

Elsevier Editorial System(tm) for Journal Of Computational Physics
Manuscript Draft

Manuscript Number:

Title: Do we really need a large number of particles to simulate bimolecular reactive transport with random walk methods? A kernel density estimation approach

Article Type: Regular Article

Keywords: Reactive transport; Particle tracking; Kernel density estimators; Geochemical processes

Corresponding Author: Mrs. Maryam Rahbaralam, M. S.

Corresponding Author's Institution: Universitat Politecnica de Catalunya

First Author: Maryam Rahbaralam, M. S.

Order of Authors: Maryam Rahbaralam, M. S.; Daniel Fernández-Garcia, PhD; Xavier Sanchez-Vila, PhD

Abstract: Groundwater reactive transport models are an essential tool for the analysis of coupled geochemical processes in Earth Systems. On a molecular level, chemical reactions are the result of collision and subsequent recombination of the molecules present in the system. One way to deal with reactive transport is using a continuum mechanics approach; for example, in a well-mixed system, reactions are a function of reactant concentrations. An alternative is to use a Lagrangian approach, by tracking all molecules and specifying reactions in terms of collision theory. A computationally efficient family of methods to solve the reactive transport problem is based on Particle Tracking (PT). While the number of particles in most realistic applications is in the order of 10^6 - 10^9 , the number of molecules even in diluted systems might be in the order of fractions of the Avogadro number. Thus, each particle actually represents a group of potentially reactive molecules. The use of a low number of particles may result not only in loss of accuracy, but also may lead to an improper reproduction of the mixing process, limited by diffusion. Recent works have used this effect as a proxy to model incomplete mixing in porous media. In this work, we propose using a Kernel Density Estimation (KDE) of the concentrations that allows getting the expected results for a well-mixed solution with a limited number of particles. The idea consists of treating each particle as a sample drawn from the pool of molecules that it represents; this way, the actual location of a tracked particle is seen as a sample drawn from the density function of the location of molecules represented by that given particle, rigorously represented by a kernel density function. The probability of reaction can be obtained by combining the kernels associated to two potentially reactive particles. We demonstrate that the observed deviation in the reaction vs time curves in numerical experiments reported in the literature could be attributed to a deficient estimation of the concentrations based on particles reconstruction, and not to the occurrence of true incomplete mixing. We further explore the evolution of the kernel size with time, linking it to the diffusion process. Our results show that KDEs are powerful tools to improve reactive transport simulations, and indicates that incomplete mixing in diluted systems should be modeled based on alternative mechanistic models and not on a limited number of particles.

Suggested Reviewers: David Benson PhD

Associate Professor, Department of Geology and Geological Engineering, Colorado School of Mines
dbenson@mines.edu

Alexandre Tartakovsky PhD

Scientist, Pacific Northwest National Laboratory
alexandre.tartakovsky@pnnl.gov

Wolfgang Nowak PhD
Head , Hydromechanics and Modelling of Hydrosystems, Institute for Modelling Hydraulic and Environmental Systems, Universität Stuttgart
Wolfgang.Nowak@iws.uni-stuttgart.de

Philippe Renard PhD
Head , The Centre for Hydrogeology and Geothermics of University of Neuchâtel
philippe.renard@unine.ch

Philippe Gouze PhD
Géosciences Montpellier, Université de Montpellier
philippe.gouze@msem.univ-montp2.fr

To: Journal of Computational Physics

Barcelona, June 10, 2015

Dear Editor:

Manuscript title: "Do we really need a large number of particles to simulate bimolecular reactive transport with random walk methods? A kernel density estimation approach"

by M. Rahbaralam, D. Fernàndez-Garcia and X. Sanchez-Vila

Hydrogeology Group, Dept. Geotechnical Engineering and Geosciences, Universitat Politècnica de Catalunya, Jordi Girona 1-3, 08034 Barcelona, Spain

Corresponding author: Rahbaralam, M.

Jordi Girona 1-3, UPC Campus Nord, 08034 Barcelona, Spain

Phone: +34 93 401 1698

Fax: +34 93 401 7251

e-mail: Maryam.rahbaralam@upc.edu

A major challenge in reactive transport modeling is to properly simulate the interaction of the reactants in dynamic geochemical systems. Random walk particle tracking methods provide a computationally efficient and adaptable approach to simulate solute transport. However, to properly simulate non-linear reactions, an unfeasible number of particles is required. On the other hand, a finite number of particles leads to a poor-mixed system limited by diffusion. Recent works have used this effect to actually model incomplete mixing in naturally occurring porous media. In this work, we demonstrate that this effect in most cases should be attributed to a deficient estimation of the concentrations and not to the occurrence of true incomplete mixing. To illustrate this, we show that a Kernel Density Estimation (KDE) of the concentrations can approach the well-mixed solution with a small number of particles. KDEs provide weighting functions of each particle mass that expands its region of influence, hence providing a wider region for chemical reactions with time. Importantly, simulation results also show that KDEs are powerful tools to improve state-of-the-art simulations of chemical reactions.

Respectfully,

Daniel Fernàndez-Garcia
Civil Engineering School, Barcelona
Technical University of Catalonia
Spain

Do we really need a large number of particles to simulate bimolecular reactive transport with random walk methods? A kernel density estimation approach

M. Rahbaralam, D. Fernàndez-Garcia and X. Sanchez-Vila

Hydrogeology Group, Dept. Geotechnical Engineering and Geosciences, Universitat Politecnica de Catalunya, Jord Girona 1-3, 08034 Barcelona, Spain

Abstract

Groundwater reactive transport models are an essential tool for the analysis of coupled geochemical processes in Earth Systems. On a molecular level, chemical reactions are the result of collision and subsequent recombination of the molecules present in the system. One way to deal with reactive transport is using a continuum mechanics approach; for example, in a well-mixed system, reactions are a function of reactant concentrations. An alternative is to use a Lagrangian approach, by tracking all molecules and specifying reactions in terms of collision theory. A computationally efficient family of methods to solve the reactive transport problem is based on Particle Tracking (PT). While the number of particles in most realistic applications is in the order of $10^6 - 10^9$, the number of molecules even in diluted systems might be in the order of fractions of the Avogadro number. Thus, each particle actually represents a group of potentially reactive molecules. The use of a low number of particles may result not only in loss of accuracy, but also may lead to an improper reproduction of the mixing process, limited by diffusion.

Recent works have used this effect as a proxy to model incomplete mixing in porous media. In this work, we propose using a Kernel Density Estimation (KDE) of the concentrations that allows getting the expected results for a well-mixed solution with a limited number of particles. The idea consists of treating each particle as a sample drawn from the pool of molecules that it represents; this way, the actual location of a tracked particle is seen as a sample drawn from the density function of the location of molecules represented by that given particle, rigorously represented by a kernel density function. The probability of reaction can be obtained by combining the kernels associated to two potentially reactive particles. We demonstrate that the observed deviation in the reaction vs time curves in numerical experiments reported in the literature could be attributed to a deficient estimation of the concentrations based on particles reconstruction, and not to the occurrence of true incomplete mixing. We further explore the evolution of the kernel size with time, linking it to the diffusion process. Our results show that KDEs are powerful tools to improve reactive transport simulations, and indicates that incomplete mixing in diluted systems should be modeled based on alternative mechanistic models and not on a limited number of particles.

Keywords: Reactive transport, Particle tracking, Kernel density estimators, Geochemical processes

1. Introduction

The complexity associated with the simulation of contaminants in polluted groundwater systems, including geochemical and microbiological effects, requires developing reliable, as well as predictive, solute transport mod-

els. A major challenge in this area is the incorporation of the interaction of the reactants facilitated by pure diffusion into a dynamic geochemical system. This problem has traditionally been addressed in porous media using a diffusion-reaction equation. The main interest has typically relied on finding or fitting the proper parameters that describe the experimental or field observations. This generic problem, posed in terms of the spatial and temporal distribution of reactants and products, has practical applications especially in the groundwater remediation context, where the choice of remediation strategy is based on accurate modeling of the anticipated degradation rate of the contaminants.

In most natural systems, the rate of reactions between the chemical species is limited by mixing because the molecules can only react when they come into physical contact to allow the corresponding bonds to occur. In well-mixed systems, like a stirred laboratory beaker, the spatial fluctuations in concentrations of the chemical species are negligible, and the rate of reaction is spatially uniform. However, if the mixing process is slow in comparison with the reaction rate, a mixing-limited regime will appear. In this regime, and when the initial spatial distribution of reactants is heterogeneous, the magnitude of concentrations as a function of time fluctuate significantly and segregated areas (islands) of one reacting species start to form in the domain [3, 25]. This segregation of the species is dictated by the reduced area of contact between reactants and was described from theoretical and numerical viewpoints several decades ago [15, 13]. Thus, it comes as no surprise that the prediction of reactions based on the well-mixed assumption associated with the classical transport and reaction equations within systems ranging

from homogeneous systems [10, 20] to heterogeneous systems [27] are bound to failure.

The study of solute transport driven by diffusion is one of the classical problems where the dichotomy between Eulerian and Lagrangian approaches has been traditionally well explained. An Eulerian approach implies setting the problem in terms of concentration of reactant species, and then postulating some macroscopic laws (either empirical or deduced from thermodynamic considerations) to describe the reaction rate based on such concentrations. The problem is then formulated from macroscopic mass balance equations such as, for example, the advection-dispersion-reaction equation (ADRE). Concentration is a macroscopic quantity (mass per unit volume) and thus it has no sense when the volume reduces to zero. An alternative is to study the process at the molecular level, theoretically tracking the movement and fate of each individual molecule, within a Lagrangian framework [7, 8, 9]. This alternative is obviously unfeasible, as even minute concentrations result in a extremely large number of molecules (1 mol equals 6.02×10^{23} molecules); furthermore, equations of the continuum are no more applicable, and reactions should be assessed from collision theory.

Particle Tracking Methods (PTM) try to combine both approaches to provide an efficient solution to reactive transport simulations in the subsurface. The idea is to split the total mass into a number of particles (typical orders are $10^6 - 10^9$), many orders of magnitude lower than the actual number of molecules, and devise a way to reconstruct the spatial and temporal distribution of concentrations of the different species from the location of particles at any given time. This methodology has obvious drawbacks such as:

(1) the estimation of macroscopic quantities (concentrations) from discrete particle information being a non-unique process, and (2) the need to set up reactive rules applied to a finite number of particles so that, when upscaled, they properly represent reaction rules observed at the macroscale. An additional drawback would be the modeler's choice of the number of particles, and more, the meaning of the concept particle itself.

Regarding the first problem, recently an optimal method for the reconstruction of concentrations and their functionals based on Kernel Density Estimators (KDE) was developed [6]. This method has been successfully applied to estimate concentrations of conservative and reactive species [11], heavily-tailed breakthrough curves [18], and reaction rates in precipitation/dissolution problems [6]. Regarding the second point, the translation of macroscopic rules to PTM, a technique was proposed by Benson and Meerschaert [1] to study reactive transport for a bimolecular reaction, introducing an algorithm that accounts for the probability that two individual particles react as a function of their relative distance; this method was an extension of the original algorithm proposed by Gillespie [7] to simulate batch reactions in a stochastic framework at the molecular level. The method avoids the need to continuously reconstruct concentration maps during the course of the simulation for each individual time step, considered the main disadvantage of PT methods [22, 4].

Here we combine these two ideas to develop a KDE model that can be directly used to study bimolecular reaction in diffusive problems. The main idea is to use Kernels to provide a new weighting functions for particles to interact following Benson and Meerschaert [1] approach. In this case, when

particles are transformed into another species or mineral, the shape of the Kernel automatically adjust, expanding its region of influence and the region of interaction between particles [6]. This method increases the number of particles that could react at a given time. The introduction of this expanding Kernel provides some insights into a problem that has been recently raised in the literature. Benson and coworkers have studied the evolution of concentration with time in a bimolecular reactive problem susceptible to incomplete mixing [1, 17, 3, 2]. These authors represented incomplete mixing by setting up a low number of particles into the reactive system; the lowest the number of particles the earliest reaction became limited. While this is an intelligent approach, we would like to stress that it is only valid when the number of molecules is really limited, but not in subsurface geochemical applications where the number of molecules available for reaction is always very high.

We thus associate incomplete mixing in some applications to the combination of low number of particles combined with an estimation of concentrations based on the particles having a zero support size .e.g., [19, 22]. We contend that the limited number of particles should not be considered as a tool for generating incomplete mixing systems. To make this point we show how our approach based on the inclusion of optimal KDEs allows reproducing complete mixing up to a very small number of particles. We content then that incomplete mixing in diluted systems should be modeled based on alternative mechanistic models (see [23]) and not on a limited number of particles.

The paper is structured as follows. First, Section 2 sets out the problem and the numerical approach proposed for the analysis of species concen-

trations. Section 3 introduces the proposed chemical reaction model using KDEs. Simulation results are presented in Section 4, validating the reactive model.

2. Statement of the problem and numerical approach

2.1. Problem definition

This work deals with a simple chemical model, i.e., a single forward bimolecular irreversible reaction $A + B \rightarrow \emptyset$. Here, two species present in the dissolved phase, A and B , react kinetically and irreversibly with unitary stoichiometric coefficients. The product of the reaction is assumed to play no role in the system (i.e., it does not produce any change in porosity or permeability). The transport of the reactants is driven by a diffusion-reaction equation with the following expression:

$$\frac{\partial C_i}{\partial t} = D\nabla^2 C_i - k_f C_A C_B, \quad (1)$$

where i stands for A, B , the concentration of species i is given by $C_i = C_i(x, t)$ (in $[ML^{-3}]$), $D[L^2T^{-1}]$ is the diffusion coefficient, and $k_f[L^3M^{-1}T^{-1}]$ is the reaction rate constant.

2.2. Analytical model

From (1), we can write that the equation satisfied by $C_A - C_B$ is conservative (i.e., not affected by geochemical reactions). If we further impose a well-mixed system, we can write

$$\frac{d}{dt}(C_A - C_B) = 0, \quad (2)$$

so that $C_A - C_B = [A_0] - [B_0]$, $[A_0]$ and $[B_0]$ being the corresponding initial concentrations. We can then use this result to rewrite the governing equation for the well-mixed concentration as

$$\frac{\partial C_A}{\partial t} = -k_f C_A (C_A + [B_0] - [A_0]). \quad (3)$$

By integrating this ODE, it is possible to write an explicit solution for the concentrations of the two species

$$C_A(t) = \frac{[A_0] - [B_0]}{1 - \frac{[B_0]}{[A_0]} \exp(-k_f([A_0] - [B_0])(t - t_0))}, \quad (4)$$

and in the particular case that $[A_0] = [B_0]$, the solution is the same for both species, given by

$$C_i(t) = \frac{[A_0]}{1 + [A_0] k_f t}; \quad i = A, B, \quad (5)$$

indicating that $C_i(t)$ scales as $k_f^{-1} t^{-1}$ for large times.

2.3. The Lagrangian approach

2.3.1. Particle tracking (PT) methods

In PT methods each particle carries a given fraction of the total mass moving in dissolved state with groundwater and reacting according to some fundamental mechanisms. A large variety of mechanisms can efficiently simulate different transport phenomena. These mechanisms range from pure advection and random walk motions to complex reaction and mass transfer processes [26, 21, 14, 2]. In a bimolecular reactive problem, it is worth noticing that the particle state is binary (e.g., it has either reacted or not), while in reality only a fraction of the pool of molecules may have reacted.

We find here again a duality in the concept of representation: even though only a portion of molecules associated with one particle react, PT methods adopt the result obtained from a Bernoulli trial sampled from the probability distribution of reacted molecules. In most approaches the total mass of each reactant is evenly distributed among the particles, so that the mass carried by each particle is simply given by the total mass in the system divided by the number of particles; e.g., for reactant A

$$m_{p,A} = \frac{1}{N_0} \int_{\Omega} [A_0(x)] d\Omega, \quad (6)$$

where N_0 is the number of particles selected by the modeler, $[A_0(x)]$ is the (spatially variable) concentration within a given volume Ω . In the case of uniform initial concentration the equation (6) results in $m_{p,A} = [A_0]\Omega/N_0$. In this latter case, the initial position of each individual particle can be drawn from a multidimensional uniform distribution. For this reason, if the number of particles is small, the density of particles initially present in different locations can be slightly different, creating artificial initial fluctuations. Such difference would be reduced by increasing the number of particles.

2.3.2. Particle displacements

Without loss of generalization, we consider the diffusion-reaction equation in 1-D. The initial locations of the A and B particles are respectively denoted by X_j^A and X_j^B , j indicating the particle number. We consider a domain interval defined as Ω . The initial particle location was drawn from a statistical uniform distribution. The diffusion of a species is modeled by a Brownian random walk motion with zero mean and variance $2D\Delta t$, being Δt the time interval. For each simulation time step, PT describes the change

of the particle species taking place between t and $t+\Delta t$. The random walk is an implementation of a Langevin equation without the drift term. Thus, the location of the particle j at a given time is a Markovian process characterized by a Brownian motion term, given by

$$X_j^i(t + \Delta t) = X_j^i(t) + \xi_t \sqrt{2D\Delta t}, \quad i = A, B \quad (7)$$

where ξ_t is a random number drawn at each time step from an independent (uncorrelated in space and time), normally distributed random variable with zero mean and unit variance. It has been shown that equation (7) satisfies the non-reactive diffusion equation (i.e., (1) with $k_f = 0$) in the limit for an infinite number of particles with infinitesimal mass [26, 21]. The inclusion of the reaction term in (1) simply modifies the species state associated with that particle.

2.3.3. Diffusion-based (DB) reaction

A Reactive Particle Tracking (RPT) method was recently proposed by Benson and Meerschaert [1] (we refer to this method as the DB model). In this method, diffusion is treated by (7), and the reaction of particles is simulated through probabilistic rules,

$$P_{DB} = k_f m_p \Delta t v(r, \Delta t), \quad (8)$$

where P_{DB} is the probability that two particles, A and B , separated by a distance r react within a time interval Δt . Here, $v(r, \Delta t)$ is the co-location probability density function (pdf), defined as the probability that two particles initially separated at a distance r , one from species A and another

from species B occupy the same position after a time interval Δt . This co-location pdf is the convolution of two Gaussian densities f_A and f_B , see Fig. 1. Assuming D constant, both pdf's have the same standard deviation $\sigma = \sqrt{2D\Delta t}$. Then, σ represents the area of influence over which two particles can come into contact and react. As demonstrated later on, the proper computation of the area of influence plays a key role in approaching realistic reactive transport solutions.

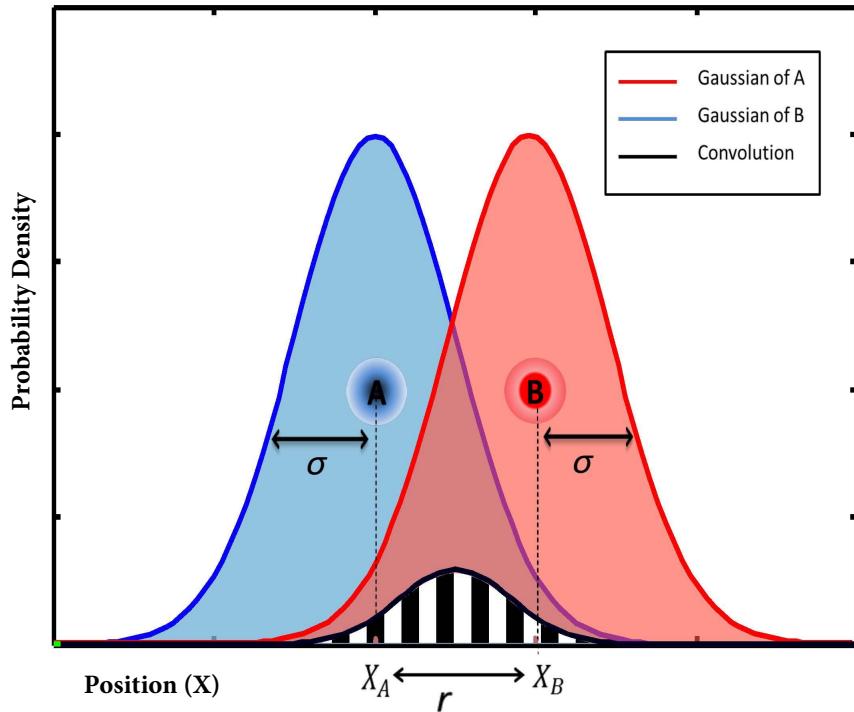


Figure 1: Convolution of two Gaussian densities with the same standard deviation σ , located at a separation distance r . The shaded area can be seen as the probability density function of the two particles A and B actually occupying the same space, so that reaction may take place according to some thermodynamic rules.

The probability of reaction P_{DB} , calculated by (8), is the probability that a pair of particles (A,B) react within a time interval Δt . Numerically, this is done by comparing a random number generated from a uniform distribution $\xi \sim U(0, 1)$ with the calculated probability. If $P_{DB} \geq \xi$, the reaction occurs and both A and B particles are removed from the system; otherwise, reaction is disregarded at that particular time step and the particle is kept. This procedure is repeated for every pair of particles (A,B) in the system. Then, the simulation continues to the next time step. At each time step, the number of remaining particles divided by the initial volume represents the normalized concentration, $[A]/[A_0]$.

2.4. Concentration reconstruction using KDE

Kernel Density Estimators (KDE) are powerful methods to reconstruct smooth concentrations and their functionals from discrete particle information in space and/or time [6, 18]. A physical way to understand kernels is to think that the one associated with a given particle represents a bunch of molecules, reflecting the pdf of the location of molecules as a function of time. Instead, the particle is just one realization of the ensemble of molecules. Since the only available information is the location of the particle, this is seen as the point estimator of the expected value of the molecule location. For this reason, even when a particle A and a particle B share the exact same location at a given time, there is still a certain probability of collocation. Thus, kernels associated with the two particles have a superposition area, and the probability of reaction can be assumed to be proportional to this area. The further away the particles are, the smaller the degree of collocation and thus the probability of reaction (similar to the approach by [1]).

Using this approach, the concentration $c(x, t)$ of a given species at a given time t can be reconstructed from the location of the particles present in the system by

$$c(x, t) = \sum_{p=1}^{N(t)} \frac{m_p}{h(t)} K\left(\frac{x - X_p(t)}{h(t)}\right), \quad (9)$$

where $N(t)$ is the number of particles at time t , X_p is the particle location, h is the bandwidth (area of influence), and K is a moving weighting function centered at the particle location chosen from functions of unitary area, i.e.,

$$\int K(\tau) d\tau = 1. \quad (10)$$

Optimal reconstruction involves a proper choice of the h parameter to account for both the diffusion process and the continuously diminishing number of particles in the system with time. In contrast to the traditional box model [24] which introduces discontinuities at the box edges, kernel estimator (9) produces smooth functions of space. These kernel functions define the region of influence of each particle. The kernel shape depends on the choice of the bandwidth or support size h and the type of kernels. A variety of kernel functions can be used, among them, Gaussian kernels are usually preferred for mathematical advantages [18].

2.4.1. The optimal area of influence

As stated before, Kernels can be seen as a way to introduce into the solution the effect of using a subset of particles to represent a very large number of molecules. The larger the degree of undersampling, the more information should be conveyed into the kernel estimator. The degree of

smoothing or the area of influence around a particle is controlled by the parameter h . If the number of particles is very large, there is no real need for kernels, and h should be very small. On the contrary, h should increase when the number of particles is not sufficient to account for a proper estimation of the spatial distribution of the concentrations. In general, a small support size combined with a finite number of particles leads to noisy estimates and the fluctuation of concentrations. Contrarily, an increase in the support size tends to oversmooth the estimated concentration distribution. Thus, an optimum choice of the support size, h^{opt} , exists. Choosing it should be based on an objective criterion about the reconstruction effort, minimizing the variance error while controlling the biasedness in the estimation. Park and Marron [16] found the optimum bandwidth as

$$h^{opt} = R \ N^{-1/5}, \quad (11)$$

where R is given as

$$R = \left(\frac{\|\mathbf{K}\|_2^2}{\|\mathbf{p}''(\mathbf{x})\|_2^2 (\mu_2(K))^2} \right)^{1/5}, \quad (12)$$

$\mu_2(K)$ is the second moment of K , $p''(x)$ is the second derivative of the concentration normalized by the total mass in the system, and the symbol $\|\cdot\|_2^2$ is the L_2 -norm operator, i.e. $\|\mathbf{p}(\mathbf{x})\|^2 = \int p(x)^2 dt$. It is clear that the optimal bandwidth h^{opt} in (11) grows inversely proportional to the number of particles; the smaller the number of particles, the larger the bandwidth size. This feature is particularly beneficial for developing a chemical reaction model, as mentioned later in the paper. The second derivative of the density function $p''(x)$ needs to be also estimated. Several methods can be used for

this purpose: (1) rule-of-thumb methods, (2) cross-validation methods, and (3) plug-in methods (see recent reviews of these methods in Ref. [12, 16]). Following a recent work by Fernàndez-Garcia and Sanchez-Vila [6], we hereby employ the plug-in method to estimate the optimal bandwidth h^{opt} .

3. The KDE-based chemical reaction model

The DB model is characterized by a constant in time area of influence characterized by a constant diffusion, see Fig. (1). Therefore, when the number of particles is very low (occurring at late times), the probability of collocation is also very low and the segregation of particles of a given type (either A or B) start occurring [1]. Alternatively, the KDE based method provides an area of influence that increases when particles are sparsely distributed in space, see equation (11). This results in an enhanced probability of reaction and in an overall reaction rate approaching to the well-mixed solution, even with a small number of particles.

If the initial concentrations of the two reactants are equal, it is also logical to assign the same standard deviation to all of them, actually given by $\sigma = h^{opt}$. This dynamic bandwidth in the KDE-based model implies a different probability of reaction at each time step, preventing the formation of segregated areas of particles. The actual expression for the probability of reaction should follow rigorously the principles of the law of mass action, and also be proportional to the thermodynamic rate k_f as well as to the co-location probability density associated with the distance between particle pairs. Based on these principles, the probability of reaction for the KDE-based model is obtained as

$$P = k_f m_p \Delta t P_{KDE}(r), \quad (13)$$

where P_{KDE} is the convolution of the two Gaussian densities of particles A and B based on h^{opt} . This convolution is a Gaussian function with zero mean and variance equal to $2(h^{opt})^2$,

$$P_{KDE}(r) = \frac{1}{2h^{opt}\sqrt{\pi}} \exp\left(\frac{-r^2}{(2h^{opt})^2}\right). \quad (14)$$

From (14), the probability of reaction decreases by increasing h^{opt} , preserving the unitary area under the co-location probability density. In other words, expanding the influential area of each particle by h^{opt} comes at the expense of decreasing the probability of reaction. On the other hand, as demonstrated later, this effect is compensated by the increase in the number of potentially reactive pairs, as each individual particle samples a larger area for the presence of particles. This is particularly beneficial to avoid the segregation of particles resulting in artificial poor mixing.

4. Simulation results

4.1. DB model

We first intend to simulate reactions with the DB model as a way to compare later on the performance of the KDE model with the DB model. A 1-D domain of size $\Omega = 64$ (normalized units) is considered. Simulations are performed for 1000 time steps. The different scenarios analyzed combine different number of initial particles N (always equal for species A and B), with two diffusion coefficients $D = 10^{-2}$ and $D = 10^{-5}$. The remaining

parameters are set as: $[A_0] = [B_0] = 1$, $k_f = 50$, and $\Delta t = 0.4$. All values throughout the paper are reported in normalized units.

Figure 2 shows the normalized concentration $[A_0]/[A]$ as a function of the normalized time t' , defined as $t' = [A_0]k_f t$. For comparison purposes, the analytical solution of the well-mixed system in (5) is also provided. At early times, due to the presence of a large number of particles in the system, the numerical and analytical solutions are identical. Therefore, the number of particles plays a negligible role and the actual concentration is closely approximated by the well-mixed solution, controlled simply by the thermodynamic rate coefficient k_f . However, at late times, a deviation from the analytical solution occurs, as chemical reactions among particles become limited by diffusion. In this diffusion-limited regime, islands of segregated particles (only one of the two species present) start to form, leading to a poor-mixed solution. A lower number of initial particles or a smaller diffusion constant enhances particle segregation and thereby the deviation from the well-mixed solution. These results are in a good agreement with those presented by Benson and Meerschaert [1].

4.2. KDE model

To illustrate the potential of our proposed KDE model, a simulation of the irreversible chemical reaction is performed with a diffusion constant of $D = 10^{-2}$ and an initial number of particles of $N = 4000$. The rest of the parameters are chosen equal to those used in the previous section 4.1. Figure 3 shows the comparison between the KDE and two DB models, as well as the analytical solution. Results demonstrate that the KDE model approaches the well-mixed analytical solution in a wider range of times, resulting in a

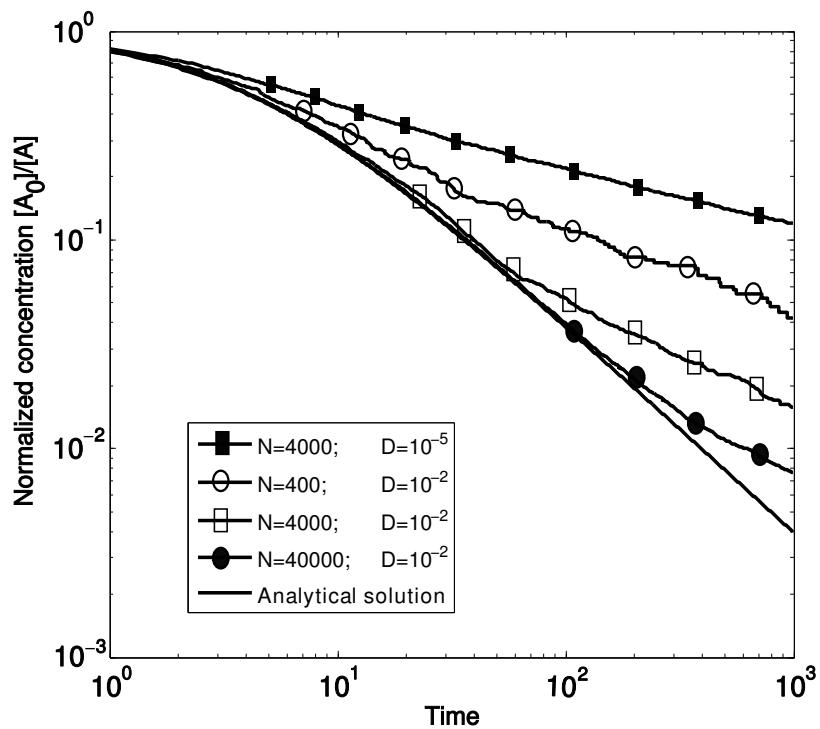


Figure 2: Normalized concentration as a function of normalized time using the DB model. Analytical solution of the well-mixed system is also plotted for comparison purposes. The time where deviation is observed depends on the number of particles that were initially in the system. Notice that when particles are used to represent a group of molecules, the actual number used in the simulations is a modeler's choice.

considerably lower dependency of the solution on the number of particles with respect to the DB model. In fact, the simulation results for $N = 4000$ using the KDE model follows closely the results of the DB model with $N = 40,000$. This is an obvious enhancement of the classical chemical reaction model, as reaction rates are predicted using a noticeable lower number of particles.

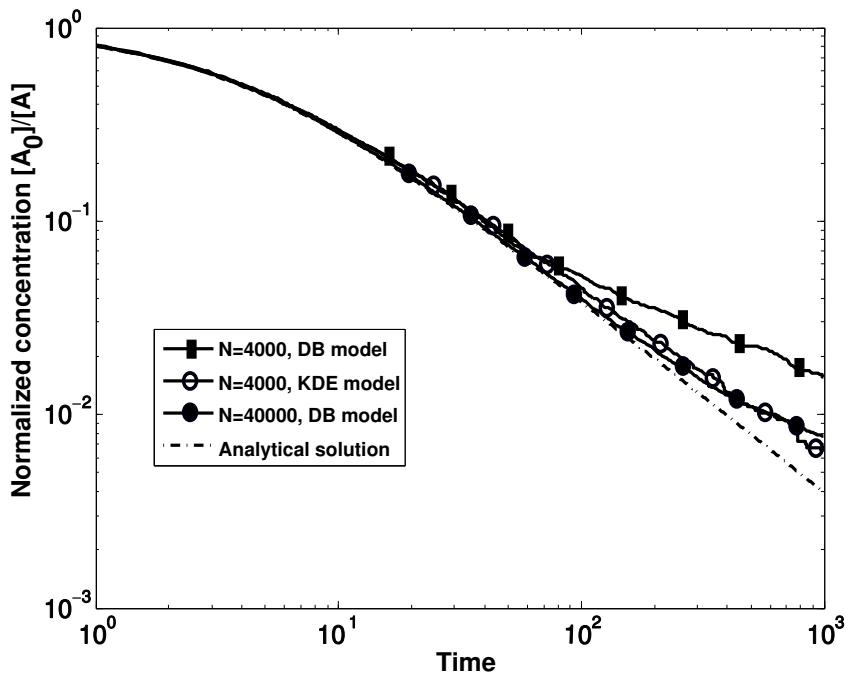


Figure 3: Normalized concentration as a function of normalized time using the KDE model, two DB models with different number of particles, and the corresponding analytical solution.

To further illustrate the behavior of the KDE model, we perform additional simulations with different number of particles being present at time zero. The results in Fig. 4 indicate that the decay of concentration of reac-

tants is significantly less sensitive to the number of particles as compared to the behavior of the DB model, c.f. Fig. 2. As the solution is almost independent of the number of particles, N can be adopted by the modeler a priori solely based on the computational effort needed for the application, rather than on its potential influence on the solution. Indirectly it also implies that it is possible to actually use a lower number of particles to achieve a given accuracy in the solution.

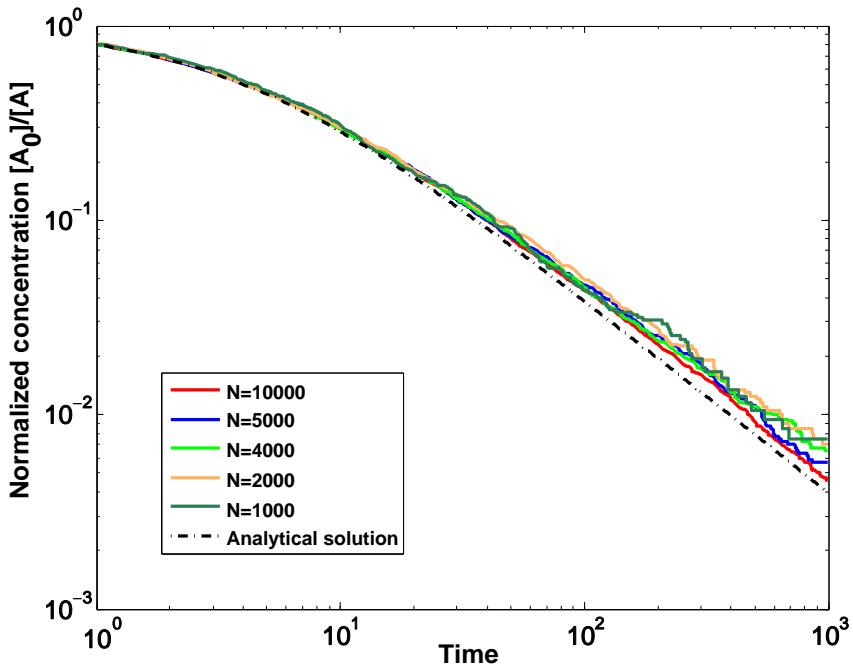


Figure 4: Normalized concentration as a function of the normalized time using the KDE model for a varying number of particles representing the initial mass. The solution is quite insensitive to the choice of number of particles being initially present in the system.

As already noted, the performance is strongly linked to the choice of the optimal bandwidth h_{opt} . This value should be continuously updated. So,

there is a risk of loosing efficiency by updating h_{opt} at each time step. Here, we propose to compute h_{opt} efficiently throughout all the simulations period by extrapolating h_{opt} from the data gathered at the initial steps (early times). Figure 5 shows the true evolution of the optimal bandwidth h_{opt} as a function of the number of particles for an initial particle number of $N = 4000$ (h_{opt} is updated at each time step). It is found that the optimal bandwidth increases as the number of particles decreases, closely following the slope of $-1/5$ (recall equation (11)). Therefore, h_{opt} as a number of the remaining particles in the system can be estimated by a linear extrapolation between the optimal values obtained at two different time steps (e.g., denoted by 1 and 2 in the figure). This extrapolation allows one to avoid the costly computation involved in solving (12) at each time step. An anomalous estimation of h_{opt} at late times is also observed, corresponding to normalized concentrations of 10^{-2} , in this problem being equivalent to the presence of only 40 particles of each species in the whole domain.

4.3. Performance analysis

To systematically compare the performances of the DB and the KDE models, we define a performance factor

$$P_f = \frac{t_{max}}{t_c} + \frac{E_{max}}{E}, \quad (15)$$

where t_c is the computational time, t_{max} is the maximum of t_c among all simulations, and E is the error of the results with respect to the analytical solution, defined as

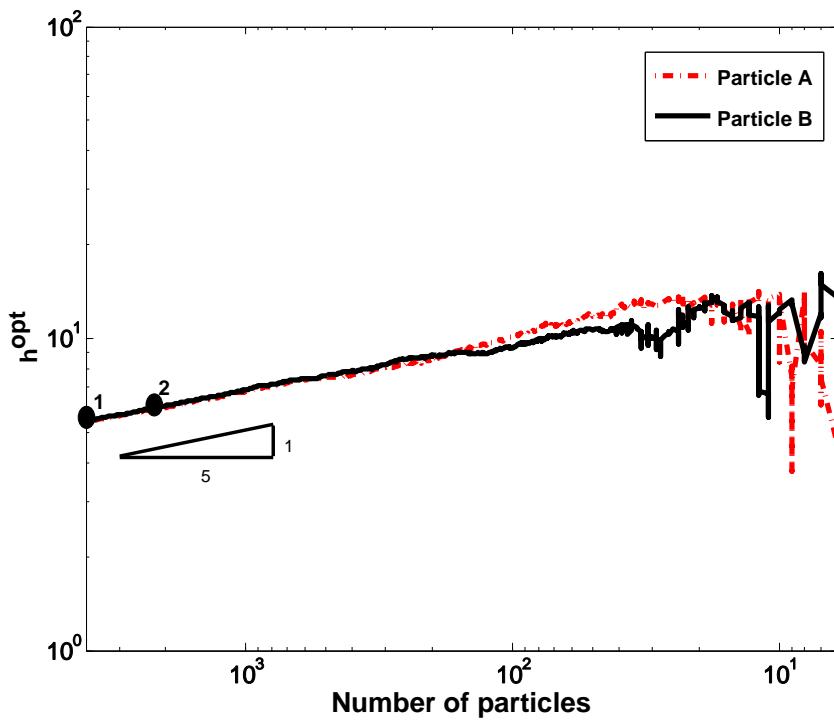


Figure 5: Optimal bandwidth as a function of the number of particles remaining in the system (not yet reacted). The points 1 and 2 represent initial steps of the simulation for the linear extrapolation of the optimal bandwidth. A slope of $1/5$ for the optimal bandwidth can be considered a very good approximation up to times corresponding to very low concentrations.

$$E = \frac{\sum_{n=1}^s \frac{[A]}{[A_0]}_{simulation} - \sum_{n=1}^s \frac{[A]}{[A_0]}_{analytical}}{\sum_{n=1}^s \frac{[A]}{[A_0]}_{analytical}}, \quad (16)$$

where E_{max} is the maximum error among all simulations. In (16), n stands for time step, and s is the maximum number of time steps considered in the numerical simulation. The performance factor P_f is obtained as a function of a dimensionless characteristic time τ , representing the ratio of the characteristic reaction time $T_R = (A_0 k_f)^{-1}$ to the diffusion time $T_D = L^2/D$,

$$\tau = \frac{T_R}{T_D}. \quad (17)$$

Figure 6 shows the performance factor associated with the DB and the KDE models for two different values of τ in a physically and computationally reasonable range.

It is found that the performance factor of the KDE model is almost insensitive to the choice of τ . This implies that the kernels can adapt themselves to any parameter change, preserving the behavior of the system. Contrarily, the behavior of the DB model exhibits a strong dependence on the system parameters. It is also found that the performance factor of the KDE model is noticeably larger than that of the DB model for low number of particles and decreases by increasing this number. The main advantage of the KDE model is that the computational time increases with the number of particles while the error E does not change considerably, i.e., the first term in equation (15) decreases while the second term remains almost constant.

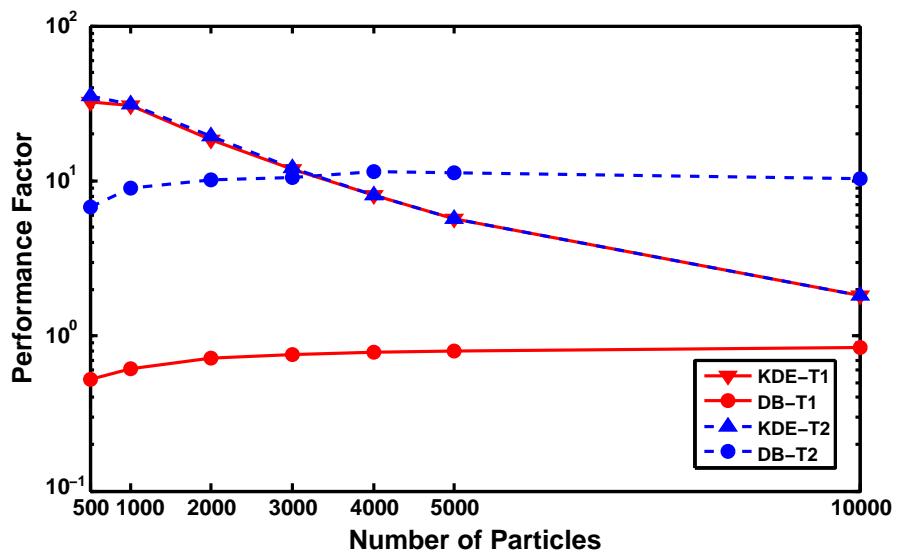


Figure 6: Performance factor as a function of the number of particles for the DB and KDE models considering two difference values of the dimensionless characteristic time $\tau_1 = 9.7 \times 10^{-6}$ and $\tau_2 = 9.7 \times 10^{-9}$.

5. Conclusions

We have proposed a chemical reaction model based on Kernel Density Estimators (KDEs) for the simulation of well-mixed systems. KDEs provide an optimal influential area that permits to accurately reconstruct concentrations with almost no artificial fluctuations. This influential area grows with decreasing number of particles, a feature which is beneficial to avoid incomplete mixing due to the segregation of particles. This problem is common in classical diffusion-based models with finite number of particles.

Simulation results of a diffusion-based (DB) model have shown that at late times of reaction, when a low number of particles remain in the system, there is a deviation from the well-mixed solution. This deviation has been solved by developing a new particle tracking (PT) model based on optimal Kernel Density Estimators. Since particles are a subset of the actual molecules present in the system, kernels can be seen as a way to account for such extreme undersampling. The actual location obtained in the PT method is the point estimator of the location, and the kernel represents the variance in the location of molecule. Collocation of molecules can be inferred from that of particles. Numerical simulations based on the KDE method have shown a noticeable enhancement of the reaction towards the well-mixed solution when this model is applied instead of a standard PT model. Furthermore, the results of the KDE model have indicated that, in contrast to the DB model, the concentration decay is almost insensitive to the initial number of particles. We have also performed a performance analysis based on the computational time and error of the models. The results have shown that the performance of the KDE model does not depend on the system parameters, highlighting

the rigorousness of our proposed approach. All these behaviors prove that the KDE model is a powerful tool for simulating chemical reactions of well-mixed systems. We content that this model is also capable of simulating incomplete mixing in porous media according to physical mechanisms rather than based on numerical artifacts.

Acknowledgements

This work has been supported by the Spanish Ministry of Science and Innovation through project FEAR (CGL2012-38120) and the European Community Framework Programme 7 through project MARSOL (FP7-619120). XS acknowledges support of Program ICREA Acadmia.

References

- [1] Benson DA, Meerschaert MM. Simulation of chemical reaction via particle tracking: Diffusion-limited versus thermodynamic rate-limited regimes. *Water Resour Res* 2008;44:W12201. doi:10.1029/2008WR007111.
- [2] Benson DA, Meerschaert MM. A simple and efficient random walk solution of multi-rate mobile/immobile mass transport equations. *Adv Water Resour* 2009;32:532-9. doi:10.1016/j.advwatres.2009.01.002.
- [3] Bolster D, de Anna P, Benson DA, Tartakovsky AM. Incomplete mixing and reactions with fractional dispersion. *Adv Water Resour* 2012;37:86-93. doi:10.1016/j.advwatres.2011.11.005.

- [4] Bosio F, Bellin A, Dumbser M. Numerical simulations of solute transport in highly heterogeneous formations: A comparison of alternative numerical schemes. *Adv Water Resour* 2013;52:178-89. doi:10.1016/j.advwatres.2012.08.006.
- [5] Dentz M, Le Borgne T, Englert A, Bijeljic B. Mixing, spreading and reaction in heterogeneous media: a brief review. *J Contam Hydrol* 2011;120-121:1-17. doi:10.1016/j.jconhyd.2010.05.002.
- [6] Fernàndez-Garcia D, Sanchez-Vila X. Optimal reconstruction of concentrations, gradients and reaction rates from particle distributions. *J Contam Hydrol* 2011;120-121:99-114. doi:10.1016/j.jconhyd.2010.05.001.
- [7] Gillespie DT. The chemical Langevin equation. *J Chem Phys* 2000;113(1):297-306.
- [8] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 1977; 81:2340-2361.
- [9] Gillespie DT. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *J Comput Phys* 1976;22:403-434.
- [10] Gramling CM, Harvey CF, Meigs LC. Reactive transport in porous media: a comparison of model prediction with laboratory visualization. *Environ Sci Technol* 2002;36:2508-14.10.doi:1021/es0157144.
- [11] Henri CV, Fernàndez-Garcia D, Toward efficiency in heterogeneous multispecies reactive transport modeling: A particle-tracking solution

- for first-order network reactions, *Water Resour Res* 2014;50:7206-7230, doi:10.1002/2013WR014956.
- [12] Jones MC, Marron JS, Sheather SJ. A brief survey of bandwidth selection for density estimation. *J Am Stat Assoc* 1996;91:401-7. doi:10.1080/01621459.1996.10476701.
- [13] Kang K, Redner S. Fluctuation-dominated kinetics in diffusion-controlled reactions. *Phys Rev A* 1985;32:435-447. doi:10.1103/PhysRevA.32.435.
- [14] Moroni M, Kleinfelter N, Cushman JH. Analysis of dispersion in porous media via matched-index particle tracking velocimetry experiments. *Adv Water Resour* 2007;30:1-15. <http://dx.doi.org/10.1016/j.advwatres.2006.02.005>.
- [15] Ovchinnikov AA, Zeldovich YB. Role of density fluctuations in bimolecular reaction kinetics. *Chem Phys* 1978;28:215-8. doi:10.1016/0301-0104(78)85052-6.
- [16] Park BU, Marron JS. Comparison of data-driven bandwidth selectors. *J Am Stat Assoc* 1990;85:66-72. doi:10.1080/01621459.1990.10475307.
- [17] Paster A, Bolster D, Benson DA. Connecting the dots: Semi-analytical and random walk numerical solutions of the diffusion-reaction equation with stochastic initial conditions. *J Comput Phys* 2014;263:91-112. doi:10.1016/j.jcp.2014.01.020.
- [18] Pedretti D, Fernández-García D. An automatic locally-adaptive method to estimate heavily-tailed breakthrough curves

- from particle distributions. *Adv Water Resour* 2013;59:52-65. doi:10.1016/j.advwatres.2013.05.006.
- [19] Pollock D. Semianalytical computation of path lines for finite-difference models. *Ground Water* 1988;26(6):743-50.
- [20] Raje DS, Kapoor V. Experimental Study of Bimolecular Reaction Kinetics in Porous Media. *Environ Sci Technol* 2000;34:1234-9. doi:10.1016/10.1021/es9908669.
- [21] Salamon P, Fernàndez-Garcia D, Gomez-Hernandez JJ. Modeling mass transfer processes using random walk particle tracking. *Water Resour Res* 2006;42. doi:10.1029/2006WR004927.
- [22] Salamon P, Fernàndez-Garcia D, Gomez-Hernandez JJ. A review and numerical assessment of the random walk particle tracking method. *J Contam Hydrol* 2006;87:277-305. doi:10.1016/j.jconhyd.2006.05.005.
- [23] Sanchez-Vila X, Fernàndez-Garcia D, Guadagnini A. Interpretation of column experiments of transport of solutes undergoing an irreversible bimolecular reaction using a continuum approximation. *Water Resour Res* 2010;46:1-7. doi:10.1029/2010WR009539.
- [24] Silverman BW, Density Estimation for Statistics and Data Analysis, London:Chapman and HallLondon 1986.
- [25] Tartakovsky AM, Meakin P. A smoothed particle hydrodynamics model for miscible flow in three-dimensional fractures and the two-dimensional Rayleigh-Taylor instability. *J Comput Phys* 2005;207:610-24. doi:10.1016/j.jcp.2005.02.001.

- [26] Tompson AFB, Dougherty DE. Particle-grid methods for reacting flows in porous media with application to Fisher's equation. *Appl Math Model* 1992;16:374-83. doi:10.1016/0307-904X(92)90071-A.
- [27] Willmann M, Carrera J, Sanchez-Vila X, Silva O, Dentz M. Coupling of mass transfer and reactive transport for nonlinear reactions in heterogeneous media. *Water Resour Res* 2010;46:W07512. doi:10.1029/2009WR007739.