Stochastic Forecasting of Algae Blooms in Lakes

P. Wang, D.M. Tartakovsky, and A.M. Tartakovsky

Abstract We consider the development of harmful algal blooms (HABs) in a lake with uncertain nutrients inflow. To quantify the impact of this uncertainty on predictions of the concentrations of various algae groups, we explore two alternative approaches based on the Fokker–Planck equation and PDF methods. Both approaches quantify predictive uncertainty by deriving deterministic equations for joint probability density functions of the algae concentrations. As an example, we study the impact of uncertain initial concentration and inflow–outflow ratio on the evolution of cyanobacteria (the blue-green algae).

1 Introduction

Anthropogenic stresses, such as discharge of wastewater, significantly accelerated eutrophication of many aquatic systems worldwide [1]. As a result, there is an explosion of harmful algae blooms (HABs) that pose serious risks to human and animal health and to ecosystem sustainability. A conservative estimate of annual economic costs of HABs and eutrophication in the USA. alone amounts to $2.2–4.6 billion [2]. Ironically, on the other end of the spectrum, recent research suggests various potential uses of algal biomass, such as biodiesel, animal feed, heating, electricity, and even pharmaceutical and cosmetic products.

Like most eco-dynamics systems, HABs involve complex interactions between different biological species and their predictions rely on mathematical models with a

---

P. Wang (✉) • A.M. Tartakovsky
Pacific Northwest National Laboratory, P.O. Box 999, MSIN K7-90, Richland, WA 99352, USA
e-mail: peng.wang@pnnl.gov; alexandre.tartakovsky@pnnl.gov

D.M. Tartakovsky
Department of Mechanical and Aerospace Engineering, University of California,
San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
e-mail: dmt@ucsd.edu

J.A. Ferreira et al. (eds.), *Modelling and Simulation in Fluid Dynamics in Porous Media*,
Springer Proceedings in Mathematics & Statistics 28, DOI 10.1007/978-1-4614-5055-9_6,

large number of uncertain parameters. A number of recent studies [3–6] employed a probabilistic framework to quantify parametric uncertainty in predictions of HABs. These analyses are typically based on simplifying assumptions and rely on the ensemble variance of concentrations to quantify predictive uncertainty. Since the concentrations of multiple competing algae species are described by a system nonlinear differential equations with multiplicative noise, their probability density functions (PDFs) are typically highly non-Gaussian. Therefore, the concentration variances do not provide information necessary to predict extreme events and to conduct risk assessments of HABs.

We present two alternative frameworks, the Fokker–Planck equation and PDF methods, that enable probabilistic forecasting of HABs in natural environments. Section 2 contains a mathematical formulation of the problem and a brief overview of their uncertain parameterizations. In Sect. 3 we derive the Fokker–Planck (Sect. 3.1) and PDF (Sect. 3.2) equations that are applicable for uncorrelated and correlated system parameters, respectively. Both deterministic equations describe the temporal evolution of the joint PDF of the concentrations of competing algae species. In Sect. 4, we use the Fokker–Planck equation to quantify the impact of uncertain initial concentration and inflow–outflow ratio on the evolution of cyanobacteria (the blue-green algae). Section 5 consists of major conclusions drawn from this study.

2 Problem Formulation

HABs typically occur when nutrients (nitrogen and phosphorous) are abundant, water is warm (>20°C) and either stagnant or quiescent, and sunlight is present [7]. It is often assumed that an aquatic system is well mixed throughout or at the top layer of water. This assumption allows one to model HABs with a system of ordinary differential equations (ODEs).

To be concrete, we base our analysis on a model of the growth of four competing algae groups: Diatoms, Chrysophyceae, nitrogen-fixing cyanobacteria, and minor species [4]. This model is generalized to account for the temporal evolution of n algae groups with biomass concentrations $c_i(t)$ $(i = 1, \dots, n)$ in a lake of volume V and average depth h. Then the model [4] consists of a system of n coupled ODEs,

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = \left(\tilde{\mu}_i - \frac{\tilde{\sigma}_i}{h} - \frac{q_{\text{out}}}{V} - f_i C_{\text{z}} \right) c_i, \qquad i = 1, \dots, n, \tag{1}$$

where $\tilde{\mu}_i$ is the natural growth rate of the ith algae group, $\tilde{\sigma}_i$ is its non-predatory loss rate, q_{out} denotes the outflow rate, and $f_i C_{\text{z}}$ is the zooplankton predator rate.

The natural growth rate $\tilde{\mu}_i$ and the non-predatory loss rate $\tilde{\sigma}_i$ vary with the average temperature in the lake, T, in accordance with

$$\tilde{\mu}_i = \mu_i \theta_i^{T - T_{\text{ref}}} \frac{I}{K_{I_i} + I} \frac{P}{K_{P_i} + P} \frac{N}{K_{N_i} + N}, \qquad \tilde{\sigma}_i = \sigma_i \theta_\sigma^{T - T_{\text{ref}}}, \tag{2}$$

Table 1 Model parameters and variables and their units (Table 3 in [4])

Parameter	Unit	Description
c_i	$(\mathrm{mg\,m^{-3}})$	Biomass concentration of the ith algae group
μ_i	$(\mathrm{day^{-1}})$	Maximum growth rate at 20°C
σ_i	$(\mathrm{day^{-1}})$	Maximum non-predatory loss rate at 20°C
θ_i		Temperature coefficients for growth rate
θ_σ		Temperature coefficients for non-predatory loss rate
K_{I_i}	$(\mathrm{W\,m^{-2}})$	Global irradiance half-saturation coefficient
K_{P_i}	$(\mathrm{mg\,m^{-3}})$	Phosphorus half-saturation coefficient
K_{N_i}	$(\mathrm{mg\,m^{-3}})$	Nitrogen half-saturation coefficient
f_iC_z	$(\mathrm{day^{-1}})$	Zooplankton rate
α_i		Relative phosphorus content of algae
β_i		Relative nitrogen content of algae
P	$(\mathrm{mg\,m^{-3}})$	Total phosphorus concentration available for the algae
P_{tot}	$(\mathrm{mg\,m^{-3}})$	Total phosphorus concentration in the lake
P_0	$(\mathrm{mg\,m^{-3}})$	Initial phosphorus concentration in the lake
c_P	$(\mathrm{mg\,m^{-3}})$	Phosphorus concentration of inflow
N	$(\mathrm{mg\,m^{-3}}]$	Total nitrogen concentration available for the algae
N_{tot}	$(\mathrm{mg\,m^{-3}})$	Total nitrogen concentration in the lake
N_0	$(\mathrm{mg\,m^{-3}})$	Initial nitrogen concentration in the lake
c_N	$(\mathrm{mg\,m^{-3}})$	Nitrogen concentration of inflow
T, T_{ref}	(°C)	Temperature and the reference temperature (20°C)
q_{out}	$(\mathrm{m^3\,day^{-1}})$	Outflow rate
Q	$(\mathrm{m^3})$	Inflow volume
$\bar{Q}$	$(\mathrm{m^3})$	Mean inflow volume
I	$(\mathrm{W\,m^{-2}})$	Global irradiance
V	$(\mathrm{m^3})$	Volume of lake
h	(m)	Depth of lake

where the rate coefficients K with various subscripts are defined in Table 1. Temperature coefficients for the growth and non-predatory loss rate are denoted by θ_i and θ_σ, respectively. Concentrations of available (nonabsorbed) nutrients (phosphorus P and nitrogen N) for algae are related to the concentrations of the algae groups by

$$P = P_{\mathrm{tot}} - \sum_{i=1}^{n} \alpha_i c_i, \qquad N = N_{\mathrm{tot}} - \sum_{i=1}^{n} \beta_i c_i, \tag{3}$$

where P_{tot} and N_{tot} are the overall nutrient concentrations in the lake; and the constants α_i and β_i denote the phosphorus and nitrogen contents of the ith algae group, respectively. The Monod form of algae growth rate (2) varies almost linearly with irradiance I and the phosphorous (P) and nitrogen (N) concentrations when these quantities are small.

Insufficient site characterization and temporal fluctuations render various parameters in (1) uncertain. The data reported in [4, 5] suggest that over the summer, temperature T, global irradiance I, outflow rate q_{out}, and predatory loss $f_i C_z$ typically exhibit much smaller variation than the fluctuations of nutrients. Consequently, we treat the total nutrients contents (P_{tot} and N_{tot}) as random functions of time t and assume the remaining parameters to be deterministic. Our goal is to compute the joint PDF of the concentrations of various algae groups, $W(\{\mathbf{C}\},t)$, where $\{\mathbf{C}\} = C_1, C_2, \ldots, C_n$ denote deterministic values (outcomes) of random algae population concentrations.

3 Stochastic Models

In many bodies of water, nutrient inflow through surface runoff and wastewater discharge is the leading factor to eutrophication. Temporal fluctuations of inflow volume $Q(t)$ is identified as the common source of uncertainty for P_{tot} and N_{tot} via relationships

$$P_{\mathrm{tot}} = P_0 + \frac{c_P Q}{V}, \qquad N_{\mathrm{tot}} = N_0 + \frac{c_N Q}{V}, \tag{4}$$

where P_0 and N_0 are the initial phosphorus and nitrogen concentrations in the lake, respectively; and c_P and c_N denote the inflow concentrations of nutrients.

Using a Reynolds decomposition to represent the runoff volume $Q(t) = \bar{Q} + Q'$ as the sum of its ensemble mean $\bar{Q}$ and zero-mean fluctuations Q', and employing a Taylor expansion of the random growth rates $\tilde{\mu}_i$ around $\bar{Q}$ yields

$$\tilde{\mu}_i = \tilde{\mu}_i(\bar{Q}) + \frac{\mathrm{d}\tilde{\mu}_i}{\mathrm{d}Q}(\bar{Q})Q' + \mathscr{O}(Q'^2). \tag{5}$$

Substitution of (5) into (1) leads to a system of n nonlinear Langevin equations with multiplicative noise $Q'(t)$,

$$\frac{\mathrm{d}c_i}{\mathrm{d}t} = h_i(\mathbf{c},t) + g_i(\mathbf{c},t)Q'(t), \qquad i = 1,\ldots,n, \tag{6}$$

where $\mathbf{c} = (c_1, c_2, \ldots, c_n)$ and

$$h_i(\mathbf{c},t) \equiv \left[\tilde{\mu}_i(\bar{Q}) - \frac{\tilde{\sigma}_i}{h} - \frac{q_{\mathrm{out}}}{V} - f_i C_z\right] c_i, \qquad g_i(\mathbf{c},t) \equiv \frac{\mathrm{d}\tilde{\mu}_i}{\mathrm{d}Q}(\bar{Q})c_i. \tag{7}$$

3.1 Fokker–Planck Equation

Following the standard procedure outlined in [8], we define the mth Kramers–Moyal expansion coefficients as

$$D^{(m)}_{i_1\ldots i_m}(\mathbf{C},t) = \frac{1}{m!}\lim_{\tau\to 0}\frac{\langle[c_{i_1}(t+\tau) - C_{i_1}]\cdots[c_{i_m}(t+\tau) - C_{i_m}]\rangle|_{c_{i_k}=C_{i_k}}}{\tau}, \tag{8}$$

where $\mathbf{C}$ is a deterministic outcome of random $\mathbf{c}$, $\langle\cdot\rangle$ denotes the ensemble mean, and $k = 1,\ldots,m$. Let us suppose that $Q'(t)$ is uncorrelated Gaussian-distributed white noise,

$$\langle Q'(t)\rangle = 0\,, \qquad \langle Q'(t_1)Q'(t_2)\rangle = 2\delta(t_1 - t_2), \tag{9}$$

where $\delta(\cdot)$ is the Dirac delta function. Then all but the first two of the Kramers–Moyal expansion coefficients vanish [8]. The nonzero coefficients are referred to as drift D_i and diffusion coefficients D_{ij}. This yields the Fokker–Planck equation for the joint PDF of the algae concentrations, $W(\mathbf{C},t)$,

$$\frac{\partial W}{\partial t} = -\sum_{i=1}^{n}\frac{\partial}{\partial C_i}\left[D_i(\mathbf{C},t)W\right] + \sum_{i,j=1}^{n}\frac{\partial^2}{\partial C_i \partial C_j}\left[D_{ij}(\mathbf{C},t)W\right] \tag{10}$$

where

$$D_i(\mathbf{C},t) = h_i(\mathbf{C},t) + g_k\frac{\partial}{\partial C_k}g_i(\mathbf{C},t), \qquad D_{ij}(\mathbf{C},t) = g_i(\mathbf{C},t)g_j(\mathbf{C},t). \tag{11}$$

3.2 PDF Method

For the correlation function $\langle Q'(t_1)Q'(t_2)\rangle$ that cannot be treated as white noise, we use the PDF method [9–11] to derive a deterministic equation for $W(\mathbf{C},t)$. We expand the concept of fine-grained single-point PDF [9–11] by introducing a fine-grained joint PDF of the concentrations of competing algae groups,

$$\Pi = \prod_{i=1}^{n}\delta[C_i - c_i(t)]. \tag{12}$$

Its ensemble average yields $W(\mathbf{C},t)$:

$$\langle\Pi\rangle = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\prod_{i=1}^{n}\delta(C_i - c_i')W(c_1',\ldots,c_n',t)\mathrm{d}c_1'\cdots\mathrm{d}c_n' = W(C_1,\ldots,C_n,t). \tag{13}$$

We show in the Appendix that the coupled system of nonlinear stochastic ODEs (6) gives rise to a linear stochastic partial differential equation (PDE) for Π,

$$\frac{\partial \Pi}{\partial t} + \sum_{i=1}^{n}\frac{\partial}{\partial C_{\mathrm{i}}}\left[\phi_i(\mathbf{C},t)\Pi\right] = 0, \qquad \phi_i(\mathbf{C},t) \equiv h_i(\mathbf{C},t) + g_i(\mathbf{C},t)Q'(t). \tag{14}$$

This PDE is subject to appropriate initial and boundary conditions.

Employing Reynolds decompositions $\Pi = W + \Pi'$ and $\phi_i = \bar{\phi}_i + \phi_i'$, and taking the ensemble average of (14), leads to a deterministic equation for W,

$$\frac{\partial W}{\partial t} + \sum_{i=1}^{n}\frac{\partial}{\partial C_i}\left[W\bar{\phi}_i(\mathbf{C},t)\right] = \sum_{i=1}^{n}\frac{\partial}{\partial C_i}\langle\Pi'\phi_i'\rangle, \tag{15}$$

which contains the unknown covariance $\langle \Pi' \phi_i' \rangle$. A closure approximation for this term can be drawn from the rich literature on stochastic averaging of linear advective transport in random velocity fields. Here we adopt the large-eddy-diffusivity (LED) approximation [11],

$$\frac{\partial W}{\partial t} + \sum_{i=1}^{n} \frac{\partial}{\partial C_i}[\bar{\phi}_i(\mathbf{C},t)W] = \sum_{i,j=1}^{n} \frac{\partial}{\partial C_i}\left(D_{ij} \frac{\partial W}{\partial C_j} \right), \tag{16}$$

where D_{ij} are components of the effective eddy-diffusivity tensor. This closure becomes exact in the limit of the correlation length of $Q'(t)$ going to zero [9]. One can verify that in this limit the PDF equation (16) reduces to the Fokker–Planck equation (10).

4 Results and Discussion

We demonstrate our approach on a relatively simple example of cyanobacteria (blue-green algae) bloom that is caused by (uncertain) inflow of nutrients. The inflow rate $Q(t)$ is modeled as white noise, so that the PDF of the cyanobacteria concentration is governed by a simplified version of the Fokker–Planck equation (10),

$$\frac{\partial W}{\partial t} = -\frac{\partial}{\partial C}[D_1(C,t)W] + \frac{\partial^2}{\partial C^2}[D_2(C,t)W], \tag{17}$$

where the drift (D_1) and diffusion (D_2) coefficients take the form (see equation (3.95) in [8])

$$D_1 = \left(\tilde{\mu}_1 - \frac{\tilde{\sigma}_1}{h} - \frac{q_{\text{out}}}{V} - f_1 C_z \right) C + \left(\frac{\partial^2 \tilde{\mu}_1}{\partial Q \partial C} C + \frac{\partial \tilde{\mu}_1}{\partial Q} \right) \frac{\partial \tilde{\mu}_1}{\partial Q} C, \tag{18a}$$

$$D_2 = \left(\frac{\partial \tilde{\mu}_1}{\partial Q} C \right)^2. \tag{18b}$$

Numerical simulations are performed with the data from previous investigations [4, 5]. A lognormal distribution $\mathscr{N}(5,1)$ is prescribed to the initial concentration. Figure 1 exhibits temporal snapshots of the PDF of the cyanobacteria concentration, $W(C,t)$, at $t = 0$, 5, and 10 days. The continuous nutrient inflow leads to rapid growth of blue-green algae from its initial mean concentration of 5–33 mg m^{-3} over a week. Widening distributions indicate rising uncertainty in the forecast. Overall, the shape of W gradually diffuses and propagates with time, as expected from the advection-diffusion (17).

Figure 2 elucidates the effects of uncertainty in the initial algae concentration. The latter is quantified in terms of the coefficient of variation (CV, standard deviation divided by mean). The PDFs $W(C,t)$ in Fig. 2 correspond to $t = 10$ days and three

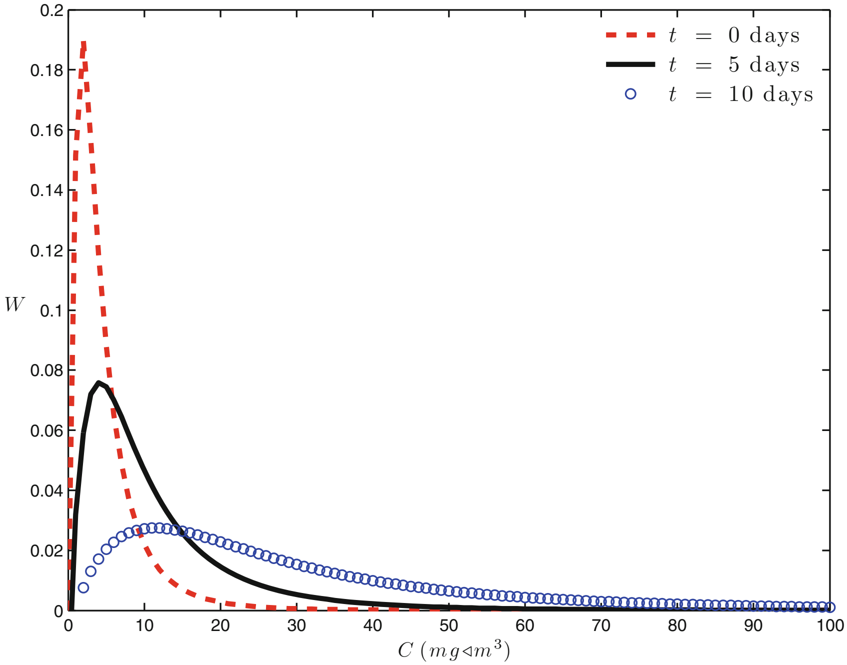

Fig. 1 Temporal snapshots of the cyanobacteria concentration PDF $W(C,t)$ at (**a**) $t = 0$ days, (**b**) $t = 5$ days, and (**c**) t 10 days

levels of CV of the initial concentration. Not surprisingly, larger initial fluctuations $(\mathrm{CV} = 2)$ lead to greater uncertainty, i.e., to longer distribution tails.

Figure 3 demonstrates the effect of average inflow on the algae growth for a fixed outflow volume. At $t = 10$ days, greater inflows $(\bar{Q}/Q_{\text{out}} = 2)$ introduce more predictive uncertainty, as indicated by a wider breadth (longer tails) of the PDF $W(C,t)$. This is to be expected, because nutrient inflow is the primary factor leading to algae bloom in lakes, and the random inflow volume is identified as the sole source of uncertainty here. Reduction of average inflow $(\bar{Q}/Q_{\text{out}} = 0.5)$ leads to smaller predictive uncertainty. However, its overall impact is limited (comparing to the time factor and initial condition) due to its small volume relative to the volume of the lake.

5 Conclusions

We present two alternative frameworks to quantify uncertainty in predictions of the concentration of various algae groups via their joint probabilistic density function (PDF). Based on a physical model routinely used for algae population dynamics in a lake, deterministic equations for the joint concentration PDF are

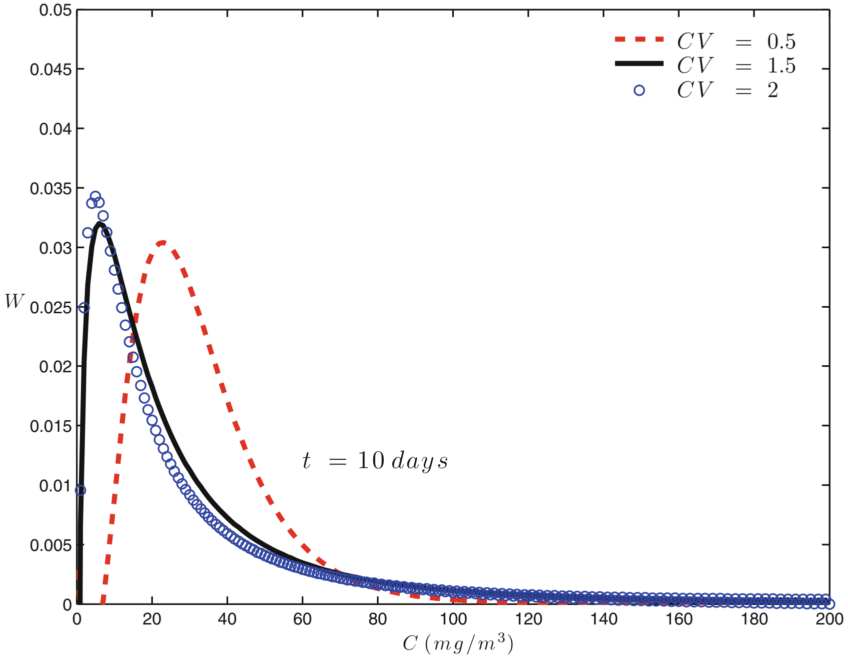

Fig. 2 The cyanobacteria concentration PDF $W(C,t)$ at $t = 10$ days for different levels of uncertainty about the initial concentration

derived by two methods, the Fokker–Planck equation and PDF method, for the uncorrelated and correlated input parameters, respectively. Our analysis leads to the following major conclusions:

1. The proposed approach provides full statistical information on the bloom of various algae species and facilitates probabilistic risk assessments by enabling computation of probabilities of rare events.
2. Uncertainty of initial population density is found to significantly affect overall predictive uncertainty.
3. Average inflow volume has limited impact on predictive uncertainty if its value is much smaller that the lake volume.

Acknowledgements This work was supported by the Office of Advanced Scientific Computing Research, the U.S. Department of Energy.

Appendix

We note that the derivatives of the raw joint PDF are

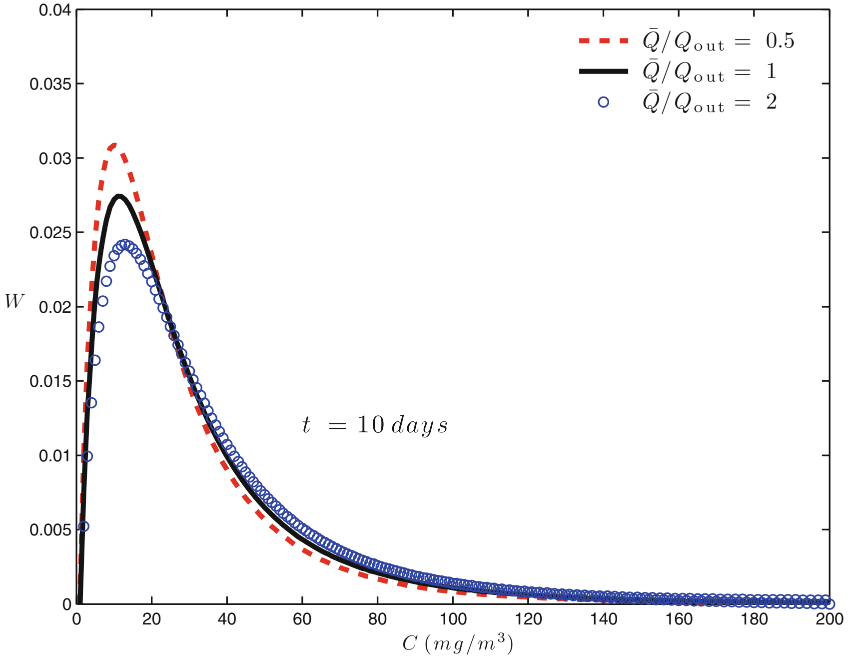

Fig. 3 Effects of average runoff $\bar{Q}$ on the cyanobacteria concentration PDF $W(C,t)$ at $t = 10$ days

$$\frac{\partial \Pi}{\partial C_i} = \frac{\partial \delta}{\partial C_i} \prod_{j=1, j\neq i}^{n} \delta(C_j - c_j), \qquad i = 1, \dots, n, \tag{19}$$

$$\frac{\partial \Pi}{\partial t} = -\sum_{i=1}^{n} \left[\frac{\partial \delta}{\partial C_i} \frac{\mathrm{d}c_i}{\mathrm{d}t} \prod_{j=1, j\neq i}^{n} \delta(C_j - c_j) \right]. \tag{20}$$

Multiplying the ith equation (6) with $\partial \Pi / \partial C_\mathrm{i}$ yields the following equations:

$$\frac{\partial \delta}{\partial C_i} \prod_{j=1, j\neq i}^{n} \delta(C_j - c_j) \frac{\mathrm{d}c_i}{\mathrm{d}t} = \frac{\partial \Pi}{\partial C_i} \phi_i(\mathbf{c}, t) = \frac{\partial [\Pi \phi_i(\mathbf{c}, t)]}{\partial C_i} = \frac{\partial [\Pi \phi_i(\mathbf{C}, t)]}{\partial C_i}. \tag{21}$$

Summation over all n equations gives

$$\frac{\partial \Pi}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial C_i} [\phi_i(\mathbf{C}, t) \Pi]. \tag{22}$$

Rearrangement of the above equation leads to (14).

References

1. Carmichael, W.: A world overview—One-hundred-twenty-seven years of research on toxic cyanobacteria—Where do we go from here? In: Hudnell, H.K. (ed.) Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs, pp. 105–125. Springer, New York (2008)
2. Dodds, W., Bouska, W., Eitzmann, J., Pilger T., Pitts, K., Riley, A., Schloesser, J., Thornbrugh, D.: Eutrophication of U.S. freshwaters: analysis of potential economic damages. Environ. Sci. Technol. **43**, 12–19 (2009)
3. Marshall, J.S., Sala, K.: A stochastic Lagrangian approach for simulating the effect of turbulent mixing on algae growth rate in photobioreactors. Chem. Eng. Sci. **66**, 384–392 (2011)
4. Malve, O., Laine, M., Haario, H., Kirkkala, T., Sarvala, J.: Bayesian modelling of algal mass occurrences: using adaptive MCMC methods with a lake water quality model. Environ. Model. Software. **22**, 966–977 (2007)
5. Haario, H., Kalachev, L., Laine, M.: Reduced models of algae growth. Bull. Math. Biol. **71**, 1626–1648 (2007)
6. Huang, D.W., Wang, H.L., Feng, J.F., Zhu, Z.W.: Modelling algal densities in harmful algal blooms (HAB) with stochastic dynamics. Appl. Math. Model. **32**, 1318–1326 (2008)
7. Hudnell, H., Jones, C., Labisi, B., Lucero, V., Hill, D.R., Eilers, J.: Freshwater harmful algal bloom (FHAB) suppression with solar powered circulation (SPC). Harmful Algae **9**, 208–217 (2010)
8. Risken, H.: The Fokker–Planck Equation: Methods of Solutions and Applications. Springer, Heidelberg (1989)
9. Kraichnan, R.H.: Eddy Viscosity and Diffusivity: Exact Formulas and Approximations. Complex Systems **1**, 805–820 (1987)
10. Pope, S.B.: Turbulent Flows. Cambridge University Press, Cambridge (2000)
11. Tartakovsky, D.M., Broyda, S.: PDF equations for advective-reactive transport in heterogeneous porous media with uncertain properties. J. Contam. Hydrol. **120–121**, 129–140 (2011)