necessarily feasible to consider this approach when studying field data. To get a better understanding of how the model compares to field data, total chlorophyll is simulated along with light transmission. We note that species that are adapted to less light have more chlorophyll per cell, so we used different weights for

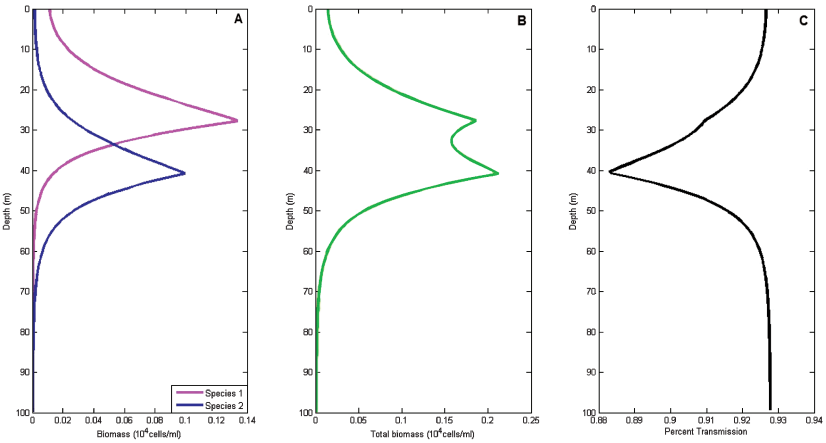

FIG. 1. (A) Species specific biomass vs. depth. (B) Simulated total chlorophyll vs. depth. (C) Simulated light transmission vs depth.

the two species to simulate chlorophyll content. Model simulations qualitatively agree with phytoplankton profiles collected in Lake Michigan [3, 21]. Current simulations are performed using parameters reported in the literature, and they may need to be adjusted for the specific environment to achieve better quantitative agreement. To analyze model sensitivity to changes in its parameter space we turn to parameter estimation and global sensitivity analysis.

4. Parameter estimation. In order to implement the model and run simulations, values for the model parameters need to be supplied. Well-known ranges for these quantities are reported in [5, 13, 21, 26, 29]. While quantities such as incident light and depth of the water column can be easily measured, other parameters are either difficult to measure or have wide reported ranges.

When experimental or field data is unavailable, the parameter estimation problem can be done on simulated data [2]. Analysis of estimation techniques and results using simulated data provide information about which parameters can be identified computationally. Further, solving estimation problems on simulated data provide insights into the type of data that is needed, as well as how frequently it must be collected in order to successfully identify unknown parameters.

4.1. *Deterministic estimation problem.* To solve the inverse problem, we set up a deterministic estimation problem using the simulated data for total biomass with random noise. Adopting the notation used in [2], let B_i be observations taken at T_n time points. We want to estimate the parameter $\mathbf{q}$ in the parameter space $Q = \mathbb{R}^{26+}$. To this end we seek to minimize

$$J(q, B) = \sum_{i=1}^{T_n} \frac{1}{|B_i|} |Y(t_i; q) - B_i| + P(q) \tag{4.1}$$

over $\mathbf{q} \in Q$, where

$$\begin{aligned} Y(t;q) =& \frac{z_b}{2N}\left[\sum_{k=1}^{N}\left(b_1(w_{k+1},t,q)+b_1(w_k,t,q)\right)+\sum_{k=1}^{N}\left(b_2(w_{k+1},t,q)+b_2(w_k,t,q)\right)\right] \\ \approx& \int_0^{z_b}\left(b_1(w,t,q)+b_2(w,t,q)\right)dw. \end{aligned} \tag{4.2}$$

$P(q)$ is a penalty function to prevent parameters from taking on negative values. In this application, $\mathbf{q}$ is a vector of the form

$$\begin{aligned} \mathbf{q} =& [R_{in_1}, R_{in_2}, h, I_{in}, a_{bg}, a_1, a_2, r_1, r_2, m_1, m_2, K_{R_1,1}, K_{R_1,2}, K_{R_2,1}, \\ & K_{R_2,2}, K_{I,1}, K_{I,2}, Y_1, Y_2, \varepsilon_{1,1}, \varepsilon_{1,2}, \varepsilon_{2,1}, \varepsilon_{2,2}, K_{swim}, \lambda, n]^T. \end{aligned}$$

Note that the diffusion parameters D_{b_1}, D_{b_2}, D_{R_1}, D_{R_2} and the swimming speeds $\nu_{1_{max}}$ and $\nu_{2_{max}}$ were excluded from the parameter vector. This exclusion was made due to numeric stability of the solution to the model. When the diffusion coefficients are large, artificial oscillations are introduced to the solution and when the swimming speeds are made too large the numerical solutions become singular.

Various algorithms exist to solve the optimization problem given by (4.1). In this discussion we utilized MATLAB's fminsearch function which uses the Nelder-Mead algorithm [23].

4.2. *Parameter estimation based on simulated data.* We analyze the sensitivity of the model to the inverse problem by solving the minimization problem using synthetic data. To generate the simulated data, we define $B_i^{\mathcal{N}}$ by

$$B_i^{\mathcal{N}} = (1+\eta_i^{\mathcal{N}}(\alpha))B_i,$$

where $\eta_i^{\mathcal{N}}$ is a normally distributed random variable with mean 0 and standard deviation of $\frac{\alpha}{3}$. This corresponds to 99.74% of $\eta_i^{\mathcal{N}}$ taking values in the interval $[-\alpha, \alpha]$. Finally, B_i is determined from equation (4.2) by evaluating $Y(t;q)$ at time points $t_i = 2i$, where $i = 0, \ldots, 400$. The initial guess for each parameter was formed by using the reported values in Table 1 and increasing it by 5%. The parameter estimation problem was solved using noise levels $\alpha = 5\%$ and 10% to analyze the impact random noise in the data set has on the model's ability to recover the original parameters.

Overall, the two parameter estimation problems give estimated parameters that are close to the true values as shown in Table 2. The parameters with the largest deviations from their true values were I_{in}, Y_1, and Y_2. While the optimized values differ from the true value, they are within the reported ranges in the literature and are biologically feasible. Certain parameters such as the incoming light can be accurately measured both in the field and in the lab eliminating the practical need of inclusion in the estimation problem. Other parameters such as the yield coefficients Y_1 and Y_2 are species dependent and can have a wide range of values [30]. Further, while there are ways to empirically determine the yield coefficients in a laboratory setting, the experiments are time consuming and involve isolating individual species and growing them in culture.

TABLE 2. Parameter estimation results for noise levels α

Parameter	True Value	$\alpha = 0.05$	$\alpha = 0.1$
R_{in_1}	150	174.3856	156.2622
R_{in_2}	100	107.9321	108.1423
h	0.01	0.009927	0.008398
I_{in}	1400	1219.377	1452.365
a_{bg}	0.35	0.156314	0.167582
a_1	0.00001	0.00000118	0.00000371
a_2	0.00001	0.00000949	0.00000989
r_1	0.4	0.37524	0.40251
r_2	0.4	0.407551	0.419206
m_1	0.2	0.20312	0.201482
m_2	0.1	0.099219	0.100393
$K_{R_1,1}$	1	0.775008	1.143783
$K_{R_1,2}$	1	1.051802	0.875298
$K_{R_2,1}$	10	13.12762	10.53918
$K_{R_2,2}$	10	7.720061	9.967949
$K_{I,1}$	50	77.52113	92.60311
$K_{I,2}$	5	5.18302	4.806114
Y_1	1000	975.9295	1108.372
Y_2	1000	1080.346	1124.755
$\varepsilon_{1,1}$	0.1	0.106558	0.112787
$\varepsilon_{2,1}$	0.1	0.086903	0.090758
$\varepsilon_{1,2}$	0.1	0.098099	0.108165
$\varepsilon_{2,2}$	0.1	0.137817	0.108417
K_{swim}	0.01	0.000877	0.001001
λ	1	1.045527	1.083861
n	4	3.9046	4.775924

5. Global sensitivity analysis. Beyond parameter identifiability, another important aspect of a model's parameter space is measuring how parameters interact to influence the dynamics of the model. Various methods such as local sensitivity analysis, global sensitivity analysis, and Bayesian algorithms are available to test how important parameters and groups of parameters are to model dynamics [33].

Global sensitivity analysis has an advantage over Bayesian methods since these methods tend to not require prior information which is needed for a Bayesian approach and are only dependent on the structure of the model itself. Two common measures for global sensitivity analysis are given by Morris screening and Sobol decomposition [33]. The Sobol decomposition is a variance based method which is sensitive to the size of the parameter space and the complexity of the model. Screening methods provide an alternative for identifying critical parameters and are able to generally provide a ranking

for parameters in terms of importance. However, unlike the variance based methods, screening methods are not able to quantify how much more important one parameter is compared to another [33].

5.1. *The Morris screening procedure.* Consider the model $y=f(q)$ where $q=[q_1,\dots,q_n]$ is the set of parameters scaled such that $q_i \in [0,1]$. To construct the Morris screening, partition $[0,1]$ into ℓ-levels. Then the elementary effect associated with the ith input is given by the difference quotient

$$d_i(q) = \frac{f(q_1,\dots,q_{i-1},q_i+\Delta,\dots,q_n) - f(q_1,\dots,q_n)}{\Delta} = \frac{f(q+\Delta e_i)-f(q)}{\Delta},$$

where the step size Δ is chosen from the set

$$\Delta \in \left\{\frac{1}{\ell-1},\dots,1-\frac{1}{\ell-1}\right\} = \Gamma_\ell.$$

The elementary effects d_i approximate large scale, local sensitivity at the point q. The step size is taken large to cover the entire parameter space. For r sample points, the sensitivity measures for q_i are the sample mean and the sample variance given by

$$\mu_i^* = \frac{1}{r}\sum_{j=1}^{r} |d_i^j(q)|$$

and

$$\sigma_i^2 = \frac{1}{r-1}\sum_{j=1}^{r}\left(d_i^j(q) - \frac{1}{r}\sum_{j=1}^{r} d_i^j(q)\right)^2,$$

where

$$d_i^j = \frac{f(q^j+\Delta e_i)-f(q^j)}{\Delta}$$

is the elementary effect associated with the ith parameter in the jth sample [33]. The mean quantifies the individual impact that a parameter has on the output, while the variance estimates the effects of interactions with other inputs.

In this application, q is a vector of the form

$$\begin{aligned} q =& [R_{in_1}, R_{in_2}, h, I_{in}, a_{bg}, a_1, a_2, D_{R_1}, D_{R_2}, D_{b_1}, D_{b_2}, \nu_{1_{max}}, \nu_{2_{max}}, \\ & r_1, r_2, m_1, m_2, K_{R_1,1}, K_{R_1,2}, K_{R_2,1}, K_{R_2,2}, K_{I,1}, K_{I,2}, Y_1, Y_2, \\ & \varepsilon_{1,1}, \varepsilon_{1,2}, \varepsilon_{2,1}, \varepsilon_{2,2}, K_{swim}, \lambda, n]^T. \end{aligned}$$

Note that unlike the parameter estimation problem, the diffusion coefficients and the swimming speeds are included in this procedure. When implementing this procedure, ℓ was chosen sufficiently large (ℓ=10) and was done 50 times before constructing μ^* and σ^2.

Examining Figure 2 we see that the most important parameters for $b_1(z,t)$ in q are q_{17}, q_4, and q_{16} which correspond to the parameters m_2, I_{in}, and m_1, respectively. From a biological perspective, I_{in} and m_1 are expected as they represent incident light and the death rate of the first phytoplankton species, respectively. For $b_2(z,t)$ we see that the most important parameters in q are q_{17}, q_5, and q_{15} which correspond to the parameters m_2, a_{bg}, and r_2, respectively.

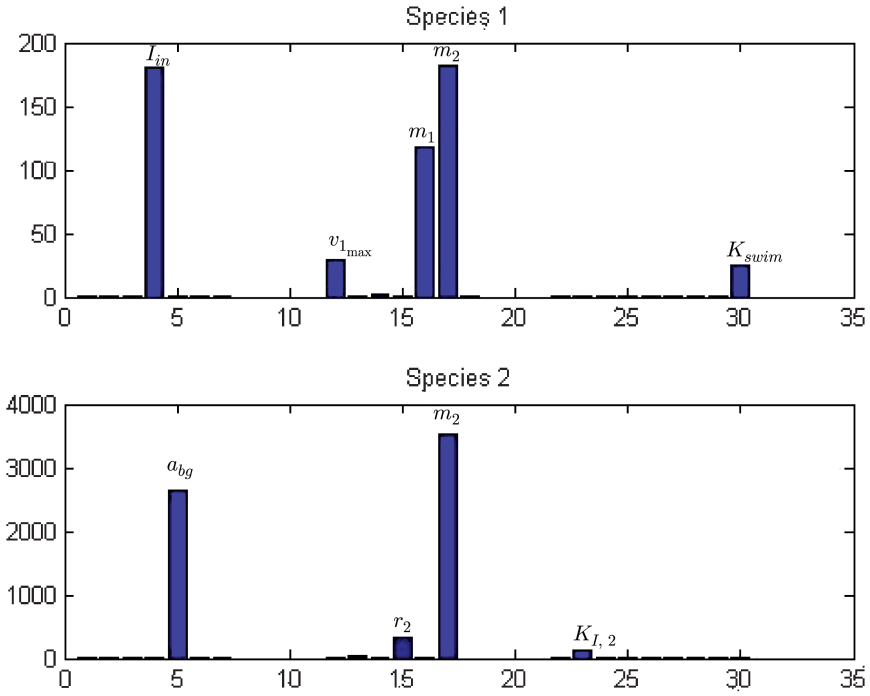

FIG. 2. Morris screening results for the phytoplankton biomass by species. Parameters in the q vector are on the horizontal axis in order as listed in the vector. The averages μ_i^* are displayed on the vertical axis.

The parameter m_2, representing the loss rate of species 2, was chosen so that the principle of competitive exclusion would not hold allowing for the coexistence of the two phytoplankton species in the model simulations. For that reason its importance for b_1 and b_2 is expected. Having both I_{in} and a_{bg} being identified by the screening procedure emphasizes the importance of understanding how phytoplankton species compete and utilize light for coexistence.

For the nutrients $R_1(z,t)$ and $R_2(z,t)$ the Morris screening procedure highlights a different group of parameters for importance. Examining Figure 3 we see that the most important parameters in q for $R_1(z,t)$ are q_{25}, q_{27}, and q_8 which correspond to Y_2, $\varepsilon_{1,2}$, and r_2, respectively. In comparison, for $R_2(z,t)$ the parameters q_{25}, q_{29}, and q_{24} which correspond to Y_2, $\varepsilon_{2,2}$, and Y_1 are the top three ranked parameters. Given the assumptions of how the biomass groups were constructed: one species favoring light, the other favoring nutrients, the dependence on the amount of nutrient in each cell and how much of the preferred nutrient released upon cell death is biologically feasible to allow coexistence.

Parameter rankings for $b_1(z,t)$, $b_2(z,t)$, $R_1(z,t)$, and $R_2(z,t)$ are given in Table 3.

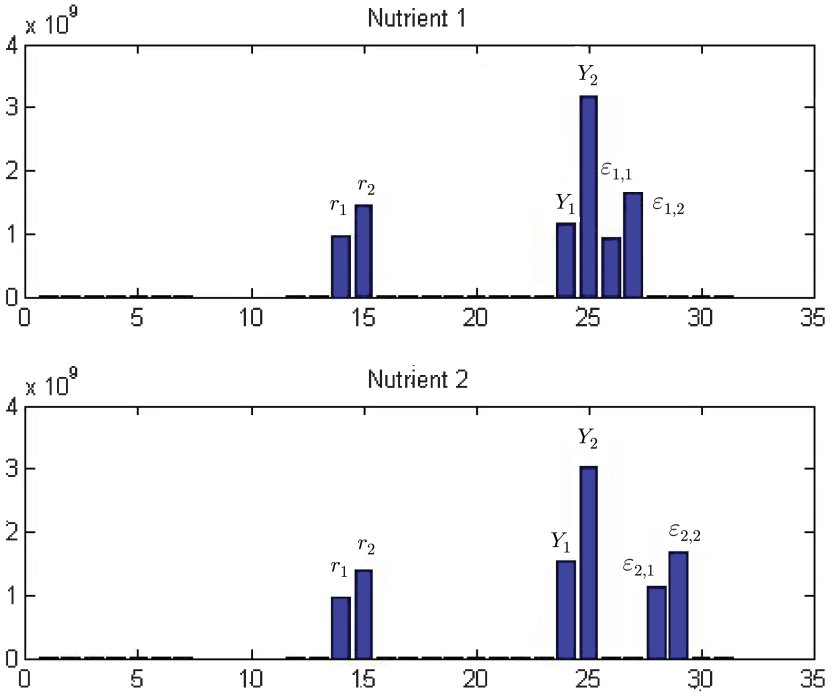

FIG. 3. Morris screening results for the nutrients. Parameters in the q vector are on the horizontal axis in order as listed in the vector. The averages μ_i^* are displayed on the vertical axis.

TABLE 3. Ranking of top 5 parameters in q from the Morris screening.

Ranking	$b_1(z,t)$	$b_2(z,t)$	$R_1(z,t)$	$R_2(z,t)$
1	m_2	m_2	Y_2	Y_2
2	I_{in}	a_{bg}	$\varepsilon_{1,2}$	$\varepsilon_{2,2}$
3	m_1	r_2	r_2	Y_1
4	$\nu_{1_{max}}$	$K_{I,2}$	Y_1	r_2
5	K_{swim}	$\nu_{2_{max}}$	r_1	$\varepsilon_{2,1}$

The diffusion parameters D_{b_1}, D_{b_2}, D_{R_1}, and D_{R_2} as well as the shape parameter n all had a mean of 0 under the Morris screening. This indicates that these parameters are not influential for the model and can be fixed in subsequent model calibration, sensitivity analysis, and uncertainty quantification to reduce model complexity [33]. Other parameters that were highlighted such as swimming speeds and light half saturation constants further demonstrate the importance of phytoplankton's abilities to regulate their position in bodies of water to adapt to various resource gradients.

5.2. *The Sobol decomposition.* The construction of the Sobol decomposition is adapted from [33] and [35]. Consider the scalar valued non-linear model $Y = f(q)$ where $q = [q_1, q_2, \ldots, q_p]$ are defined so that the q_i are independent random variables uniformly

distributed on $[0,1]$. The Sobol indices are based on a second-order High Dimensional Model Representation (HDMR) given by

$$f(q) = f_0 + \sum_{i=1}^{p} f_i(q_i) + \sum_{1\le i<j\le p} f_{ij}(q_i, q_j). \tag{5.1}$$

Since the representation is not unique, constraints are placed on f_i and f_{ij} to ensure uniqueness of the representation. This is accomplished by minimizing the functional

$$\mathcal{J} = \int_{[0,1]^p} \Big[f(q) - \Big(f_0 + \sum_{i=1}^{p} f_i(q_i) + \sum_{1\le i<j\le p} f_{ij}(q_i, q_j) \Big) \Big]^2 dq$$

subject to

$$\int_0^1 f_{i_1,\dots,i_s}(q_{i_1}, \dots, q_{i_s}) dq_{i_k} = 0$$

for $k = 1, \dots, s$ and $s = 1, \dots, p$. The component functions f_i and f_{ij} are given by

$$f_i = \int_{[0,1]^{p-1}} f(q)\ dq_{\sim i}$$

and

$$f_{ij} = \int_{[0,1]^{p-2}} f(q)\ dq_{\sim i,j},$$

where the notation $dq_{\sim i}$ represents $dq_1, \dots, dq_{i-1}, dq_{i+1}, \dots, dq_p$.

The Sobol decomposition method uses the expansion given in (5.1) to quantify the contribution of each parameter to the variance of the response. The total variance of the response Y is defined by

$$D = \mathrm{Var}(Y) = \int_{[0,1]^p} f^2(q)\ dq - f_0^2,$$

where f_0 is the mean response given by

$$f_0 = \int_{[0,1]^p} f(q)\ dq.$$

To define the Sobol indices, we rewrite D as the sum of variances due to first-order and second-order parameter interactions by

$$D = \sum_{i=1}^{p} D_i + \sum_{1\le i<j\le p} D_{ij},$$

where

$$D_i = \int_{[0,1]^p} f_i^2(q_i)\ dq_i$$

and

$$D_{ij} = \int_{[0,1]^{2p}} f_{ij}^2(q_i, q_j)\ dq_i dq_j.$$

The Sobol indices are then defined by

$$S_i = \frac{D_i}{D}, \text{ and } S_{ij} = \frac{D_{ij}}{D},$$

for $i, j = 1, \ldots, p$. The indices given by S_i are the first-order sensitivity indices and measure the contribution of the parameter q_i on the response variance while the indices given by S_{ij} measure the interaction of parameters q_i and q_j on the response variance.

The computation of first- and second-order indices requires $p + \frac{p(p-1)}{2}$ model responses. To reduce the computational cost, total sensitivity indices S_{T_i} which quantify the total impact of a parameter q_i on the response are defined by

$$S_{T_i} = S_i + \sum_{j=1}^{p} S_{ij}.$$

In this application the vector q is the same as the vector used in the Morris screening.

Unlike the Morris screening results presented in Figure 2, the Sobol decomposition results given in Figure 4 show that the loss rates m_1 and m_2 have the largest impact on $b_1(z,t)$ and $b_2(z,t)$, respectively. The other parameters in the vector q had a sensitivity index S_T of 0 indicating there was no impact on the response variance. From a computational standpoint, this result is consistent with the numerical sensitivity the model has to the mortality coefficients. Theoretical analysis of related models indicated that the mortality coefficient was a bifurcation parameter [7, 9, 10].

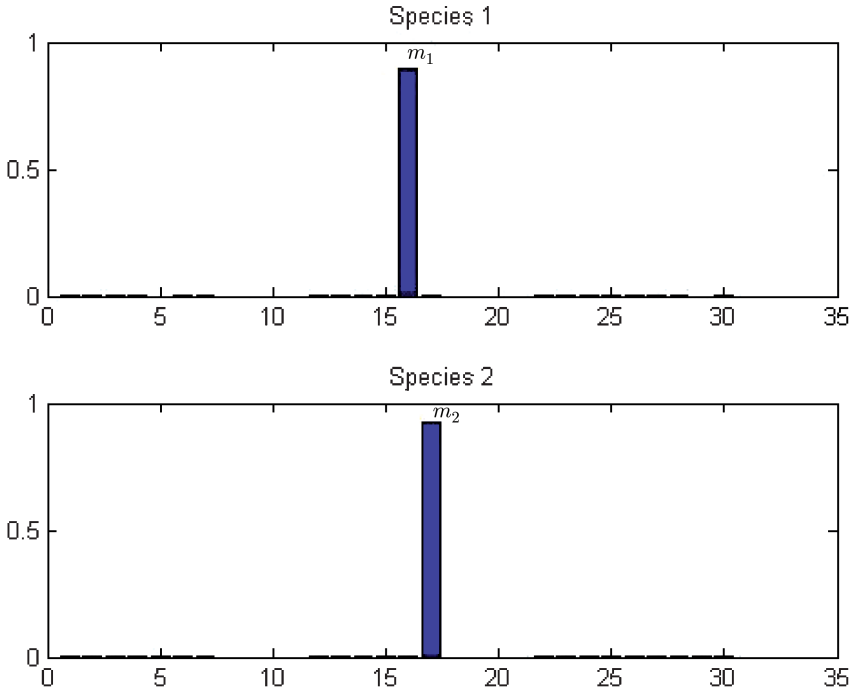

FIG. 4. Sobol decomposition results for the phytoplankton biomass by species. Parameters in the vector **q** are on the horizontal axis in the order as listed in the vector. The total sensitivity indices S_T are displayed on the vertical axis.

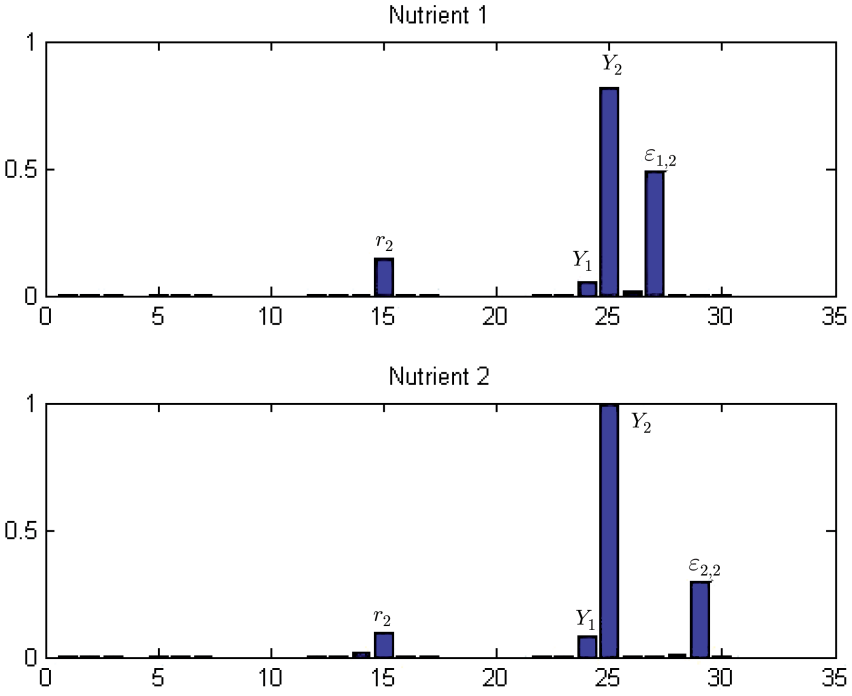

FIG. 5. Sobol decomposition results for the nutrients. Parameters in the q vector are on the horizontal axis in order as listed in the vector. The total sensitivity indices S_T are displayed on the vertical axis.

The Sobol decomposition results for $R_1(z,t)$ and $R_2(z,t)$ are given in Figure 5. As with the Morris screening, the Sobol decomposition method identified the parameters Y_2, $\varepsilon_{1,2}$, Y_1 and r_2 as having the largest impact on $R_1(z,t)$ and the parameters Y_2, $\varepsilon_{2,2}$, Y_1 and r_2 on $R_2(z,t)$. While the method does identify the parameters r_2 and Y_1 having an impact on the response variance, the impact is less than the Morris screening procedure indicates.

6. Conclusion. The two species model with preferential nutrient uptake presented here is able to phenomenologically replicate common depth profiles and offers insights into what physical and physiological parameters are important to consider in model calibration. The parameter estimation problem was successfully solved on 26 of the model's parameters. While simulated data was used in this presentation, this approach offered insights into parameters that are numerically difficult to estimate. Of these, the incident light I_{in} is the easiest to resolve with measurement, and can be fixed for simulations done on data generated from field observations. If equation (2.9) was modified to have a time dependent incident light term, then another inverse problem could be set up and solved to see how the model would respond to data that came from actual measurements.

When the model was analyzed under the Morris screening and Sobol decomposition procedures, the model's full parameter space was considered to determine a ranking of the parameters and their impact on the model's output for $b_1(z,t)$, $b_2(z,t)$, $R_1(z,t)$, and $R_2(z,t)$. The results of the screening algorithm for the biomass were consistent with biological intuition and highlighted the importance of parameters associated with how phytoplankton species use light. The Sobol decomposition method highlighted the possible numerical instability of the solutions when the loss coefficients are made too large. For the nutrient, the yield coefficients were ranked as most important in both methods. Since the yield coefficients Y_1 and Y_2 have a large range of values reported in the literature and are susceptible to being estimated with error, care should be taken with these parameters in simulations.

References

[1] R. Ashino, M. Nagase, and R. Vaillancourt, *Behind and beyond the MATLAB ODE suite*, Comput. Math. Appl. **40** (2000), no. 4-5, 491–512, DOI 10.1016/S0898-1221(00)00175-9. MR1772651

[2] H. T. Banks and K. L. Bihari, *Modelling and estimating uncertainty in parameter estimation*, Inverse Problems **17** (2001), no. 1, 95–111, DOI 10.1088/0266-5611/17/1/308. MR1818494

[3] A. S. Brooks and B. G. Torke, *Vertical and seasonal distribution of Chlorophyll a in Lake Michigan*, Journal of the Fisheries Research Board of Canada (1977), 34:2280-2287.

[4] A. S. Brooks and D. N. Edginton, *Biogeochemical control of phosphorus cycling and primary production in Lake Michigan*, Limnology and Oceanography (1994), 39-4: 961-968.

[5] J. Chattopadhyay, S. Pal, and R. R. Sarkar, *Mathematical modeling of harmful algal blooms supported by experimental findings*, Ecological Complexity (2004), 1: 225-235.

[6] Q. Dortch, *The interaction between ammonium and nitrate uptake in phytoplankton*, Marine ecology press series, Oldendorf (1990), 61-1: 183-201.

[7] Y. Du and S.-B. Hsu, *Concentration phenomena in a nonlocal quasi-linear problem modelling phytoplankton. I. Existence*, SIAM J. Math. Anal. **40** (2008), no. 4, 1419–1440, DOI 10.1137/07070663X. MR2466162

[8] Y. Du and S.-B. Hsu, *Concentration phenomena in a nonlocal quasi-linear problem modelling phytoplankton. II. Limiting profile*, SIAM J. Math. Anal. **40** (2008), no. 4, 1441–1470, DOI 10.1137/070706641. MR2466163

[9] Y. Du and S.-B. Hsu, *On a nonlocal reaction-diffusion problem arising from the modeling of phytoplankton growth*, SIAM J. Math. Anal. **42** (2010), no. 3, 1305–1333, DOI 10.1137/090775105. MR2653252

[10] Y. Du and L. Mei, *On a nonlocal reaction-diffusion-advection equation modelling phytoplankton dynamics*, Nonlinearity **24** (2011), no. 1, 319–349, DOI 10.1088/0951-7715/24/1/016. MR2746150

[11] P. M. Gilbert, F. P. Wilkerson, R. C. Dugdale, J. A. Raven, C. L. Dupont, P. R. Leavitt, and T. M. Kana, *Pluses and minuses of ammonium and nitrate uptake and assimilation by phytoplankton and implications for productivity and community composition, with emphasis on nitrogen-enriched conditions*, Limnology and Oceanography (2015), 61-1: 165-197.

[12] G. C. Hays, A. J. Richardson, and C. Robinson, *Climate change and marine plankton*, Trends in Ecology & Evolution (2005), 20: 337-334.

[13] A. Howard, *Modeling movement patterns of the cyanobacterium Microcystis*, Ecological Applications (2001), 11-1: 304-310.

[14] J. Huisman and B. Sommeijer, *Simulation techniques for the population dynamics of sinking phytoplankton in light-limited environments*, Modeling, Analysis and Simulation (2002), 1-17.

[15] J. Huisman and F. J. Weissing, *Light-limited growth and competition for light in well-mixed aquatic environments: an elementary model*, Ecology (1994), 75-2: 507-520.

[16] J. Huisman and F. J. Weissing, *Competition for nutrients and light in a mixed water column: a theoretical analysis*, The American Naturalist (1995), 146-4: 536-564.

[17] G. E. Hutchinson, *The paradox of the plankton*, The American Naturalist (1961), 95-882: 137-145.

[18] A. D. Jassby and T. Pratt, *Mathematical formulation of the relationship between photosynthesis and light for phytoplankton*, Limnology and Oceanography (1976), 21-4: 540-547.

[19] E. F. Keller and L. A. Segel, *Travelling bands of chemotactic bacteria: A theeoretical analysis*, Journal of Theoretical Biology (1971), 30: 235-248.
[20] J. T. O. Kirk, *Theoretical-analysis of contribution of algal cells to attenuation of light within natural waters 1. General treatment of suspensions of pigmented cells*, New Phytologist (1975), 75: 11-20.
[21] C. A. Klausmeier and E. Lichtman, *Algal games: the vertical distribution of phytoplankton in poorly mixed water columns*, Limnology and Oceanography (2001), 46-8: 1998-2007.
[22] D. Krause-Jensen and K. Sand-Jensen, *Light attenuation and photosynthesis of aquatic plant communities*, Limnology and Oceanography (1998), 43-3: 396-407.
[23] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, *Convergence properties of the Nelder-Mead simplex method in low dimensions*, SIAM J. Optim. **9** (1999), no. 1, 112–147, DOI 10.1137/S1052623496303470. MR1662563
[24] M. A. Leibold, *Resources and predatorss can affect the vertical distributions of zooplankton*, Limnology and Oceanography (1990), 35: 33-45.
[25] R. MacArthur and R. Levins, *Competition, habitat selection, and character displacement in a patchy environment*, Proceedings of the National Academy of Sciences (USA) (1964), 51: 1207-1210.
[26] J. P. Mellard, K. Yoshiyama, E. Litchman, and C. A. Klausmeier, *The vertical distribution of phytoplankton in stratified water columns*, J. Theoret. Biol. **269** (2011), 16–30, DOI 10.1016/j.jtbi.2010.09.041. MR2974462
[27] J. A. Mortonson and A. S. Brooks, *Occurrence of a deep nitrate maximum in Lake Michigan*, Canadian Journal of Fisheries and Aquatic Science (1980), 37:1025-1027.
[28] A. Okubo, *Diffusion and ecological problems: mathematical models*, Biomathematics, vol. 10, Springer-Verlag, Berlin-New York, 1980. An extended version of the Japanese edition, *Ecology and diffusion*; Translated by G. N. Parker. MR572962
[29] O. M. Phillips, *The equilibrium and stability of simple marine biological systems I. Primary nutrient consumers*, The American Naturalist (1973), 107-953: 73-93.
[30] C. S. Reynolds, *The ecology of freshwater phytoplankton*, Cambridge University Press, 1984.
[31] G-Y. Rhee and I. J. Gotham, *The effect of environment factors on phytoplankton growth: light and the interactions of light with nitrate limitation*, Limnology and Oceanography (1981), 26-4: 649-659.
[32] T. J. Smayda, *The suspension and sinking of phytoplankton in the sea*, Annual Review of Oceanography and Marine Biology (1970), 8: 353-414.
[33] R. C. Smith, *Uncertainty quantification*, Computational Science & Engineering, vol. 12, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2014. Theory, implementation, and applications. MR3155184
[34] F. M. Stewart and B. R. Levin, *Partitioning of resources and the outcome of interspecific competition: a model and some general considerations*, The American Naturalist (1973), 107-954: 171-198.
[35] M. T. Wentworth, R. C. Smith, and H. T. Banks, *Parameter selection and verification techniques based on global sensitivity analysis illustrated for an HIV model*, SIAM/ASA J. Uncertain. Quantif. **4** (2016), no. 1, 266–297, DOI 10.1137/15M1008245. MR3479703