

## Quantification of uncertainty in geochemical reactions

Gowri Srinivasan,<sup>1,2</sup> Daniel M. Tartakovsky,<sup>1,3</sup> Bruce A. Robinson,<sup>4</sup>  
and Alejandro B. Aceves<sup>2</sup>

Received 28 February 2007; revised 12 July 2007; accepted 13 September 2007; published 28 December 2007.

[1] Predictions of reactive transport in the subsurface are routinely compromised by both model (structural) and parametric uncertainties. We present a set of computational tools for quantifying these two types of uncertainties. The model uncertainty is resolved at the molecular scale where epistemic uncertainty incorporates aleatory uncertainty. The parametric uncertainty is resolved at both molecular and continuum (Darcy) scales. We use the proposed approach to quantify uncertainty in modeling the sorption of neptunium through a competitive ion exchange. This radionuclide is of major concern for various high-level waste storage projects because of its relatively long half-life and its high-solubility and low-sorption properties. We demonstrate how parametric and model uncertainties affect one's ability to estimate the distribution coefficient. The uncertainty quantification tools yield complete probabilistic descriptions of key parameters affecting the fate and migration of neptunium in the subsurface rather than the lower statistical moments. This is important, since these distributions are highly skewed.

**Citation:** Srinivasan, G., D. M. Tartakovsky, B. A. Robinson, and A. B. Aceves (2007), Quantification of uncertainty in geochemical reactions, *Water Resour. Res.*, 43, W12415, doi:10.1029/2007WR006003.

### 1. Introduction

[2] Subsurface modeling is notoriously plagued by uncertainties that are both epistemic (i.e., reducible through observations) and aleatory (i.e., irreducible because of inherent stochasticity) in nature. While this distinction is somewhat artificial in most hydrologic applications (see the relevant discussion by *Tartakovsky* [2007, and references therein]), biogeochemistry is an area where both types of uncertainties are present. This is because multiple biological and/or chemical reactions occurring simultaneously form a complex nonlinear system, whose dynamics are neither continuous nor entirely deterministic.

[3] Epistemic uncertainty can be subdivided into parametric and structural (model) uncertainty, both of which have been the subject of recent investigations. Parametric uncertainty in geochemistry reflects our partial knowledge of the key parameters governing reactive transport, including distribution coefficient, reaction rate constants and, more fundamentally, the surface areas of reacting species. These and other parameters are uncertain not only because of measurement errors and spatial heterogeneity, but because of a significant scale effect. The latter refers to a discrepancy between parameter values obtained from laboratory and field experiments. A typical example is a difference of several orders of magnitude between field and

laboratory estimates of reaction rate constants [e.g., *Paces*, 1983; *Velbel*, 1986]. Additional flaws in the estimation of field weathering reaction rates involve the failure to properly account for various factors in the chemical rate laws [*Lichtner*, 1993]. Time- and scale-dependent effective kinetic rate constants have been related to grain size distributions [*Lichtner and Tartakovsky*, 2003] and have been variously analyzed via volumetric [*Kechagia et al.*, 2002] and stochastic [*Dagan and Indelman*, 1999] averaging. Recent assessments of parametric uncertainty in reactive transport include *Sanchez-Vila and Rubin* [2003] and *Dentz and Berkowitz* [2005].

[4] Model (structural, conceptual) uncertainty in geochemistry manifests itself on a multiplicity of scales, ranging from field scale, to laboratory scale, to pore scale, to molecular scale. On macroscopic scales, including both field and laboratory scales, the validity of standard mathematical descriptions that are based on the advection-dispersion equation with reactive terms is being challenged because of their failure to capture anomalous transport and other observed phenomena. The recently proposed alternatives include stochastic nonlocal transport models [*Rubin*, 2003, and references therein], continuous time random walk models [*Berkowitz et al.*, 2006], fractional derivative transport equations [*Schumer et al.*, 2003], residence time transfer function approach [*Robinson et al.*, 2003] and delayed diffusion models [*Dentz and Tartakovsky*, 2006]. While in principle one can quantify this source of uncertainty by either selecting the “best” model via the Bayesian maximum entropy approach [*Christakos*, 1990] or averaging over several alternative models via the maximum likelihood Bayesian averaging [*Neuman*, 2003], no attempt to do so in the context of geochemistry has been reported.

[5] Model uncertainty is further exacerbated by the studies [*Tartakovsky et al.*, 2007] that suggest that proper modeling of many geochemical phenomena, including precipitation

<sup>1</sup>Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

<sup>2</sup>Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico, USA.

<sup>3</sup>Department of Mechanical and Aerospace Engineering, University of California, San Diego, La Jolla, California, USA.

<sup>4</sup>Yucca Mountain and Nevada Test Site Environmental Programs, Los Alamos National Laboratory, Los Alamos, New Mexico, USA.

and dissolution, should explicitly account for pore-scale processes. Moreover, the theoretical [Kapoor *et al.*, 1997] and experimental [Raje and Kapoor, 2000; Gramling *et al.*, 2002] investigations showed that it might be necessary to go down to the molecular scale to capture adequately mixing processes that affect the rate of a reaction involving two or more reactants.

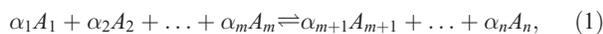
[6] It is at the molecular level that epistemic (model) uncertainty in coupled chemical reactions is intimately intertwined with aleatory uncertainty. The latter is caused by the randomness in collisions of molecules of reacting species and the nonuniformity of molecular mixtures, especially in such complex systems as porous media. It plays a crucial role in selecting a particular reaction path from many possible reaction paths at any given time  $t$ . A proper identification of reaction rate equations is an integral part of a modeling process that is prone to uncertainty. To quantify this source of epistemic and aleatory uncertainties, we employ the stochastic simulation algorithm of Gillespie [1977] that is used extensively in biochemistry. This algorithm provides probabilistic predictions of the number of reacting molecules of a given species that participate in the chemical reaction process at a given time.

[7] Our main goals are to introduce computational tools for quantification of parametric and model uncertainties in geochemical reactions and to investigate the impact of these types of uncertainty on one's ability to predict reaction rates, distribution coefficients, etc. In section 2, we formulate a set of equations describing geochemical reactions between multiple species and analyze them deterministically, i.e., ignoring uncertainty. Sections 3 and 4 present computational tools for quantifying parametric and model uncertainties respectively. These tools are used in section 5 to ascertain the importance of the two types of uncertainties by considering a relatively simple reactive system involving two reactants. Uncertainty in a more challenging problem, sorption of neptunium through the competitive ion exchange, is quantified in section 6.

## 2. Chemical Reactions

[8] To analyze epistemic and aleatory uncertainties affecting reactive transport in the subsurface, we focus on chemical reactions while disregarding other transport mechanisms, including advection, diffusion and hydrodynamic dispersion. (This approach is justified in the concluding remarks in section 7).

[9] Consider a (reversible) chemical reaction between  $n$  species  $A_1, A_2, \dots, A_n$  that can be represented as



where  $\{\alpha_i\}_{i=1}^n$  are stoichiometric coefficients. The concentration  $C_i(t) \equiv [A_i]$  of a reacting species  $A_i$  at time  $t$  is described by a nonlinear rate equation,

$$\frac{dC_i}{dt} = F_i(C_1, C_2, \dots, C_n), \quad i = 1, \dots, n, \quad (2)$$

subject to appropriate initial conditions. Model (and aleatory) uncertainty refers to the imperfect knowledge about the functional forms of  $F_i$  ( $i = 1, \dots, n$ ). Parametric

uncertainty refers to the imperfect knowledge about the coefficients entering the functions  $F_i$  ( $i = 1, \dots, n$ ) and/or initial concentrations.

[10] A simple example of (1)–(2) consists of an irreversible chemical reaction involving three reacting species, two reactants  $A_1$  and  $A_2$  forming a species  $A_3$ , such that



A deterministic framework for modeling this reaction postulates reaction rate equations for the concentrations  $C_i$  of each species, e.g.,

$$\frac{dC_i}{dt} = -kC_1C_2, \quad i = 1, 2, \quad \frac{dC_3}{dt} = kC_1C_2, \quad (4)$$

prescribes a reaction rate constant  $k$ , and specifies initial conditions, e.g.,

$$C_1(0) = C, \quad C_2(0) = C, \quad C_3(0) = 0, \quad (5)$$

where the initial concentration of reactants has a precisely determined value  $C$ . The initial concentrations of  $C_1$  and  $C_2$  do not have to be the same. We set  $C_1(0) = C_2(0) = C$  to simplify the presentation.

[11] The system of reaction equations (4)–(5) implies that  $C_1(t) = C_2(t)$ . This yields analytical solutions for the concentration of each species,

$$C_1(t) = C_2(t) = \frac{C}{Ckt + 1}, \quad C_3(t) = \frac{Ckt}{Ckt + 1}. \quad (6)$$

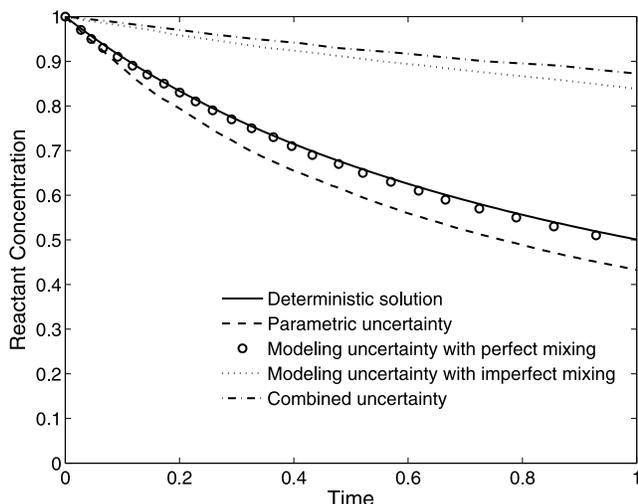
These solutions are represented by the solid lines in Figures 1 and 2 for  $C = 1$  and  $k = 1$ , both defined in consistent units.

[12] By its very nature, this deterministic analysis ignores uncertainty in both reaction rate constant  $k$  (parametric uncertainty) and the choice of reaction rate equations (4) (modeling uncertainty). In the following sections, we introduce mathematical tools to analyze these two sources of uncertainty and demonstrate their importance.

## 3. Quantification of Parametric Uncertainty

[13] Incomplete knowledge of the reaction rate constants  $k$  in (4) at the field scale gives rise to parametric uncertainty. Among various factors affecting the values of reaction rate constants are the heterogeneity of a subsurface environment and the presence of other reacting species whose minute concentrations might not be of immediate concern but affect the rate of the reactions.

[14] The latter source of uncertainty introduces temporal fluctuations of reaction rate constants, which cannot be accurately ascertained under field conditions. This uncertainty can be quantified by treating the reaction rate constant  $k$  as a random process  $k(t)$  with a prescribed distribution (e.g., Gaussian or lognormal), mean  $\bar{k}$ , variance  $\sigma_k^2$ , and a correlation structure (e.g., white noise used in our simulations). This renders the reaction rate equations (4) stochastic, so that their solutions are given in terms of probability density functions.



**Figure 1.** Predictions (ensemble means) of the reactant's concentration reflecting various sources of uncertainty.

### 3.1. Solutions of Fokker-Planck Equation

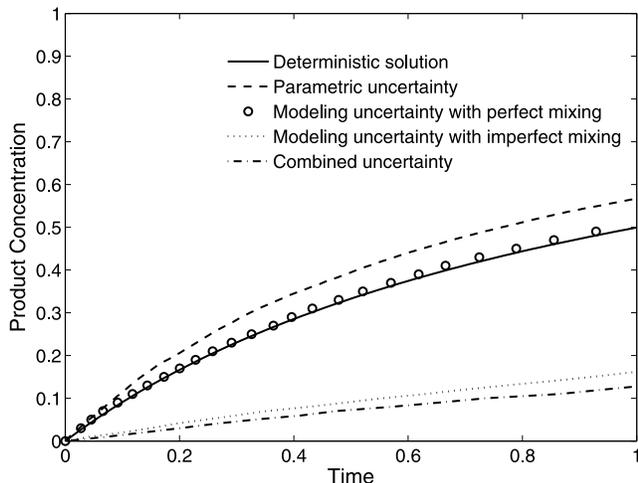
[15] Close inspection of the Langevin-type equations (4) reveals that they have only one independent variable, say concentration  $C_1$ . Its probability density function  $p(c_1; t)$  satisfies the Fokker-Planck equation [Risken, 1989],

$$\frac{\partial p}{\partial t} = -\frac{\partial v p}{\partial c_1} + \frac{\partial}{\partial c_1} \left( D \frac{\partial p}{\partial c_1} \right), \quad (7)$$

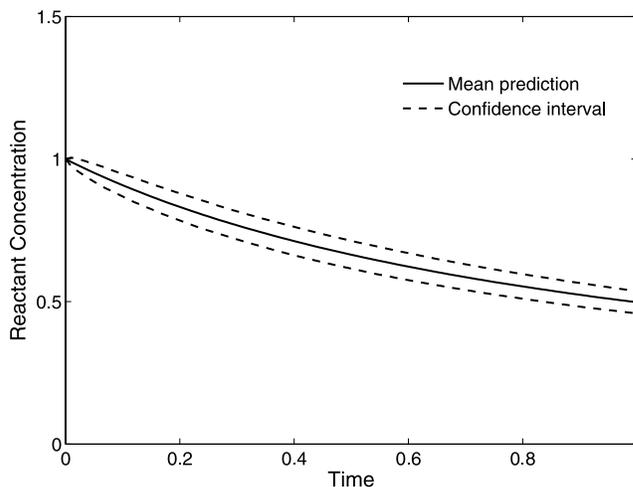
where the drift and diffusion coefficients according to Itô are given by

$$v = -\bar{k}c_1^2 \quad \text{and} \quad D = \frac{1}{2}c_1^4, \quad (8)$$

respectively. An initial condition for (7) reflects the degree of certainty in the initial concentration  $C$  and, in general, is expressed by a probability density function,  $p(c_1; 0) = p_0(c_1)$ . If it is known precisely, the initial condition takes the form  $p(c_1; 0) = \delta(c_1 - C)$ , where  $\delta(x)$  is the Dirac delta



**Figure 2.** Predictions (ensemble means) of the product's concentration reflecting various sources of uncertainty.



**Figure 3.** Confidence interval (CI) resulting from quantification of parametric uncertainty for Gaussian reaction rate constant  $k$ .

function. In the simulations below, we employ the latter initial condition.

[16] We solved the Fokker-Planck equation (7)–(8) both analytically and numerically with a forward Euler algorithm. Once  $p(c_1; t)$  is found, the  $n$ th statistical moment of concentrations can be computed as

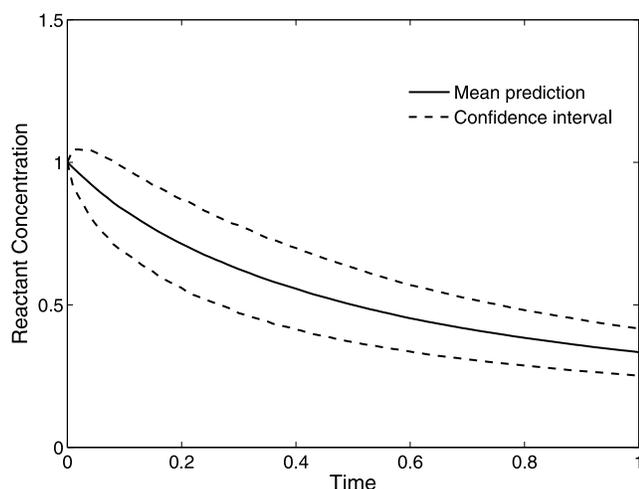
$$\langle C_1^n(t) \rangle = \int c_1^n p(c_1; t) dc_1. \quad (9)$$

Numerical and analytical solutions for the mean (the dashed lines in Figures 1 and 2) are identical. These results correspond the reaction rate constant  $k$  that is modeled as a lognormal process with mean 0 and variance 0.1. This corresponds to the geometric mean  $\bar{k} = 1$ , with possible values of  $k$  spanning 3 orders of magnitude. A comparison of the dashed and solid lines in Figures 1 and 2 show that incorporating parametric uncertainty leads to faster reaction rate predictions than in the deterministic case.

[17] Next, we compute confidence intervals for the concentration estimates  $\bar{C}_i$ , which are defined as  $\bar{C}_i \pm 2\sigma_{C_i}$ . If  $C_i$  were Gaussian, that would correspond to the 95% confidence interval. To investigate the impact of the parametric distributional assumption on uncertainty quantification, we alternatively treat  $k$  as either Gaussian or lognormal process. In both cases, the mean value of  $k$  is set to  $\bar{k} = 1$  and its standard deviation is set to  $\sigma_k = 0.33$ . The corresponding confidence intervals for the estimates of  $C_1$  are shown by the dashed lines in Figures 3 and 4. One can see that the contribution of parametric uncertainty to the overall predictive uncertainty is quite significant, with larger parametric uncertainty translating into larger predictive uncertainty.

### 3.2. Direct Solutions of Stochastic Partial Differential Equations

[18] For general, fully coupled chemical reactions with  $n$  reacting species (1), the dimensionality of the Fokker-Planck equations can be as high as  $n + 1$ . The computational expense associated with their solutions is often so high as to render this approach impractical. In this case, one can solve stochastic differential equations (2) or (4) directly.



**Figure 4.** Confidence interval (CI) resulting from quantification of parametric uncertainty for lognormal reaction rate constant  $k$ .

[19] An Itô formulation of the Langevin-type equations (4) can be solved numerically by using the following stochastic Euler discretization scheme [Kloeden and Platen, 1992],

$$C_1(t + \Delta t) = C_1(t) - \bar{k}C_1(t)C_2(t)\Delta t - C_1(t)C_2(t)\Delta W, \quad (10)$$

where  $\Delta t$  is a time step and  $\Delta W$  is a Gaussian variable with zero mean and the variance proportional to the variance of  $k$  and the time step  $\Delta t$ . This is in effect the same as varying the reaction rate constant at each time step by generating a random number with mean  $\bar{k}$  and a given variance. Similar equations can be written for the concentrations  $C_2$  and  $C_3$ .

[20] To compute the statistical moments of the concentrations  $C_i$  ( $i = 1, \dots, 3$ ), one should average the results of these calculations over many realizations. In our simulations, 1000 realizations were sufficient to reproduce accurately the mean. The corresponding curves coincide with their analytical and numerical counterparts computed by solving the Fokker-Plank equations, so that the dashed lines in Figures 1 and 2 represent the mean concentrations computed with (10).

#### 4. Quantification of Model Uncertainty

[21] Reaction rate equations (4) provide a macroscopic description (or a mathematical abstraction) of chemical reactions that occur at the molecular level. To quantify the uncertainty associated with the selection of this particular model, one can attempt to derive it from “first principles” by accounting for the interactions of molecules of the species  $A_1$  and  $A_2$ . This can be achieved by deriving the Master equation for the probability distribution of the collisions [Risken, 1989]. Solving the Master equation can become prohibitively expensive as the number of species increases.

[22] The stochastic simulation algorithm of Gillespie [1977] provides a computationally efficient alternative to the Master equation approach. At the heart of this algorithm

lies the concept of a reaction probability density function  $P(\tau, \mu)$ , which is the joint probability for reaction  $\mu$  to occur in the infinitesimal time interval  $[t + \tau, t + \tau + \Delta\tau]$  given a certain state at time  $t$ . A particular reaction  $\mu$  is selected randomly from a set of all possible reactions, which is determined by the molecular populations of reactants for each individual reaction.

[23] The time  $\tau$ , during which no reactions occur, depends upon the total molecular population of all reacting species and is conceptually similar to the residence time reflecting the randomness of collisions. Assigning a constant deterministic value to  $\tau$  yields standard deterministic reaction rate equations, such as (4), which are discretized in time with a time step  $\tau$ .

[24] The Gillespie [1977] algorithm assumes that molecular populations of reactants are uniform in space, which results in reaction rate constants  $k$  that are known with certainty (deterministic). This assumption breaks down in complex environments, such as porous media. To account for the variability in mixing, we modify  $\tau$  (see below) by a random parameter  $m > 1$ , such that  $m = 1$  corresponds to uniform mixing and the larger  $m$  the longer it takes for reacting molecules to collide. More complex mixing laws [e.g., Robinson and Viswanathan, 2003] or mechanistic simulations designed to inform the selection of  $m$  could be incorporated as well.

[25] For the reacting system used as an example in section 2,  $\mu$  represents the only possible reaction (3) and  $\tau$  is the only random variable. The modified Gillespie algorithm we propose consists of the following steps.

[26] 1. Compute the total number of the reacting pairs of available molecules,  $a_0 = kC_1C_2$ .

[27] 2. Generate random numbers  $r_1$  on the uniform unit interval and  $m$  uniformly random on the interval  $[1, 10]$ .

[28] 3. Compute  $\tau$  as

$$\tau = -ma_0^{-1} \ln r_1. \quad (11)$$

[29] 4. Increase time by  $\tau$ , decrease the number of reacting molecules of  $A_1$  and  $A_2$  by 1 each, and increase the number of molecules of the product  $A_3$  by 1.

[30] 5. Repeat steps 1–4 until the reactant population goes to zero.

[31] The result of these simulations is a stochastic realization of concentration profiles. The statistics of concentrations are computed in the Monte Carlo fashion by averaging over an ensemble of realizations. The circles in Figures 1 and 2 represent the mean concentrations of reacting species  $A_1$  and  $A_2$  and their product  $A_3$ , obtained from 1000 realizations of  $r_1$  for deterministic  $m = 1$ . Since these simulations reproduce the conditions of uniform mixing and deterministic reaction rate constants, they coincide with the deterministic concentration predictions, a slight discrepancy between the two stemming from the finite size of Monte Carlo sampling. The dotted lines in Figures 1 and 2 represent the mean of 1000 realizations of the random pair  $(r_1, m)$ . They demonstrate how the heterogeneity of molecular mixture (random  $m > 1$ ) affects the mean concentrations of the reacting species by predicting significantly slower rates for the chemical reaction. The

imperfectly mixed reacting mixture results in molecules of the reacting species taking longer to come into contact, thus causing the reaction to proceed at a slower rate.

## 5. Parametric and Model Uncertainties

[32] The algorithm described above does not explicitly account for parametric uncertainty, i.e., uncertainty about the actual value of reaction rate constant  $k$ . To account for both sources of uncertainty, we modify the first step of the algorithm by replacing the constant value  $k$  with its random counterpart whose mean and variance are  $\bar{k}$  and  $\sigma_k^2$ , respectively. The statistics of concentrations are computed in the Monte Carlo fashion by averaging over an ensemble of realizations. The dash-dotted lines in Figures 1 and 2 represent the mean concentrations of the reacting species  $A_1$  and  $A_2$  and their product  $A_3$ , obtained from 1000 realizations, where each realization draws one random triplet from the set  $(r_1, m, k)$ .

[33] The tools developed above can be used to quantify the relative effects of parametric and model uncertainty on one's ability to model geochemical reactions. This section provides such an analysis for the relatively simple irreversible chemical reaction (3) involving two reactants and one product. A more challenging reactive system, dealing with neptunium ion exchange, is analyzed in section 6.

[34] Figures 1 and 2 compare the predictions of the evolution of the concentration of the reactants  $A_1$  or  $A_2$  and the product  $A_3$ , respectively. These predictions alternatively ignore uncertainty (the solid lines), account for parametric uncertainty (the dotted line), account for model uncertainty (the dashed lines), and account for both sources of uncertainty (the dash-dotted lines). The total time of 1 unit corresponds to the time it takes the reactant concentration to decrease by 50% in the deterministic scenario.

[35] The deterministic predictions are obtained by replacing the time fluctuating reaction rate constant  $k$  with its mean value  $\bar{k}$ , i.e., by setting  $k = \bar{k} = 1$ . To facilitate the comparison, numerical simulations to quantify combined parametric and model uncertainties (the generalized stochastic simulation algorithm) used the same random numbers as those generated to solve the stochastic differential equations (10) for quantification of parametric uncertainty and to implement the Gillespie algorithm for quantification of model uncertainty.

[36] Parametric uncertainty leads to predictions of reaction rates that are faster than those obtained either by ignoring uncertainty or by accounting for model uncertainty only or by accounting for both sources of uncertainty. The nonlinearity of the reactions (4) and the skewness of the lognormal distribution of  $k$  ensures that even though the randomly generated values of  $k$  had the same mean value  $\bar{k}$  as that used in deterministic simulations, the resulting reaction is always faster than the reaction predicted by deterministic simulations.

[37] Model uncertainty results in predicted reactions that are slower than their deterministic counterparts. Accounting for both parametric and model uncertainties leads to the slowest predicted reactions. This behavior is caused by the nonlinearity of the reacting system, in which the effect of both parametric and model uncertainties is to predict reaction rates that are slower than those predicted by quantifying model uncertainty only.

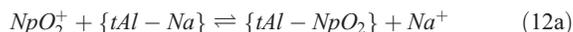
[38] Figures 3 and 4 provide probabilistic estimates of the reactant concentration and the corresponding confidence intervals for Gaussian and lognormal reactive rate constants  $k$ , respectively. The confidence intervals are defined here as  $\bar{C}_i \pm 2\sigma_{C_i}$ . The normal distribution has the mean  $\bar{k} = 1$  and standard deviation  $\sigma_k = 0.33$ , and the lognormal distribution has the mean  $\log \bar{k} = 0$  and standard deviation  $\log \sigma_k = 0.33$ . The higher variability of  $k$  in the latter case translates into the wider confidence intervals, i.e., into the higher predictive uncertainty.

[39] The analysis above reveals that the impact of both sources of uncertainty on geochemical reactions is significant and cannot be ignored.

## 6. Neptunium Ion Exchange Example

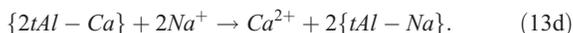
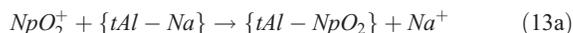
[40] A more challenging example of a coupled chemical reaction system (1) is the neptunium competitive ion exchange. Neptunium-237 is one of the key radionuclides investigated by several high-level waste storage projects because of its long half-life and its high-solubility and low-sorption properties. Migration of neptunium-237 from the proposed repository site could be significantly retarded because of sorption by zeolitic minerals that lie beneath the repository. Minerals such as sodium, calcium and magnesium ions that are naturally present in the groundwater compete with neptunium ions for the same sorption sites. In other words, the geochemical processes affecting the fate and transport of neptunium-237 are complex and highly uncertain.

[41] One of these processes is a competitive ion exchange mechanism, where the positively charged sodium  $Na^+$  and calcium  $Ca^{2+}$  ions are in competition with neptunium ions  $NpO_2^+$ ,



Reaction (12a) describes a neptunium ion displacing a sorbed sodium ion from a zeolite  $tAl^-$  that acts as a sorption site. Reaction (12b) describes the displacement of a calcium ion by two sodium ions.

[42] This set of reversible coupled chemical reactions, which involves six reacting species, can be represented by four irreversible coupled reactions,



Let  $C_1$  and  $C_2$  denote the concentrations of sorbed ( $\{tAl-Na\}$ ) and aqueous ( $Na^+$ ) sodium ions, respectively;  $C_3$  and  $C_4$  the concentrations of sorbed ( $\{tAl-NpO_2\}$ ) and

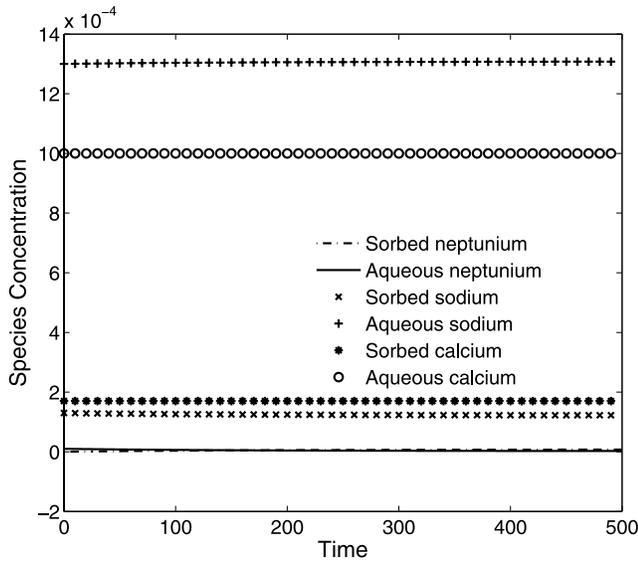


Figure 5. Log concentrations of the reacting species.

aqueous ( $NpO_2^+$ ) neptunium ions, respectively; and  $C_5$  and  $C_6$  the concentrations of sorbed ( $\{2Al-Ca\}$ ) and aqueous ( $Ca^{2+}$ ) calcium ions, respectively.

[43] In addition to these concentrations, one is also interested in computing transport parameters that affect the fate and migration of neptunium-237 in groundwater. The distribution coefficient  $K_d$  is defined as the ratio of the concentrations of sorbed neptunium ions ( $\{Al-NpO_2^+\}$ ) and those present in the aqueous phase ( $NpO_2^+$ ),

$$K_d = C_3/C_4. \tag{14}$$

### 6.1. Deterministic Analysis

[44] A deterministic analysis of the reacting system (13) relies on the rate equations

$$\frac{dC_1}{dt} = -k_1 C_1 C_4 + k_2 C_2 C_3 - 2k_3 C_1^2 C_6 + 2k_4 C_2^2 C_5 \tag{15a}$$

$$\frac{dC_2}{dt} = k_1 C_1 C_4 - k_2 C_2 C_3 + 2k_3 C_1^2 C_6 - 2k_4 C_2^2 C_5 \tag{15b}$$

$$\frac{dC_3}{dt} = k_1 C_1 C_4 - k_2 C_2 C_3 \tag{15c}$$

$$\frac{dC_4}{dt} = -k_1 C_1 C_4 + k_2 C_2 C_3 \tag{15d}$$

$$\frac{dC_5}{dt} = k_3 C_1^2 C_6 - k_4 C_2^2 C_5 \tag{15e}$$

$$\frac{dC_6}{dt} = -k_3 C_1^2 C_6 + k_4 C_2^2 C_5, \tag{15f}$$

where  $k_1$ ,  $k_2$ ,  $k_3$  and  $k_4$  represent the reaction rate constants for the reactions (13a), (13b), (13c), and (13d), respectively. The average values of these coefficients,  $k_1 = 10^{1.58}$ ,  $k_2 = 1$ ,  $k_3 = 10^{1.5}$  and  $k_4 = 1$ , have been determined experimentally in a report cited by *Viswanathan et al.* [1998]. It is worthwhile to point out that these experiments measured the average equilibrium constants  $K$  defined as the ratio of forward to backward reaction rates,  $K_1 = k_1/k_2$  and  $K_2 = k_3/k_4$ . Determination of the values of reactions rate constants from such experiments is, obviously, not unique.

[45] Initial concentrations,  $C_1 = 1.3 \times 10^{-4}$ ,  $C_2 = 1.3 \times 10^{-3}$ ,  $C_3 = 0$ ,  $C_4 = 10^{-5}$ ,  $C_5 = 1.7 \times 10^{-4}$ , and  $C_6 = 10^{-3}$  (units are mol/L), have been determined [*Viswanathan et al.*, 1998] from the average concentrations of sodium ions in the groundwater by sampling at various locations. Since the solubility of neptunium in water at a given pH is known, one can determine the concentration of neptunium in groundwater. Assuming that initially all neptunium is dissolved in the groundwater, one obtains initial conditions for (15).

[46] Solutions of the rate equations (15) subject to these initial conditions are shown in Figure 5. Since the concentrations of sodium and calcium ions are 2 orders of magnitude higher than the initial concentration of neptunium (aqueous), and since the concentrations of sodium and calcium ions (both sorbed and aqueous) do not change significantly over time, the log plots of these concentrations look superficially constant. Figure 6 elucidates the temporal variability of the concentrations of the sorbed and aqueous neptunium ions. Both concentrations reach steady state, at which point they were substituted into (14) to compute the distribution coefficient  $K_d = 3.53$ .

[47] A host of uncertainties, both structural and parametric, are unaccounted for in this deterministic analysis. For example, while the initial concentrations of sodium and calcium ions are not known exactly, they have a profound effect on the distribution coefficient. [*Viswanathan et al.*, 1998] compared the values of  $K_d$  computed numerically for various initial sodium ion concentrations with the experimental values recorded by *Tait et al.* [1996]. They found that  $K_d$  varies from 0.2 to 5.5 when the initial sodium ion

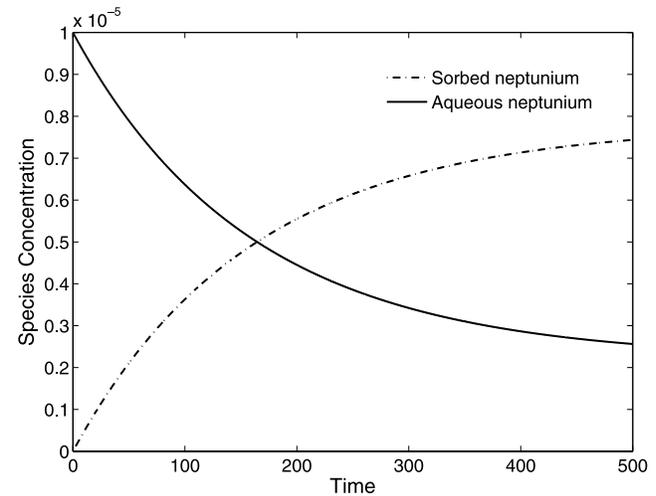
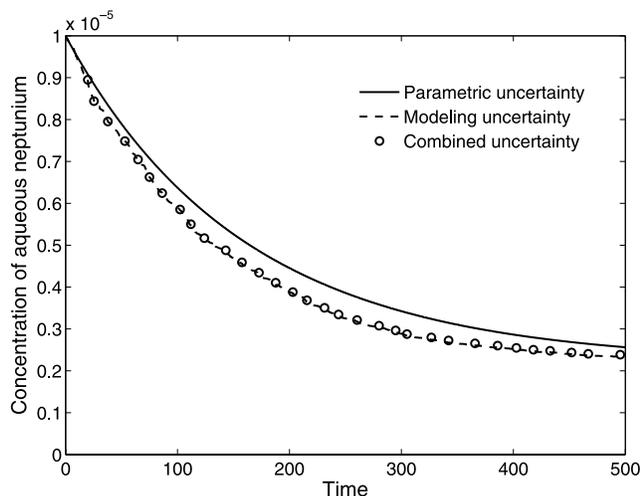


Figure 6. Log concentration of sorbed and aqueous neptunium ions.



**Figure 7.** Impact of various sources of uncertainty on the predicted (log) concentration of aqueous neptunium.

concentration varies from 0.1 mol/L to 0.002 mol/L, respectively.

### 6.2. Quantification of Parametric Uncertainty

[48] To account for uncertainty in the reaction rate constants  $k_1$ ,  $k_2$ ,  $k_3$ , and  $k_4$ , we treat them as normal white noise, whose mean values were determined from the experimental data in a report cited by *Viswanathan et al.* [1998]. Reasonable variance values were assumed for the purpose of illustration. This renders the rate equations (15) stochastic. Since the corresponding Fokker-Planck equation is seven-dimensional, solving it either analytically or numerically is not straightforward.

[49] Instead, we use the discretization scheme of the Itô formulation of the Langevin-type equations (15), as described in section 3.2. The mean concentrations of aqueous and sorbed neptunium computed with this algorithm are shown by the dashed lines in Figures 7 and 8, respectively.

### 6.3. Quantification of Model Uncertainty

[50] We use the modified stochastic simulation algorithm (section 4) to quantify model uncertainty in the ion exchange reactions. Now the reaction probability density function  $P(\tau, \mu)$  is truly a joint probability for reaction  $\mu$  to occur in the infinitesimal time interval  $[t + \tau, t + \tau + \Delta\tau]$ , since there is more than one reaction involved. The algorithm is similar to the single irreversible reaction case described in section 4, with an additional step of determining which reaction  $\mu$  occurs.

[51] 1. Compute the total number of reacting pairs of molecules available for each reaction,

$$a_1 = k_1 C_1 C_4, \quad a_2 = k_2 C_2 C_3 \quad (16a)$$

$$a_3 = k_3 C_1^2 C_6, \quad a_4 = k_4 C_2^2 C_5, \quad (16b)$$

and compute their sum  $a_0 = a_1 + a_2 + a_3 + a_4$ .

[52] 2. Generate random numbers  $r_1$  and  $r_2$  on the uniform unit interval and  $m$  uniformly random on the interval  $[1, 10]$ .

[53] 3. Compute  $\tau$  as

$$\tau = -ma_0^{-1} \ln r_1. \quad (17)$$

[54] 4. Determine which reaction  $\mu$  occurs by taking  $\mu$  to be that integer for which

$$\sum_{j=1}^{\mu-1} a_j < r_2 a_0 \leq \sum_{j=1}^{\mu} a_j. \quad (18)$$

[55] 5. Update time by  $\tau$  and molecular levels for reaction  $\mu$  (decrease reactants by 1 and increase products by 1).

[56] 6. Repeat steps 1–5 until either of the reactant population goes to zero or steady state is reached, in this case, the latter.

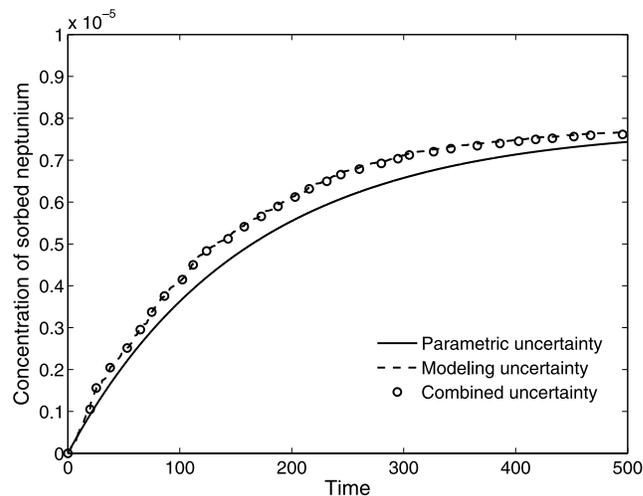
[57] The result of these simulations is a stochastic realization of the concentration profiles. The statistics of concentrations are computed in the Monte Carlo fashion by averaging over an ensemble of realizations. The circles in Figures 7 and 8 represent the mean concentrations of aqueous and sorbed neptunium, respectively, obtained from 1000 realizations of the parametric triplet  $(r_1, r_2, m)$ .

### 6.4. Joint Uncertainty Quantification

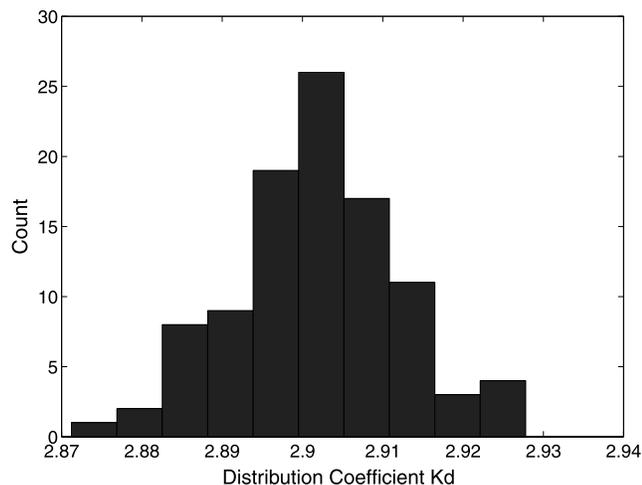
[58] To quantify both parametric and model uncertainties, we modified the stochastic simulation algorithm in section 6.3 to allow for random fluctuations of reaction rate constant (see also section 5). To facilitate the comparison of the effects of various sources of uncertainty, we used the same set of random numbers in all three computations.

[59] Figures 7 and 8 reveal that parametric uncertainty plays a smaller role than the model uncertainty in the overall ability to predict chemical reactions. Model uncertainty leads to predictions of reactions that are slower than their deterministic counterparts. This is because it takes into account the possibility of imperfect collisions and mixing levels of reactants.

[60] Accounting for both model and parametric uncertainties leads to predictions of reactions that are similar to those that incorporate only model uncertainty, and are



**Figure 8.** Impact of various sources of uncertainty on the predicted (log) concentration of sorbed neptunium.



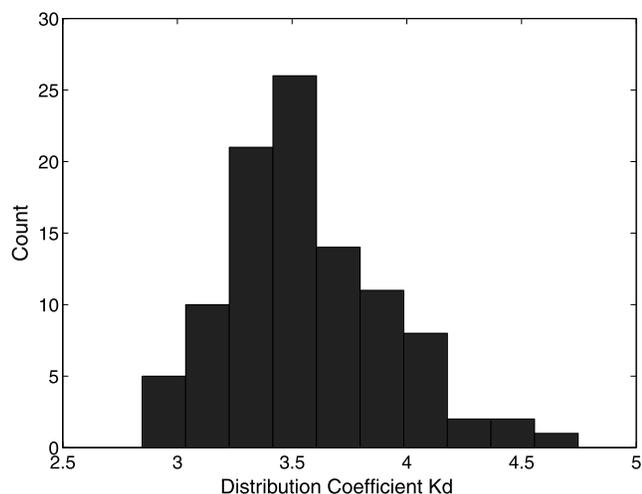
**Figure 9.** Histogram of the distribution coefficient values accounting for parametric uncertainty.

slower than those that account for parametric uncertainty only. This finding is to be expected, because of the Gaussian distribution of reaction rate constants, which ensures that the effects of higher reaction rate constants in predicting faster reaction rates are similar to the effect of the corresponding lower reaction rate constants.

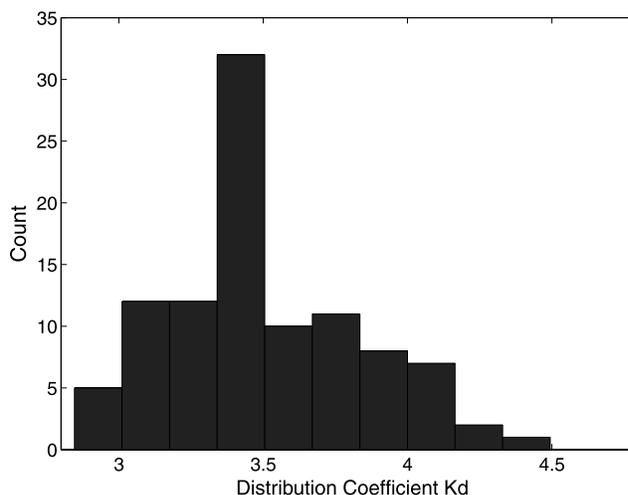
[61] To compute the distribution coefficient  $K_d$ , we allow the reactions to reach steady state. The corresponding concentrations of neptunium in sorbed and aqueous phases are substituted into (14) to compute  $K_d$ . The values of the distribution coefficient  $K_d$  computed for every realization of the numerical algorithms that account for parametric uncertainty, model uncertainty, and both types of uncertainty are reported as histograms in Figures 9, 10, and 11, respectively.

[62] The computed values of  $K_d$  ranged from 2.5 to 5.0, with the mean values of 3.53, 3.51 and 3.39 obtained by accounting for parametric, model, and both sources of uncertainty, respectively.

[63] To test the robustness of our results and conclusions, we considered two random models for the reaction rate



**Figure 10.** Histogram of the distribution coefficient values accounting for model uncertainties.



**Figure 11.** Histogram of the distribution coefficient values accounting for both sources of uncertainty.

constants and mixing parameters (Gaussian and lognormal) with a wide range of variances. While not shown here graphically, the results of these simulations reveal that the range of the distribution coefficient  $K_d$  was not significantly affected by the variances of the input parameters. When the reaction rate constants were allowed to vary by 2 orders of magnitude according to the lognormal distribution, the joint uncertainty model yielded slightly higher values of  $K_d$ , which ranged from 4.2 to 4.7.

## 7. Conclusions

[64] We presented a set of mathematical tools to quantify parametric and model uncertainty in coupled geochemical reactions. These were used to study both a simple irreversible reaction involving two reacting species and more realistic reversible reactions involving neptunium ion exchange. Our analysis leads to the following major conclusions.

[65] 1. The rate at which a reaction occurs varies greatly because of the inclusion of various sources of uncertainty.

[66] 2. The range of distribution coefficients obtained for the neptunium competitive ion exchange process considered underscores the need for uncertainty quantification. It is expected that the same behavior will be present in more complex, multicomponent systems involving more chemical species.

[67] 3. The proposed approach yields a complete probabilistic description of the reaction rates and distribution coefficient, key parameters affecting the fate and migration of neptunium in the subsurface. This is important, since these distributions are highly skewed.

[68] 4. The uncertainty quantification tools employed in this study are fully scalable and can be used to investigate any number of coupled reversible or irreversible geochemical reactions.

[69] 5. Transport of radionuclides such as neptunium could be greatly affected by the reactions that take place in the subsurface and hence quantification of modeling and parametric uncertainties is crucial in describing the overall transport.

[70] The uncertainty quantification of geochemical reactions carried out in this study is important in its own right. It

can also be incorporated into reactive transport models by employing operator splitting techniques [Kirkner and Reeves, 1988].

[71] **Acknowledgments.** This research was performed under the auspices of the Los Alamos National Laboratory. It was partially supported by the DOE Office of Science Advanced Scientific Computing Research (ASCR) program in Applied Mathematical Sciences and by the Office of Science (BER), Cooperative Agreement DE-FC02-07ER64324. The authors are thankful to Demetris Koutsoyiannis, Associate Editor, and two anonymous reviewers for their insightful reviews of the manuscript.

## References

- Berkowitz, B., A. Cortis, M. Dentz, and H. Scher (2006), Modeling non-Fickian transport in geological formations as a continuous time random walk, *Rev. Geophys.*, *44*, RG2003S, doi:10.1029/2005RG000178.
- Christakos, G. (1990), A Bayesian maximum entropy view to the spatial estimation problem, *Math. Geol.*, *22*(7), 763–777.
- Dagan, G., and P. Indelman (1999), Reactive solute transport inflow between a recharging and a pumping well in a heterogeneous aquifer, *Water Resour. Res.*, *35*(12), 3639–3647.
- Dentz, M., and B. Berkowitz (2005), Exact effective transport dynamics in a one-dimensional random environment, *Phys. Rev. E*, *72*, 031110, doi:10.1103/PhysRevE.72.031110.
- Dentz, M., and D. M. Tartakovsky (2006), Delay mechanisms of non-Fickian transport in heterogeneous media, *Geophys. Res. Lett.*, *33*, L16406, doi:10.1029/2006GL027054.
- Gillespie, D. T. (1977), Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.*, *81*(25), 2340–2361.
- Gramling, C. M., C. F. Harvey, and L. C. Meigs (2002), Reactive transport in porous media: A comparison of model prediction with laboratory visualization, *Environ. Sci. Technol.*, *36*, 2508–2514.
- Kapoor, V., L. W. Gelhar, and F. Miralles-Wilhelm (1997), Bimolecular second order reactions in spatially varying flows: Segregation induced scale dependent transformation rates, *Water Resour. Res.*, *33*(4), 527–536.
- Kechagia, P., I. N. Tsimpanogiannis, Y. C. Yortsos, and P. C. Lichtner (2002), On the upscaling of reaction-transport processes in porous media with fast or finite kinetics, *Chem. Eng. Sci.*, *57*, 2565–2577.
- Kirkner, D. J., and H. Reeves (1988), Multicomponent mass transport with homogeneous and heterogeneous chemical reactions: The effect of the chemistry on the choice of numerical algorithm: 1. Theory, *Water Resour. Res.*, *24*(10), 1719–1729.
- Kloeden, P. E., and E. Platen (1992), *Numerical Solution of Stochastic Differential Equations*, Springer, New York.
- Lichtner, P. C. (1993), Scaling properties of time-space kinetic mass transport equations and the local equilibrium limit, *Am. J. Sci.*, *293*, 257–296.
- Lichtner, P. C., and D. M. Tartakovsky (2003), Stochastic analysis of effective rate constant for heterogeneous reactions, *Stochastic Environ. Res. Risk Assess.*, *17*, 419–429.
- Neuman, S. P. (2003), Maximum likelihood Bayesian averaging of uncertain model predictions, *Stochastic Environ. Res. Risk Assess.*, *17*(5), 291–305.
- Paces, T. (1983), Rate constant of dissolution derived from the measurements of mass balances in catchments, *Geochim. Cosmochim. Acta*, *47*, 1855–1863.
- Raje, D. S., and V. Kapoor (2000), Experimental study of bimolecular reaction kinetics in porous media, *Environ. Sci. Technol.*, *34*, 1234–1239.
- Risken, H. (1989), *The Fokker Planck Equation: Methods of Solution and Applications*, 2nd ed., Springer, New York.
- Robinson, B. A., and H. S. Viswanathan (2003), Application of the theory of micromixing to groundwater reactive transport models, *Water Resour. Res.*, *39*(11), 1313, doi:10.1029/2003WR002368.
- Robinson, B. A., C. Li, and C. K. Ho (2003), Performance assessment model development and analysis of radionuclide transport in the unsaturated zone, Yucca Mountain, Nevada, *J. Contam. Hydrol.*, *62–63*, 249–268.
- Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, Oxford Univ. Press, New York.
- Sanchez-Vila, X., and Y. Rubin (2003), Travel time moments for sorbing solutes in heterogeneous domains under nonuniform flow conditions, *Water Resour. Res.*, *39*(4), 1086, doi:10.1029/2002WR001399.
- Schumer, R., D. A. Benson, M. M. Meerschaert, and B. Baeumer (2003), Fractal mobile/immobile solute transport, *Water Resour. Res.*, *39*(10), 1296, doi:10.1029/2003WR002141.
- Tait, C. D., S. A. Ekberg, D. L. Clark, P. D. Palmer, I. R. Triay, and C. R. Cotter (1996), Sorption as a function of groundwater chemistry to elucidate sorption mechanisms, Yucca Mountain Site Characterization Proj. Milestone 3219, technical report, Los Alamos Natl. Lab, Los Alamos, N. M.
- Tartakovsky, A. M., P. Meakin, T. D. Scheibe, and B. D. Wood (2007), A smoothed particle hydrodynamics model for reactive transport and mineral precipitation in porous and fractured porous media, *Water Resour. Res.*, *43*, W05437, doi:10.1029/2005WR004770.
- Tartakovsky, D. M. (2007), Probabilistic risk analysis in subsurface hydrology, *Geophys. Res. Lett.*, *34*, L05404, doi:10.1029/2007GL029245.
- Velbel, M. (1986), The mathematical basis for determining rates of geochemical and geomorphic processes in small forested watersheds by mass balance. Examples and implications, in *Rates of Chemical Weathering of Rocks and Minerals*, edited by S. Coleman and D. Dethier, pp. 439–451, Academic, New York.
- Viswanathan, H. S., B. A. Robinson, A. J. Valocchi, and I. R. Triay (1998), A reactive transport model of Neptunium migration from the potential repository at Yucca Mountain, *J. Contam. Hydrol.*, *209*, 251–280.

A. B. Aceves, Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM 87131, USA.

B. A. Robinson, Yucca Mountain and NTS Environmental Programs, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

G. Srinivasan, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. (gowri@lanl.gov)

D. M. Tartakovsky, Department of Mechanical and Aerospace Engineering, University of California, San Diego, 9500 Gilman Drive, Mail Code 0411, La Jolla, CA 92093, USA. (dmt@ucsd.edu)