

Water Resources Research

RESEARCH ARTICLE

10.1029/2018WR022525

Key Points:

- We provide uncertainty estimates for predicted concentration of nitrogen compounds in riverine environments
- We obtain joint probability density functions for a multicomponent reactive system
- The method yields prior distribution approximation for data assimilation and efficient parameter estimation

Correspondence to:

D. M. Tartakovsky,
tartakovsky@stanford.edu

Citation:

Boso, F., Marzadri, A., & Tartakovsky, D. M. (2018). Probabilistic forecasting of nitrogen dynamics in hyporheic zone. *Water Resources Research*, 54, 4417–4431. <https://doi.org/10.1029/2018WR022525>

Received 2 JAN 2018

Accepted 29 MAY 2018

Accepted article online 7 JUN 2018

Published online 4 JUL 2018

Probabilistic Forecasting of Nitrogen Dynamics in Hyporheic Zone

Francesca Boso¹ , Alessandra Marzadri² , and Daniel M. Tartakovsky¹ 

¹Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA, ²Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

Abstract Nitrification-denitrification processes in the hyporheic zone control the dynamics of dissolved inorganic nitrogen (DIN) species and can lead to production of nitrous oxide, which contributes to the greenhouse effect. We consider DIN dynamics in an advection-dominated regime, wherein transport and reactions occur along streamlines crossing hyporheic sediments. Our focus is on the impact of uncertainty in both stream water quality and rate constants of the subsurface reactions on predictions of DIN concentrations. We derive equations for a joint probability density function (PDF), and corresponding marginal PDFs and cumulative distribution functions (CDFs), of the species concentrations. Their derivation requires a novel closure, which depends on the mean and (co)variance of the species concentrations. We use streamline coordinates to reduce the dimensionality of the PDF/CDF equations and, hence, the computational effort of solving them. For the sake of completeness, we also present similar equations in Cartesian coordinates. By providing a complete probabilistic description of species concentrations at the bedform scale, our PDF/CDF equations allow one to evaluate the impact of random spatiotemporal variability in the inputs on the DIN dynamics. They yield physically based prior distributions for data assimilation and can be deployed to guide measurement campaigns by identifying regions with the largest predictive uncertainty.

1. Introduction

The hyporheic zone (HZ) is an important hot spot for biochemical processes that transform, decompose, and uptake solutes, thus modifying water quality (Gooseff, 2010, and references therein). Channel morphology induces near-bed pressure variations at the water sediment interface, which give rise to a so-called pumping process, one of the main mechanisms responsible for hyporheic flow (e.g., Elliott & Brooks, 1997; Tonina & Buffington, 2007). This mechanism causes surface water to enter the streambed sediment in downwelling areas (high-pressure zones) and to exit it in upwelling areas (low-pressure zones), thus creating a constantly cycling connection between surface and subsurface riverine environments (Gooseff, 2010, and references therein).

Numerous investigations analyzed hyporheic flow induced by “pumping” in two- and three-dimensional bedform morphologies (e.g., Tonina, 2012, and references therein). These studies were used to predict the fate and transport of key biogeochemical species, such as dissolved oxygen (DO), carbon (C), phosphorous (P), and nitrogen (N), under constant and periodic boundary conditions (e.g., Boano et al., 2014, and references therein). We supplement such analyses by focusing on dissolved inorganic nitrogen (DIN) species, that is, ammonium NH_4^+ and nitrate NO_3^- , whose dynamics determines both stream eutrophication conditions (Smith et al., 1999) and nitrous oxide (N_2O) emissions (Beaulieu et al., 2011). Nutrient enrichment in the form of NH_4^+ and NO_3^- increases metabolic activities of stream-hyporheic water. The latter, in turn, accelerates DO consumption (Peterson et al., 2001), thus controlling the spatial arrangement of oxic and anoxic areas in the HZ (Briggs et al., 2013). This influences nitrification-denitrification processes: Nitrification converts NH_4^+ to NO_3^- under oxic conditions, and denitrification converts NO_3^- to N_2O under both oxic and anoxic conditions (Beaulieu et al., 2011). N_2O is one of the most important greenhouse gases contributing, for example, to the stratospheric ozone destruction (Ravishankara et al., 2009).

Current models of DO and DIN dynamics in the HZ assume either constant or periodically varying in-stream loads and constant reaction rate coefficients (Tonina et al., 2015), thus ignoring the impact of their variability

on the overall dynamics. Uncertainty about the role of these driving forces (boundary and initial conditions) on the overall N_2O emissions, together with paucity of available data (field measurements are often time-consuming and costly; Baulch et al., 2011), motivates the development of quantitative methods that leverage the knowledge of appropriate physical and chemical laws to statistically characterize the global N_2O budget.

We adopt a probabilistic framework to quantify the impact of uncertainty in both the (possibly time variant) composition of river water and the rate constants controlling the key chemical reactions on predictions of the nitrogen cycle and, in particular, on predictions of N_2O emissions at the bedform scale (with NO_3^- being the primary source of N_2O). In so doing, we treat the rate constants and the uncertain initial/boundary conditions as random variables, so that the governing transport equations become stochastic. Their solutions are given in terms of joint probability density functions (PDFs) of the species concentrations or, equivalently, in terms of corresponding cumulative distribution functions (CDFs).

Monte Carlo simulations (MCS) can be used to estimate the PDFs/CDFs of species concentrations or, as is usually done, their ensemble moments such as mean and variance. MCS are conceptually straightforward, easy to implement, and robust; yet they are computationally intensive especially when they are used to compute full PDFs. The computational cost of MCS increases with the number of uncertain parameters and their variance. Moment differential equations (MDEs), that is, deterministic equations for statistical moments (mean and covariance) of solute concentration, provide an alternative to MCS. An extensive body of literature devoted to MDEs for transport of passive and reactive solutes includes Cushman (1997), Dagan and Neuman (1997), and Hu et al. (2002). (We are not aware of the use of MDEs in the context of biochemical transformations in the HZ.) MDEs can be more efficient than MCS but provide only partial probabilistic information (e.g., mean and variance, rather than a full PDF).

The method of distributions (Tartakovsky & Gremaud, 2016) is another alternative to MCS, which overcomes this limitation of MDEs by deriving a deterministic differential equation for either PDF or CDF of solute concentration, rather than its first two moments. The PDF/CDF equations are exact for advection-reaction single-species transport in deterministic velocity fields regardless of the linear or nonlinear nature of the reaction term (Boso et al., 2014; Lichtner & Tartakovsky, 2003; Venturi et al., 2013). The presence of diffusion and/or dispersion requires a closure approximation (Boso & Tartakovsky, 2016).

We further develop this approach by deriving a partial-differential equation for the joint PDF of the concentrations of the reacting species involved in DIN transformations in the HZ. Our closures are constructed to guarantee that the resulting PDFs have the same mean and variance as those computed with the corresponding MDEs. In section 2, we present a general formulation of the PDF equations for multicomponent reactive transport, alternatively written in the streamline (section 2.2) and Cartesian (Appendix D) coordinates. This approach is used to model DIN transformations in the HZ (section 3) and compared with MCS in section 4 for a linear reaction case and a given flow field. Section 5 summarizes major conclusions drawn from our analysis.

2. Materials and Methods

Governing equations for water flow and transport of dissolved interacting compounds in a saturated alluvial pack are formulated in section 2.1. In section 2.2, we present equations for the joint PDF of these species and for the marginal CDF of each compound.

2.1. Flow and Reactive Transport Within the HZ

2.1.1. Fluid Flow

Under steady-state conditions, hyporheic flow is described by a groundwater equation

$$\nabla \cdot \mathbf{j} = 0, \quad \mathbf{j} = -\mathbf{k}\nabla h, \quad (1)$$

where $\mathbf{j}(\mathbf{x})$ is the Darcy velocity, $h(\mathbf{x})$ is the hydraulic head, and $\mathbf{k}(\mathbf{x})$ is the hydraulic conductivity tensor of the alluvium. For spatially heterogeneous conductivity fields, (1) must be solved numerically (Hester et al., 2013). However, sand-bedded streams are typically characterized by low heterogeneity (Tonina et al., 2016) and, hence, can be treated as homogeneous and isotropic so that $\mathbf{k} = K\mathbf{I}$, where \mathbf{I} is the identity matrix. With these assumptions, analytical solutions of (1) have been derived for a two-dimensional dune (Elliott & Brooks, 1997; Marzadri et al., 2016; Packman & Brooks, 2001) or a ripple (Stonedahl et al., 2010) and for three-dimensional

pool-riffle morphology (Marzadri et al., 2010). Such solutions yield analytical expressions for macroscopic flow velocity $\mathbf{u}(\mathbf{x}) = \mathbf{j}(\mathbf{x})/\theta$, where θ is the streambed porosity.

2.1.2. Solute Transport

Concentration $c_i(\mathbf{x}, t)$ of an i th dissolved species satisfies an advection-dispersion-reaction equation

$$\frac{\partial c_i}{\partial t} + \nabla \cdot (\mathbf{u}c_i) = \nabla \cdot (\mathbf{D}\nabla c_i) + r_i(\mathbf{c}; \mathbf{p}), \quad i = 1, \dots, N_c, \quad (2)$$

where N_c is the number of reacting species, the dispersion tensor $\mathbf{D} = \mathbf{D}(D_m, u; \alpha)$ accounts for molecular diffusion (D_m) and hydrodynamic dispersion that depends on the magnitude of macroscopic velocity $u = |\mathbf{u}|$ and dispersivity tensor α . The reaction term r_i can involve any subset of the species concentrations set $\mathbf{c} = \{c_1, \dots, c_{N_c}\}$ and is parameterized by a set of kinetic rate constants \mathbf{p} . When transverse dispersion is negligible, (2) is recast in streamline coordinates as (e.g., Gelhar & Collins, 1971)

$$\frac{\partial c_i}{\partial t} + v \frac{\partial c_i}{\partial \tau} = d_L \frac{\partial^2 c_i}{\partial \tau^2} + r_i(\mathbf{c}; \mathbf{p}), \quad v = 1 + (d_L + d_m) \frac{\partial u}{\partial s}, \quad \tau = \int_0^{s(\mathbf{a})} \frac{ds'}{u(s')}; \quad i = 1, \dots, N_c, \quad (3)$$

where the travel time τ is specific of each streamline originating at point \mathbf{a} on the water-sediment interface; s is the coordinate along the streamline, at which variations of the velocity module u and its gradient $\partial u/\partial s$ are tracked; and $d_L = D_L/u^2$ and $d_m = D_m/u^2$ are the transformed longitudinal dispersion (D_L) and molecular diffusion (D_m) coefficients, respectively. The longitudinal dispersion coefficient is defined as $D_L = D_m + \alpha_L u(s)$, with longitudinal dispersivity α_L . Each streamline is characterized by a total residence time τ_R , which represents the time taken by a particle to travel from the injection point $\mathbf{x} = \mathbf{a}$ (downwelling) to the exit point $\mathbf{x} = \mathbf{b}$ (upwelling). Equation (3) is subject to initial and boundary conditions (ICs and BCs)

$$c_i(t = 0, \tau) = c_{i,\text{in}}(\tau), \quad c_i(t, \tau = 0) = c_i(t, \tau = \tau_R) = c_i^{(a)}(t), \quad i = 1, \dots, N_c. \quad (4)$$

The equal Dirichlet BCs at both the downwelling and upwelling ends of each streamline ($\tau = 0$ and $\tau = \tau_R$) represent the i th species concentration in the river-water column, $c_i^{(a)}(t)$, or, more precisely, at the river-water/sediment interface. This choice of BCs is suitable for advection-dominated transport in the HZ zone, wherein the dynamics within a bedform is significantly slower than the dynamics in river water. For other applications, the Dirichlet BCs in (4) can be replaced with appropriate Neumann or Robin BCs at the downwelling and upwelling boundaries.

2.2. Method of Distributions: Joint PDF of Multiple Reacting Species

Our objective is to quantify the impact of uncertainty in parameters \mathbf{p} and the initial and boundary conditions $c_{i,\text{in}}(\tau)$ and $c_i^{(a)}(t)$ on predictions of the concentrations of dissolved reacting species. We assume flow velocity \mathbf{u} to be deterministic, that is, known with certainty from either an analytical solution (as done in the subsequent application), a numerical solution of (1), or in situ observations. The use of (3) to model reactive transport relies on the assumption of negligibly small transverse dispersion and is generally applicable to advection-dominated phenomena. That is justifiable in HZ studies, since bedform sediments are characterized by high hydraulic conductivity (Salehin et al., 2004; Sawyer & Cardenas, 2009). When this assumption is not valid, solute transport has to be described by (2) instead of (3). For such transport scenarios, we derive a general PDF equation in Appendix D but do not pursue its analysis here.

We show in Appendix A that the joint PDF $f(C_1, \dots, C_{N_c}; t, \tau)$ of species concentrations c_1, \dots, c_{N_c} at time-space point (t, τ) satisfies

$$\frac{\partial f}{\partial t} + v(s) \frac{\partial f}{\partial \tau} = d_L \frac{\partial^2 f}{\partial \tau^2} + \sum_{i=1}^{N_c} \left\{ \alpha_i(t, \tau) \frac{\partial}{\partial C_i} [(C_i - \bar{c}_i) f] + \beta_i(t, \tau) \frac{\partial f}{\partial C_i} \right\}, \quad (5a)$$

with

$$\alpha_i = \frac{\chi_i}{2} - \frac{d_L}{\sigma_i^2} \left(\frac{\partial \bar{c}_i}{\partial \tau} \right)^2 - \frac{S_i}{2\sigma_i^2}, \quad \beta_i = -r_i(\bar{\mathbf{c}}, \bar{\mathbf{p}}) - \rho_i. \quad (5b)$$

Here $\bar{\mathbf{p}}$ indicates the ensemble mean of the rate constants \mathbf{p} ; ρ_i , S_i , and χ_i are the closure terms in the MDEs (see Appendices B and C) for the ensemble mean, $\bar{c}_i(t, \tau) = \int C_i f(\mathbf{C}; t, \tau) d\mathbf{C}$, and variance, $\sigma_i^2(t, \tau) = \int C_i^2 f(\mathbf{C}; t, \tau) d\mathbf{C} - \bar{c}_i^2$, of concentration $c_i(t, \tau)$. Since the coefficients in (5) are given in terms of the integrals of f , it is a nonlinear integro-differential equation.

To expedite its solution, one can precompute these coefficients, which turns (5) into a linear advection-diffusion equation. This can be accomplished with MCS. In this context, the reliance on PDF equation (5)

is beneficial because the lower moments, such as \bar{c}_i and σ_i^2 (and, possibly, the mixed second moments like ρ_{ij} , S_i and χ_i), converge with the number of MC realizations, N_{MCS} , at the rate of $1/\sqrt{N_{MCS}}$ (Fishman, 2013, Ch. 2, provides guidelines on the choice of N_{MCS} for a selected accuracy level). The number of MC realizations needed to estimate higher statistical moments with a given accuracy level increases with the moment's order (Bierig & Chernov, 2016). Hence, accurate MCS computation of non-Gaussian PDFs, such as f , requires significantly more realizations/computational effort than MCS computation of the lower moments that serve as inputs in (5).

To accelerate computation even further, we compute \bar{c}_i and σ_i^2 as solutions of MDEs

$$\begin{aligned} \frac{\partial \bar{c}_i}{\partial t} + v(s) \frac{\partial \bar{c}_i}{\partial \tau} &= d_L \frac{\partial^2 \bar{c}_i}{\partial \tau^2} + r_i(\bar{\mathbf{C}}; \bar{\mathbf{p}}) + \rho_i(t, \tau), \\ \frac{\partial \sigma_i^2}{\partial t} + v(s) \frac{\partial \sigma_i^2}{\partial \tau} &= d_L \frac{\partial^2 \sigma_i^2}{\partial \tau^2} + S_i(t, \tau) - \chi_i \sigma_i^2. \end{aligned} \quad (6)$$

These are solved numerically once closures approximations are defined in Appendix B.

Equation (5) is defined on an augmented space $\tilde{\Omega} = \Omega \times \Omega_c$ that consists of the travel-time domain Ω and the N_c -dimensional concentration coordinates of the ("event") space Ω_c . Its solution is a *joint* PDF, $f(\mathbf{C}; t, \tau)$, of *all* the reactive species. A probabilistic description of individual (e.g., *i*th) species is provided by the marginal PDF, $f_i(C_i; t, \tau)$. The latter satisfies a PDF equation

$$\frac{\partial f_i}{\partial t} + v(s) \frac{\partial f_i}{\partial \tau} = d_L \frac{\partial^2 f_i}{\partial \tau^2} + \alpha_i(t, \tau) \frac{\partial}{\partial C_i} [(C_i - \bar{c}_i) f_i] + \beta_i(t, \tau) \frac{\partial f_i}{\partial C_i}. \quad (7)$$

It is obtained by integrating (5) over $C_1, \dots, C_{i-1}, C_{i+1}, \dots, C_{N_c}$ and using the definition of a marginal PDF, $f_i(C_i) = \int f(C_1, \dots, C_{N_c}) dC_1, \dots, dC_{i-1}, dC_{i+1}, \dots, dC_{N_c}$. Several properties of PDF equations (5) and (7) are noteworthy. First, these PDF equations are self-consistent in the sense that both give rise to MDEs (6); for example, the equation for mean concentration $\bar{c}_i(t, \tau) = \int C_i f_i(C_i; t, \tau) dC_i = \int C_i f(\mathbf{C}; t, \tau) d\mathbf{C}$ in (6) can be obtained by multiplying either (5) or (7) with C_i and integrating the former over \mathbf{C} or the latter over C_i (see Appendix C). Second, while f_i provides a probabilistic description of the *i*th species concentration c_i only, it is not independent from the other species concentrations $c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_{N_c}$. That is because mean concentration \bar{c}_i and functions α_i and β_i , which serve as input parameters in (7), depend on the remaining species concentrations via (6).

Equations (5) and (7) are subject to initial and boundary conditions at the boundaries of the augmented domain $\tilde{\Omega} = \Omega \times \Omega_c$ and $\tilde{\Omega}_i = \Omega \times \Omega_{c,i}$, respectively. The initial conditions and the boundary conditions with respect to \mathbf{x} or τ (i.e., on the bounding surface of flow domain Ω) depend on the uncertainty about the initial and boundary conditions of the original transport problem. This uncertainty is quantified in terms of a (joint, possibly space-time varying) PDF of $c_{i,\text{in}}$ and $c_i^{(0)}$. If these driving forces are known with certainty, their PDFs are Dirac delta functions; for example, the PDF of a known-with-certainty initial concentration $c_{i,\text{in}}$ is $f_{i,\text{in}}(C_i; \tau) = \delta(C_i - c_{i,\text{in}})$ and the initial condition for (7) takes the form $f_i(C_i; t = 0, \tau) = \delta(C_i - c_{i,\text{in}})$. (The presence of Neumann and/or Robin boundary segments of Ω would translate into BCs that contain a spatial derivative of f_i .) Unlike MCS and MDEs, spatiotemporal variability of uncertain initial and boundary conditions or correlation between initial and boundary conditions increase neither the complexity of the PDF solution nor its computational burden.

Boundary conditions for f (or f_i) on the boundary of the event space Ω_c (or $\Omega_{c,i}$) are less straightforward to assign. One way of dealing with this difficulty is to be agnostic about possible values of solute concentration c_i by allowing it vary between $-\infty$ and $+\infty$, in which case $\Omega_i = (-\infty, +\infty)$ and the corresponding boundary conditions become either $f_i(\pm\infty; t, \tau) = 0$ or $\partial f_i / \partial C_i(\pm\infty; t, \tau) = 0$. Another way is to reformulate (7) in terms of CDF $F_i(C_i; t, \tau) = \int_0^{C_i} f_i(C_i; t, \tau) dC_i$ (Tartakovsky & Broyda, 2011),

$$\frac{\partial F_i}{\partial t} + v(s) \frac{\partial F_i}{\partial \tau} = d_L \frac{\partial^2 F_i}{\partial \tau^2} + [\alpha_i(t, \tau) (C_i - \bar{c}_i) + \beta_i(t, \tau)] \frac{\partial F_i}{\partial C_i}. \quad (8)$$

Specification of boundary conditions for this CDF equation is straightforward. For example, if c_i is known with certainty to vary between $C_{i,\text{min}}$ (e.g., 0) and $C_{i,\text{max}}$ (e.g., ∞), then $\Omega_i = [C_{i,\text{min}}, C_{i,\text{max}}]$ and the boundary conditions in the event space are $F_i(C_{i,\text{min}}; t, \tau) = 0$ and $F_i(C_{i,\text{max}}; t, \tau) = 1$. If needed, PDF f_i can then be recovered from CDF F_i via differentiation.

Equation (8) yields the CDF of the concentration of each (i th, with $i = 1, \dots, N_c$) reacting species as a function of time and travel time. In other words, for each point along a specific streamline, $F_i(C_i; t, \tau) \equiv \mathbb{P}[c_i(t, \tau) \leq C_i]$ quantifies the probability, \mathbb{P} , that a value of the i th species' concentration, $c_i(t, \tau)$, does not exceed a threshold C_i . This information allows one to make probabilistic predictions with relevant thresholds (as opposed to, say, estimating mean concentrations). Relevant thresholds include anoxic levels for oxygen, eutrophication limits for nitrate, or large nitrous oxide productions; an outcome of such a prediction is a spatial map (e.g., throughout the bedform) of the probability of threshold exceedance. Equations (5), (7), or (8) can also be used to compute prior distributions to be updated via data assimilation/Bayesian inference to reduce uncertainty about the concentration estimates or to identify parameter values.

2.2.1. Purely Advective Transport

If transport is either purely advective or strongly advection dominated, then d_L and d_m are negligible, transport equation (3) and PDF/CDF equations (5), (7), and (8) become hyperbolic, and their solutions depend only on upstream boundary conditions (along the downwelling zone, i.e., at $\tau = 0$). In this case, a solution of the single one-dimensional equation for a semi-infinite domain is enough to map the solution throughout the whole bedform.

2.2.2. Deterministic Reaction Rates

If the kinetic reaction rates \mathbf{p} are known with certainty, that is, deterministic, the closed (joint) PDF equation is derived without approximating the reaction term, with $\beta_i = r_i(\mathbf{c}, \mathbf{p})$ in (5). If longitudinal dispersion and molecular diffusion are negligible as well, then the joint PDF equation becomes exact,

$$\frac{\partial f}{\partial t} + \frac{\partial f}{\partial \tau} = - \sum_{i=1}^{N_c} \frac{\partial r_i(\mathbf{C}; \mathbf{p}) f}{\partial C_i}, \quad (9)$$

regardless of the type of nonlinearity of the reaction terms r_i . Equations for both the marginal PDFs and the moments can be derived from (9). For a general form of r_i , this step would require a closure approximation.

3. Probabilistic Forecasting of DIN Dynamics

Hydrologic interactions between surface and subsurface riverine environments induce biogeochemical activity within the HZ, where solutes, nutrients, contaminants, and pathogens react with each other depending on the HZ redox conditions and the time that water spent traveling throughout the sediments (Battin et al., 2003). We focus on DIN dynamics associated with nitrification-denitrification processes in the HZ. The latter is the main biogeochemical processes controlling N_2O production (Beaulieu et al., 2011; Hall et al., 2009) along the entire length of the streamlines due to the presence of anoxic microsites (Briggs et al., 2015). In this scenario, the dynamics of DO controls the distribution of bulk oxic and anoxic areas, wherein denitrification takes place (Harvey et al., 2013).

3.1. Fluid Flow

Deterministic velocity $u(s)$, its gradient $\partial u / \partial s$ along individual streamlines, residence time, and an equation for streamlines, serve as inputs for the PDF/CDF equations developed in section 2. All these quantities, along with a mapping between streamline and Cartesian coordinates for the bedform, can be computed either analytically or numerically (as discussed in section 1).

For illustration purposes, we employ the advective pumping model or APM (Elliott & Brooks, 1997). It represents an idealized representation of hyporheic exchange in an isolated dune bedform, where hydraulic interactions are driven by a sinusoidal pressure wave at the horizontal interface between a stream and a homogeneous, isotropic bedform of length L . The APM provides an analytical solution to steady-state two-dimensional Darcy flow equation (1) in a rectangular domain $\mathcal{D} = \{\mathbf{x} : -L/2 \leq x_1 \leq L/2, 0 \leq x_2 \leq \infty\}$, with hydraulic head $h(x_1, x_2)$ satisfying boundary conditions

$$h(-L/2, x_2) = h(L/2, x_2), \quad \frac{\partial h}{\partial x_2}(x_1, -\infty) = 0, \quad h(x_1, 0) = h_m \sin(\lambda x_1), \quad (10)$$

where h_m is the amplitude of the head variation and $\lambda = 2\pi/L$ is the dune wavelength. This solution yields a closed-form expression for macroscopic flow velocity $\mathbf{u}(\mathbf{x}) = \mathbf{j}(\mathbf{x})/\theta$,

$$\mathbf{u}(\mathbf{x}) = \begin{pmatrix} -u \cos(\lambda x_1 e^{\lambda x_2}) \\ -u \sin(\lambda x_1 e^{\lambda x_2}) \end{pmatrix}, \quad (11)$$

Table 1
Values of the Parameters Used in Our Simulations of the DIN Transformations in a Sandy Dune

Parameter	Value	Source
<i>Hydraulic properties</i>		
Hydraulic conductivity	$K = 172.8 \text{ m/d}$	Quick et al. (2016)
Porosity	$\theta = 0.38$	Quick et al. (2016)
Dune length	$L = 1 \text{ m}$	Quick et al. (2016)
Dune height	$H_d = 0.09 \text{ m}$	Quick et al. (2016)
Average stream velocity	$U = 0.39 \text{ m/s}$	Quick et al. (2016)
Average flow depth	$Y = 0.3 \text{ m}$	Quick et al. (2016)
<i>Transport properties</i>		
Rate of biomass respiration	$\bar{k}_0 = 16.4 \text{ 1/d}$	Beaulieu et al. (2011)
Rate of nitrification	$\bar{k}_1 = 629 \text{ 1/d}$	Beaulieu et al. (2011)
Rate of denitrification	$\bar{k}_2 = 5.54 \text{ 1/d}$	Beaulieu et al. (2011)
Longitudinal dispersion	$D_L = 1.1 \cdot 10^{-6} \text{ m}^2/\text{s}$	Marzadri et al. (2012)
Longitudinal dispersivity	$\alpha_L = 0.004 \text{ m}$	Bear (1972, p. 609)
Median grain size	$d_{50} = 0.0022 \text{ m}$	Quick et al. (2016)
N_2O yield	$N_y = 0.84\%$	Beaulieu et al. (2011)
<i>Boundary conditions</i>		
Concentration of dissolved oxygen	$\bar{c}_0^{(a)} = 10 \text{ mg/L}$	Rutherford (1994)
Concentration of dissolved ammonium	$\bar{c}_1^{(a)} = 13.7 \text{ }\mu\text{g/L}$	Marzadri et al. (2017)
Concentration of dissolved nitrate	$\bar{c}_2^{(a)} = 324.3 \text{ }\mu\text{g/L}$	Marzadri et al. (2017)
Concentration of dissolved dinitrogen + nitrous oxide	$\bar{c}_3^{(a)} = 0.284 \text{ }\mu\text{g/L}$	Marzadri et al. (2017)
Amplitude of in-stream nitrate signal	$A = 0.25$	
Phase of in-stream nitrate signal	$P = 1.0 \text{ d}$	

Note. DIN = dissolved inorganic nitrogen.

where $u = K\lambda h_m$ is the (constant) velocity magnitude. The coordinates of a fluid particle's trajectory along the streamline originating at $\mathbf{a} = (a, 0)^T$ are

$$\mathbf{x}(\tau) = \begin{pmatrix} a - \lambda K h_m \tau \theta^{-1} \cos(\lambda a) \\ -\lambda^{-1} \ln \left\{ \cos \left[\lambda a - \lambda^2 K h_m \tau \theta^{-1} \cos(\lambda a) \right] \cos^{-1}(\lambda a) \right\} \end{pmatrix} \quad (12)$$

where τ is the travel time. The wave amplitude (Shen et al., 1990),

$$h_m = 0.28 \frac{U^2}{2g} \left(\frac{H_d/Y}{0.34} \right)^p, \quad p = \begin{cases} 3/8 & \text{if } H_d/Y \leq 0.34 \\ 3/2 & \text{otherwise} \end{cases}, \quad (13)$$

depends on both the stream water characteristics (i.e., mean flow depth Y and mean flow velocity U) and the bedform morphology (i.e., dune height H_d). The residence time for a streamline originating on the bedform's surface at point $\mathbf{a} = (a, 0)^T$ and exiting the bedform in the upwelling region at point $\mathbf{b} = (b, 0)^T$ is

$$\tau_R(a) = \frac{\theta b}{\lambda K h_m \cos(\lambda b)}. \quad (14)$$

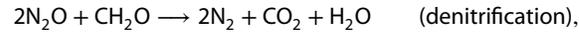
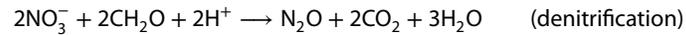
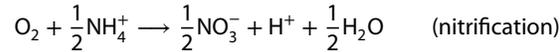
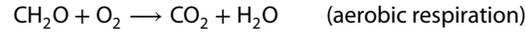
Finally, since the velocity is constant along the streamline, $\partial u / \partial s = 0$. Values of the hydraulic parameters in (11)–(14) are collated in Table 1.

Despite its simplifications (e.g., the omission of contribution of turbulent dispersion and turnover on hyporheic exchange), the APM was successfully used to detect surface-subsurface water exchange in field (Wörman et al., 2002) and flume (Elliott & Brooks, 1997) experiments with fixed or slowly varying geometry. Transport regimes that violate the foundational assumptions of the APM (e.g., when “pumping” is not the dominant process) might be treated by replacing (11)–(14) with their more evolved counterparts. For example, the shallow section of a streambed, wherein high bed porosity and sediment size induce large slip velocities at the water sediment interface, can be described with analytical solutions of Packman et al. (2004);

or high stream flows, which mobilize the upper part of the streambed sediment, call for the use of analytical solutions of Packman and Brooks (2001).

3.2. Solute Transport

To be specific, we consider a kinetic system



in which DO is consumed because of aerobic respiration, ammonium acts as a source of nitrate during nitrification, while denitrification occurs sequentially as nitrate reduction (i.e., NO_3^- is converted to N_2O) and N_2O reduction to N_2 . The overall production of nitrogen gases (N_{gas}) in the form of both N_2O and N_2 is regulated by the denitrification kinetics (Marzadri et al., 2011), while the proportion of produced N_2O (as opposed to N_2) is given by the N_2O yield, N_y (Beaulieu et al., 2011; Quick et al., 2016). These biogeochemical processes are microbially mediated by an idealized microbial community uniformly distributed within the hyporheic sediment. We consider linearized reaction kinetics, so that the reaction terms in (3) take the form $r_i(\mathbf{c}; \mathbf{p}) = p_i c_i + p_{ij} c_j$, $\mathbf{p} = \{k_0, k_1, k_2\}$ and (3) becomes

$$c_0 \equiv \text{DO} : \frac{\partial c_0}{\partial t} + \frac{\partial c_0}{\partial \tau} = d_L \frac{\partial^2 c_0}{\partial \tau^2} - k_0 c_0 \quad (15a)$$

$$c_1 \equiv [\text{NH}_4^+] : \frac{\partial c_1}{\partial t} + \frac{\partial c_1}{\partial \tau} = d_L \frac{\partial^2 c_1}{\partial \tau^2} - k_1 c_1 \quad (15b)$$

$$c_2 \equiv [\text{NO}_3^-] : \frac{\partial c_2}{\partial t} + \frac{\partial c_2}{\partial \tau} = d_L \frac{\partial^2 c_2}{\partial \tau^2} - k_2 c_2 + k_1 c_1 \quad (15c)$$

$$c_3 \equiv [\text{N}_{\text{gas}}] : \frac{\partial c_3}{\partial t} + \frac{\partial c_3}{\partial \tau} = d_L \frac{\partial^2 c_3}{\partial \tau^2} + k_2 c_2. \quad (15d)$$

The linearity approximation is common in the HZ literature (e.g., Butturini et al., 2000; Zarnetske et al., 2012). It is compatible with the goal of our investigation: to analyze the impact of uncertain reaction rates with an agile and parsimonious tool capable of capturing the basic dynamics of nitrification-denitrification and oxygen consumption. Our general PDF equations are capable of handling nonlinear reaction terms. Doing so would require either development of more complex closures for MDEs or the use of MCS to compute the lower statistical moments, as discussed in section 2.2. The (random) kinetic rates k_0 , k_1 , and k_2 correspond to, respectively, the first-order reactions for biomass respiration, nitrification, and denitrification. We treat them as lognormal random variables (although any other distribution can be chosen instead). This randomness reflects uncertainty about values of these parameters, which arises from transfer of laboratory measurements to the field, inaccuracies in direct measurements of reaction rates, and possible temperature dependence. The values for the kinetic rates, \bar{k}_i , and the N_2O yield reported in Table 1 represent the averages of the relevant data collected during the Lotic Intersite Nitrogen Experiments: LINXI (Mulholland et al., 2008) and LINXII (Beaulieu et al., 2011). To be specific, we assign a coefficient of variation $CV \equiv \sigma_{k_i} / \bar{k}_i = 0.4$ for each kinetic reaction rate, where σ_{k_i} denotes the corresponding standard deviation.

We also allow for uncertainty (randomness) in spatially varying initial conditions, which represent the steady state with given uncertain (random) boundary conditions (BC SS in Table 1). The latter, in-stream concentrations $c_i^{(a)}$ of DO ($i = 0$), NH_4^+ ($i = 1$), NO_3^- ($i = 2$), and N_{gas} ($i = 3$), are treated as lognormal random variables. Their ensemble means, reported in Table 1, represent either the commonly observed daily value (Rutherford, 1994; in the case of $c_0^{(a)}$) or the average values observed along the Manistee River (MI, USA) during typical summer baseflow conditions (Marzadri et al., 2017; in the case of $c_i^{(a)}$ with $i = 1, 2, 3$). Following Marzadri et al. (2011), we assume that $\text{N}_2(t, 0) = 0$ and, consequently, $c_3^{(a)} = \text{N}_{\text{gas}}(t, 0) = \text{N}_2\text{O}(t, 0)$. All four boundary concentrations are assigned the coefficient of variation $CV = \sigma_i / \bar{c}_i^{(a)} = 0.4$ with variance σ_i^2 .

The steady-state distributions of the reacting species are constructed via MCS with $N_{\text{MCS}} = 100$ realizations, which facilitates the subsequent comparison between the DIN dynamics predicted alternatively with the MCS

and the method of distributions. The transient simulations use these steady-state solutions as the initial concentrations $c_{i,in}$ and keep the same boundary conditions $c_i^{(a)}$, except for the boundary condition for $c_2 = [\text{NO}_3^-]$. The latter is replaced first with $c_2^{(a, \text{tr})} = c_2^{(a)} + c_2^{(a)} \mathcal{H}(t)$ where $\mathcal{H}(\cdot)$ is the Heaviside function, and then with $c_2^{(a, \text{tr})} = c_2^{(a)} [1 + A \sin(n\pi t/P)]$, where A and P are amplitude and period of the in-stream fluctuations of $[\text{NO}_3^-]$. The first boundary condition represents a sudden increase of nitrate in the stream due to an accidental release, while the second one mimicks a daily release of nitrate from an upstream wastewater treatment plant. Since NO_3^- dynamics controls N_2O production, different modes of injection of NO_3^- along the water/sediment interface are expected to affect production of N_2O in the streambed. Uncertainty in in-stream concentrations stems from data scarcity and spatiotemporal variability of in-stream loads along the river, which depend on land use and hydrologic conditions.

We solve CDF equation (8), whose coefficients are computed with the MDEs (6) in which the closure variables are set to $\rho_i = 0$, $S_i = 2\bar{\rho}_i\sigma_i^2 + 2\bar{\rho}_{ij}\sigma_{ij}^2$, $\chi_i = 0$, $S_{ij} = \bar{\rho}_i\sigma_{ij}^2\bar{\rho}_{ij}\sigma_{ij} + \bar{\rho}_i\sigma_{ij}^2 + \bar{\rho}_{im}\sigma_{im}^2$, $\chi_{ij} = 0$. CDF equation (8) is subject to initial and boundary conditions

$$\begin{aligned} F_i(C_j; \tau, t = 0) &= F_{i,ss}(C_j; \tau), & F_i(C_j; \tau = 0, t) &= F_i(C_j; \tau = \tau_R, t) = \text{Ln} \left\{ \tilde{c}_i^{(a, \text{tr})}, \sigma_i^{(a, \text{tr})} \right\}, \\ F_i(C_j = C_{i, \text{min}}; \tau, t) &= 0, & F_i(C_j = C_{i, \text{max}}; \tau, t) &= 1, \end{aligned} \quad (16)$$

where $F_{i,ss}$ is the CDF of the i th species at the steady state, and $\text{Ln}\{\cdot\}$ is the lognormal distribution with mean $\tilde{c}_i^{(a, \text{tr})}$ and standard deviation $\sigma_i^{(a, \text{tr})} = 0.4\tilde{c}_i^{(a, \text{tr})}$. In the simulations reported below we set $C_{i, \text{min}} = 0$ for all species, while $C_{i, \text{max}}$ is chosen appropriately depending on the dynamics of the i th species. The numerical solution of (8) is obtained via finite differences with $\Delta\tau = 1.7 \cdot 10^{-3}$ d in the τ direction, $\Delta C_j = (C_{j, \text{max}} - C_{j, \text{min}})/N_{\text{cell}}$ with $N_{\text{cell}} = 200$ cells in the C_j direction, and time step $\Delta t = 0.01$ d. The solution of CDF equation (8) is a map of the bedform-scale probability of exceedance for each of the reacting species.

We also conduct MCS with $N_{\text{MCS}} = 1,000$ realizations to assess the impact of our selection of the (trivial) closures on the accuracy of the resulting estimates of the concentration moments and CDFs.

4. Results

We focus on quantification of uncertainty in predictions of the nitrous oxide concentration, $\tilde{c}_3(t, \tau) \equiv [\text{N}_2\text{O}] = N_y c_3$, because of its significance as a greenhouse gas. We consider a deterministic (i.e., known with certainty) N_2O yield, although uncertainty (randomness) in this value can be considered as well by either (i) including a two-step model of denitrification with random N_2O and N_2 production rates or (ii) regarding the PDF of \tilde{c}_3 to be conditioned on the value of the N_2O yield. In the latter case, the multiplication of this conditional PDF by the PDF of N_y , $f_{N_y}(N'_y)$, and integration over N'_y would yield the (marginal) PDF of N_2O , $f_3(\tilde{c}_3; \tau, t)$. Two environmental scenarios of NO_3^- input to the HZ, as quantified by the two alternative boundary conditions in the previous section, are considered.

For the step load of NO_3^- , the CDF $F_3(C_3; t, \tau)$ is computed with (8). Contour plots of its rescaled counterpart $F_3(\tilde{c}_3; t, \tau)$, the CDF of $\tilde{c}_3 = N_y c_3$, are shown in Figure 1 (left column) at times $t = 0.5$ day (top row) and 2.5 d (bottom row), as function of threshold value \tilde{c}_3 and travel time τ , along the deepest streamline characterized by residence time $\tau_R = 2.6$ days. The increase in $[\text{NO}_3^-]$ at the downwelling end of the streamline increases production of N_2O that travels along the streamline due to advection, with mild dispersive spreading. Since $F_3(\tilde{c}_3; t, \tau) \equiv \mathbb{P}[\tilde{c}_3(t, \tau) \leq \tilde{c}_3]$, the contour plots of CDF represent maps of the probability of the nitrous oxide concentration $\tilde{c}_3(t, \tau)$ not exceeding a threshold value \tilde{c}_3 . As such, they identify regions with the highest probability of finding high concentrations of nitrous oxide. The wider the CDF at a given τ , the larger the predictive uncertainty. Uncertainty increases away from the streamline's edges, towards the center of the bedform.

Our CDF equation (8) is based on a set of closures, whose accuracy and robustness cannot be ascertained a priori. By way of validation, we show in Figure 1 CDF $F_3(\tilde{c}_3; \cdot, \tau)$ resulting from the numerical solution of (8) (left column) and MCS (middle column). These two alternative methods yield solutions for $F_3(\tilde{c}_3; \cdot, \tau)$, which are virtually indistinguishable in terms of an "eyeball measure." To make this comparison more quantitative, we use the Earth Mover's metric,

$$E_3(t, \tau) = \int_{\tilde{c}_{3, \text{min}}}^{\tilde{c}_{3, \text{max}}} |F_3^{\text{MD}}(\tilde{c}_3; t, \tau) - F_3^{\text{MCS}}(\tilde{c}_3; t, \tau)| d\tilde{c}_3, \quad (17)$$

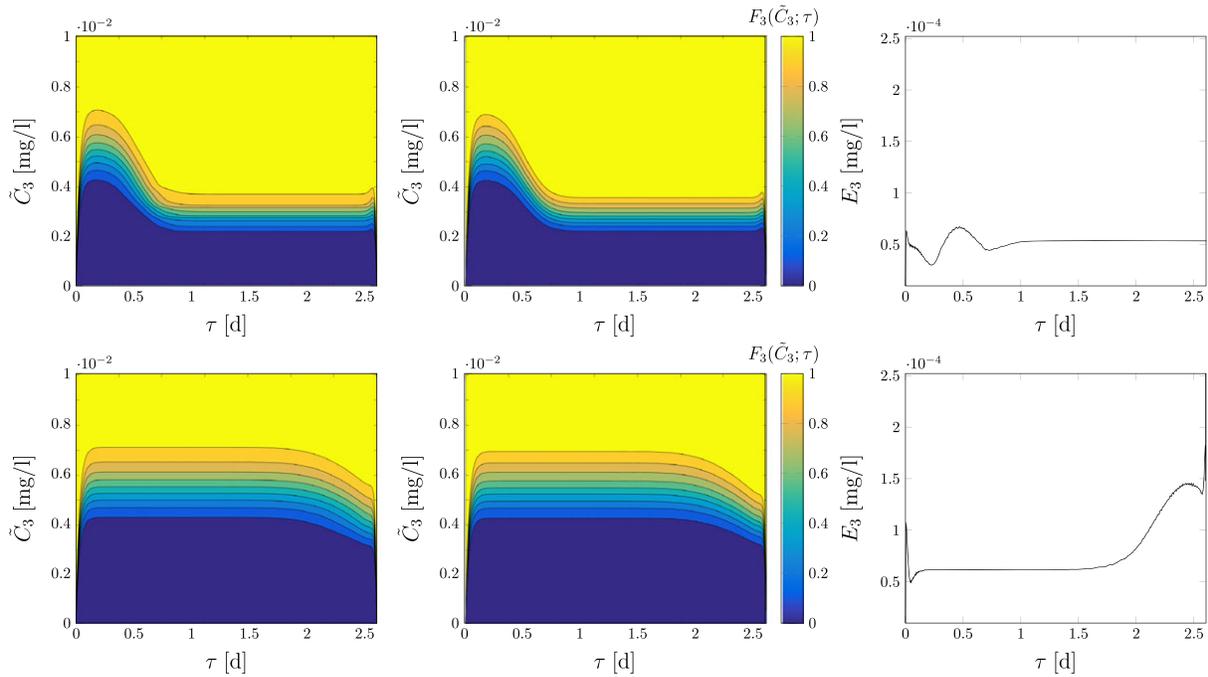


Figure 1. Step load of NO_3^- : Temporal snapshots, at times $t = 0.5$ day (top row) and 2.5 days (bottom row), of CDF $F_3(\tilde{C}_3; t, \tau)$ computed with the method of distributions (MD, left column) and MCS (middle column), and of error $E_3(t, \tau)$ between the two (right column). The contour plots of CDF $F_3(\tilde{C}_3; \tau)$ are for the streamline with residence time $\tau_R = 2.6$ days; the isocontour lines differ by 10%. A contour plot of $F_3(\tilde{C}_3; \tau) \equiv \mathbb{P}[\tilde{c}_3(\cdot, \tau) \leq \tilde{C}_3]$ is a map of the bedform-scale probability, \mathbb{P} , of nitrous oxide concentration, $\tilde{c}_3 = [N_2O]$, not exceeding a threshold value \tilde{C}_3 . The discrepancy between the CDFs estimated with MD and MCS, E_3 , does not exceed 1% for $t = 0.5$ day and 2.5% for $t = 2.5$ days. CDF = cumulative distribution function; MCS = Monte Carlo simulations.

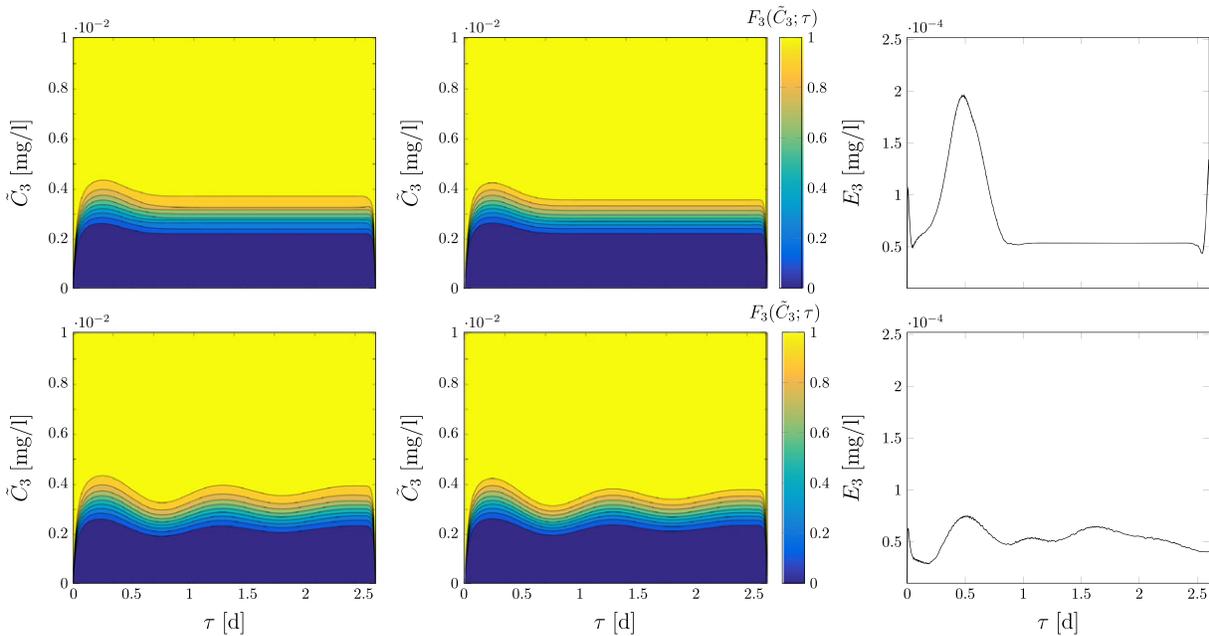


Figure 2. Periodically fluctuating load of NO_3^- : Temporal snapshots, at times $t = 0.5$ day (top row) and 2.5 days (bottom row), of CDF $F_3(\tilde{C}_3; t, \tau)$ computed with the MD (left column) and MCS (middle column), and of error $E_3(t, \tau)$ between the two (right column). The contour plots of CDF $F_3(\tilde{C}_3; \tau)$ are for the streamline with residence time $\tau_R = 2.6$ days; the isocontour lines differ by 10%. The discrepancy between the CDFs estimated with MD and MCS, E_3 , does not exceed 2.5% for $t = 0.5$ day and 1% for $t = 2.5$ days.

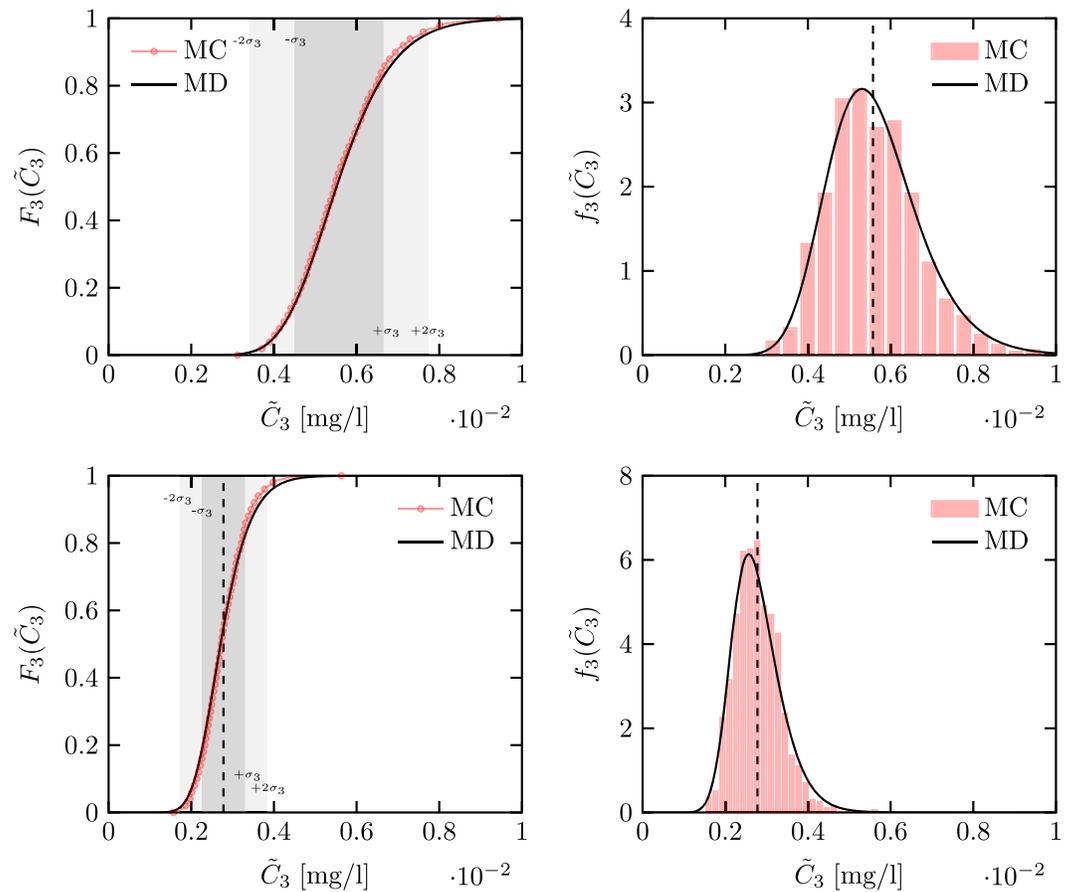


Figure 3. CDF $F_3(\tilde{C}_3)$ (left column) and PDF $f_3(\tilde{C}_3)$ (right column) of $\tilde{c}_3(t, \tau) = [\text{N}_2\text{O}]$ along the streamline with $\tau_R = 2.6$ d, at $t = 2.5$ d and $\tau = 1$ d, for the step (top row) and periodically fluctuating (bottom row) loads of N_3O^- . These distributions are computed with the MD approach (solid lines) and MCS (lines with circles for CDFs and histogram bars for PDFs). The dashed vertical line and shaded intervals represent the mean $\langle \tilde{c}_3 \rangle = N_y \bar{c}_3$ and standard deviation obtained from MCS; they highlight the asymmetry of the distributions. CDF = cumulative distribution function; MCS = Monte Carlo simulations; PDF = probability density function.

as a measure of discrepancy between the CDFs estimated with the method of distributions, $F_3^{\text{MD}}(\tilde{C}_3; t, \tau)$, and MCS, $F_3^{\text{MCS}}(\tilde{C}_3; t, \tau)$. By this metric, the difference between the two methods is less than 1% at $t = 0.5$ day and does not exceed 2.5% at $t = 0.5$ day (Figure 1, right column). This small discrepancy is mostly due to numerical error resulting from the discretization with cell size $\Delta\tilde{C}_3 = 5 \cdot 10^{-5}$ in the \tilde{C}_3 direction. It can be reduced by increasing the spatial resolution, that is, decreasing $\Delta\tilde{C}_3$, and/or reducing numerical diffusion.

Figure 2 exhibits results of the analogous analysis for the periodically fluctuating load of NO_3^- . A similar degree of agreement between the CDFs estimated with the MD and MCS demonstrates the robustness of the proposed method of distributions.

Picking a point (t, τ) , that is, selecting a vertical cross section of the maps in Figures 1 and 2, yields a full probabilistic description of nitrous oxide concentration, $\tilde{c}_3(t, \tau) = [\text{N}_2\text{O}]$, at the space-time point of interest. Figure 3 provides such description in terms of both CDF $F_3(\tilde{C}_3)$ (left column) and corresponding PDF $f_3(\tilde{C}_3)$ (right column). This figure provides yet another visual confirmation of the close agreement between the CDFs and PDFs estimated with the MD and MCS. For the two NO_3^- load scenarios considered, the distributions are asymmetric, that is, non-Gaussian. This finding reveals that the mean and variance of \tilde{c}_3 , the two quantities routinely computed with MDE-based stochastic analyses, are not sufficient to capture low probability events described by the distribution's tails.

The mean $\langle \tilde{c}_3 \rangle$ and variance $\langle \tilde{c}_3^2 \rangle$ of \tilde{c}_3 , which are used as input for CDF equation (8), vary with time t and travel time τ , as shown in Figure 4. Both quantities are computed, alternatively, by using MDEs (6) and MCS.

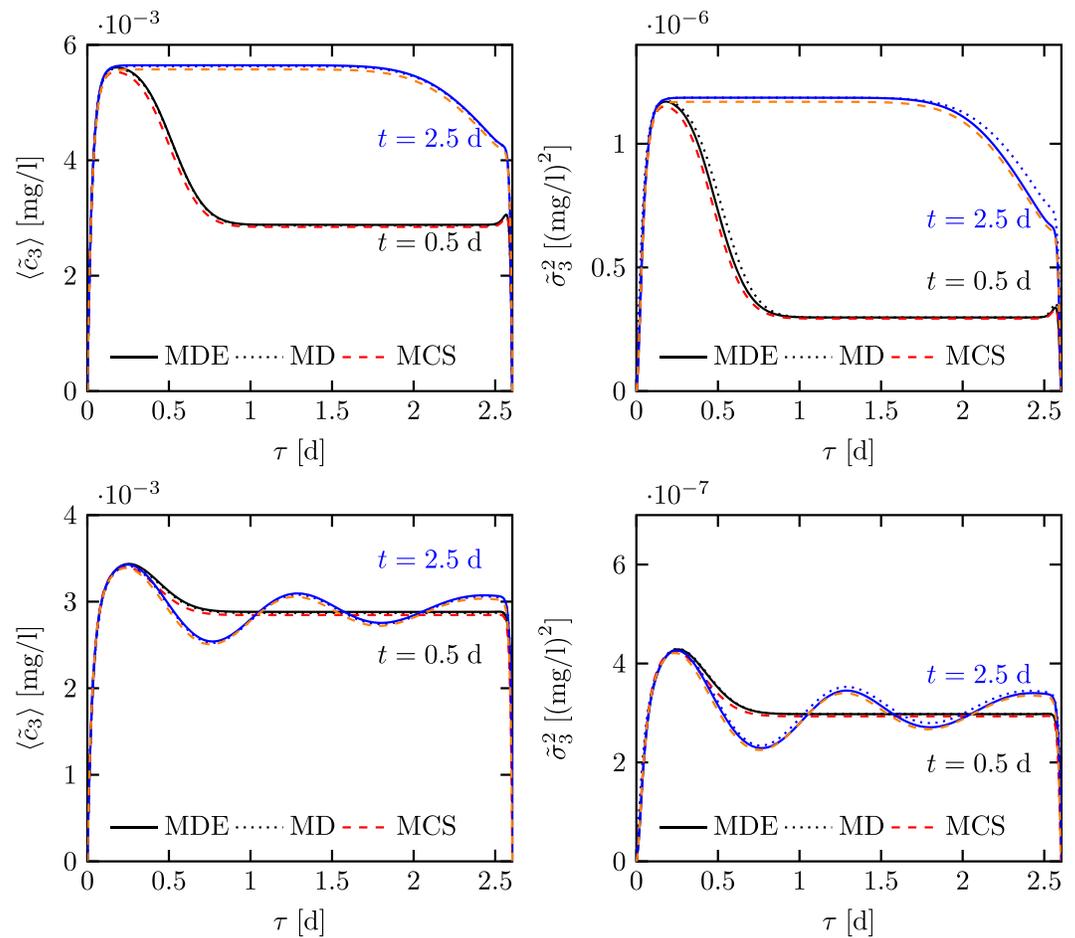


Figure 4. Mean $\langle \tilde{c}_3 \rangle$ (left) and variance σ_3^2 (right) of $\tilde{c}_3 = [\text{N}_2\text{O}]$ computed, alternatively, with the MDEs (solid line), the method of distributions (dotted line), and MCS (dashed line) at $t = 0.5$ and 2.5 days for the step (top row) and periodically fluctuating (bottom row) loads of N_3O^- .

The close agreement between the two methods is remarkable considering that we have used a “mean-field” (zeroth-order) approximation to close MDEs (6). By construction, our solution for CDF $F_3(\tilde{C}_3)$ has to preserve the $\langle \tilde{c}_3 \rangle$ and σ_3^2 predicted with MDEs. Figure 4 demonstrates that this is indeed the case, up to a small numerical error.

Finally, we repeat this analysis for each streamline to construct exceedance probability maps in the physical space with coordinates (x_1, x_2) by plotting $\mathbb{P}[\tilde{c}_3(t, \tau) > \tilde{C}_3] = 1 - F_3(\tilde{C}_3; t, \tau)$ with $\tau = \tau(\mathbf{x})$ in (12). Figure 5 maps out the probability, $1 - F_3(C^*; \mathbf{x}, t)$, of the local value of concentration $\tilde{c}_3(\mathbf{x}, t)$, at time $t = 0.5$ day, exceeding the thresholds $C^* = 4.2 \cdot 10^{-3}$ and $3.4 \cdot 10^{-3}$ for the step and periodic loads of N_3O^- , respectively. The streamline color represents the probability of the nitrous oxide concentration, $\tilde{c}_3 = [\text{N}_2\text{O}]$, along the streamline exceeding C^* . The shallow streamlines have a residence time that is too short to allow for consistent production of N_2O , while larger production and accumulation occur along the deeper streamlines. The latter are also regions with largest concentrations and concentration gradients in the upwelling zone, and thus with largest amounts of N_2O discharge into the stream.

5. Summary and Future Work

We investigated the impact of parametric uncertainty on one’s ability to predict geochemical transformations of the key species involved in DIN circulation in the HZ. Uncertain reaction rate constants and driving forces (initial and boundary states) were treated as random variables, giving rise to a system of coupled stochastic advection-dispersion-reaction equations for species concentrations. Since the latter are known to have

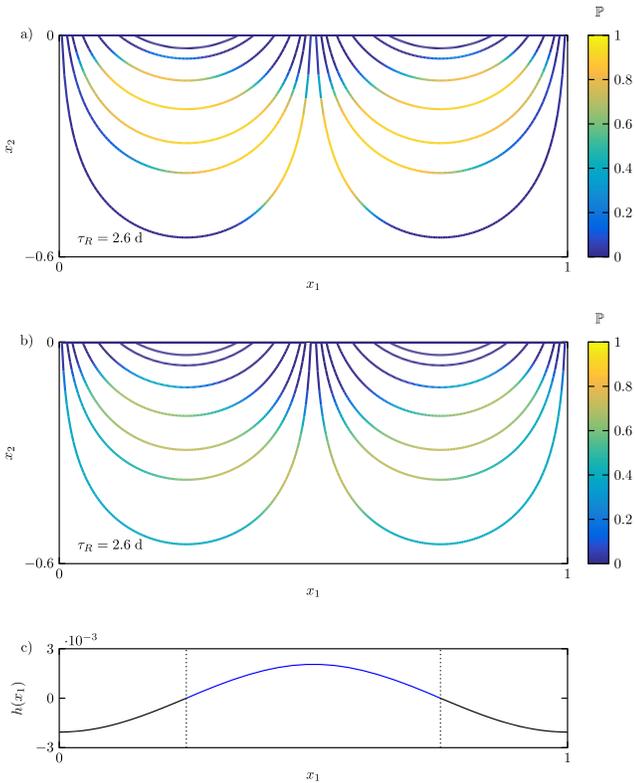


Figure 5. Streamlines, (a) and (b), resulting from the pressure profile (c) at the boundary. The blue section in (c) represents the downwelling region, according to the APM model. The exceedance probability maps $1 - F_3(C^*; t, \tau)$ at time $t = 0.5$ day are represented in (a) for $C^* = 4.2 \cdot 10^{-3}$ and the step load of nitrate, and (b) for $C^* = 3.4 \cdot 10^{-3}$ and the periodically fluctuating load. The streamlines are characterized by different characteristic residence time τ_R , which is a function of the entry point along the downwelling area. APM = advective pumping model.

non-Gaussian (i.e., asymmetric and long tailed) distributions, their probability of exceedance cannot be captured with lower moments (mean and variance). While Monte Carlo (MC) simulations allow one, in principle, to estimate such probabilities, large numbers of MC runs needed to make these estimates reliable/accurate render MC impractical.

Instead, we developed deterministic equations for the joint PDF for all the species concentrations, as well as equations for corresponding marginal PDFs and CDFs for concentrations of individual species. The solution of a CDF equation, such as (8), is a map of the bedform-scale exceedance probability for each of the reacting species. Such maps provide quantitative answers to practical questions: How likely is the concentration of a solute of interest to exceed its legal limit? What is the confidence level of a particular prediction of the spatiotemporal distribution of solute concentrations? What is the environmental risk posed by a release of a certain chemical? And so forth. CDF or PDF maps can also be used to guide sensor placement: measurements taken in regions with highest predictive uncertainty have a bigger impact on the reduction of the overall uncertainty once the data are assimilated into the model. In this context, PDF solutions serve as prior distributions for Bayesian data assimilation. Varying the levels and/or sources of uncertainty provides direct information on the sensitivity of the estimates in terms of their PDFs, rather than just lower moments. Finally, solutions to joint PDF equations, such as (5) or its marginals, allow one to investigate correlations between different species and might facilitate system identifiability.

In the future, we will extend our analysis to account for uncertainty in streamline geometry (Pryshlak et al., 2015). The latter stems from spatial variability of hydraulic conductivity, from spatiotemporal fluctuations in hydrodynamic conditions of a river and bedform morphology, and from interactions with baseflow. Our ultimate goal is to develop tools for probabilistic forecasting of the overall production of nitrous oxide both at the bedform scale and in a river reach and a river network.

Appendix A: Derivation of PDF Equation

We start by defining a joint “raw” PDF for all the species c_1, \dots, c_{N_c} as

$$\Pi(\mathbf{C}; t, \tau) = \prod_{i=1}^{N_c} \delta [C_i - c_i(t, \tau)]. \quad (\text{A1})$$

Its ensemble mean gives the joint PDF $f(\mathbf{C})$:

$$\langle \Pi(\mathbf{C}) \rangle \equiv \int_{\Omega_c} \left(\prod_{i=1}^{N_c} \delta_i (C_i - \tilde{C}_i) \right) f(\tilde{\mathbf{C}}) d\tilde{\mathbf{C}} = f(\mathbf{C}), \quad (\text{A2})$$

where Ω_c is the domain for the concentration coordinates $\mathbf{C} = \{C_1, \dots, C_{N_c}\}$. Multiplying each equation in (3) by $-\partial\Pi/\partial C_i$, while taking into account that

$$\frac{\partial\Pi}{\partial t} = \sum_{i=0}^3 -\frac{\partial\Pi}{\partial C_i} \frac{\partial c_i}{\partial t}, \quad \frac{\partial\Pi}{\partial \tau} = \sum_{i=0}^3 -\frac{\partial\Pi}{\partial C_i} \frac{\partial c_i}{\partial \tau}, \quad -\frac{\partial\Pi}{\partial C_i} r_i(\mathbf{c}) = -\frac{\partial\Pi r_i(\mathbf{C})}{\partial C_i},$$

yields

$$\frac{\partial\Pi}{\partial t} + v \frac{\partial\Pi}{\partial \tau} = \sum_i -d_L \frac{\partial\Pi}{\partial C_i} \frac{\partial^2 c_i}{\partial \tau^2} - \frac{\partial\Pi r_i(\mathbf{C}; \mathbf{p})}{\partial C_i}, \quad (\text{A3})$$

where $v(s)$ is defined in (3). Ensemble averaging (A3) yields

$$\frac{\partial f}{\partial t} + v \frac{\partial f}{\partial \tau} = d_L \frac{\partial^2 f}{\partial \tau^2} + \sum_i \frac{\partial}{\partial C_i} \left(d_L \left\langle \frac{\partial \Pi}{\partial \tau} \frac{\partial c_i}{\partial \tau} \right\rangle \right) - \frac{\partial \langle \Pi(\mathbf{C}) r_i(\mathbf{C}; \mathbf{p}) \rangle}{\partial C_i}. \quad (\text{A4})$$

In general, it is necessary to find a suitable closure for the terms $M_i = d_L \langle \frac{\partial \Pi}{\partial \tau} \frac{\partial c_i}{\partial \tau} \rangle$ and $R_i = \langle \Pi r_i(\mathbf{C}) \rangle$. Boso and Tartakovsky (2016) developed a closure for advective-dispersive transport of a single dissolved species. We generalize their approach to derive deterministic equations for marginal PDFs $f_i(C_i) = \int f(C_1, \dots, C_{N_c}) dC_1, \dots, dC_{i-1}, dC_{i+1}, \dots, dC_{N_c}$. These equations are closed by enforcing their consistency with the lower moments of c_i (i.e., mean and variance) to obtain (5). In the special case of deterministic reaction rates \mathbf{p} and zero dispersion $d_L = 0$, (A4) is exact and leads to (9).

Appendix B: Moments Equations

Ensemble average of (3) yields the first of equation (6), wherein the closure ρ_i retains all higher-order fluctuation terms for both concentrations and reaction rates. Subtracting the equation for the mean from (3) yields an equation for the concentration fluctuations $c'_i = c_i - \bar{c}_i$,

$$\frac{\partial c'_i}{\partial t} + v \frac{\partial c'_i}{\partial \tau} = d_L \frac{\partial^2 c'_i}{\partial \tau^2} + r'_i, \quad (\text{B1})$$

where $r'_i = r_i - \bar{r}_i$ and $\bar{r}_i = r_i(\bar{\mathbf{c}}; \bar{\mathbf{p}}) + \rho_i$. The equation for the covariances $\sigma_{ij}^2 = \langle c'_i c'_j \rangle$ is obtained by multiplying (B1) with c'_j and taking the ensemble mean. After algebraic manipulations that yields

$$\frac{\partial \sigma_{ij}^2}{\partial t} + v \frac{\partial \sigma_{ij}^2}{\partial \tau} = d_L \frac{\partial^2 \sigma_{ij}^2}{\partial \tau^2} + S_{ij} - \chi_{ij} \sigma_{ij}^2, \quad (\text{B2})$$

where $S_{ij} = \langle c'_j r'_i \rangle + \langle c'_i r'_j \rangle$ and $\chi_{ij} = 2d_L \langle \frac{\partial c'_i}{\partial \tau} \frac{\partial c'_j}{\partial \tau} \rangle$. Setting $i = j$, we obtain an equation for variance σ_i^2 ,

$$\frac{\partial \sigma_i^2}{\partial t} + v \frac{\partial \sigma_i^2}{\partial \tau} = d_L \frac{\partial^2 \sigma_i^2}{\partial \tau^2} + S_i - \chi_i \sigma_i^2. \quad (\text{B3})$$

For the sake of simplicity, we consider linear reactions in the form $r_i(\mathbf{c}) = p_i c_i + p_{ij} c_j$ and a trivial (zeroth-order) approximation for the closures: $\rho_i = 0$, $S_{ij} = \bar{p}_i \sigma_{ij}^2 + \bar{p}_{ij} \sigma_{ij}^2 + \bar{p}_i \sigma_{il}^2 + \bar{p}_{im}^2 \sigma_{im}^2$. Their accuracy is assessed by comparison with MCS.

Appendix C: Consistency Between PDF Equations and MDEs in Streamline Coordinates

Consistency between (5) and MDEs (6) is imposed in order to define α_i and β_i . The closure for the diffusive term is structurally analogous to IEM closures in the turbulence literature, corrected for the correct reproduction of both mean and variance of the concentration of each species (Boso & Tartakovsky, 2016). The reaction-related closure approximates the conditional mean that would arise with an expansion around the nonconditional mean with corrections that depend on variances and covariances.

First, we obtain equations for the marginal PDF f_i by integrating (5) with respect to all \mathbf{c}_j for which $j \neq i$,

$$\frac{\partial f_i}{\partial t} + v \frac{\partial f_i}{\partial \tau} = d_L \frac{\partial^2 f_i}{\partial \tau^2} + \frac{\partial}{\partial C_i} [\alpha_i (C_i - \bar{c}_i) f_i + \beta_i f_i]. \quad (\text{C1})$$

Multiplying (C1) by C_i and integrating over Ω_{τ} , we obtain the equation for \bar{c}_i regardless of the expression assigned to α_i and assuming $\beta_i = -(r_i(\bar{\mathbf{c}}; \bar{\mathbf{p}}) + \rho_i)$. By multiplying (C1) by C_i^2 and considering $\sigma_i^2 = \int C_i^2 dC_i - \bar{c}_i^2$, we set $\alpha_i = \frac{\chi_i}{2} - \frac{d_L}{\sigma_i^2} \frac{\partial \bar{c}_i}{\partial \tau} \frac{\partial \bar{c}_i}{\partial \tau} - \frac{S_i}{2\sigma_i^2}$.

Appendix D: Consistency Between PDF Equations and MDEs in Cartesian Coordinates

An equation for the raw PDF $\Pi = \prod_{i=1}^{N_c} \delta_i$ is rewritten for (2) with random inputs and parameters and deterministic velocity as

$$\frac{\partial \Pi}{\partial t} + \mathbf{u} \cdot \nabla \Pi = \nabla \cdot (\mathbf{D} \nabla \Pi) + \sum_{i=1}^{N_c} \frac{\partial}{\partial C_i} (D \nabla c \cdot \nabla \Pi) - \frac{\partial \Pi r_i(\mathbf{C}; \mathbf{p})}{\partial C_i}. \quad (D1)$$

An equation for the joint PDF f is obtained via ensemble averaging,

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f = \nabla \cdot (\mathbf{D} \nabla f) + \sum_{i=1}^{N_c} \frac{\partial}{\partial C_i} \langle D \nabla c \cdot \nabla \Pi \rangle - \frac{\partial \langle \Pi r_i(\mathbf{C}; \mathbf{p}) \rangle}{\partial C_i}. \quad (D2)$$

Its closure is performed via the introduction of two closure terms $\tilde{\alpha}_i$ and β_i :

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f = \nabla \cdot (\mathbf{D} \nabla f) + \sum_{i=1}^{N_c} \frac{\partial}{\partial C_i} [\tilde{\alpha}_i (C_i - \bar{c}_i) f + \beta_i f], \quad (D3)$$

where $\tilde{\alpha}_i = \frac{\chi}{2} - \frac{d_i}{\sigma_i^2} \nabla \bar{c}_i \cdot \nabla \bar{c}_i - \frac{S_i}{2\sigma_i^2}$ and β_i is defined as in the streamline case.

Acknowledgments

This work was supported in part by the National Science Foundation under grant DMS-1620103. There are no data sharing issues since all of the numerical information is provided in the figures produced by solving the equations in the paper. We thank the Editor, the Associate Editor, and the three reviewers for their constructive comments.

References

- Battin, T. J., Kaplan, L., Newbold, J. D., & Hendricks, S. P. (2003). A mixing model analysis of stream solute dynamics and the contribution of a hyporheic zone to ecosystem function. *Freshwater Biology*, 48(6), 995–1014. <https://doi.org/10.1046/j.1365-2427.2003.01062.x>
- Baulch, H. M., Schiff, S. L., Maranger, R., & Dillon, P. J. (2011). Nitrogen enrichment and the emission of nitrous oxide from streams. *Global Biogeochemical Cycles*, 25, GB4013. <https://doi.org/10.1029/2011GB004047>
- Bear, J. (1972). *Dynamics of fluids in porous media*. New York: Dover Publications, Inc.
- Beaulieu, J. J., Tank, J. L., Hamilton, S. K., Wollheim, W. M., Hall, R. O., & Mulholland, P. J. (2011). Nitrous oxide emission from denitrification in stream and river networks. *PNAS Proceeding of the National Academy of Sciences of the United States of America*, 108(1), 214–219.
- Bierig, C., & Chernov, A. (2016). Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method. *Stochastics and Partial Differential Equations: Analysis and Computations*, 4(1), 3–40.
- Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, 52, 603–679. <https://doi.org/10.1002/2012RG000417>
- Boso, F., Broyda, S. V., & Tartakovsky, D. M. (2014). Cumulative distribution function solutions of advection-reaction equations with uncertain parameters. *Proceedings of the Royal Society A*, 470(2166), 20140189. <https://doi.org/10.1098/rspa.2014.0189>
- Boso, F., & Tartakovsky, D. M. (2016). The method of distributions for dispersive transport in porous media with uncertain hydraulic properties. *Water Resources Research*, 52, 4700–4712. <https://doi.org/10.1002/2016WR018745>
- Briggs, M. A., Day-Lewis, F. D., Zarnetske, J. P., & Harvey, J. W. (2015). A physical explanation for the development of redox microzones in hyporheic flow. *Geophysical Research Letters*, 42, 4402–4410. <https://doi.org/10.1002/2015GL064200>
- Briggs, M. A., Lautz, L., & Hare, D. K. (2013). Residence time control on hot moments on net nitrate and uptake in hyporheic zone. *Hydrological Processes*, 28, 3741–3751.
- Butturini, A., Battin, T. J., & Sabater, F. (2000). Nitrification in stream sediment biofilms: The role of ammonium concentration and doc quality. *Water Research*, 34(2), 629–639.
- Cushman, J. H. (1997). *The physics of fluids in hierarchical porous media: Angstroms to miles*. New York: Kluwer Academic Publishers.
- Dagan, G., & Neuman, S. P. (Eds.) (1997). *Subsurface flow and transport: A stochastic approach*. Cambridge, New York: Cambridge University Press.
- Elliott, A. H., & Brooks, N. H. (1997). Transfer of nonsorbing solutes to a streambed with bedforms: Theory. *Water Resources Research*, 33(1), 123–136.
- Fishman, G. (2013). *Monte Carlo: Concepts, algorithms, and application*. New York: Springer Science & Business Media.
- Gelhar, L., & Collins, M. (1971). General analysis of longitudinal dispersion in nonuniform flow. *Water Resources Research*, 7(6), 1511–1521.
- Gooseff, M. N. (2010). Defining hyporheic zones—Advancing our conceptual and operational definitions of where stream water and groundwater meet. *Geography Compass*, 4(8), 945–955.
- Hall, R. O., Tank, J. L., Sobota, D. J., Mulholland, P. J., O'Brien, J. M., Dodds, W. K., et al. (2009). Nitrate removal in stream ecosystems measured by 15N addition experiments: Total uptake. *Limnology and Oceanography*, 54(3), 653–665. <https://doi.org/10.4319/lo.2009.54.3.0653>
- Harvey, J. W., Böhlke, J. K., Voytek, M. A., Scott, D., & Tobias, C. R. (2013). Hyporheic zone denitrification: Controls on effective reaction depth and contribution to whole-stream mass balance. *Water Resources Research*, 49, 6298–6316. <https://doi.org/10.1002/wrcr.20492>
- Hester, E. T., Young, K. I., & Widdowson, M. A. (2013). Mixing of surface and groundwater induced by riverbed dunes: Implications for hyporheic zone definitions and pollutant reactions. *Water Resources Research*, 49, 5221–5237. <https://doi.org/10.1002/wrcr.20399>
- Hu, B. X., Huang, H., Hassan, A. E., & Cushman, J. H. (2002). Stochastic reactive transport in porous media: Higher-order closures. *Advances in Water Resources*, 25(5), 513–531.
- Lichtner, P. C., & Tartakovsky, D. (2003). Stochastic analysis of effective rate constant for heterogeneous reactions. *Stochastic Environmental Research and Risk Assessment*, 17(6), 419–429.
- Marzadri, A., Dee, M. M., Tonina, D., Bellin, A., & Tank, J. L. (2017). Role of surface and subsurface processes in scaling N₂O emissions along riverine networks. *PNAS Proceeding of the National Academy of Sciences of the United States of America*, 114(17), 4330–4335. <https://doi.org/10.1073/pnas.1617454114>
- Marzadri, A., Tonina, D., & Bellin, A. (2011). A semianalytical three-dimensional process-based model for hyporheic nitrogen dynamics in gravel bed rivers. *Water Resources Research*, 47, W11518. <https://doi.org/10.1029/2011WR010583>

- Marzadri, A., Tonina, D., & Bellin, A. (2012). Morphodynamic controls on redox conditions and on nitrogen dynamics within the hyporheic zone: Application to gravel bed rivers with alternate-bar morphology. *Journal of Geophysical Research*, *117*, G00N10. <https://doi.org/10.1029/2012JG001966>
- Marzadri, A., Tonina, D., Bellin, A., & Valli, A. (2016). Mixing interfaces, fluxes, residence times and redox conditions of the hyporheic zones induced by dune-like bedforms and ambient groundwater flow. *Advances in Water Resources*, *88*, 139–151.
- Marzadri, A., Tonina, D., Bellin, A., Vignoli, G., & Tubino, M. (2010). Semi-analytical analysis of hyporheic flow induced by alternate bars. *Water Resources Research*, *46*, W07531. <https://doi.org/10.1029/2009WR008285>
- Mulholland, P. J., Helton, A. M., Poole, G. C., Hall, R. O., Hamilton, S. K., Peterson, B. J., et al. (2008). Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature*, *452*, 202–206.
- Packman, A. I., & Brooks, N. H. (2001). Hyporheic exchange of solutes and colloids with moving bed forms. *Water Resources Research*, *37*(10), 2591–2605.
- Packman, A. I., Salheim, M., & Zaramella, M. (2004). Hyporheic exchange with gravel beds: Basic hydrodynamic interactions and bedform-induced advective flows. *Journal of Hydraulic Engineering*, *130*(7), 647–656.
- Peterson, B., Wollheim, W. M., Mulholland, P. J., Webster, J. R., Meyer, J. L., Tank, J. L., et al. (2001). Control of nitrogen export from watershed by headwater streams. *Nature*, *292*, 86–89. <https://doi.org/10.1126/science.1056874>
- Pryshlak, T. T., Sawyer, A. H., Stonedahl, S. H., & Soltanian, M. R. (2015). Multiscale hyporheic exchange through strongly heterogeneous sediments. *Water Resources Research*, *51*, 9127–9140. <https://doi.org/10.1002/2015WR017293>
- Quick, A. M., Reeder, W. J., Farrell, T. B., Tonina, D., Feris, K. P., & Benner, S. G. (2016). Controls on nitrous oxide emissions from the hyporheic zones of streams. *Environmental Science & Technology*, *50*(21), 11,491–11,500. <https://doi.org/10.1021/acs.est.6b02680>
- Ravishankara, A. R., Daniel, J. S., & Portmann, R. W. (2009). Nitrous oxide (N_2O): The dominant ozone-depleting substance emitted in the 21st century. *Science*, *326*, 123–125. <https://doi.org/10.1126/science.1176985>
- Rutherford, J. C. (1994). *River Mixing*. Chichester, England: John Wiley and Sons.
- Salehin, M., Packman, A. I., & Paradis, M. (2004). Hyporheic exchange with heterogeneous streambeds: Laboratory experiments and modeling. *Water Resources Research*, *40*, W11504. <https://doi.org/10.1029/2003WR002567>
- Sawyer, A. H., & Cardenas, M. B. (2009). Hyporheic flow and residence time distributions in heterogeneous cross-bedded sediment. *Water Resources Research*, *45*, W08406. <https://doi.org/10.1029/2008WR007632>
- Shen, H. V., Fehlman, H. M., & Mendoza, C. (1990). Bed form resistances in open channel flows. *Journal of Hydraulic Engineering*, *116*(6), 799–815.
- Smith, V. H., Tilman, G., & Nekola, J. (1999). Eutrophication: Impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. *Environmental Pollution*, *100*, 179–196. [https://doi.org/10.1016/S0269-7491\(99\)00091-3](https://doi.org/10.1016/S0269-7491(99)00091-3)
- Stonedahl, S. H., Harvey, J. W., Wörman, A., Salehin, M., & Packman, A. I. (2010). A multiscale model for integrating hyporheic exchange from ripples to meanders. *Water Resources Research*, *46*, W12539. <https://doi.org/10.1029/2009WR008865>
- Tartakovsky, D. M., & Broyda, S. (2011). PDF equations for advective–reactive transport in heterogeneous porous media with uncertain properties. *Journal of Contaminant Hydrology*, *120*, 129–140.
- Tartakovsky, D. M., & Gremaud, P. A. (2016). Method of distributions for uncertainty quantification. In R. Ghanem, D. Higdon, & O. Houman (Eds.), *Handbook of Uncertainty Quantification* (pp. 763–783). New York: Springer. https://doi.org/10.1007/978-3-319-12385-1_27
- Tonina, D. (2012). Surface water and streambed sediment interaction: The hyporheic exchange. In *Fluid Mechanics of Environmental Interfaces* (Chap. 9, pp. 255–294). Boca Raton, FL: CRC Press Taylor and Francis Group.
- Tonina, D., & Buffington, J. M. (2007). Hyporheic exchange in gravel bed rivers with pool-riffle morphology: Laboratory experiment and three-dimensional modeling. *Water Resources Research*, *43*, W01421. <https://doi.org/10.1029/2005WR004328>
- Tonina, D., de Barros, F., Marzadri, A., & Bellin, A. (2016). Does streambed heterogeneity matter for hyporheic residence time distribution in sand-bedded streams? *Advances in Water Resources*, *96*, 120–126. <https://doi.org/10.1016/j.advwatres.2016.07.009>
- Tonina, D., Marzadri, A., & Bellin, A. (2015). Benthic uptake rate due to hyporheic exchange: The effects of streambed morphology for constant and sinusoidally varying nutrient loads. *Water*, *7*, 398–419. <https://doi.org/10.3390/w7020398>
- Venturi, D., Tartakovsky, D. M., Tartakovsky, A. M., & Karniadakis, G. E. (2013). Exact PDF equations and closure approximations for advective–reactive transport. *Journal of Computational Physics*, *243*, 323–343. <https://doi.org/10.1016/j.jcp.2013.03.001>
- Wörman, A., Packman, A. I., Johansson, H., & Jonsson, K. (2002). Effect of flow-induced exchange in hyporheic zones on longitudinal transport of solutes in stream and rivers. *Water Resources Research*, *38*, 2–1–2-15. <https://doi.org/10.1029/2001WR000769>
- Zarnetske, J. P., Haggerty, R., Wondzell, S. M., Bokil, V. A., & González-Pinzón, R. (2012). Coupled transport and reaction kinetics control on the nitrate sources-sink function of hyporheic zones. *Water Resources Research*, *48*, W11508. <https://doi.org/10.1029/2012WR011894>