Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR023033

Lagrangian Modeling of Reactive Transport in Heterogeneous Porous Media With an Automatic Locally Adaptive Particle Support Volume

Guillem Sole-Mari[1,2] **and Daniel Fernàndez-Garcia**[1,2]

[1]Department of Civil and Environmental Engineering (DECA), Universitat Politècnica de Catalunya, Barcelona, Spain, [2]Hydrogeology Group, UPC-CSIC, Barcelona, Spain

Key Points:
- We present a locally adaptive kernel density estimator for computation of chemical reactions in random walk models of solute transport
- The support volume adapts to the nearby particle distribution, adjusting its size and stretching along the direction of minimum curvature
- The local approach outperforms the existing methods when estimating concentrations and reaction rates in a heterogeneous porous medium

Correspondence to:
G. Sole-Mari,
guillem.sole.mari@upc.edu

Citation:
Sole-Mari, G., & Fernàndez-Garcia, D. (2018). Lagrangian modeling of reactive transport in heterogeneous porous media with an automatic locally adaptive particle support volume. *Water Resources Research*, *54*, 8309–8331. https://doi.org/10.1029/2018WR023033

Received 28 MAR 2018
Accepted 15 SEP 2018
Accepted article online 1 OCT 2018
Published online 27 OCT 2018

Abstract The particle support volume is crucial for simulating reactive transport with Lagrangian methods as it dictates the interaction among particles. Assuming that it is constant in space, the particle support volume can be selected by means of kernel density estimation theory, an approach that has been shown to provide accurate estimates in simple setups. However, the particle support volume should intuitively vary with the particle position and evolve with time so as to mimic the local behavior of the solute plume. In this paper, we present a new approach to select a locally optimal particle support volume in reactive transport simulations. We consider that each particle has a different support volume that can locally adapt its shape and size with time based on the nearby particle distribution. By introducing a new optimality criterion, closed-form expressions of the particle support volume are presented under certain assumptions. In advection-dominated transport, we propose to orient the support volume along the local velocities. Numerical simulations of solute transport in a randomly heterogeneous porous medium demonstrate that the new approach can substantially increase accuracy with a more rapid convergence to the true solution with the number of particles. The error reduction seen in local approaches is particularly important in regions with extreme (high and low) density of particles. The method is shown to be computationally efficient, displaying better results than traditional histogram or global kernel methods for the same computational effort.

1. Introduction

Numerical models that deal with multiple species and chemical reactions are typically based on Eulerian approaches (e.g., Lichtner et al., 2015; Saaltink et al., 2004; Xu et al., 2014; Yeh et al., 2004). However, the incorporation of physical and biochemical heterogeneities into grid-based Eulerian codes that solve the multicomponent advection-dispersion equation coupled with reactions suffers from numerical problems stemming from the need to fulfill a small grid-Peclet number $\Delta x/a$ to properly simulate reactions, being a the dispersivity and Δx the size of the numerical spatial discretization. Knowing that in porous media many chemical reactions are driven by mixing (De Simoni et al., 2005, 2007; Dentz et al., 2011; Gramling et al., 2002; Martinez-Landa et al., 2012) and controlled by very small values of transverse dispersivities (Cirpka et al., 1999, 2015), these numbers are seldom achieved in practical applications, and numerical simulations tend to overpredict the total amount of reaction produced (Benson et al., 2017; Sanchez-Vila & Fernàndez-Garcia, 2016). This renders reactive transport modeling a major challenge in hydrogeology nowadays.

Lagrangian approaches that simulate transport by moving particles with simple mechanisms constitute an attractive technique to overcome some of these problems. These methods are well established for nonreactive chemical systems. Different authors have compared Eulerian and Lagrangian approaches for nonreactive transport. Salamon et al. (2006a) compared the total variation diminishing scheme in MT3D (Zheng & Wang, 1999) with different Lagrangian approaches. These authors found that the total variation diminishing scheme overestimates dispersion even when the velocity contrast between grid cells is relatively small. This effect was significantly less important in the standard random walk method based on the Ito-Fokker-Planck equation. They also concluded that the random walk method is most advantageous when high resolution or many model runs (stochastic modeling) are needed in advective-dominated problems. Along the same line, Boso et al. (2013) compared several popular Eulerian and Lagrangian schemes. These authors showed that the Eulerian schemes overestimate mixing due to numerical dispersion and truncation errors, while Lagrangian approaches, such as the random walk method (e.g., Salamon et al., 2006a) and the smoothed particle hydrodynamics (e.g., Herrera et al., 2009; Tartakovsky & Meakin, 2005), are virtually free

of numerical dispersion. The smoothed particle hydrodynamics was found to be computationally demanding compared to other Lagrangian approaches.

In addition to these numerical features, several authors have demonstrated that the random walk method can also efficiently deal with non-Fickian transport phenomena (Berkowitz et al., 2006; Cvetkovic & Haggerty, 2002; Delay & Bodin, 2001; Dentz & Castro, 2009; Zhang & Benson, 2008) and multiple porosity systems (Benson & Meerschaert, 2009; Huang et al., 2003; Salamon et al., 2006b; Tsang & Tsang, 2001; Willmann et al., 2013). Linear reactive transport problems including network reactions can also be easily handled (Henri & Fernàndez-Garcia, 2014, 2015). However, these Lagrangian methods have not been sufficiently developed for simulating complex reactive transport problems, even though several advantages are reported in the literature (Sanchez-Vila & Fernàndez-Garcia, 2016). For instance, the recent comparison of Eulerian and Lagrangian reactive transports reported by Benson et al. (2017) showed that Lagrangian simulations in a heterogeneous velocity field with a bimolecular reaction of the type $A + B \rightarrow P$ can outperform the most popular Eulerian schemes. These authors reported twice as much excess reaction (due to advection error) in the first-order Eulerian simulations compared to the Lagrangian, using 1 million cells with a grid Peclet number of 10. These results indicated that a very strict constraint on the grid Peclet number is required in Eulerian reactive transport simulations.

However, the random walk method is not free of disadvantages. The discrete nature of particles can lead to statistical fluctuations of the concentrations, which are only eliminated by using an infinite number of particles. In practice, this is not feasible, and one needs to deal with a limited number of particles. This causes a subsampling problem which calls for the definition of the particle support volume. The formal interpretation of the particle support volume is crucial for reactive transport as it defines the interaction of one particle with nearby particles of different species. Different interpretations of the particle support volume can be found in the literature. Some authors (e.g., Benson & Meerschaert, 2008; Edery et al., 2009; Engdahl et al., 2017; Hansen et al., 2014; Tompson, 1993; Tompson et al., 1996) have considered that the particle support volume is a fixed quantity defined based on the numerical discretization of the transport problem or the particle motion. The latter approach was proposed by Benson and Meerschaert (2008) who determined the particle support volume based on the probability of collocation of two particles due to local dispersion. This method has been largely extended in different directions (Bolster et al., 2012, 2016; Paster et al., 2013, 2014) and recently applied to simulate biodegradation in a real field application (Ding et al., 2017). A different approach was introduced by Fernàndez-Garcia and Sanchez-Vila (2011) who defined the particle support volume based on the kernel density estimation (KDE) theory. This approach expresses the particle support volume as a function of the concentration distribution and the number of particles used. Since concentrations are affected not only by local dispersion but also by the model-scale fluid dynamics, this particle support volume also changes with time depending on the deformation of the particle plume. This kernel-based approach has been demonstrated to largely minimize the errors involved in the reconstruction of concentrations and in quantifying mixing and human health risk (Fernàndez-Garcia & Sanchez-Vila, 2011; Pedretti & Fernàndez-Garcia, 2013; Siirila-Woodburn et al., 2015). Moreover, this approach has also been extended to simulate nonlinear reactive transport (Rahbaralam et al., 2015; Schmidt et al., 2017). Recently, Sole-Mari et al. (2017) extended the kernel-based approach to simulate any kind of reactions other than bimolecular monovalent reactions. Until now, in these kernel-based methods, the particle support volume associated with the particles of a given chemical species has been allowed to evolve in time but not in space.

In this paper we explore the potential benefits of allowing the particle support volume to vary not only in time but also in space. We develop a novel approach capable of estimating an optimal locally adaptive kernel function at the particle location based on the minimization of a local error criterion. Then, we compare the new approach with the time-dependent global optimal kernel approach and the traditional binning (histogram) method. To do so, we evaluate the estimation accuracy of the concentrations and the reaction rates in a 2-D heterogeneous reactive transport model.

2. Background

2.1. KDE and Random Walk Models

KDE is a nonparametric technique that has become popular in the field of data analytics. It is typically used to estimate the underlying probability density function (pdf) of a continuous random variable given a sample

$\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$. Its superiority over other estimators such as histograms (i.e., counting values that fall into a given support volume or bins) has been theoretically shown in the literature (e.g., Härdle, 1991; Silverman, 1986). In the classical multivariate KDE approach, the pdf $p(\mathbf{x})$ is approximated as

$$p(\mathbf{x}) \cong \widehat{p}(\mathbf{x}) = \frac{1}{N} \sum_{\alpha=1}^{N} W_{\mathrm{H}}(\mathbf{x} - \mathbf{X}_\alpha, \mathbf{H}), \tag{1}$$

where the bandwidth matrix $\mathbf{H}$ is a symmetric positive definite $d \times d$ matrix, being d the dimensionality of $\mathbf{x}$, and W_{H} is a scaled kernel function defined as

$$W_{\mathrm{H}}(\mathbf{u}, \mathbf{H}) = |\mathbf{H}|^{-\frac{1}{2}} W\left(\mathbf{H}^{-\frac{1}{2}} \mathbf{u}\right). \tag{2}$$

Here the kernel function W integrates to 1, has zero-mean, and satisfies

$$\boldsymbol{\mu}_2(W) = \int_{\mathbb{R}^d} \mathbf{u}\, \mathbf{u}^{\mathsf{T}}\, W(\mathbf{u}) \mathrm{d}\mathbf{u} = \mu_2(W) \mathbf{I}_d, \tag{3}$$

where $\boldsymbol{\mu}_2(W)$ is the second moment matrix of the kernel W, and $\mathbf{I}_d$ is the $d \times d$ identity matrix. The order of a kernel is defined as the lowest positive integer k for which the kth moment is nonzero. For simplicity, we considered a second-order kernel function, meaning that $\mu_2(W) \neq 0$. A common choice for W is the Gaussian kernel,

$$W(\mathbf{u}) = (2\pi)^{-\frac{d}{2}} \exp\left(-\frac{1}{2}\, \mathbf{u}^{\mathsf{T}} \mathbf{u}\right). \tag{4}$$

If the kernel function is Gaussian, $\mathbf{H}$ is the covariance matrix of W_{H}.

In the context of random walk particle tracking (RWPT) models of solute transport in porous media, $p(\mathbf{x})$ represents the continuum-scale distribution of the mass of a specific solute in a porous medium at the $\mathbf{x}$ location, and $\mathbf{X}_\alpha$ is the position of the αth numerical particle. Then, $p(\mathbf{x})$ can be directly linked with resident solute concentrations $c(\mathbf{x})$ by

$$c(\mathbf{x}) = M \frac{p(\mathbf{x})}{\phi(\mathbf{x})}, \tag{5}$$

where $c(\mathbf{x})$ is expressed as amount of substance per unit volume of solution, $\phi(\mathbf{x})$ is volume of solution per unit d-dimensional volume of medium (effective porosity in the case of a fully saturated 3-D porous medium), and M is the total amount of substance of the solute,

$$M = \int_{\mathbb{R}^d} \phi(\mathbf{x}) c(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{6}$$

The bandwidth matrix $\mathbf{H}$ of the kernel function can be seen here as the support volume of a particle. This parameter is crucial in reactive transport simulations as it defines the potential interaction among particles. To illustrate this point, let us consider a chemical reaction A + B → C with a reaction rate described as $r = k_f c_{\mathrm{A}} c_{\mathrm{B}}$, where the subscripts A and B refer to the chemical species and k_f is the reaction rate coefficient. In this case, Sole-Mari et al. (2017) suggested that the probability of reaction of particle A_α in a given time interval $[t, t + \Delta t]$ is

$$P\left(\mathrm{A}_\alpha \rightarrow \mathrm{C}_\gamma\right) = \frac{k_f}{\phi\left(\mathbf{X}^{\mathrm{AB}}_{\alpha\beta}\right)} \Delta t \sum_{\beta=1}^{N_{\mathrm{B}}} m^{\mathrm{B}}_\beta W\left(\mathbf{X}^{\mathrm{A}}_\alpha - \mathbf{X}^{\mathrm{B}}_\beta, \mathbf{H}^{\mathrm{A}}_\alpha + \mathbf{H}^{\mathrm{B}}_\beta\right), \tag{7}$$

where $\mathbf{X}^{\mathrm{A}}_\alpha$, $\mathbf{X}^{\mathrm{B}}_\beta$, $\mathbf{H}^{\mathrm{A}}_\alpha$, and $\mathbf{H}^{\mathrm{B}}_\beta$ are the positions and kernel bandwidth matrices for particles A_α and B_β, respectively, and the porosity ϕ is evaluated at a weighted midpoint:

$$\mathbf{X}^{\mathrm{AB}}_{\alpha\beta} = \left(\mathbf{H}^{\mathrm{A}^{-1}}_\alpha + \mathbf{H}^{\mathrm{B}^{-1}}_\beta\right)^{-1} \left(\mathbf{H}^{\mathrm{A}^{-1}}_\alpha \mathbf{X}^{\mathrm{A}}_\alpha + \mathbf{H}^{\mathrm{B}^{-1}}_\beta \mathbf{X}^{\mathrm{B}}_\beta\right). \tag{8}$$

This formulation can be extended to more complex stoichiometries and nonlinear reaction rates by simply adding a compensation function (Sole-Mari et al., 2017), as described in Appendix A. In any case, the important point here is that the probability of reaction directly depends on the distance between particles associated with different species and their corresponding bandwidth matrices.

2.2. Selection of the Optimal Bandwidth Matrix

The resulting estimates of concentrations and reactions strongly depend on the choice of the bandwidth matrix **H**. If |**H**| is too large, the estimated concentrations will be oversmoothed and reactions overpredicted. If |**H**| is too small, the estimated concentrations will display fluctuations leading to incomplete mixing (Benson & Meerschaert, 2008). Consequently, an optimal bandwidth matrix exists. A common optimality criterion in statistics for the selection of **H** is the Mean Integrated Squared Error, formally written as

$$MISE(\mathbf{H}) = \int_{\mathbb{R}^d} \mathrm{E}\left\{ \left(\widehat{p}(\mathbf{x}) - p(\mathbf{x})\right)^2 \right\} \mathrm{d}\mathbf{x}, \tag{9}$$

where E{·} is the expected value operator. It can be shown that in the limit, when $N|\mathbf{H}|^{1/2} \to \infty$ and $|\mathbf{H}| \to 0$, the *MISE* approaches an asymptotic value, typically denoted as *AMISE* (e.g., Härdle, 1991), given by

$$AMISE(\mathbf{H}) \approx \frac{R(W)}{N|\mathbf{H}|^{\frac{1}{2}}} + \frac{1}{4}\mu_2^2(W) T(p, \mathbf{H}), \tag{10}$$

where

$$R(W) = \int_{\mathbb{R}^d} W^2(\mathbf{y}) \mathrm{d}\mathbf{y}, \tag{11}$$

$$T(p, \mathbf{H}) = \int_{\mathbb{R}^d} \mathrm{tr}^2 \left(\mathbf{H} \frac{\partial^2 p(\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^{\mathsf{T}}} \right) \mathrm{d}\mathbf{x}. \tag{12}$$

The first term on the right-hand side of (10) is the asymptotic integral of the mean variance, and the second term is the asymptotic integral of the mean squared bias. tr indicates the trace of a matrix. For completeness, the derivation of expression (10) is given in Appendix B. The optimal bandwidth matrix is the one that minimizes the *AMISE* with respect to **H**. Closed-form solutions to this minimization problem only exist for certain parametrizations of **H** (Wand & Jones, 1993). For $d > 1$, the unconstricted case (without a priori assumptions) does not have a closed-form solution, although it can be solved numerically (e.g., Gramacki, 2018). The simplest approach is what we refer to as the Global Isotropic (GI) method, which parametrizes the bandwidth matrix using only one parameter: $\mathbf{H} = h_{\mathrm{GI}}^2 \mathbf{I}_d$. In this case, the optimal h_{GI} (obtained by imposing $\partial AMISE / \partial h_{\mathrm{GI}} = 0$) can be written as (e.g., Silverman, 1986)

$$h_{\mathrm{GI}} = \left(\frac{d\, R(W)}{\mu_2^2(W)\, N\, T(p, \mathbf{I}_d)} \right)^{\frac{1}{d+4}}. \tag{13}$$

Since h_{GI} in (13) depends on the unknown second spatial derivatives of the true pdf through $T(p, \mathbf{I}_d)$, a variety of methods have been developed for selecting the bandwidth matrix. The general consensus is that the most useful methods are the plug-in and the cross validation (Jones et al., 1996; Park & Marron, 1990). In this article, we will use the plug-in method (Botev et al., 2010; Sheather & Jones, 1991), which consists in determining $T(p, \mathbf{I}_d)$ with another kernel estimator. The bandwidth of this kernel depends in turn on the squared integral of higher-order derivatives of $p(\mathbf{x})$, which can also be estimated by other kernel estimators. Recursively, this leads to an infinite sequence, which is typically solved by assuming a normal distribution of $p(\mathbf{x})$ in the estimation of some high-order integral squared derivatives (Sheather & Jones, 1991).

As presented so far, the minimization of *AMISE* provides a globally optimal bandwidth matrix, which is assumed constant in space. However, spatially variable degrees of smoothing may be needed in real situations, most particularly at the edges of the particle distributions where fewer data points exist. To overcome this problem, two different types of locally adaptive kernel estimators have been proposed in the literature that allow the bandwidth matrix **H** to vary as a function of the estimation location **x** (balloon estimator) or the particle position $\mathbf{X}_\alpha$ (sample-point estimation). We refer to Sain (2002) for a detailed review on locally adaptive multivariate KDE approaches. The most popular balloon estimator is the nearest neighbor density estimator introduced by Loftsgaarden and Quesenberry (1965). Assuming an isotropic parametrization $\mathbf{H} = h^2 \mathbf{I}_d$, this method considers that h is proportional to the distance from **x** to the kth nearest data point (which is asymptotically equivalent to considering that $h \propto 1/p$). When applied pointwise, the balloon estimator behaves just like the fixed bandwidth estimator (Terrell & Scott, 1992). However, this method is not mass-conservative (the pdf does not need to integrate to 1)

and suffers from severe bias problems as well as discontinuities (Hall, 1983; Mack & Rosenblatt, 1979). In contrast, sample-point estimators (Breiman et al., 1977) associate a different bandwidth matrix with each sample point. Most of these approaches set the bandwidth as proportional to h_{GI} multiplied by some negative power of a pilot density estimate evaluated at $\mathbf{X}_\alpha$ (e.g., Abramson, 1982; Breiman et al., 1977; Pedretti & Fernàndez-Garcia, 2013; Silverman, 1986). For example, Abramson's (1982) *square root law* suggests that $h \propto p^{-1/2}$. In these sample-point methods, the local bandwidth is not independent from the global one, and it is restricted to be only a function of the magnitude of the density (Sain, 2002). Hence, the effect of more complex features of the local density distribution is neglected. Moreover, these methodologies do not determine the kernel shape in multivariate KDE. A more general approach is the *diffusion kernel* by Botev et al. (2010), who also define a priori the dependence of the kernel function on a pilot estimate of p. Sain and Scott (1996) proposed dividing the domain of a univariate distribution in a number of bins (5–30) and applying a piecewise constant bandwidth to each bin. They optimized the set of bin bandwidths numerically to minimize the MISE. The method renders a considerable gain in accuracy with respect to the fixed bandwidth estimator. Sain (2002) extended this methodology to implement it on bivariate distributions.

3. A Locally Adaptive Kernel Density Estimator

In this section we develop a sample-point estimator that relies on a new local definition of the estimation error used to select the locally optimal kernel bandwidth in 2-D and 3-D problems. As opposed to balloon estimators, a sample-point estimator is inherently mass-conservative (the estimated pdf always integrates to 1). We do not rely on simple rules such as $h \propto p^{-1/2}$, but the local variations of the kernel in space occur naturally to satisfy a minimum error criterion. Unlike the approach proposed by Sain and Scott (1996), the support volume is not piecewise constant in a set of bins but varies smoothly with the particle location. Under different assumptions on the kernel shape, we give the corresponding expressions of the bandwidth matrix. Then, in section 3.2, we present the implementation algorithm.

3.1. Theoretical Development

Given a random distribution of particles $\{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ obtained from a RWPT simulation at a given time t, the kernel density estimate $\widehat{p}(\mathbf{x})$ of $p(\mathbf{x})$ is defined as

$$\widehat{p}(\mathbf{x}) = \frac{1}{N}\sum_{\alpha=1}^{N} W_{\mathrm{H}}(\mathbf{x}-\mathbf{X}_\alpha, \mathbf{H}_\alpha), \tag{14}$$

where $\mathbf{H}_\alpha$ is the bandwidth matrix associated with the αth particle at the $\mathbf{X}_\alpha$ position, that is, $\mathbf{H}_\alpha = \mathbf{H}(\mathbf{X}_\alpha)$. The bandwidth matrix $\mathbf{H}_\alpha$ is determined based on the minimization of the Local Asymptotic Mean Integrated Squared Error centered at the particle position, which we define as

$$LMISE_\alpha(\mathbf{H}_\alpha) = \int_{\mathbb{R}^d} E\left\{\left(\widehat{p}(\mathbf{x}) - p(\mathbf{x})\right)^2\right\} W_{\mathrm{H}}(\mathbf{x}-\mathbf{X}_\alpha, \mathbf{H}_0)\, \mathrm{d}\mathbf{x}. \tag{15}$$

The kernel function $W_{\mathrm{H}}(\mathbf{x}-\mathbf{X}_\alpha, \mathbf{H}_0)$ is used to penalize the error distance away from the particle position where the approximation is centered. The bandwidth matrix $\mathbf{H}_0$ controls the strength of this penalty, which is assumed constant for all particles. Assuming that the bandwidth matrix does not substantially vary near the particle position $\mathbf{X}_\alpha$ (i.e., within the support volume defined by $\mathbf{H}_0$), the $LMISE_\alpha$ approaches to an asymptotic expression, in the limit when $N|\mathbf{H}_\alpha|^{1/2} \to \infty$ and $|\mathbf{H}_\alpha| \to 0$, denoted here as $ALMISE_\alpha$. This is demonstrated in Appendix B. The $ALMISE_\alpha$ can be formally written as

$$ALMISE_\alpha(\mathbf{H}_\alpha) \approx \frac{1}{N^2}\left(n_\alpha(\rho) R(W) |\mathbf{H}_\alpha|^{-\frac{1}{2}} + \frac{1}{4}\mu_2^2(W) T_\alpha(\rho, \mathbf{H}_\alpha)\right), \tag{16}$$

where ρ is the density of particles defined as the number of particles per unit volume, that is, $\rho(\mathbf{x}) = Np(\mathbf{x})$, and

$$T_\alpha(\rho, \mathbf{H}_\alpha) = \int_{\mathbb{R}^d} W_{\mathrm{H}}(\mathbf{x}-\mathbf{X}_\alpha, \mathbf{H}_0)\, \mathrm{tr}^2\left(\mathbf{H}_\alpha \frac{\partial^2 \rho}{\partial \mathbf{x} \partial \mathbf{x}^T}\right) \mathrm{d}\mathbf{x}, \tag{17}$$

$$n_\alpha(\rho) = \int_{\mathbb{R}^d} W_{\mathrm{H}}(\mathbf{x}-\mathbf{X}_\alpha, \mathbf{H}_0)\ \rho(\mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{18}$$

The first term on the right-hand side of (16) is the variance. It directly depends on the smoothed particle density n_α and the volume of the kernel bandwidth, $|\mathbf{H}_\alpha|^{1/2}$, meaning that excessively small kernel bandwidths can generate noise in the estimation. The second term is the squared bias. It is proportional to the value of the net roughness T_α, which can be seen as a measure of the degree of curvature of the hills and valleys of the density distribution near the particle position $\mathbf{X}_\alpha$ weighted by $\mathbf{H}_\alpha$. Strong spatial variations of the density gradients combined with excessively big or wrongly oriented kernel bandwidths can cause oversmoothing.

The optimal bandwidth matrix, which balances the variance and the bias terms, can be determined from the solution of the following minimization problem:

$$\mathbf{H}_\alpha = \arg \min_{\mathbf{H}'_\alpha} ALMISE_\alpha\left(\mathbf{H}'_\alpha\right). \tag{19}$$

When a full bandwidth matrix parametrization is used, the minimization of $ALMISE_\alpha$ involves $d(1 + d)/2$ parameters. Similar to the global case, a closed-form solution does not exist for $d > 1$. Although numerical optimization would still be possible, it would entail an unpractical amount of calculations. However, it is possible to derive explicit solutions of $\mathbf{H}_\alpha$ under certain assumptions. Let us first decompose the bandwidth matrix into three components (similar to eigenvalue decomposition),

$$\mathbf{H}_\alpha = h_\alpha^2 \mathbf{V}_\alpha \mathbf{S}_\alpha \mathbf{V}_\alpha^{\mathrm{T}}, \tag{20}$$

where $\mathbf{V}_\alpha$ is the rotation matrix, $\mathbf{S}_\alpha$ is the shape matrix that defines the relative elongation of the ellipsoidal kernel, and h_α is the scaling parameter of the kernel. At this point, $\mathbf{V}_\alpha$ and $\mathbf{S}_\alpha$ are assumed different from $\mathbf{I}_d$. At the end of this section we introduce possible assumptions that lead to different variants of the local approach. The shape matrix is a diagonal matrix defined as $\mathbf{S}_\alpha = \mathrm{diag}\,(s_1, \ldots, s_d)$ with $|\mathbf{S}_\alpha| = 1$. Knowing that $|\mathbf{S}_\alpha| = 1$, the last component of $\mathbf{S}_\alpha$ can be written as $s_d = \left(\prod_{i=1}^{d-1} s_i\right)^{-1}$. In order to reduce the number of parameters, we considered that the principal directions of the bandwidth matrix $\mathbf{H}_\alpha$ are predetermined, that is, the rotation matrix $\mathbf{V}_\alpha$ is known. From this, $ALMISE_\alpha$ is a function of the following d parameters: $\{h_\alpha, s_1, \ldots, s_{d-1}\}$. Substituting (20) into (16) and defining the projected Hessian matrix as

$$\mathbf{G}_\alpha = \mathbf{V}_\alpha^{\mathrm{T}} \frac{\partial^2 \rho}{\partial \mathbf{x} \partial \mathbf{x}^{\mathrm{T}}} \mathbf{V}_\alpha, \tag{21}$$

we have

$$ALMISE_\alpha(h_\alpha, s_1, \ldots, s_{d-1}) \approx \frac{1}{N^2}\left(\frac{n_\alpha(\rho)\, R(W)}{h_\alpha^d} + \frac{h_\alpha^4}{4}\mu_2^2(W)\, T_\alpha(\rho, \mathbf{S}_\alpha)\right), \tag{22}$$

where

$$T_\alpha(\rho, \mathbf{S}_\alpha) = \sum_{i=1}^{d}\sum_{j=1}^{d} s_i s_j \psi_{ij}(\rho), \tag{23}$$

$$\psi_{ij}(\rho) = \int_{\mathbb{R}^d} W_{\mathrm{H}}(\mathbf{x} - \mathbf{X}_\alpha, \mathbf{H}_0)\, G_{ii}(\mathbf{x}) G_{jj}(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{24}$$

Here the coefficients G_{ii} are the diagonal components of $\mathbf{G}_\alpha$, which are the projections of the Hessian along the eigenvectors of the local bandwidth matrix $\mathbf{H}_\alpha$, and ψ_{ij} form a symmetric matrix $\boldsymbol{\psi}_\alpha$, designated here as the roughness matrix.

Taking $\partial ALMISE/\partial h_\alpha = 0$ and $\partial ALMISE/\partial s_k = 0$ for $k = 1, \ldots, d - 1$, we obtain that the optimal kernel scaling parameter can be written as

$$h_\alpha = \left(\frac{d\, R(W) n_\alpha(\rho)}{\mu_2^2(W) T_\alpha(\rho, \mathbf{S}_\alpha)}\right)^{\frac{1}{d+4}}, \tag{25}$$

and the shape matrix components can be written as

$$s_1 = \left(\frac{\psi_{22}}{\psi_{11}}\right)^{\frac{1}{4}}, s_2 = \left(\frac{\psi_{11}}{\psi_{22}}\right)^{\frac{1}{4}}, \quad \text{for} \quad d = 2, \tag{26}$$

$$s_1 = \left(\frac{\psi_{22}\psi_{33}}{\psi_{11}^2}\right)^{\frac{1}{6}}, \quad s_2 = \left(\frac{\psi_{11}\psi_{33}}{\psi_{22}^2}\right)^{\frac{1}{6}}, \quad s_3 = \left(\frac{\psi_{11}\psi_{22}}{\psi_{33}^2}\right)^{\frac{1}{6}}, \quad \text{for} \quad d = 3. \tag{27}$$

The derivations of these expressions are given in Appendix C. It should be noted that the solution for the case $d = 3$ is an approximation. Substituting (26) and (27) into (23), we have

$$T_\alpha = 2(\psi_{11}\psi_{22})^{\frac{1}{2}} + 2\psi_{12}, \quad \text{for} \quad d = 2, \tag{28}$$

$$\begin{aligned} \mathrm{T}_\alpha = {} & 3(\psi_{11}\psi_{22}\psi_{33})^{\frac{1}{3}} + 2\psi_{23}\left(\psi_{22}\psi_{33}/\psi_{11}^2\right)^{-\frac{1}{6}} + 2\psi_{13}\left(\psi_{11}\psi_{33}/\psi_{22}^2\right)^{-\frac{1}{6}} \\ & + 2\psi_{12}\left(\psi_{11}\psi_{22}/\psi_{33}^2\right)^{-\frac{1}{6}}, \quad \text{for } d = 3. \end{aligned} \tag{29}$$

Note from (25) that smaller values of the optimal scaling parameter h_α will occur for higher values of the net roughness T_α. This elucidates the mimetic nature of the local adaption: The kernel becomes larger in those areas where concentrations are smoother, and it becomes smaller in the presence of abrupt hills and valleys of the concentrations. Similarly, note from (26) and (27) that the smallest eigenvalue of the kernel shape corresponds to the local dominant orientation of the curvature of the particle densities. That is to say that the local kernel is more elongated along the minimum curvature, where fluctuations are smaller. Based on this, we propose two different ways to choose the rotation matrix $\mathbf{V}_\alpha$ in RWPT simulations. The first one considers a rotation matrix that aligns the kernel bandwidth matrix along the mean flow direction or any other foreseeable direction of marked anisotropy in the velocity field. This has the advantage that $\mathbf{V}_\alpha$ becomes a constant matrix that does not depend on the particle position $\mathbf{X}_\alpha$. It is therefore computationally more efficient. We call this parametrization the Local Diagonal (LD) approach. The other consists of orienting the kernel bandwidth matrix along the local flow velocity vector, estimated as

$$\mathbf{v}_\alpha = \int_{\mathbb{R}^d} W_{\mathrm{H}}(\mathbf{x} - \mathbf{X}_\alpha, \mathbf{H}_0)\, \mathbf{v}(\mathbf{x}) \mathrm{d}\mathbf{x}, \tag{30}$$

where $\mathbf{v}(\mathbf{x})$ is the flow velocity at position $\mathbf{x}$. This is most appropriate in advection-dominated transport. We call this parametrization the Local Rotated (LR) approach. The rotation matrix in this case can be written as

$$\mathbf{V}_\alpha = \frac{1}{w_2}\begin{bmatrix} v_1 & -v_2 \\ v_2 & v_1 \end{bmatrix}, \quad \text{for} \quad d = 2, \tag{31}$$

$$\mathbf{V}_\alpha = \frac{1}{w_2 w_3}\begin{bmatrix} v_1 w_2 & -v_2 w_3 & v_1 v_3 \\ v_2 w_2 & v_1 w_3 & v_2 v_3 \\ v_3 w_2 & 0 & -w_2 \end{bmatrix}, \quad \text{for} \quad d = 3, \tag{32}$$

where v_i is the ith component of $\mathbf{v}_\alpha$, $w_2 = \sqrt{v_1^2 + v_2^2}$, and $w_3 = \sqrt{v_1^2 + v_2^2 + v_3^2}$.

In addition to the LD and LR, we also considered a third variant of the method, the Local Isotropic (LI), in which $\mathbf{S}_\alpha$ is assumed unitary instead of adopting the optimal shape described by (26) and (27). Table 1 summarizes the relevant equations for the three approaches, namely, those defining the factors of the optimal kernel matrix (20) and the main parameters involved in the optimization.

Figure 1 exemplifies the locally adaptive behavior of the three proposed kernels, graphically represented by the 95% confidence region of the Gaussian kernel determined by $\mathbf{H}_\alpha$, for an illustrative distribution of particles.

3.2. Implementation

By examining expressions (21) to (29), we see that the asymptotically optimal bandwidth matrix (20) depends on the second spatial derivatives of the unknown density of particles defined by the Hessian matrix of $\rho(\mathbf{x})$. Following the basic ideas of the plug-in method, we estimated these second spatial derivatives with another kernel estimator, defined as

Table 1
Summary of Relevant Equations for the Different Local Approaches

	Rotation (V_α)	Shape (S_α)	Scaling (h_α)	Net roughness (T_α)	Roughness (ψ_α)	Projected Hessian (G_α)
Isotropic	$\mathbf{I}_d$	$\mathbf{I}_d$	equation (25)	$\sum_{i=1}^{d}\sum_{j=1}^{d}\psi_{ij}$	equation (24)	$\frac{\partial^2\rho}{\partial\mathbf{x}\partial\mathbf{x}^T}$
Diagonal	$\mathbf{I}_d$	equations (26) and (27)	equation (25)	equations (28) and (29)	equation (24)	$\frac{\partial^2\rho}{\partial\mathbf{x}\partial\mathbf{x}^T}$
Rotated	equations (31) and (32)	equations (26) and (27)	equation (25)	equations (28) and (29)	equation (24)	equation (21)

$$\frac{\partial^2\rho}{\partial x_i\partial x_j}(\mathbf{x})\cong\sum_{\alpha=1}^{N}W_{ij}^{(2)}\left(\mathbf{x}-\mathbf{X}_\alpha,\mathbf{H}_{ij}^{(2)}\right), \tag{33}$$

where

$$W_{ij}^{(2)}=\frac{\partial^2 W_{\mathrm{H}}}{\partial x_i\partial x_j}. \tag{34}$$

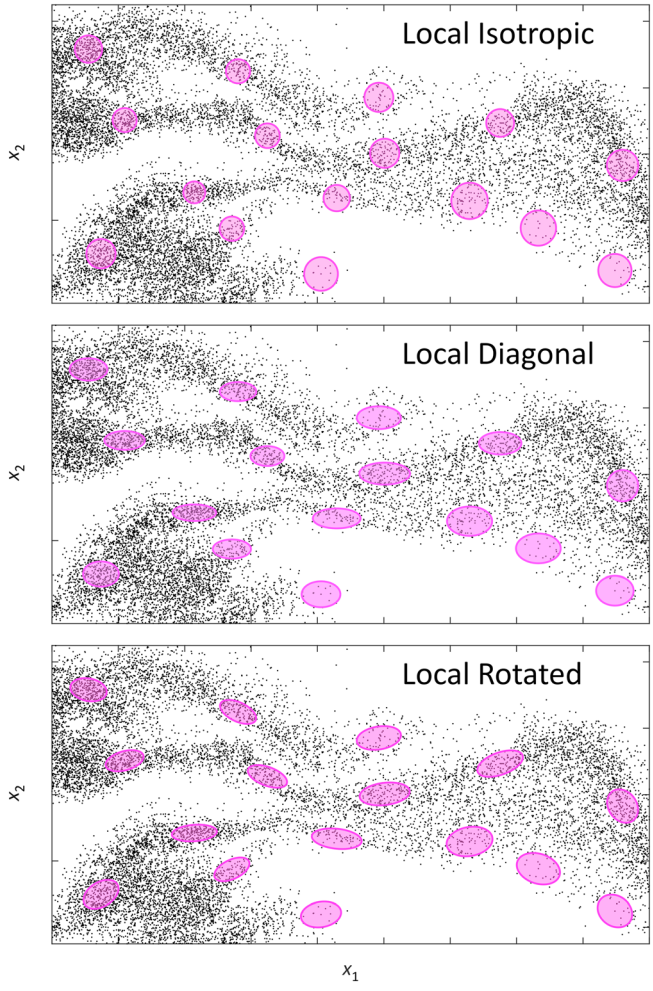

Figure 1. Illustrative representation of the local kernel adaptation at different positions of a particle plume, for the three proposed local approaches. The ellipsoids cover the 95% confidence region of the kernel.

This estimator can be directly obtained by taking partial derivatives of $\hat{p}(\mathbf{x})$ in (1). The bandwidth matrix of this estimator is not necessarily the same as $\mathbf{H}_\alpha$. In fact, obtaining the optimal $\mathbf{H}_{ij}^{(2)}$ (by the plug-in method) entails the estimation of fourth-order density derivatives, which in turn requires determining other bandwidth matrices that involve the estimation of sixth-order density derivatives and so on (e.g., Botev et al., 2010; Gramacki, 2018). In this work, we use the Improved Sheather and Jones (ISJ) algorithm developed by Botev et al. (2010) for the computation of $\mathbf{H}_{ij}^{(2)}$.

The LR approach involves the estimation of $\mathbf{v}_\alpha$, which is computed with a simple quadrature. To do this, the domain is divided into regular bins of size λ centered at the following positions: $\{\mathbf{x}_1, \ldots, \mathbf{x}_\eta\}$. Then

$$\mathbf{v}_\alpha \cong \lambda^d \sum_{\xi=1}^{\eta} W_{\mathrm{H}}(\mathbf{x}_\xi - \mathbf{X}_\alpha; \mathbf{H}_0)\, \mathbf{v}(\mathbf{x}_\xi). \tag{35}$$

Expressions (33) and (35) allow for the computation of $\mathbf{G}_\alpha(\mathbf{x})$ at any point by (21), which enables the estimation of ψ_{ij}:

$$\psi_{ij} \cong \lambda^d \sum_{\xi=1}^{\eta} W_{\mathrm{H}}(\mathbf{x}_\xi - \mathbf{X}_\alpha; \mathbf{H}_0) G_{ii}(\mathbf{x}_\xi) G_{jj}(\mathbf{x}_\xi). \tag{36}$$

We found that $\lambda = h_{\mathrm{GI}}$ provides an adequate discretization of these integrals (further decreasing the value of λ did not significantly alter the results), being h_{GI} the optimal bandwidth size of the GI method. The density of particles can be easily computed by introducing (14) into (18), which gives the following expression

$$\begin{aligned} n_\alpha &= \int_{R^d} W_H(\mathbf{x} - \mathbf{X}_\alpha, \mathbf{H}_0) \rho(\mathbf{x}) d\mathbf{x} \cong \sum_{\omega=1}^{N} \int_{R^d} W_H(\mathbf{x} - \mathbf{X}_\omega, \mathbf{H}_0) W_H(\mathbf{x} - \mathbf{X}_\alpha, \mathbf{H}_\alpha) \mathrm{d}\mathbf{x} \\ &= \sum_{\omega=1}^{N} W_H(\mathbf{X}_\omega - \mathbf{X}_\alpha, \mathbf{H}_0 + \mathbf{H}_\alpha). \end{aligned} \tag{37}$$

A rough estimate of n_α can be obtained here by setting $\mathbf{H}_\alpha$ as the optimal bandwidth size of the GI method, $\mathbf{H}_\alpha = h_{\mathrm{GI}}^2 \mathbf{I}_d$. Note that the local analysis window defined by $\mathbf{H}_0$ is also an input parameter. We found that $\mathbf{H}_0$ should be as small as possible while complying with $\mathbf{H}_0 > \mathbf{H}_\alpha$ for all α and all principal directions. Note that this constitutes an implicit rule. In this study, setting $\mathbf{H}_0 = (3h_{\mathrm{GI}})^2 \mathbf{I}_d$ yielded adequate results.

Algorithm for local optimization:

1. Compute $\mathbf{H}_{ij}^{(2)}$ by using a plug-in method such as ISJ.
2. On each evaluation point $\mathbf{x}_\xi = \{\mathbf{x}_1, \ldots, \mathbf{x}_\eta\}$, compute $\frac{\partial^2 \rho}{\partial \mathbf{x} \partial \mathbf{x}^\intercal}(\mathbf{x}_\xi)$ by (33).
3. On each integration point $\mathbf{X}_\alpha = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$,

a If local rotation (LR method), compute $\mathbf{v}_\alpha$ by (35) and apply rotation (21) to compute $\mathbf{G}_\alpha$ on the neighboring $\mathbf{x}_\xi$. Otherwise, $\mathbf{G}_\alpha(\mathbf{x}_\xi) = \frac{\partial^2 \rho}{\partial \mathbf{x} \partial \mathbf{x}^\intercal}(\mathbf{x}_\xi)$.
b Compute $\boldsymbol{\psi}_\alpha$ by (36).
c Compute n_α by (37).
d Use the relevant expressions given in section 3.2 to compute $\mathbf{H}_\alpha$.

Since performing step 3 on every $\mathbf{X}_\alpha$ can be computationally demanding, one can select a reduced number of integration points $\left\{\mathbf{X}_1', \ldots, \mathbf{X}_\mu'\right\}$ to calculate the corresponding local bandwidth matrices and compute $\mathbf{H}_\alpha$ for all particles by interpolation. In this work, we took the set $\left\{\mathbf{X}_1', \ldots, \mathbf{X}_\mu'\right\}$ to be the same as $\{\mathbf{x}_1, \ldots, \mathbf{x}_\eta\}$.

The equivalent global approaches can be obtained by replacing $W_{\mathrm{H}}(\mathbf{x}_\xi - \mathbf{X}_\alpha; \mathbf{H}_0)$ in (18), (24), and (30) by $1/\Omega$, where Ω is the total d-dimensional volume of the model domain. In this case, we have that $n_\alpha = N/\Omega$ and $\mathbf{v}_\alpha = \langle \mathbf{v} \rangle$, where $\langle \mathbf{v} \rangle$ is the mean flow velocity. Note that the latter condition renders the Global Diagonal (GD) and Global Rotated approaches identical. The functionals ψ_{ij} in the global case can be written as

$$\psi_{ij} = \frac{1}{\Omega} \int_{\mathbb{R}^d} \frac{\partial^2 \rho}{\partial x_i^2} \frac{\partial^2 \rho}{\partial x_j^2} \mathrm{d}\mathbf{x}. \tag{38}$$

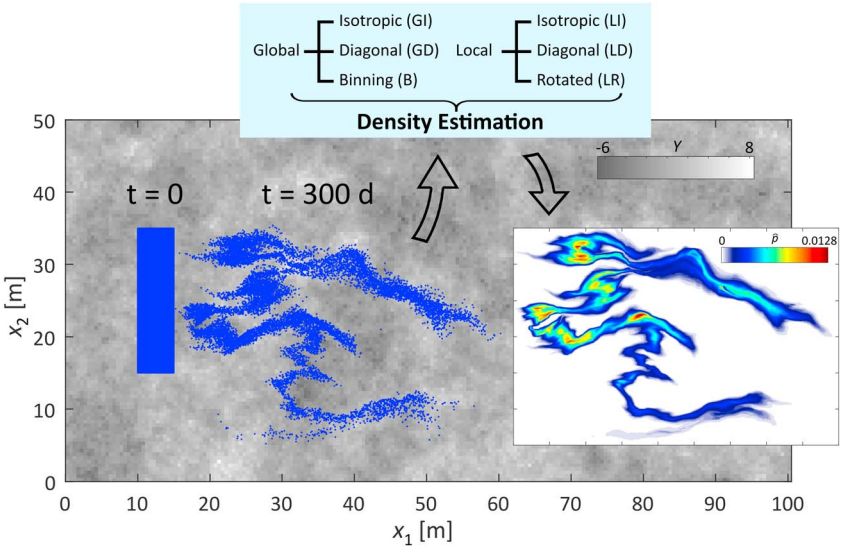

Figure 2. Setup of the conservative transport simulation and density estimation problem addressed in section 4.1. In the background, the log-conductivity field $Y(\mathbf{x})$. On top, the different density estimation approaches tested.

These integrals can be solved either by means of a simple quadrature as done here for the local approaches or by direct computation of the kernel convolutions between pairs of individual particles. We refer to Silverman (1986), Härdle (1991), Botev et al. (2010), or Gramacki (2018) for details on possible algorithms to solve these integrals. In this work, we used the ISJ method presented by Botev et al. (2010), which relies on the discrete cosine transform to compute (38).

4. Numerical Investigations

In this section we compare the performance of the six different kernel density estimators listed in Figure 2, which include the presented global and local approaches and the binning method. The latter represents the traditional approach typically used to estimate concentrations from RWPT simulations. It consists in counting particles falling into bins defined based on the discretization of the flow problem. In essence, each method estimates the support volume of a particle $\mathbf{H}_\alpha$ in a different way. We analyze then the effect that this particle interpretation has on the estimates of concentrations and reaction rates. Sections 4.1 and 4.2 examine the performance of the different approaches in terms of concentrations and reaction rates, respectively.

The numerical investigations simulate solute transport in a heterogeneous confined aquifer with dimensions 100.5×50 m^2. The distribution of the natural logarithm of the hydraulic conductivity $Y(\mathbf{x}) = \ln K(\mathbf{x})$ represents one realization of a Generalized Sub-Gaussian stochastic process (Panzeri et al., 2016; Riva, Neuman, et al., 2015; Riva, Panzeri, et al., 2015). This model assumes that $Y(\mathbf{x}) = U(\mathbf{x})G(\mathbf{x})$, where $G(\mathbf{x})$ is a Gaussian random field and the subordinator $U(\mathbf{x})$ is a nonnegative random process independent of $G(\mathbf{x})$. We consider that $U(\mathbf{x})$ follows a lognormal distribution of zero mean and variance $\sigma_U^2 = (2-\kappa)^2$ and that $G(\mathbf{x})$ is characterized by an isotropic exponential covariance function with an integral scale of $I_G = 3$ m. When $\kappa \to 2$, this model renders a Gaussian field, $Y(\mathbf{x}) \to G(\mathbf{x})$. In our simulations we set $\kappa = 1.8$, meaning that $Y(\mathbf{x})$ is slightly non-Gaussian. The mean and variance of $Y(\mathbf{x})$ are given as $\langle Y \rangle = 0$ and $\sigma_Y^2 = 2.2$. The resulting integral scale of the non-Gaussian field is $I_Y = 2.88$ m. The porosity and the aquifer thickness are considered constant with a value of 0.25 and 10 m, respectively. The head is prescribed on the left and right boundaries, forcing a rightward flow with mean velocity $\langle v \rangle = 0.0747$ m/day. The upper and lower boundaries are impermeable. The flow model is discretized in square cells of size 0.5×0.5 m^2. The intercell transmissivities are computed by the harmonic mean of the adjacent cell transmissivities. Groundwater flow is assumed at steady state and solved by MODFLOW 2005 (Harbaugh, 2005).

Solute particles are injected uniformly in a rectangular area at $t = 0$ and transported according to the standard random walk method based on the Ito-Fokker-Planck equation (e.g., Salamon et al., 2006b). Figure 2 shows a sketch of the simulation setup. We adopt a local longitudinal dispersivity of $a_\ell = 5$ cm and a local transverse

dispersivity of $a_t = 0.5$ cm. The hybrid interpolation method (linear for advection and bilinear for dispersion) suggested by LaBolle et al. (1996) is used to compute the particle displacements. All particles share the same mass. The time step is dynamically estimated to satisfy a small Courant number and a high Peclet number at each nonempty grid cell, that is, $Cu = v\Delta t/\Delta x < 0.1$ and $Pe = \Delta x^2/2D\Delta t > 10$, where Δt is the time step, Δx is the cell size, v is the norm of the particle velocity, and D is the dominant eigenvalue of the particle dispersion tensor ($D = a_\ell v$).

4.1. Estimation of Concentrations

In this case, we simulate solute transport of a conservative (nonreactive) species moving through the $Y(\mathbf{x})$ heterogeneous porous medium. The distribution of particles obtained at $t = 300$ days (when the solute plume moved about eight integral scales) is used to analyze the convergence of the concentrations as a function of the number of particles N. At this time, the numerical particle cloud has developed a complex shape (Figure 2), which is seen as a worse-case scenario for reconstructing concentrations and their functionals. For comparison purposes, we consider that the concentrations obtained with $N = 1.3 \times 10^7$ and reconstructed with the LR approach are the true solution. We note that all methods lead to this true solution when $N \approx 1.3 \times 10^7$.

The convergence to this solution is then analyzed based on the Integrated Squared Error (ISE_p) of the normalized concentrations, defined as

$$ISE_p = \int_{\mathbb{R}^d} (\widehat{p}(\mathbf{x}) - p(\mathbf{x}))^2 \mathrm{d}\mathbf{x}, \tag{39}$$

where $\widehat{p}$ is the estimated normalized concentration, and p is the reference value. Figure 3 offers a graphical comparison of the estimated concentrations obtained with the B, GI, and LR methods in a specific part of the domain for $N = 1.3 \times 10^4$ particles. Note that the part of the domain shown and the simulation time in Figures 1 and 3 are the same. Even though the reduction of noise and oversmoothing from B to GI is visually more obvious than from GI to LR, the LR method resulted in a stronger reconstruction of details. We attribute this gain to the adaptability of the kernel function, which can adapt its size and locally elongate its shape along the direction of minimum curvature.

Figure 4a shows the ISE_p as a function of the number of particles for the different methods employed. For comparison purposes, the ISE_p is presented as a relative quantity, that is, the ratio of ISE_p to the values obtained with the GI method, denoted as ISE_p^{GI}. Therefore, the convergence of the GI method cannot be directly seen from this figure. Yet we note that the GI method was found to converge to the true solution with a power law behavior of type $ISE_p \propto N^{-0.56}$, which agrees roughly with the theoretical convergence of the $AMISE$ given by $AMISE \propto N^{-4/(d+4)}$ (e.g., Härdle, 1991). Results shown in Figure 4a demonstrate that the ISE_p was substantially reduced by the use of local kernel approaches. In general, this reduction improved with the complexity of the kernel approach. Most importantly, the convergence rates of the local approaches (LI, LD, and LR) were faster than their global equivalents, leading to strong differences when $N > 10^4$. The B method is also shown in this figure to illustrate the contrast against the traditional method. In this case, since an arbitrary-fixed support volume was used, results display the existence of a number of particles N for which that specific bin size was optimal at that moment in time. In any case, the estimation error was always significantly larger than the one given by all kernel approaches.

To identify where this error reduction is taking place in the domain, we analyze the cumulative integrated squared errors, $CISE_p(q)$, obtained at different thresholds of concentrations q, defined as

$$CISE_p(q) = \int_{\mathbb{R}^d} (\widehat{p}(\mathbf{x}) - p(\mathbf{x}))^2 \, \mathscr{H}(q - p(\mathbf{x})) \, \mathrm{d}\mathbf{x}, \tag{40}$$

where $\mathscr{H}$ is the Heaviside step function. $CISE_p(q)$ can be seen as the partial integration of the squared errors in the region of concentrations below the threshold, q. This permits to identify the regions of the concentrations with larger errors.

Figure 4b shows the ratio $CISE_p/CISE_p^{\mathrm{GI}}$ as a function of $q/p_{\max}$, for $N = 1.6 \times 10^6$. $CISE_p^{\mathrm{GI}}$ refers to the results obtained with the GI method, and $p_{\max}$ is the maximum density of the reference distribution. The ratio

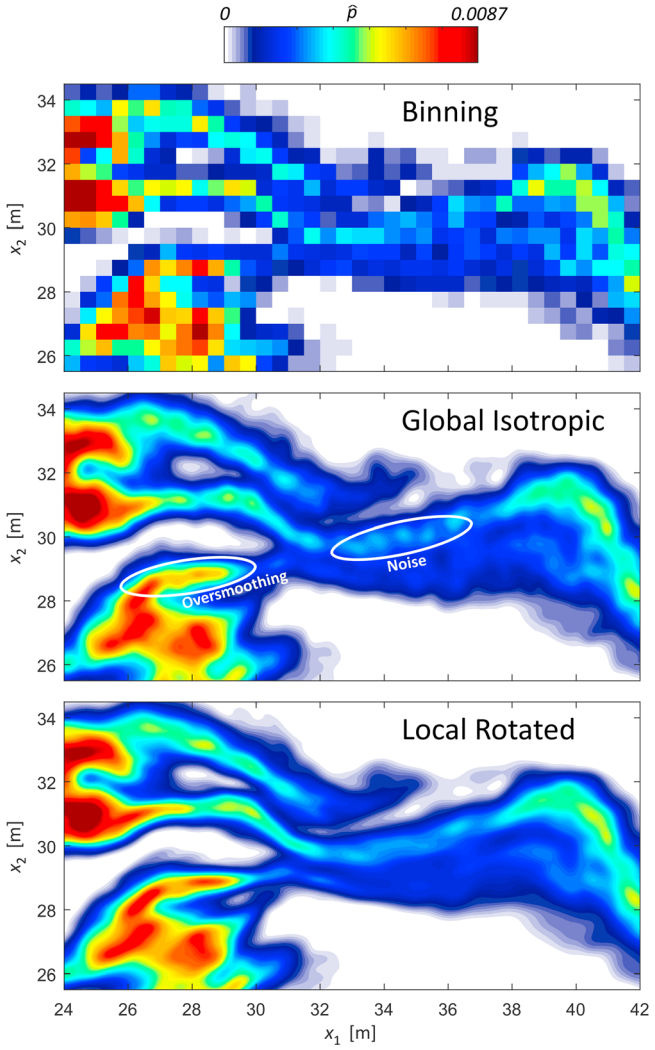

Figure 3. Graphical comparison of the estimation of $p(\boldsymbol{x})$ in a specific region of the model domain at $t = 300$ days, obtained by three different density estimators. The Local Rotated method (bottom) is proposed in this work.

$CISE_p/CISE_p^{GI}$ provided by the local kernel approaches displayed an important increase after $q/p_{max} \approx 0.1$, then reached a local maximum at about $q/p_{max} \approx 0.38$, and afterward slightly decreased with higher threshold values. This indicates that the error reduction seen in local approaches was particularly important in regions with high and low density of particles. We conclude then that local kernels can better reproduce the extreme values of concentrations, including those that typically occur at the plume edges, where most of the mixing-driven reactions are taking place in reactive transport. On the other hand, for the binning approach, errors in low density areas were particularly high with respect to the GI method. This result illustrates the poor capability of the B method to estimate low concentrations, and hence, front-mixing-driven reactions.

We then analyze the influence of the Peclet number, $Pe_\ell = I_Y/a_\ell$, the degree of heterogeneity, σ_Y^2, and the normalized mean travel distance, $t\langle v\rangle/I_Y$, one at a time, on the particle support volume estimated with the LD and GD approaches. The number of particles and the ratio of local dispersivities are maintained constant as $N = 1.6 \times 10^6$ and $a_t/a_\ell = 0.1$, respectively. The impact of σ_Y^2 is analyzed by rescaling the random field by a constant factor. The differences are quantified through two normalized error indexes,

$$\epsilon_h = \frac{\left[\sum_{\alpha=1}^{N}\left(h_{\alpha,\mathrm{LD}} - h_{\mathrm{GD}}\right)^2\right]^{\frac{1}{2}}}{\sum_{\alpha=1}^{N} h_{\alpha,\mathrm{LD}}}, \quad \epsilon_s = \frac{\left[\sum_{\alpha=1}^{N}\left(s_{1,\alpha,\mathrm{LD}} - s_{1,\mathrm{GD}}\right)^2\right]^{\frac{1}{2}}}{\sum_{\alpha=1}^{N} s_{1,\alpha,\mathrm{LD}}}. \tag{41}$$

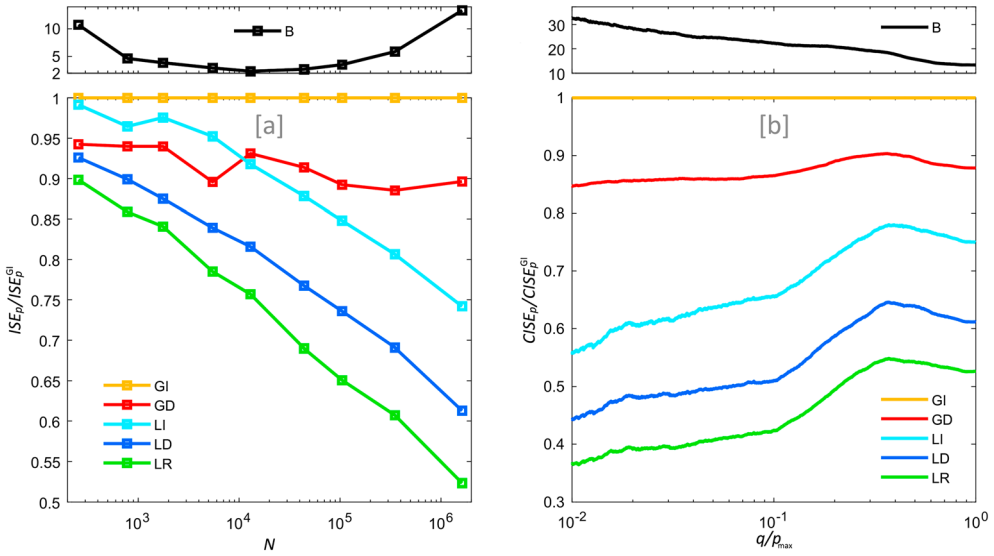

Figure 4. (a) ISE in the density estimation as defined in (39), relative to the Global Isotropic method, for different numbers of particles N and (b) the Cumulative ISE in the density estimation as defined in (40), relative to the Global Isotropic method, as a function of the threshold q, relative to the domain's maximum true density. The legend shows the acronyms of the tested approaches (see Figure 2). ISE = Integrated Squared Error.

Figure 5 compares the distribution of the scaling and the elongation parameters $\{h, s_1\}$ estimated within the solute plume by the LD and GD methods for different problem conditions. The spatial distribution of particles and the error indexes are also shown for completeness.

In general, results indicate that the net roughness (i.e., the degree of curvature of the concentrations; see equations (23)–(25)) controls the scale of the particle support volume. At high Pe_ℓ and σ_Y^2, solute transport is strongly dominated by advection, displaying complex and irregular concentration structures. As a result, h becomes small to properly represent the details. As the solute plume grows with time, resulting from a uniform injection in a heterogeneous medium, the particle support volume (scale and elongation) also grows with time to better represent the deformation of the solute bodies under advection-dominated transport.

The spatial variability of the net roughness controls the spatial variability of h_α within the solute plume in the LD approach. This can be seen by noticing that ϵ_h increases with σ_Y^2 and travel times. On the other hand, the dependency of ϵ_h on Pe_ℓ is complex, resulting from two sources of spatial variability; one attributed to the differences in local dispersion and another to the differences in solute plume stretching due to advection.

The elongations s_1 of the kernel function estimated with the GD and the LD methods show a clear systematic discrepancy. The LD method can identify the local longitudinal fingers of the solute plume, whereas the GD method can only describe the general behavior of the solute plume concentrations but not their details. For this reason, ϵ_s was particularly high at early stages of the plume deformation.

4.2. Estimation of Reaction Rates

We simulate solute transport with biodegradation through the previously generated $Y(\mathbf{x})$ field. The setup of the flow and transport problem is the same as before, but we consider two chemical species undergoing an irreversible bimolecular reaction of type A + B → C. Transport is described by the advection-dispersion equation with a degradation term,

$$\frac{\partial c_s}{\partial t} = -\mathbf{v}\nabla c_s + \nabla\cdot(\mathbf{D}\nabla c_s) - r, \quad s = \mathrm{A}, \mathrm{B}, \tag{42}$$

$$\frac{\partial c_C}{\partial t} = -\mathbf{v}\nabla c_C + \nabla\cdot(\mathbf{D}\nabla c_C) + r. \tag{43}$$

The initial solute distribution is now divided into two halves, each containing 100 mol (0.8 mol/m^3) of one of the reactants (A and B), to simulate a sharp interface at $t = 0$ (see Figure 6). Initially, the concentration of C is 0,

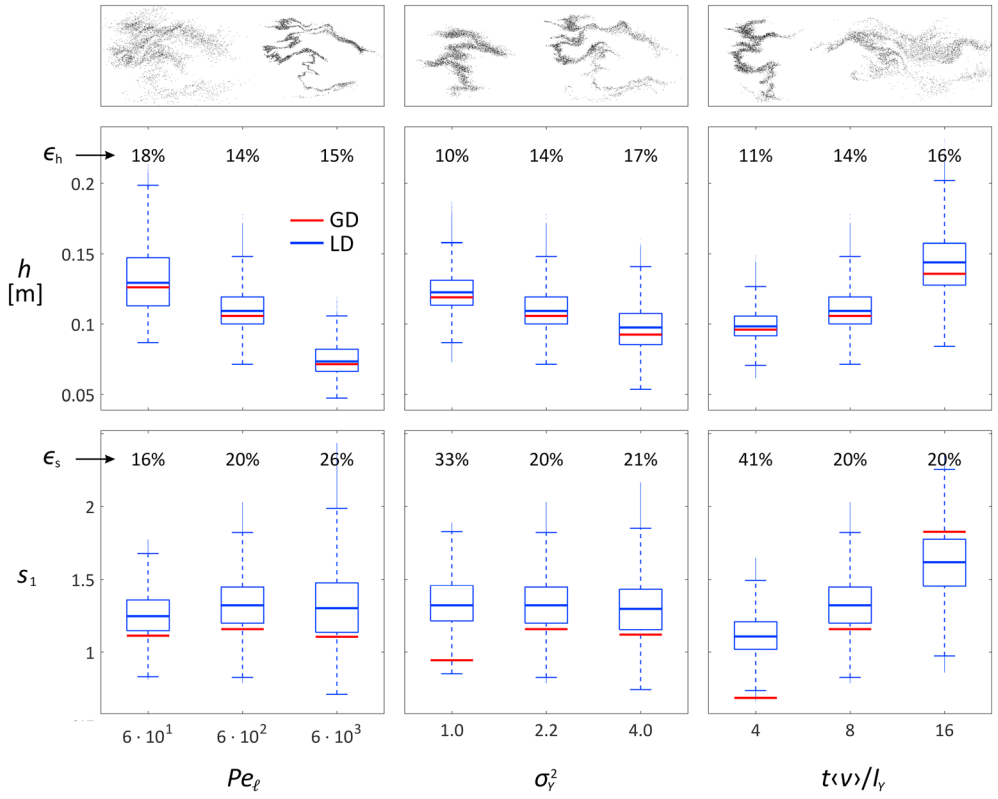

Figure 5. Box plots of the distribution of h and s_1 in the LD (blue) and GD (red) approaches for different problem conditions. Discrepancy indexes and the distribution of particles associated with extreme cases are also provided.
GD = Global Diagonal; LD = Local Diagonal.

and the same number of A and B particles are uniformly distributed with equal mass in both halves, that is, $N_A = N_B = N/2$. The total simulated time is 830 days. The reaction rate follows a double Monod kinetic model with limited conditions caused by simultaneously high concentrations of the electron donor (A) and the electron acceptor (B),

$$r = k_f \frac{c_A}{K_A + c_A} \frac{c_B}{K_B + c_B}, \tag{44}$$

where the subscripts A, B, and C denote the chemical species. We set the kinetic constants as K_A = 1.67 mol/m^3, K_B = 0.0156 mol/m^3, and k_f = 0.003 mol·m^{-3}·day^{-1}. For each time step Δt of the random walk simulation, the probability of reaction of a given particle A_α during the time interval $[t, t + \Delta t]$ is determined by

$$P_t\left(A_\alpha \rightarrow C_\gamma | \Delta t\right) = k_f \, \Delta t \, c_B\left(\mathbf{X}_\alpha^A, t\right) g\left(c_A\left(\mathbf{X}_\alpha^A, t\right), c_B\left(\mathbf{X}_\alpha^A, t\right)\right), \tag{45}$$

where $c_A\left(\mathbf{X}_\alpha^A, t\right)$ and $c_B\left(\mathbf{X}_\alpha^A, t\right)$ are the concentrations of the species A and B obtained at the particle position $\mathbf{X}_\alpha^A$ (here the chemical species is indicated as a superscript) at time t, estimated from the aforementioned methods, and g is the compensation function, written here as

$$g(c_A, c_B) = k_f \frac{1}{K_A + c_A} \frac{1}{K_B + c_B}. \tag{46}$$

Equation (45) provides an efficient simplification of the probability of reaction previously determined by Sole-Mari et al. (2017). Further details of this simplification are given in the Appendix A.

We note that significant changes in $\boldsymbol{\psi}_\alpha$ and n_α in (22) typically occur at a much slower pace than solute transport. Consequently, the optimization of the support volume (19) does not need to be performed at every time

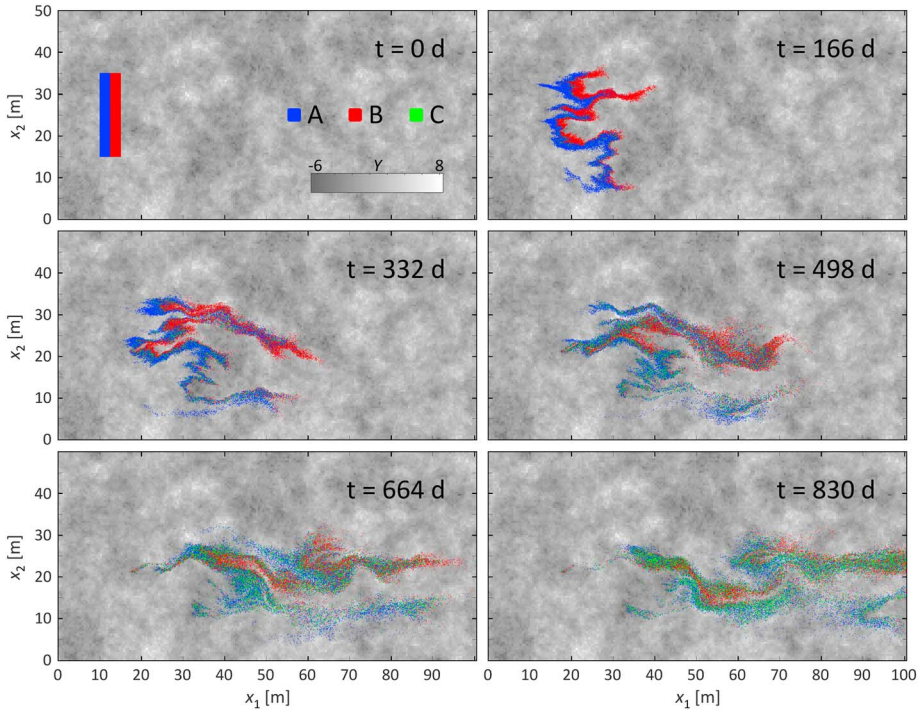

Figure 6. Particle plumes corresponding to species A, B, and C at different times in the reactive simulation addressed in section 4.2. In the background, the log-conductivity field $Y(\mathbf{x})$.

step. Here, in order to improve computational efficiency, we found that the following time frequency Δt_H for kernel optimization with $\gamma = 2.5$ provided accurate results,

$$\Delta t_H = \gamma \min\left(\frac{\Delta x}{v}, \frac{\Delta x^2}{2D}\right), \tag{47}$$

where v is the highest particle velocity norm and $D = a_\ell v$.

The subsequent analysis focuses on the estimation error of the probability of reaction. For this, we use the normalized reaction rate,

$$r_P = -\frac{1}{c_A}\frac{\partial c_A}{\partial t}, \tag{48}$$

This quantity can be estimated at each particle location $\mathbf{X}_\alpha^A$ at any given time t,

$$r_P \cong \frac{1}{\Delta t} P_t\left(A_\alpha \rightarrow C_\gamma | \Delta t\right), \tag{49}$$

which is directly proportional to the probability of reaction of an A-particle.

Then, we compare the time evolution of the cumulative distribution function of r_P obtained with the different estimation techniques as a function of N. A global measure of error is defined based on the Relative Root Integrated Squared Error of the reaction rate ($RRISE_r$):

$$RRISE_r = \left(\frac{\int_0^\infty \int_0^1 (\widehat{r}_P(\beta,t) - r_P(\beta,t))^2 d\beta dt}{\int_0^\infty \int_0^1 (r_P(\beta,t))^2 d\beta dt}\right)^{\frac{1}{2}}, \tag{50}$$

where $r_P(\beta, t)$ and $\widehat{r}_P(\beta, t)$ are, respectively, the reference and estimated r_P values for which the corresponding cumulative distribution of r_P obtained at time t is equal to β, for $0 < \beta < 1$. The value of $RRISE_r$ can be

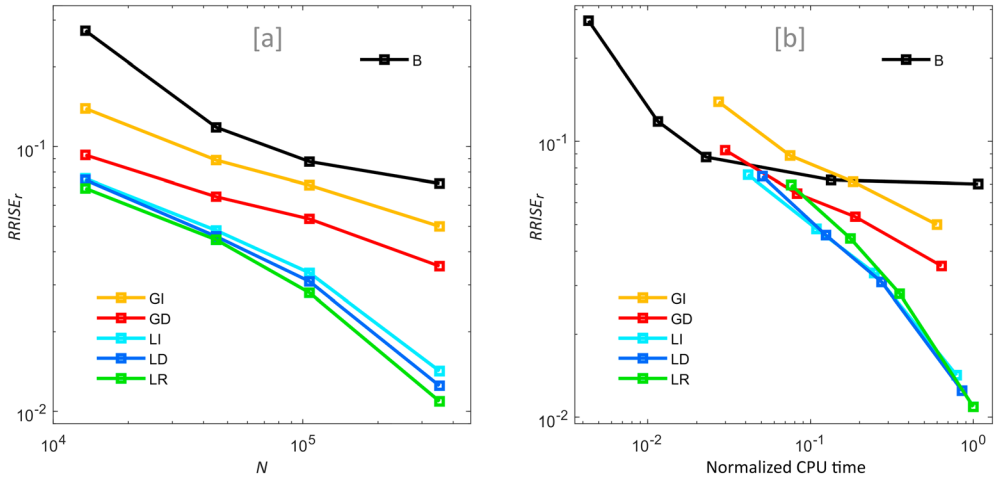

Figure 7. RRISE of the reaction rates as defined in (50), with respect to (a) the number of particles N and (b) the normalized CPU time (where "1" corresponds to the CPU time obtained using the LR method with 3.2×10^5 particles). The legend shows the acronyms of the tested approaches (see Figure 2). RRISE = Relative Root Integrated Squared Error.

understood as the relative standard error involved in the estimation of the particle reaction probabilities. Here the reference r_P values are calculated from the random walk solution obtained with $N = 1.6 \times 10^6$ and interpreted with the LR approach.

Figure 7 shows $RRISE_r$ as a function of N and the central processing unit (CPU) time. The local kernel approaches exhibited a substantial increase in accuracy and a faster convergence rate with N and CPU time compared with their global equivalents. As expected, the estimation error of the Binning method was the highest, displaying a decreasing rate of convergence with increasing N. As a result, the method became highly inefficient for simulating reactions. Importantly, for a typical number of particles in RWPT simulations ($N > 5 \times 10^4$), the presented local kernel approaches were computationally efficient. That is, the local kernel approaches required less CPU time to reach the same level of accuracy, because fewer particles were needed to obtain similar results.

5. Conclusions

The definition of the particle support volume is crucial for simulating reactive transport with Lagrangian methods as it dictates the interaction among particles. In this paper, we have presented a novel approach to select the optimal particle support volume, represented by the bandwidth matrix of a kernel function, in random walk models of reactive transport. In our method, each particle is equipped with a different support volume that locally adapts its shape and size based on the neighboring particle distribution. For this, we have adapted the classical principles of bandwidth selection in KDE theory to find the optimal particle support volume in a local window environment. The kernel bandwidth matrix $\mathbf{H}_\alpha$ defining the αth particle support has been decomposed into a rotation matrix $\mathbf{V}_\alpha$, a shape matrix $\mathbf{S}_\alpha$, and a scaling parameter h_α. Under different assumptions of $\mathbf{V}_\alpha$ and parametrizations of $\mathbf{S}_\alpha$, we have obtained closed-form expressions of h_α and $\mathbf{S}_\alpha$. In advection-dominated transport, we have proposed to align $\mathbf{V}_\alpha$ along the local velocities. Numerical simulations in a randomly heterogeneous porous medium have shown that this method results in a strong reconstruction of details, which is particularly important at the edges of the solute plumes where a small density of particles coexist with abrupt changes in concentrations. This gain has been mainly attributed to the local adaptability of the support volume, which can adjust its size and locally elongate its shape along the direction of minimum curvature where small changes in concentration gradients occur. These local kernel approaches have been demonstrated to substantially increase accuracy and improve convergence to the true solution with the number of particles. Importantly, in reactive transport RWPT simulations performed with a typical number of particles ($N > 10^4$), these local kernel methods have been shown to be computationally efficient, yielding better results than traditional histogram or global kernel methods for the same computational effort.

Appendix A: Kernel-Based Computation of Chemical Reactions in RWPT

In this appendix, we review and simplify the kernel-based random walk methodology for reactive transport presented in Sole-Mari et al. (2017). We consider an irreversible chemical reaction involving reactants A and B which turn into product C, with stoichiometric coefficients χ_A, χ_B, χ_C:

$$\chi_A A + \chi_B B \rightarrow \chi_C C, \tag{A1}$$

with a reaction rate $r(\mathbf{x}, t)$ that depends on the reactants' concentrations:

$$r(\mathbf{x},t) = \frac{1}{\chi_C}\frac{\partial c_C(\mathbf{x},t)}{\partial t} = k_f c_A(\mathbf{x},t) c_B(\mathbf{x},t) g(c_A(\mathbf{x},t), c_B(\mathbf{x},t)), \tag{A2}$$

where k_f is the forward reaction coefficient, and $g(c_A, c_B)$ is a function that describes the nonbilinearity of the reaction. In pure bimolecular reactions, $g = 1$. Sole-Mari et al. (2017) suggested that the expression for the probability of a reaction event for particle A_α at time t and for a time step Δt is

$$\begin{aligned} &P_t\left(A_\alpha \rightarrow C_\gamma | \Delta t\right) \\ &= \frac{\chi_A k_f}{\phi\left(\mathbf{X}^{AB}_{\alpha\beta}\right)} \Delta t \sum_{\beta=1}^{N_B} m^B_\beta W\left(\mathbf{X}^A_\alpha - \mathbf{X}^B_\beta, \mathbf{H}^A_\alpha + \mathbf{H}^B_\beta\right) g\left(c_A\left(\mathbf{X}^{AB}_{\alpha\beta}, t\right) c_B\left(\mathbf{X}^{AB}_{\alpha\beta}, t\right)\right), \end{aligned} \tag{A3}$$

where $\mathbf{H}^A_\alpha$ and $\mathbf{H}^B_\beta$ being the optimal kernel bandwidth matrices for A_α and B_β and

$$\mathbf{X}^{AB}_{\alpha\beta} = \left(\mathbf{H}^{A^{-1}}_\alpha + \mathbf{H}^{B^{-1}}_\beta\right)^{-1} \left(\mathbf{H}^{A^{-1}}_\alpha \mathbf{X}^A_\alpha + \mathbf{H}^{B^{-1}}_\beta \mathbf{X}^B_\beta\right). \tag{A4}$$

An analogous expression defines the probability of reaction of B_β. Values for $c_A\left(\mathbf{X}^{AB}_{\alpha\beta}, t\right)$ and $c_B\left(\mathbf{X}^{AB}_{\alpha\beta}, t\right)$ are obtained by KDE. In order to ensure deterministic compliance with the stoichiometry, as described in Sole-Mari et al. (2017), the probabilities of reaction are only estimated for reactant A. Then, upon occurrence of reaction for particle A_α, the closest B_β particle reacts with it. The mass assigned to the resulting C_γ particle must fulfill total mass conservation. As an alternative to stochastic particle annihilation (which is the approach used in this paper), expression (A3) can be used for the calculation of deterministic mass variations on particles as in Bolster et al. (2016), by computing:

$$\Delta m^A_\alpha = P\left(A_\alpha \rightarrow C_\gamma | \Delta t\right) m^A_\alpha. \tag{A5}$$

We refer to the original paper for the complete derivation of expression (A3) and for implementation details.

For this work, we simplify the method to improve efficiency by representing the A_α particle with a Dirac Delta kernel function (i.e., setting $\mathbf{H}^A_\alpha \rightarrow \mathbf{0}$) in (A3). This is equivalent to assuming that the position of particle A_α is deterministic in the reaction step. Then $\mathbf{X}^{AB}_{\alpha\beta} \rightarrow \mathbf{X}^A_\alpha$ and expression (A3) simplifies to

$$P_t\left(A_\alpha \rightarrow C_\gamma | \Delta t\right) = \chi_A\, k_f\, \Delta t\, c_B\left(\mathbf{X}^A_\alpha, t\right) g\left(c_A\left(\mathbf{X}^A_\alpha, t\right), c_B\left(\mathbf{X}^A_\alpha, t\right)\right). \tag{A6}$$

Appendix B: Derivation of the *AMISE* and *ALMISE*$_\alpha$ Expressions

In this appendix we derive expressions (10) and (16), corresponding to the Asymptotic Mean Integrated Squared Error and the Asymptotic Local Mean Integrated Squared Error, respectively. To do so, we first determine the Asymptotic Mean Squared Error $AMSE(\mathbf{x}, \mathbf{H})$ given a constant $\mathbf{H}$:

Starting from (1),

$$\widehat{p}(\mathbf{x}) = \frac{1}{N}\sum_{\alpha=1}^{N} W_H(\mathbf{x} - \mathbf{X}_\alpha, \mathbf{H}), \tag{B1}$$

with each $\mathbf{X}_\alpha$ being a sampled value from the true distribution $p(\mathbf{x})$. The expected value of $\widehat{p}(\mathbf{x})$ is

$$\begin{aligned} \mathrm{E}\{\widehat{p}(\mathbf{x})\} = \mathrm{E}\{W_H(\mathbf{x} - \mathbf{X}_\alpha, \mathbf{H})\} &= \int_{\mathbb{R}^d} W_H(\mathbf{x} - \mathbf{y}, \mathbf{H}) p(\mathbf{y}) \mathrm{d}\mathbf{y} = \\ &= \int_{\mathbb{R}^d} |\mathbf{H}|^{-\frac{1}{2}} W\left(\mathbf{H}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{y})\right) p(\mathbf{y}) \mathrm{d}\mathbf{y}. \end{aligned} \tag{B2}$$

We perform a change of variable with $\mathbf{u}\equiv\mathbf{H}^{-\frac{1}{2}}(\mathbf{y}-\mathbf{x})$:

$$\int_{\mathbb{R}^d}|\mathbf{H}|^{-\frac{1}{2}}W\left(\mathbf{H}^{-\frac{1}{2}}(\mathbf{x}-\mathbf{y})\right)p(\mathbf{y})\mathrm{d}\mathbf{y}=\int_{\mathbb{R}^d}|\mathbf{H}|^{-\frac{1}{2}}W(\mathbf{u})p\left(\mathbf{x}+\mathbf{H}^{\frac{1}{2}}\mathbf{u}\right)\left|\mathbf{J}_{\mathbf{u}}\left(\mathbf{H}^{\frac{1}{2}}\mathbf{u}\right)\right|\mathrm{d}\mathbf{u}= \tag{B3}$$
$$\int_{\mathbb{R}^d}W(\mathbf{u})p\left(\mathbf{x}+\mathbf{H}^{\frac{1}{2}}\mathbf{u}\right)\mathrm{d}\mathbf{u},$$

where $\left|\mathbf{J}_{\mathbf{u}}\left(\mathbf{H}^{\frac{1}{2}}\mathbf{u}\right)\right|=|\mathbf{H}|^{\frac{1}{2}}$ is the Jacobian determinant of $\mathbf{H}^{\frac{1}{2}}\mathbf{u}$ w.r.t. $\mathbf{u}$. Expanding around $\mathbf{x}$:

$$\int_{\mathbb{R}^d}W(\mathbf{u})p\left(\mathbf{x}+\mathbf{H}^{\frac{1}{2}}\mathbf{u}\right)\mathrm{d}\mathbf{u}= \tag{B4}$$
$$=\int_{\mathbb{R}^d}W(\mathbf{u})\left[p(\mathbf{x})+\frac{\partial p(\mathbf{x})}{\partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}\mathbf{u}+\frac{1}{2}\mathbf{u}^{\mathsf{T}}\mathbf{H}^{\frac{1}{2}}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\ \partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}\mathbf{u}+0_3\right]\mathrm{d}\mathbf{u}=$$
$$=p(\mathbf{x})+\frac{1}{2}\int_{\mathbb{R}^d}W(\mathbf{u})\mathbf{u}^{\mathsf{T}}\mathbf{H}^{\frac{1}{2}}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\ \partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}\mathbf{u}+0_4,$$

where O_k is the error associated with the truncation of the Taylor expansion terms of order $\geq k$. Therefore, we have the following asymptotic expression for $\mathrm{E}\{\widehat{p}(\mathbf{x})\}$ as $|\mathbf{H}|\rightarrow 0$:

$$\mathrm{E}\{\widehat{p}(\mathbf{x})\}\approx p(\mathbf{x})+\frac{1}{2}\int_{\mathbb{R}^d}W(\mathbf{s})\mathbf{u}^{\mathsf{T}}\mathbf{H}^{\frac{1}{2}}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}\mathbf{u}, \tag{B5}$$

and the estimation bias can be written as

$$\mathrm{Bias}\{\widehat{p}(\mathbf{x})\}=\mathrm{E}\{\widehat{p}(\mathbf{x})\}-p(\mathbf{x})\approx\frac{1}{2}\int_{\mathbb{R}^d}W(\mathbf{u})\mathbf{u}^{\mathsf{T}}\mathbf{H}^{\frac{1}{2}}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}\mathbf{u}\mathrm{d}\mathbf{u}. \tag{B6}$$

Defining

$$\mathbf{F}\equiv\mathbf{H}^{\frac{1}{2}}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}, \tag{B7}$$

we obtain

$$\frac{1}{2}\int_{\mathbb{R}^d}W(\mathbf{u})\mathbf{u}^{\mathsf{T}}\mathbf{H}^{\frac{1}{2}}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}\mathbf{u}\mathrm{d}\mathbf{u}=\frac{1}{2}\int_{\mathbb{R}^d}W(\mathbf{u})u_iF_{ij}u_j d\mathbf{u}= \tag{B8}$$
$$\frac{1}{2}F_{ij}\int_{\mathbb{R}^d}u_iu_jW(\mathbf{u})\mathrm{d}\mathbf{u}=\frac{1}{2}\mu_2(W)F_{ii}=\frac{1}{2}\mu_2(W)\mathrm{tr}(\mathbf{F}).$$

Repeated subindices in (B8) follow the Einstein notation. Introducing (B7) into (B8),

$$\frac{1}{2}\mu_2(W)\mathrm{tr}(\mathbf{F})=\frac{1}{2}\mu_2(W)\mathrm{tr}\left(\mathbf{H}^{\frac{1}{2}}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathsf{T}}}\mathbf{H}^{\frac{1}{2}}\right)=\frac{1}{2}\mu_2(W)\mathrm{tr}\left(\mathbf{H}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathsf{T}}}\right), \tag{B9}$$

we obtained an asymptotic expression of the estimation bias:

$$\mathrm{Bias}\{\widehat{p}(\mathbf{x})\}\approx\frac{1}{2}\mu_2(W)\mathrm{tr}\left(\mathbf{H}\frac{\partial^2 p(\mathbf{x})}{\partial\mathbf{x}\partial\mathbf{x}^{\mathsf{T}}}\right). \tag{B10}$$

Now, for the estimation variance, we have that

$$\mathrm{Var}\{\widehat{p}(\mathbf{x})\}=\frac{1}{N}\mathrm{Var}\{W_{\mathrm{H}}(\mathbf{x}-\mathbf{X}_\alpha,\mathbf{H})\}=\frac{1}{N}\left(\mathrm{E}\{W_{\mathrm{H}}^2(\mathbf{x}-\mathbf{X}_\alpha,\mathbf{H})\}-\mathrm{E}^2\{\widehat{p}(\mathbf{x})\}\right). \tag{B11}$$

Introducing (B5) into (B11), we have

$$\mathrm{Var}\{\widehat{p}(\mathbf{x})\}=\frac{1}{N}\int_{\mathbb{R}^d}W_{\mathrm{H}}^2(\mathbf{x}-\mathbf{y},\mathbf{H})p(\mathbf{y})\mathrm{d}\mathbf{y}-\frac{1}{N}(p(\mathbf{x})+O_2)^2. \tag{B12}$$

Performing a change of variable, $\mathbf{u} \equiv \mathbf{H}^{-\frac{1}{2}}(\mathbf{y}-\mathbf{x})$, on the integral in (B12):

$$\int_{\mathbb{R}^d} W_H^2(\mathbf{x}-\mathbf{y},\mathbf{H})p(\mathbf{y})\mathrm{d}\mathbf{y} = \int_{\mathbb{R}^d} |\mathbf{H}|^{-\frac{1}{2}} W^2(\mathbf{u}) p\left(\mathbf{x}+\mathbf{H}^{\frac{1}{2}}\mathbf{u}\right)\mathrm{d}\mathbf{u} = \tag{B13}$$
$$= \int_{\mathbb{R}^d} |\mathbf{H}|^{-\frac{1}{2}} W^2(\mathbf{u}) \left[p(\mathbf{x}) + \frac{\partial p(\mathbf{x})}{\partial \mathbf{x}^T}\mathbf{H}^{\frac{1}{2}}\mathbf{u} + 0_2\right]\mathrm{d}\mathbf{u} = |\mathbf{H}|^{-\frac{1}{2}} R(W) p(\mathbf{x}) + 0_2.$$

Introducing (B13) into (B12), we obtain

$$\mathrm{Var}\{\widehat{p}(\mathbf{x})\} = \frac{1}{N}\left(|\mathbf{H}|^{-\frac{1}{2}} R(W) p(\mathbf{x}) + O_2\right) - \frac{1}{N}(p(\mathbf{x}) + O_2)^2. \tag{B14}$$

As $|\mathbf{H}| \to 0$ and $N|\mathbf{H}|^{\frac{1}{2}} \to \infty$, the asymptotic expression for the estimation variance becomes

$$\mathrm{Var}\{\widehat{p}(\mathbf{x})\} \approx \frac{R(W)}{N|\mathbf{H}|^{1/2}} p(\mathbf{x}). \tag{B15}$$

Given that $MSE(\mathbf{x},\mathbf{H}) = \mathrm{Var}\{\widehat{p}(\mathbf{x})\} + \mathrm{Bias}^2\{\widehat{p}(\mathbf{x})\}$ and knowing (B10) and (B15), we can write the Asymptotic Mean Squared Error expression as

$$AMSE(\mathbf{x},\mathbf{H}) \approx \frac{R(W)}{N|\mathbf{H}|^{\frac{1}{2}}} p(\mathbf{x}) + \frac{1}{4}\mu_2^2(W)\mathrm{tr}^2\left(\mathbf{H}\frac{\partial^2 p(\mathbf{x})}{\partial \mathbf{x}\partial \mathbf{x}^T}\right). \tag{B16}$$

Integration over $\mathbf{x}$ yields the *AMISE* expression given in (10):

$$AMISE(\mathbf{H}) = \int_{\mathbb{R}^d} AMSE(\mathbf{x},\mathbf{H})\mathrm{d}\mathbf{x} = \frac{R(W)}{N|\mathbf{H}|^{\frac{1}{2}}} + \frac{1}{4}\mu_2^2(W)\int_{\mathbb{R}^d}\mathrm{tr}^2\left(\mathbf{H}\frac{\partial^2 p(\mathbf{x})}{\partial \mathbf{x}\partial \mathbf{x}^T}\right)\mathrm{d}\mathbf{x}. \tag{B17}$$

Assuming that $\mathbf{H}_\alpha$ does not change substantially over the local environment defined by $\mathbf{H}_0$, we can rewrite the weighted integration defined in (15) in terms of (B16) and integrate to obtain the expression given in (16):

$$ALMISE_\alpha(\mathbf{H}_\alpha) = \int_{\mathbb{R}^d} AMSE(\mathbf{x},\mathbf{H}_\alpha) W_H(\mathbf{x}-\mathbf{X}_\alpha,\mathbf{H}_0)\mathrm{d}\mathbf{x} = \tag{B18}$$
$$= \frac{1}{N^2}\left(n_\alpha(\rho) R(W)|\mathbf{H}_\alpha|^{-\frac{1}{2}} + \frac{1}{4}\mu_2^2(W) T_\alpha(\rho,\mathbf{H}_\alpha)\right),$$

where $\rho(\mathbf{x}) = Np(\mathbf{x})$ and

$$T_\alpha(\rho,\mathbf{H}_\alpha) = \int_{\mathbb{R}^d} W_\mathrm{H}(\mathbf{x}-\mathbf{X}_\alpha,\mathbf{H}_0)\ \mathrm{tr}^2\left(\mathbf{H}_\alpha\frac{\partial^2 \rho}{\partial \mathbf{x}\partial \mathbf{x}^T}\right)\mathrm{d}\mathbf{x}, \tag{B19}$$

$$n_\alpha(\rho) = \int_{\mathbb{R}^d} W_\mathrm{H}(\mathbf{x}-\mathbf{X}_\alpha,\mathbf{H}_0)\ \rho(\mathbf{x})\ \mathrm{d}\mathbf{x}. \tag{B20}$$

Appendix C: Derivation of S_α for the LD and LR case

In this appendix we derive expressions (26) and (27) given in section 3.1. Starting from expression (22):

$$ALMISE(h_\alpha, s_1, \ldots, s_{d-1}) \approx \frac{1}{N^2}\left(\frac{n_\alpha(\rho)\ R(W)}{h_\alpha^d} + \frac{h_\alpha^4}{4}\mu_2^2(W)\ T_\alpha(\rho,\mathbf{S}_\alpha)\right), \tag{C1}$$

where

$$T_\alpha(\rho,\mathbf{S}_\alpha) = \sum_{i=1}^{d}\sum_{j=1}^{d} s_i s_j \psi_{ij}(\rho), \tag{C2}$$

$$\psi_{ij}(\rho) = \int_{\mathbb{R}^d} W_\mathrm{H}(\mathbf{x}-\mathbf{X}_\alpha,\mathbf{H}_0)\ G_{ii}(\mathbf{x}) G_{jj}(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{C3}$$

By taking $\partial ALMISE/\partial h_\alpha = 0$ and $\partial ALMISE/\partial s_k = 0$ for $k = 1, \ldots, d-1$, we obtain the system of equations:

$$\frac{\partial ALMISE}{\partial h_\alpha} = \frac{1}{N^2}\left(-d\frac{n_\alpha(\rho)\, R(W)}{h_\alpha^{d+1}} + h_\alpha^3 \mu_2^2(W)\, T_\alpha(\rho, \mathbf{S}_\alpha)\right) = 0, \tag{C4}$$

$$\frac{\partial ALMISE}{\partial s_k} = \frac{h_\alpha^4\, \mu_2^2(W)}{2N^2}\left[\left(\sum_{i=1}^{d} s_i \psi_{ik}\right) - \frac{1}{s_k}\left(\prod_{i=1}^{d-1}\frac{1}{s_i}\right)\left(\sum_{i=1}^{d} s_i \psi_{id}\right)\right] = 0, \tag{C5}$$

which simplifies to

$$h_\alpha = \left(\frac{d\, R(W) n_\alpha(\rho)}{\mu_2^2(W) T_\alpha(\rho, \mathbf{S}_\alpha)}\right)^{\frac{1}{d+4}}, \tag{C6}$$

$$s_k \sum_{i=1}^{d} s_i \psi_{ik} = \left(\prod_{i=1}^{d-1}\frac{1}{s_i}\right)\sum_{i=1}^{d} s_i \psi_{id}, \quad \text{for } k = 1, \ldots, d-1. \tag{C7}$$

Expression (C6) is given in (25), and expression (C7) is handled differently for each value of $d > 1$.

Case $d = 2$.

In this case, (C7) can be rewritten as

$$s_1^2 \psi_{11} = \frac{1}{s_1^2}\psi_{22}. \tag{C8}$$

The unique positive solution is

$$s_1 = \left(\frac{\psi_{22}}{\psi_{11}}\right)^{\frac{1}{4}}, \quad s_2 = \left(\frac{\psi_{11}}{\psi_{22}}\right)^{\frac{1}{4}}, \tag{C9}$$

which is the one given in (26).

Case $d = 3$.

In this case, (C7) can be rewritten as

$$\begin{cases} s_1^2\psi_{11} + s_1 s_2 \psi_{12} = \dfrac{1}{s_1}\psi_{23} + \dfrac{1}{s_1^2 s_2^2}\psi_{33}, \\ s_2^2\psi_{22} + s_1 s_2 \psi_{12} = \dfrac{1}{s_2}\psi_{13} + \dfrac{1}{s_1^2 s_2^2}\psi_{33}. \end{cases} \tag{C10}$$

In principle, this system does not have a simple solution, but it becomes much easier to handle if we assume that the "cross" terms ψ_{ij}, $i \neq j$, have a much smaller influence on the resulting optimal bandwidth shape compared to ψ_{ii}. When that condition is fulfilled, we have

$$s_1 = \left(\frac{\psi_{22}\psi_{33}}{\psi_{11}^2}\right)^{\frac{1}{6}}, \quad s_2 = \left(\frac{\psi_{11}\psi_{33}}{\psi_{22}^2}\right)^{\frac{1}{6}}, \quad s_3 = \left(\frac{\psi_{11}\psi_{22}}{\psi_{33}^2}\right)^{\frac{1}{6}}. \tag{C11}$$

which is the result given in (27). Note that, while $\mathbf{S}_\alpha$ in this case may not be the exact optimal bandwidth shape, h_α will still be the exact optimal scale for that given $\mathbf{S}_\alpha$.

An alternative approach for $d = 3$ would be to parametrize $\mathbf{S}_\alpha$ with only one parameter s_1 (which implies choosing a preferential elongation direction), and then $s_2 = s_3 = s_1^{-1/2}$. It can be easily shown that the minimization of the $ALMISE_\alpha$ in this case yields

$$s_1 = \left(\frac{\left[8\psi_{11}(\psi_{22} + \psi_{33} + 2\psi_{23}) + (\psi_{12} + \psi_{13})^2\right]^{\frac{1}{2}} - (\psi_{12} + \psi_{13})}{4\psi_{11}}\right)^{\frac{2}{3}}, \tag{C12}$$

$$s_2 = s_3 = \left(\frac{4\psi_{11}}{\left[8\psi_{11}(\psi_{22} + \psi_{33} + 2\psi_{23}) + (\psi_{12} + \psi_{13})^2 \right]^{\frac{1}{2}} - (\psi_{12} + \psi_{13})} \right)^{\frac{1}{3}}. \tag{C13}$$

And substituting into (C2),

$$T_\alpha = \psi_{11} s_1^2 + 2(\psi_{12} + \psi_{13}) s_1^{\frac{1}{2}} + (\psi_{22} + \psi_{33} + 2\psi_{23}) s_1^{-1}. \tag{C14}$$

It is worth noting that, under the same assumption of (C12), that is, by neglecting the effect of ψ_{ij}, $i \neq j$, on the optimal shape matrix $\mathbf{S}_\alpha$, we would get a very similar expression for s_1:

$$s_1 = \left(\frac{\psi_{22} + \psi_{33}}{2\psi_{11}} \right)^{\frac{1}{3}}. \tag{C15}$$

Acknowledgments
Financial support was provided by the Spanish Ministry of Economy and Competitiveness, through projects WE-NEED, PCIN-2015-248; and INDEMNE, CGL2015-69768-R (MINECO/FEDER). G. S. acknowledges financial support by AGAUR. The paper provides all the information needed to replicate the results. The output data from the simulations are freely available from the authors upon request.

References

Abramson, I. S. (1982). On bandwidth variation in kernel estimates—A square root law. *The Annals of Statistics*, *10*(4), 1217–1223. https://doi.org/10.1214/aos/1176345986

Benson, D. A., Aquino, T., Bolster, D., Engdahl, N., Henri, C. V., & Fernàndez-Garcia, D. (2017). A comparison of Eulerian and Lagrangian transport and non-linear reaction algorithms. *Advances in Water Resources*, *99*, 15–37. https://doi.org/10.1016/j.advwatres.2016.11.003

Benson, D. A., & Meerschaert, M. M. (2008). Simulation of chemical reaction via particle tracking: Diffusion-limited versus thermodynamic rate-limited regimes. *Water Resources Research*, *44*, W12201. https://doi.org/10.1029/2008WR007111

Benson, D. A., & Meerschaert, M. M. (2009). A simple and efficient random walk solution of multi-rate mobile/immobile mass transport equations. *Advances in Water Resources*, *32*(4), 532–539. https://doi.org/10.1016/j.advwatres.2009.01.002

Berkowitz, B., Cortis, A., Dentz, M., & Scher, H. (2006). Modeling non-Fickian transport in geological formations as a continuous time random walk. *Reviews of Geophysics*, *44*, RG2003. https://doi.org/10.1029/2005RG000178

Bolster, D., de Anna, P., Benson, D. A., & Tartakovsky, A. M. (2012). Incomplete mixing and reactions with fractional dispersion. *Advances in Water Resources*, *37*, 86–93. https://doi.org/10.1016/J.ADVWATRES.2011.11.005

Bolster, D., Paster, A., & Benson, D. A. (2016). A particle number conserving Lagrangian method for mixing-driven reactive transport. *Water Resources Research*, *52*, 1518–1527. https://doi.org/10.1002/2015WR018310

Boso, F., Bellin, A., & Dumbser, M. (2013). Numerical simulations of solute transport in highly heterogeneous formations: A comparison of alternative numerical schemes. *Advances in Water Resources*, *52*, 178–189. https://doi.org/10.1016/j.advwatres.2012.08.006

Botev, Z. I., Grotowski, J. F., & Kroese, D. P. (2010). Kernel density estimation via diffusion. *The Annals of Statistics*, *38*(5), 2916–2957. https://doi.org/10.1214/10-AOS799

Breiman, L., Meisel, W., & Purcell, E. (1977). Variable kernel estimates of multivariate densities. *Technometrics*, *19*(2), 135–144. https://doi.org/10.2307/1268623

Cirpka, O. A., Chiogna, G., Rolle, M., & Bellin, A. (2015). Transverse mixing in three-dimensional nonstationary anisotropic heterogeneous porous media. *Water Resources Research*, *51*, 241–260. https://doi.org/10.1002/2014WR015331

Cirpka, O. A., Frind, E. O., & Helmig, R. (1999). Numerical simulation of biodegradation controlled by transverse mixing. *Journal of Contaminant Hydrology*, *40*(2), 159–182. https://doi.org/10.1016/S0169-7722(99)00044-3

Cvetkovic, V., & Haggerty, R. (2002). Transport with multiple-rate exchange in disordered media. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, *65*(5). https://doi.org/10.1103/PhysRevE.65.051308

Delay, F., & Bodin, J. (2001). Time domain random walk method to simulate transport by advection-dispersion and matrix diffusion in fracture networks. *Geophysical Research Letters*, *28*(21), 4051–4054. https://doi.org/10.1029/2001GL013698

Dentz, M., & Castro, A. (2009). Effective transport dynamics in porous media with heterogeneous retardation properties. *Geophysical Research Letters*, *36*, L03403. https://doi.org/10.1029/2008GL036846

Dentz, M., Le Borgne, T., Englert, A., & Bijeljic, B. (2011). Mixing, spreading and reaction in heterogeneous media: A brief review. *Journal of Contaminant Hydrology*, *120-121*, 1–17. https://doi.org/10.1016/j.jconhyd.2010.05.002

Ding, D., Benson, D. A., Fernàndez-Garcia, D., Henri, C. V., Hyndman, D. W., Phanikumar, M. S., & Bolster, D. (2017). Elimination of the reaction rate "scale effect": Application of the Lagrangian reactive particle-tracking method to simulate mixing-limited, field-scale biodegradation at the Schoolcraft (MI, USA) site. *Water Resources Research*, *53*, 10,411–10,432. https://doi.org/10.1002/2017WR021103

Edery, Y., Scher, H., & Berkowitz, B. (2009). Modeling bimolecular reactions and transport in porous media. *Geophysical Research Letters*, *36*, L02407. https://doi.org/10.1029/2008GL036381

Engdahl, N. B., Benson, D. A., & Bolster, D. (2017). Lagrangian simulation of mixing and reactions in complex geochemical systems. *Water Resources Research*, *53*, 3513–3522. https://doi.org/10.1002/2017WR020362

Fernàndez-Garcia, D., & Sanchez-Vila, X. (2011). Optimal reconstruction of concentrations, gradients and reaction rates from particle distributions. *Journal of Contaminant Hydrology*, *120–121*(C), 99–114. https://doi.org/10.1016/j.jconhyd.2010.05.001

Gramacki, A. (Ed.) (2018). *Bandwidth selectors for kernel density estimation BT—Nonparametric kernel density estimation and its computational aspects* (pp. 63–83). Cham: Springer International Publishing.

Gramling, C. M., Harvey, C. F., & Meigs, L. C. (2002). Reactive transport in porous media: A comparison of model prediction with laboratory visualization. *Environmental Science & Technology*, *36*(11), 2508–2514. https://doi.org/10.1021/es0157144

Hall, P. (1983). On near neighbour estimates of a multivariate density. *Journal of Multivariate Analysis*, *13*(1), 24–39. https://doi.org/10.1016/0047-259X(83)90003-9

Hansen, S. K., Scher, H., & Berkowitz, B. (2014). First-principles derivation of reactive transport modeling parameters for particle tracking and PDE approaches. *Advances in Water Resources*, *69*, 146–158. https://doi.org/10.1016/J.ADVWATRES.2014.04.007

Harbaugh, A. W. (2005). MODFLOW-2005, the U. S. Geological survey modular ground-water model—The ground-water flow process, U.S. Geol. Surv. Tech. Methods, 253. U.S. Geological Survey Techniques and Methods 6-A16.
Härdle, W. (1991). Kernel density estimation. In *Smoothing techniques: With implementation in S* (pp. 43–84). New York, NY: Springer. https://doi.org/10.1007/978-1-4612-4432-5_2
Henri, C. V., & Fernàndez-Garcia, D. (2014). Toward efficiency in heterogeneous multispecies reactive transport modeling: A particle-tracking solution for first-order network reactions. *Water Resources Research*, *50*, 7206–7230. https://doi.org/10.1002/2013WR014956
Henri, C. V., & Fernàndez-Garcia, D. (2015). A random walk solution for modeling solute transport with network reactions and multi-rate mass transfer in heterogeneous systems: Impact of biofilms. *Advances in Water Resources*, *86*, 119–132. https://doi.org/10.1016/j.advwatres.2015.09.028
Herrera, P. A., Massabó, M., & Beckie, R. D. (2009). A meshless method to simulate solute transport in heterogeneous porous media. *Advances in Water Resources*, *32*(3), 413–429. https://doi.org/10.1016/j.advwatres.2008.12.005
Huang, H., Hassan, A. E., & Hu, B. X. (2003). Monte Carlo study of conservative transport in heterogeneous dual-porosity media. *Journal of Hydrology*, *275*(3–4), 229–241. https://doi.org/10.1016/S0022-1694(03)00045-3
Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, *91*(433), 401–407. https://doi.org/10.1080/01621459.1996.10476701
LaBolle, E. M., Fogg, G. E., & Tompson, A. F. B. (1996). Random-walk simulation of transport in heterogeneous porous media: Local mass-conservation problem and implementation methods. *Water Resources Research*, *32*(3), 583–593. https://doi.org/10.1029/95WR03528
Lichtner, P., Satish, K., Hammond, G., Lu, C., Bisht, G., Kumar, J., et al. (2015). PFLOTRAN user manual: A massively parallel reactive flow and transport model for describing surface and subsurface processes. Technical Report.
Loftsgaarden, D. O., & Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, *36*(3), 1049–1051. https://doi.org/10.1214/aoms/1177700079
Mack, Y., & Rosenblatt, M. (1979). Multivariate k-nearest neighbor density estimates. *Journal of Multivariate Analysis*, *9*(1), 1–15. https://doi.org/10.1016/0047-259X(79)90065-4
Martinez-Landa, L., Carrera, J., Dentz, M., Fernàndez-Garcia, D., Nardí, A., & Saaltink, M. W. (2012). Mixing induced reactive transport in fractured crystalline rocks. *Applied Geochemistry*, *27*(2), 479–489. https://doi.org/10.1016/J.APGEOCHEM.2011.09.016
Panzeri, M., Riva, M., Guadagnini, A., & Neuman, S. P. (2016). Theory and generation of conditional, scalable sub-Gaussian random fields. *Water Resources Research*, *52*, 1746–1761. https://doi.org/10.1002/2015WR018348
Park, B. U., & Marron, J. S. (1990). Comparison of data-driven bandwidth selectors. *Journal of the American Statistical Association*, *85*(409), 66–72. https://doi.org/10.1080/01621459.1990.10475307
Paster, A., Bolster, D., & Benson, D. A. (2013). Particle tracking and the diffusion-reaction equation. *Water Resources Research*, *49*, 1–6. https://doi.org/10.1029/2012WR012444
Paster, A., Bolster, D., & Benson, D. A. (2014). Connecting the dots: Semi-analytical and random walk numerical solutions of the diffusion-reaction equation with stochastic initial conditions. *Journal of Computational Physics*, *263*, 91–112. https://doi.org/10.1016/j.jcp.2014.01.020
Pedretti, D., & Fernàndez-Garcia, D. (2013). An automatic locally-adaptive method to estimate heavily-tailed breakthrough curves from particle distributions. *Advances in Water Resources*, *59*, 52–65. https://doi.org/10.1016/j.advwatres.2013.05.006
Rahbaralam, M., Fernàndez-Garcia, D., & Sanchez-Vila, X. (2015). Do we really need a large number of particles to simulate bimolecular reactive transport with random walk methods? A kernel density estimation approach. *Journal of Computational Physics*, *303*, 95–104. https://doi.org/10.1016/j.jcp.2015.09.030
Riva, M., Neuman, S. P., & Guadagnini, A. (2015). New scaling model for variables and increments with heavy-tailed distributions. *Water Resources Research*, *51*, 4623–4634. https://doi.org/10.1002/2015WR016998
Riva, M., Panzeri, M., Guadagnini, A., & Neuman, S. P. (2015). Simulation and analysis of scalable non-Gaussian statistically anisotropic random functions. *Journal of Hydrology*, *531*, 88–95. https://doi.org/10.1016/J.JHYDROL.2015.06.066
Saaltink, M. W., Batlle, F., Ayora, C., Carrera, J., & Olivella, S. (2004). RETRASO, a code for modeling reactive transport in saturated and unsaturated porous media. *Acta Geologica*, *2*(3), 235–251. https://doi.org/10.1344/105.000001430
Sain, S. R. (2002). Multivariate locally adaptive density estimation. *Computational Statistics and Data Analysis*, *39*(2), 165–186. https://doi.org/10.1016/S0167-9473(01)00053-6
Sain, S. R., & Scott, D. W. (1996). On locally adaptive density estimation. *Journal of the American Statistical Association*, *91*(436), 1525–1534. https://doi.org/10.1080/01621459.1996.10476720
Salamon, P., Fernàndez-Garcia, D., & Gómez-Hernández, J. J. (2006a). Modeling mass transfer processes using random walk particle tracking. *Water Resources Research*, *42*, W11417. https://doi.org/10.1029/2006WR004927
Salamon, P., Fernàndez-Garcia, D., & Gómez-Hernández, J. J. (2006b). A review and numerical assessment of the random walk particle tracking method. *Journal of Contaminant Hydrology*, *87*(3–4), 277–305. https://doi.org/10.1016/j.jconhyd.2006.05.005
Sanchez-Vila, X., & Fernàndez-Garcia, D. (2016). Debates-stochastic subsurface hydrology from theory to practice: Why stochastic modeling has not yet permeated into practitioners? *Water Resources Research*, *52*, 9246–9258. https://doi.org/10.1002/2016WR019302
Schmidt, M. J., Pankavich, S., & Benson, D. A. (2017). A kernel-based Lagrangian method for imperfectly-mixed chemical reactions. *Journal of Computational Physics*, *336*, 288–307. https://doi.org/10.1016/j.jcp.2017.02.012
Sheather, S. J., & Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *53*(3), 683–690. https://doi.org/10.2307/2345597
Siirila-Woodburn, E. R., Fernàndez-Garcia, D., & Sanchez-Vila, X. (2015). Improving the accuracy of risk prediction from particle-based breakthrough curves reconstructed with kernel density estimators. *Water Resources Research*, *51*, 4574–4591. https://doi.org/10.1002/2014WR016394
Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, *Monographs on statistics and applied probability*. London: Chapman and Hall. https://doi.org/10.1007/978-1-4899-3324-9
De Simoni, M., Carrera, J., Sánchez-Vila, X., & Guadagnini, A. (2005). A procedure for the solution of multicomponent reactive transport problems. *Water Resources Research*, *41*, W11410. https://doi.org/10.1029/2005WR004056
De Simoni, M., Sanchez-Vila, X., Carrera, J., & Saaltink, M. W. (2007). A mixing ratios-based formulation for multicomponent reactive transport. *Water Resources Research*, *43*, W07419. https://doi.org/10.1029/2006WR005256
Sole-Mari, G., Fernàndez-Garcia, D., Rodríguez-Escales, P., & Sanchez-Vila, X. (2017). A KDE-based random walk method for modeling reactive transport with complex kinetics in porous media. *Water Resources Research*, *53*, 9019–9039. https://doi.org/10.1002/2017WR021064
Tartakovsky, A. M., & Meakin, P. (2005). A smoothed particle hydrodynamics model for miscible flow in three-dimensional fractures and the two-dimensional Rayleigh–Taylor instability. *Journal of Computational Physics*, *207*(2), 610–624. https://doi.org/10.1016/j.jcp.2005.02.001

Terrell, G. R., & Scott, D. W. (1992). Variable kernel density estimation. *The Annals of Statistics*, *20*(3), 1236–1265. https://doi.org/10.1214/aos/1176348768

Tompson, A. F. B. (1993). Numerical simulation of chemical migration in physically and chemically heterogeneous porous media. *Water Resources Research*, *29*(11), 3709–3726. https://doi.org/10.1029/93WR01526

Tompson, A. F. B., Schafer, A. L., & Smith, R. W. (1996). Impacts of physical and chemical heterogeneity on cocontaminant transport in a sandy porous medium. *Water Resources Research*, *32*(4), 801–818. https://doi.org/10.1029/95WR03733

Tsang, Y. W., & Tsang, C. F. (2001). A particle-tracking method for advective transport in fractures with diffusion into finite matrix blocks. *Water Resources Research*, *37*(3), 831–835. https://doi.org/10.1029/2000WR900367

Wand, M. P., & Jones, M. C. (1993). Comparison of Smoothing Parameterizations in Bivariate Kernel Density Estimation. *Journal of the American Statistical Association*, *88*(422), 520–528. Retrieved from http://www.jstor.org/stable/2290332

Willmann, M., Lanyon, G. W., Marschall, P., & Kinzelbach, W. (2013). A new stochastic particle-tracking approach for fractured sedimentary formations. *Water Resources Research*, *49*, 352–359. https://doi.org/10.1029/2012WR012191

Xu, T., Sonnenthal, E., Spycher, N., & Zheng, L. (2014). *TOUGHREACT V3.0-OMP reference manual: A parallel simulation program for non-isothermal multiphase geochemical reactive transport*. Berkeley: Lawrence Berkeley National Laboratory, University of California.

Yeh, G. T., Sun, J., Jardine, P. M., Burgos, W. D., Fang, Y., Li, M. H. & Siegel, M. D. (2004). HYDROGEOCHEM 5.0: A three-dimensional model of coupled fluid flow, thermal transport, and hydrogeochemical transport through variably saturated conditions. Version 5.0. Tennessee: National Laboratory Oak Ridge.

Zhang, Y., & Benson, D. A. (2008). Lagrangian simulation of multidimensional anomalous transport at the MADE site. *Geophysical Research Letters*, *35*, L07403. https://doi.org/10.1029/2008GL033222

Zheng, C., & Wang, P. P. (1999). MT3DMS. A modular three-dimensional multispecies transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater systems; documentation and user's guide. ed. U.A.C.o. Engineers.