AGU PUBLICATIONS

Water Resources Research

TECHNICAL REPORTS: METHODS

10.1002/2017WR021810

Key Points:

- We introduce a novel method to parameterize the SMM from breakthrough curve measurements
- Our methodology naturally allows us to calculate upper and lower bound matrices, allowing uncertainty quantification
- The SMM, parameterized using our method, accurately predicts breakthrough curves at multiple downstream locations

Correspondence to:
T. Sherman,
Thomas.J.Sherman.46@nd.edu

Citation:
Sherman, T., Fakhari, A., Miller, S., Singha, K., & Bolster, D. (2017). Parameterizing the spatial Markov model from breakthrough curve data alone. *Water Resources Research*, *53*, 10,888–10,898. https://doi.org/10.1002/2017WR021810

Received 4 SEP 2017
Accepted 18 NOV 2017
Accepted article online 27 NOV 2017
Published online 6 DEC 2017

Parameterizing the Spatial Markov Model From Breakthrough Curve Data Alone

Thomas Sherman[1], Abbas Fakhari[1], Savannah Miller[2], Kamini Singha[2], and Diogo Bolster[1]

[1]Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN, USA, [2]Hydrologic Science and Engineering, Colorado School of Mines, Golden, CO, USA

Abstract The spatial Markov model (SMM) is an upscaled Lagrangian model that effectively captures anomalous transport across a diverse range of hydrologic systems. The distinct feature of the SMM relative to other random walk models is that successive steps are correlated. To date, with some notable exceptions, the model has primarily been applied to data from high-resolution numerical simulations and correlation effects have been measured from simulated particle trajectories. In real systems such knowledge is practically unattainable and the best one might hope for is breakthrough curves (BTCs) at successive downstream locations. We introduce a novel methodology to quantify velocity correlation from BTC data alone. By discretizing two measured BTCs into a set of arrival times and developing an inverse model, we estimate velocity correlation, thereby enabling parameterization of the SMM in studies where detailed Lagrangian velocity statistics are unavailable. The proposed methodology is applied to two synthetic numerical problems, where we measure all details and thus test the veracity of the approach by comparison of estimated parameters with known simulated values. Our results suggest that our estimated transition probabilities agree with simulated values and using the SMM with this estimated parameterization accurately predicts BTCs downstream. Our methodology naturally allows for estimates of uncertainty by calculating lower and upper bounds of velocity correlation, enabling prediction of a range of BTCs. The measured BTCs fall within the range of predicted BTCs. This novel method to parameterize the SMM from BTC data alone is quite parsimonious, thereby widening the SMM's practical applicability.

1. Introduction

The heterogeneous structure of natural porous media, ranging from pore to aquifer scales (e.g., Cushman, 2013) leads to so-called anomalous transport, which is transport that cannot be adequately described in an upscaled manner by an effective Fickian advection-dispersion equation. Such anomalous transport is ubiquitous in hydrologic settings and has been observed in subsurface (e.g., Benson & Wheatcraft, 2001) and surface flow environments (e.g., Aubeneau et al., 2014; Haggerty et al., 2001). It is typically characterized by a broad range of spatial and/or temporal scales that must be represented in any modeling framework that is to be successful in replicating observed behaviors.

To this end, several anomalous transport frameworks have emerged (e.g., Benson et al., 2000; Berkowitz et al., 2006; Haggerty & Gorelick, 1995). Here we will focus on a model that belongs to the broader family of continuous time random walks (CTRW). Within a CTRW, a solute plume is discretized into a large number of particles, whose trajectories are represented by jumps that have distance and/or time increments randomly chosen from probability distributions. Most CTRW methods, as applied in hydrologic settings to date, assume particle jumps are independent and identically distributed. However, increasing evidence shows that at scales of interest, the independence assumption may be questionable and that correlation effects between successive jumps play an important role (Le Borgne et al., 2008a). In particular, particle trajectories display intermittency (de Anna et al., 2013; Kang et al., 2014), which can be observed as a particle's velocity and acceleration alternates between durations of high and low variability. This intermittent behavior cannot be adequately described by the aforementioned models.

One model that accounts for correlation is the spatial Markov model (SMM; Le Borgne et al., 2008a, 2008b). It describes a plume with a large number of discrete particles that transition through space and time following probabilistic rules; however, unlike other random walk models, the SMM fixes particle spatial jump size and then enforces correlations between successive transitions, a step essential to accurately capturing intermittency (de Anna et al., 2013) and which is particularly important for advection-dominated systems (Bolster et al., 2014). The SMM has been successful in modeling transport across a diverse range of synthetic systems of interest, including flows through highly heterogeneous permeability fields (Le Borgne et al., 2008a, 2008b), complex pore scale systems (Kang et al., 2014), inertially dominated systems (Bolster et al., 2014; N. Sund et al., 2015) as well as both synthetic and field-scale fractured media (Kang et al., 2011, 2015, 2016). Recently, it was also extended to predict reactive transport (N. Sund et al., 2017; Sund et al., 2016) as well as mixing and dilution effects (N. L. Sund et al., 2017).

A key ingredient of the SMM is the so-called transition matrix, which encodes correlation effects. Accurately parameterizing this is paramount to the model's success. In most studies to date, high-resolution particle-tracking methods enable direct measurement of the transition matrix by simulating particle trajectories. Note that in this paper, "measured" refers to simulated data. This particle-tracking approach is computationally cumbersome and, without simplification, restricts application of the SMM to numerically simulated systems. A more practical approach assumes a simplified diagonal structure for the transition matrix that depends on only one parameter, which can be fit to data. This was done by Kang et al. (2015) to successfully model field-scale push-pull data from an experimental fractured media experiment. While a very efficient and practical approach, the assumed structure of the transition matrix may be limiting. The strength of such an approach is parameterization of the SMM without the need of Lagrangian statistics from high-resolution numerical simulations and only the use of breakthrough curve (BTC) data, which in any practical context is the best one can hope for.

To detect correlation effects, one must at a minimum have two BTCs measured at different downstream locations. Here our goal is to propose an approach by which to parameterize the SMM using BTC data alone. We will do so within the context of synthetic numerical data for two idealized flows, which have previously been studied to test the SMM (N. Sund et al., 2015; Sund et al., 2016). The benefit of this is that we can know exactly what the actual transition matrix is and compare it to the estimates with our proposed method. Additionally, our method allows for the calculation of uncertainty in the transition matrix, which can be propagated through the SMM, enabling predictions of BTCs with quantifiable ranges of uncertainty. This paper is organized as follows. Section 2 describes the flow fields under consideration, the simulation framework for conservative transport, and the SMM as used in this study. Section 3 proposes a method for quantifying velocity correlation from two measured BTCs. Section 4 compares predicted BTCs using correlation as estimated by our approach with actual BTCs. Section 5 provides a summary and conclusions.

2. Methods

2.1. Model Setup

2.1.1. Flow Fields

In this study, we simulate conservative solute transport through two simple steady state laminar flows in 2-D. These are deliberately chosen as they have previously served as benchmarks for studying and validating the SMM (N. Sund et al., 2015; Sund et al., 2016). In those studies, the transition matrix for the SMM was measured by high-resolution direct numerical simulation of Lagrangian particle trajectories. Here we will estimate the transition matrix only through the use of BTC data. Thus, these simple flows serve as ideal conceptual systems that will act as proof of concept that velocity correlation can be quantified using BTC data alone.

The first system is a horizontal Poiseuille flow in a channel of dimensionless height, $H=1$, bounded by $0<y<H$ with unit mean velocity ($\bar{u}=\frac{1}{6}$). The velocity field is given by

$$u(y)=y(1-y). \tag{1}$$

For this flow, we consider transport in an advection-dominated system such that the Péclet number, Pe, equals 1,000. We define $\mathrm{Pe}=\frac{\bar{u}H}{D_m}$, where D_m is the molecular diffusion coefficient. This case is chosen because it has been shown that correlation effects are important (N. Sund et al., 2015).

The second system is Stokes flow through a periodic array of impermeable cylindrical obstacles (Figure 1). The flow field is pressure driven from left to right and is periodic in both x and y directions. The diameter (D) of each cylinder is chosen such that $D=0.4L_0$, where L_0 is the length of the periodic cell. For this problem

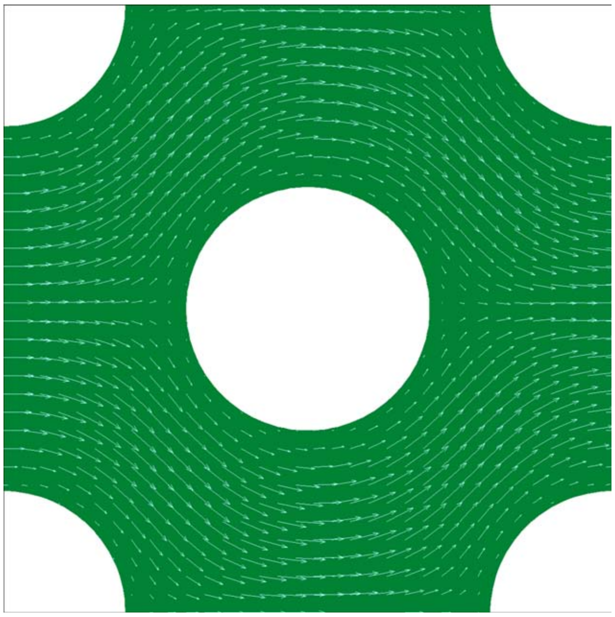

Figure 1. Schematic of the second system. Only one periodic computational cell is depicted.

setup, the Péclet number is defined as $\mathrm{Pe}=\frac{\bar{u}L_0}{D_m}$. We test two cases, Pe $=200$ and $\mathrm{Pe}=1,000$. The incompressible version of the lattice Boltzmann method (LBM) is employed to obtain the velocity field for this system (He & Luo, 1997).

2.1.2. Simulation Transport

We assume that transport in these systems is governed by the advection-diffusion equation, which we model using a standard Lagrangian random walk method. In all cases we impose a flux-weighted pulse initial condition at $x=0$. The solute plume is discretized into N particles, whose trajectories are defined by the Langevin equation:

$$\begin{aligned} x_i(t+\Delta t)&=x_i(t)+u_i\Delta t+\xi_i\sqrt{2D_m\Delta t} \\ y_i(t+\Delta t)&=y_i(t)+v_i\Delta t+\eta_i\sqrt{2D_m\Delta t}, \\ i&=1\ldots N \end{aligned} \tag{2}$$

where x_i and y_i are longitudinal and vertical positions, u_i and v_i are velocity components in the x and y directions, Δt is a fixed time step, and ξ and η are independent identically distributed random numbers sampled from a standard normal distribution. No flux boundary conditions at solid boundaries are enforced by elastic reflection. Note that N was chosen to be sufficiently large, such that results converged and noise effects were minimized. For systems 1 and 2, $N=10^6$ and 10^5, respectively.

2.2. Spatial Markov Model

2.2.1. Transport and Breakthrough Curves

Particle motion in the SMM is governed by the following Langevin equation:

$$x_{n+1}=x_n+L, \tag{3}$$

$$t_{n+1}=t_n+\tau_n. \tag{4}$$

At every model iteration, each particle jumps a fixed distance L with random times, τ_n values, which are chosen from a distribution, $\psi(\tau)$. The systems under consideration can be divided into spatially homogeneous cells of distance L. Hence, in each iteration, we model the time taken for a particle to pass through one cell. The time it takes each particle to reach the outlet of the nth cell is the sum of the travel times for each cell, $\sum_{i=1}^{n}\tau_i$. Tracking travel times for all particles gives the distribution of arrival times at the outlet of each cell and allows for modeling of BTCs, which show the temporal change in concentration of a solute at a fixed location. The SMM assumes that the distribution of travel times and correlation effects are stationary over each cell (N. L. Sund et al., 2015). Thus, for our two systems, we choose L sufficiently large, such that L is a characteristic jump length which satisfies the velocity field stationary requirement. Note that the chosen cell length for the Poiseuille and Stokes systems are $L=5H$ and $2.5D$, respectively.

If subsequent travel times, i.e., subsequent τ_n values, are independent random samples, then the SMM becomes an uncorrelated CTRW with fixed spatial increments. However, a strength of the SMM is that for each particle, τ_n is correlated to τ_{n-1}. In the systems under consideration, diffusion is the only mechanism by which a particle can change streamlines. Hence, a particle that traverses a cell on a fast (slow) streamline is more likely to stay on a fast (slow) streamline in the subsequent cell, particularly in a strongly advection-dominated flow (Bolster et al., 2014), such as in the $\mathrm{Pe}=1,000$ case. Correlating successive time steps enables the SMM to capture this behavior. In heterogeneous porous media, the intermittent pore structure and incompressibility of the fluid result in 3-D anomalous transport and similar persistence of particle velocities (Kang et al., 2014).

2.2.2. Transition Matrix

The probability transition matrix is the means by which the SMM accounts for correlation of travel times between neighboring cells. The distribution of travel times for each cell is sorted from fastest to slowest and divided into k classes of equal probability, i.e., class 1 contains the fastest and class k contains the slowest times. In this study, travel times are divided into 10 classes, which is estimated to be the minimum number of classes necessary to capture the entire range of correlation effects (Le Borgne et al., 2011; N. L. Sund et al., 2015). Each transition matrix element, $T_{i,j}$, is the probability that a particle has a travel time in class j in the nth cell, given its travel time was in class i in the $n-1$th cell.

$$T_{i,j}=P(\tau_{n+1} \in \text{class}\, j | \tau_n \in \text{class}\, i). \tag{5}$$

In numerical simulations, where each particle is tracked, each element in the transition matrix can be calculated directly from the Lagrangian particle statistics (Le Borgne et al., 2008b, 2011). For a large number of particles, a steady flow, and spatially periodic system, the transition matrix between any two subsequent cells is identical. Tracking particle travel times for only the first two cells enables calculation of the transition matrix, which results in knowledge of BTCs at every cell outlet in the system (e.g., Bolster et al., 2014; N. Sund et al., 2015). The transition matrices for both systems considered here, generated by applying this brute force method, are measured for comparison purposes (Figures 3 and 5). Details on the implementation of the transition matrix into the SMM and into (3) are available in many previous papers (e.g., Bolster et al., 2014; Le Borgne et al., 2011).

3. Estimating Correlation From BTCs Alone

Using the SMM is contingent on knowledge of velocity correlation. From a simplicity standpoint, this makes the uncorrelated CTRW more attractive, as one BTC provides enough information to create a *PDF* of arrival times and run the model. However, in highly correlated systems the uncorrelated CTRW fails to accurately predict BTCs downstream, making it necessary to account for correlation. Since correlation is a key ingredient of the SMM, two BTCs are needed to infer correlation strength. In this section, we introduce a novel method for estimating velocity correlation by relating BTCs measured at neighboring cells.

Our proposed methodology is valid when the assumptions of the SMM hold. These assumptions are that the distribution of travel times and correlation effects for each cell are stationary. The periodic geometries in our idealized conceptual steady flows satisfy these assumptions. However, periodicity is not required. The SMM had been successfully applied to a diverse range of systems, including nonperiodic highly heterogeneous porous media (Le Borgne et al., 2008a, 2008b) and field-scale fractured media (Kang et al., 2015). In systems that violate these assumptions, such as turbulent flows (N. Sund et al., 2015), an alternative model will be better suited. Note that our proposed method is valid when BTCs are measured at distances L and $2L$ from the particle source.

3.1. Finding PDFs of Arrival Times

Discretizing BTCs into a finite set of arrival times allows for calculation of a discrete probability distribution (Figure 2: Step 1). We measure two BTCs, which will be denoted *BTC*1 and *BTC*2. Note that the initial pulse particle injection occurs at $x=0$, and *BTC*1 and *BTC*2 are measured at $x=L$ and $x=2L$, respectively.

We divide each measured BTC into a set of finite arrival times, τ_i values, $i=1,\ldots,M$. In this study, M was approximately 500 for each discretized *BTC*. For each τ_i, its corresponding area of the BTC, A_i, can be

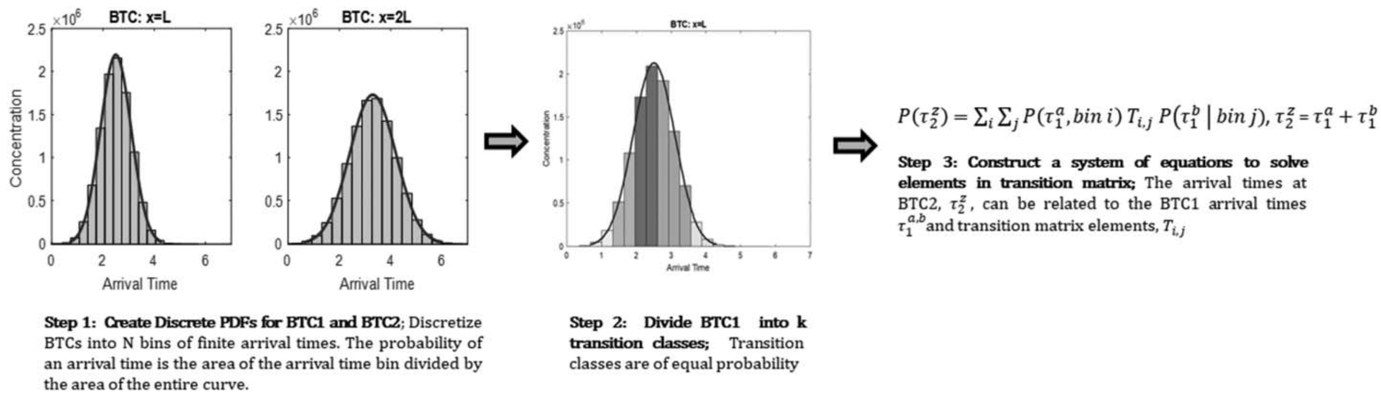

Figure 2. Schematic outlining the procedure to estimate the transition matrix.

expressed as the area under the BTC in the range, $[\frac{\tau_i-\tau_{i-1}}{2}, \frac{\tau_{i+1}-\tau_i}{2}]$. It follows that the probability of a particle arriving with time τ_i is the ratio of A_i to the area of the entire BTC:

$$P(\tau_i)=\frac{A_i}{\sum_{j=1}^{M} A_j}. \tag{6}$$

The procedure described above must be completed for both *BTC*1 and *BTC*2. The travel time distribution for cell 2 is also needed, but the spatial Markov model assumes the velocity statistics are stationary across cells. This implies the arrival time distribution from $x=0$ to $x=L$ is equivalent to the travel time distribution from $x=L$ to $x=2L$. *BTC*2 gives the arrival time distribution at $x=2L$, starting from $x=0$.

3.2. Classifying Arrival Times and Construction of a System of Equations

The transition matrix classifies particle arrival times from fastest to slowest and quantifies the probability that a particle changes classes after traversing successive spatial cells. Therefore, it is necessary to create time arrival classes. The distribution of arrival times can be sorted into equiprobable bins, where for k classes each bin has probability $\frac{1}{k}$ (Figure 2: Step 2). Bin 1 contains the fastest arrival times and bin k contains slowest arrival times. It is then possible to calculate the conditional probability, $P(\tau_i|\text{bin } k)$, of a τ_i arrival time in a given bin k.

The discretized *PDFs* obtained from *BTC*1 and *BTC*2 as well as the conditional probabilities for each arrival time and corresponding bin give sufficient information to estimate each element in the transition matrix. Recall that the *PDF* obtained for *BTC*2 represents the probabilities of times it takes a particle to travel from $x=0$ to $x=2L$. Additionally, the *PDF* of *BTC*1 is the probabilities of times it takes a particle to travel from $x=0$ to $x=L$, which assuming stationarity is equivalent to the distribution of times for a particle to travel from $x=L$ to $x=2L$. Let τ_2^z be an arrival time at $x=2L$, τ_1^a be an arrival time at $x=L$, and τ_1^b be the time it takes a particle to travel from $x=L$ to $x=2L$. Then the probability of a τ_2^z arrival time is the sum of all combinations of τ_1^a and τ_1^b such that $\tau_1^a+\tau_1^b=\tau_2^z$.

$$P(\tau_2^z)=\sum_{\tau_1^a+\tau_1^b=\tau_2^z} P(\text{cell } 1=\tau_1^a)P(\text{cell } 2=\tau_1^b|\text{cell } 1=\tau_1^a). \tag{7}$$

There is not enough information to solve equation (7) from the measurements of two BTCs as τ_1^b is dependent on τ_1^a. However, equation (7) can be written in terms of the elements of the transition matrix and probabilities of arrival times:

$$P(\tau_2^z)=\sum_i \sum_{j_{\tau_1^a+\tau_1^b=\tau_2^z}} P(\text{cell } 1=\tau_1^a, \text{bin } i)T_{i,j}P(\text{cell } 2=\tau_1^b|\text{bin } j). \tag{8}$$

Let *BTC*2 be discretized into Q points and recall that the transition matrix contains k classes. Then if $Q \geq k^2$, equation (8) can be used to construct a system of equations, where the number of equations is greater than or equal to the number of unknowns.

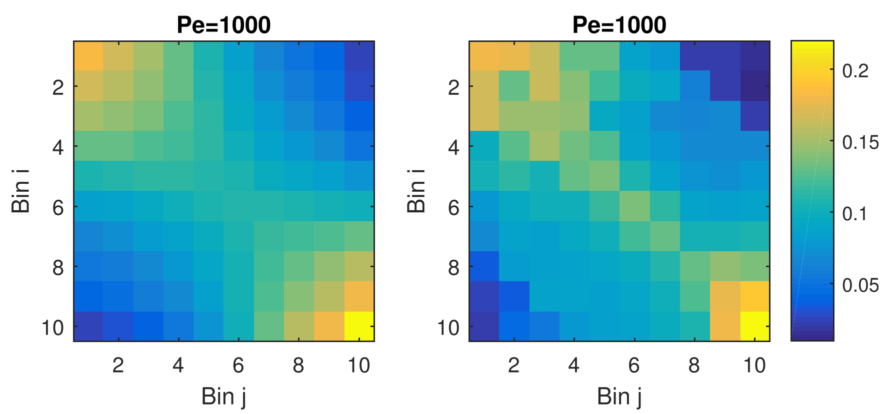

Figure 3. (left) The measured transition matrix calculated from Lagrangian statistics in the Poiseuille system. (right) The estimated transition matrix using the outlined estimation method.

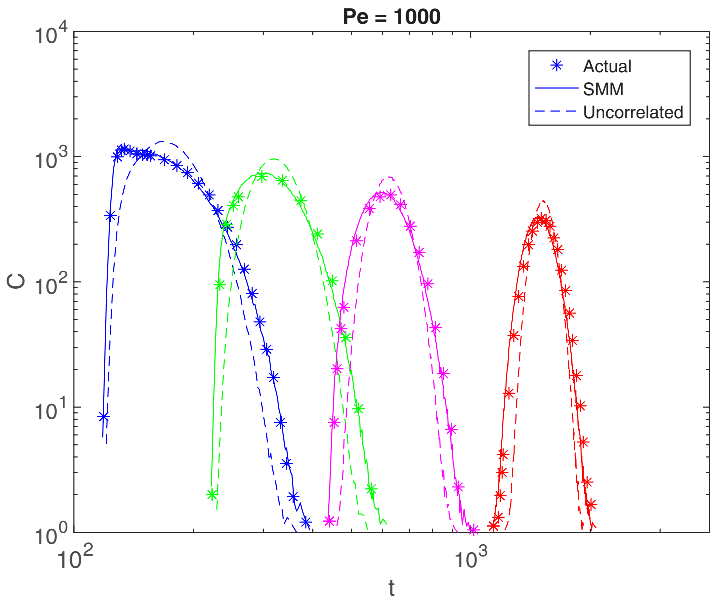

Figure 4. A comparison of predicted BTCs using an uncorrelated CTRW, predicted BTCs using the spatial Markov model with estimated correlation, and the actual BTCs at $x=5L$ (blue), $10L$ (green), $20L$ (magenta), and $50L$ (red) in the Poiseuille system.

$$\begin{pmatrix} m_{1,(1,1)} & \cdots & m_{1,(k,k)} \\ \vdots & \ddots & \vdots \\ m_{Q,(1,1)} & \cdots & m_{Q,(k,k)} \end{pmatrix} \times \begin{pmatrix} T_{1,1} \\ \vdots \\ T_{k,k} \end{pmatrix} = \begin{pmatrix} P(\tau_2^1) \\ \vdots \\ P(\tau_2^Q) \end{pmatrix}. \tag{9}$$

Here row z corresponds to the zth discretized arrival time on $BTC2$, where $z=1, Q$ is the fastest/slowest arrival times, respectively. Element $m_{z,(i,j)}=\sum_{\tau_1^a+\tau_1^b=\tau_2^z} P(\text{cell } 1=\tau_1^a, \text{bin } i)P(\text{cell } 2=\tau_1^b|\text{bin } j)$ can be interpreted as the contribution of probability for all the combinations of arrival times in class i after cell 1 that transition to class j in cell 2 and sum to τ_2^z. All elements in the probability contribution matrix are known, as well as the probability for each arrival time for $BTC2$, thus enabling the systems of equations to be solved and the transition matrix to be recovered.

3.3. Transition Matrix Statistics: Upper and Lower Bounds

In theory, as Q, the number of points $BTC2$ is discretized into, approaches infinity, the procedure outlined above will yield the actual transition matrix. However, discretizing $BTC2$ into a finite set induces error. In practice, only an estimation of the transition matrix can be achieved. As long as Q equals the number of elements in the transition matrix, then in principle we should be able to estimate correlation effects. For each τ_2^z, we can use equation (9) to write

$$P(\tau_2^z)=\sum_i^k \sum_j^k m_{z,(i,j)} T_{i,j}. \tag{10}$$

A least squares method is employed to minimize the $norm(P(\tau_2^z)-\sum_i^k \sum_j^k m_{z,(i,j)} T_{i,j})$, providing an estimate for each $T_{i,j}$ where $m_{z,(i,j)} > 0$. If $m_{z,(i,j)}=0$, then $T_{i,j}=0$ minimizes the *norm*. A $T_{i,j}$ estimated as zero indicates that no combination of class i and class j times can be combined to equal time τ_z^2. After iterating this process for all τ_z^2 arrival times, a range of estimated $T_{i,j}$ values is obtained for each matrix element. The number of estimates obtained for each element is equal to the number of discrete times in BTC2s *PDF*. Elements are typically overestimated when $m_{z,(i,j)}$ is orders of magnitude smaller than $P(\tau_z^2)$, as can be the case for late times on heavy tails. For cases such as these, alternative binning procedures such as logarithmic weighting may be preferable (e.g., Bolster et al., 2014; Le Borgne et al., 2011). For each $T_{i,j}$ set, we account for overestimates by filtering elements that exceed a threshold, 0.4 and 0.2 for Pe=1, 000 and 200 respectively, thereby allowing for accurate estimation of transition matrix values. These thresholds were determined as follows. A typical set of estimated $T_{i,j}$ values contains a subset of clearly wrong estimates that correspond to estimates made using small $m_{z,(i,j)}$ values in the probability contribution matrix, as well as a cluster of similar estimates that correspond to estimates made when $m_{z,(i,j)}$ values are large. By inspecting these estimated $T_{i,j}$ sets, we set a threshold that is greater than the values contained in the cluster of estimates, but smaller than values which are clearly artifacts due to

Table 1
The Mean Squared Error Is Calculated for All Estimated BTCs in Each Experiment

Mean squared error	Stokes: Pe = 200	Stokes: Pe = 1,000	Poiseuille: Pe = 1,000
Estimated SMM: 5*L*	56	305	905.2
Uncorrelated CTRW: 5*L*	365	11,806	2,899
Estimated SMM: 10*L*	46	78	1,010.4
Uncorrelated CTRW: 10*L*	248	4,227	6,879
Estimated SMM: 20*L*	42	27	1,137.2
Uncorrelated CTRW: 20*L*	165	1,456	15,906
Estimated SMM: 50*L*	38	11	1,580.4
Uncorrelated CTRW: 50*L*	104	1,301	41,733

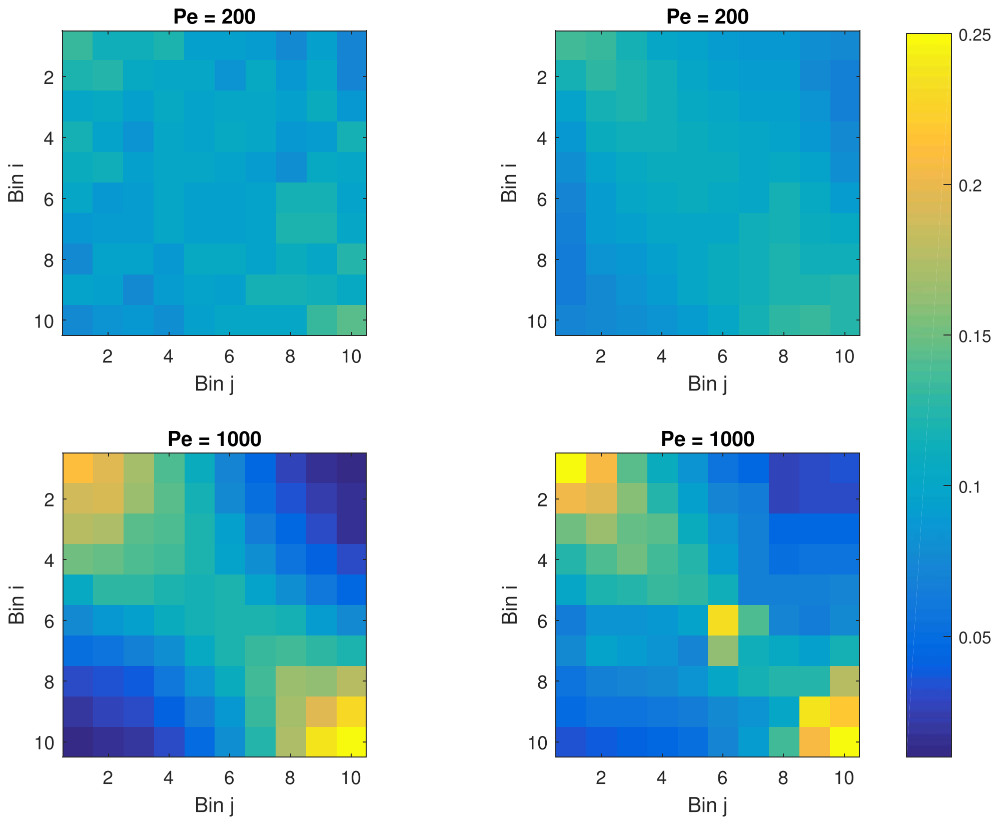

Figure 5. Actual versus estimated matrices in the Stokes system. (left) The measured transition matrices calculated using Lagrangian statistics. (right) Their estimated counterparts.

small $m_{z,(i,j)}$ values. This process thereby filters the estimated values down to the elements in which we have the most confidence, elements with large values in the probability contribution matrix.

For each $T_{i,j}$ set, the mean and standard deviation is calculated. This enables the construction of three matrices, a matrix with the mean estimated element values and matrices one standard deviation above and below the mean. Each matrix is then normalized so that the sum of each row equals 1. After this normalization procedure, the lower and upper bound matrices can be used to propagate uncertainty, yielding a range of concentrations for predicted BTCs rather than just a fixed value. This allows the spatial Markov model to be run while incorporating a degree of uncertainty, which may be useful from the perspective of risk assessment (Bolster et al., 2009; de Barros et al., 2011).

4. Results and Discussion

The proposed method for estimating the transition matrix is tested on simulation data from the Poiseuille and Stokes flow systems. BTCs are measured at $x=L$ and $x=2L$ and the mean estimated transition matrix is used to predict BTCs at cell outlets further downstream.

Table 2
Correlation Strength, IC, Is Calculated for the Measured and Estimated Transition Matrices Using $IC=\frac{\overline{T_{max}}-\overline{T_{min}}}{\overline{T_{max}}}$, *Where* $\overline{T_{max}}$ *Is the Average of the Maximum Values From Each Row*

Transition matrix correlation strength: IC	*Measured IC*	*Mean matrix IC*	*Lower matrix IC*	*Upper matrix IC*
Stokes: Pe = 200	0.3268	0.3664	0.2910	0.4880
Stokes: Pe = 1,000	0.8324	0.7889	0.6618	0.9406
Poiseuille: Pe = 1,000	0.6829	0.7423	0.6348	0.8776

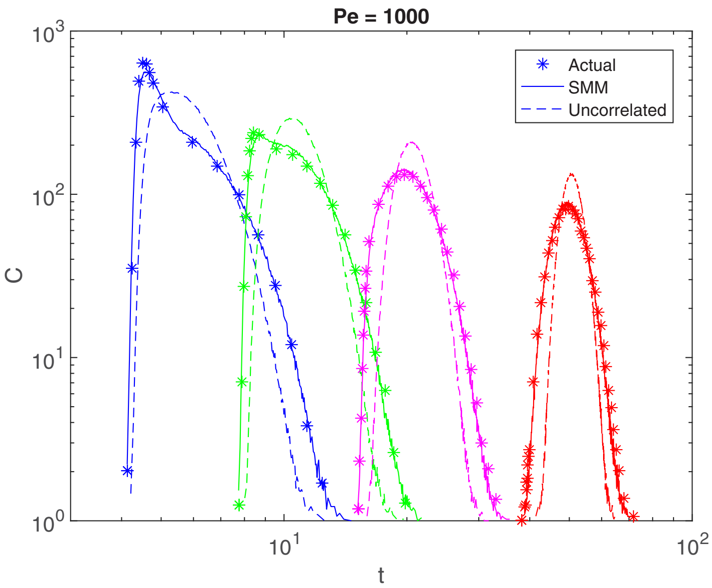

Figure 6. A comparison of predicted BTCs using an uncorrelated CTRW, predicted BTCs using the spatial Markov model with estimated correlation, and the actual BTCs at $x=5L$ (blue), $10L$ (green), $20L$ (magenta), and $50L$ (red) in the Stokes system with Pe = 1,000.

Figure 3 compares the estimated and actual transition matrices for the Poiseuille system with Pe$=1,000$. Notice the strong diagonal correlation and low probabilities in the upper right and lower left corners on the measured transition matrix. This indicates a particle's current travel time is highly dependent on its travel time in the previous cell. The estimated transition matrix captures this diagonally dominant feature. The estimated matrix overestimates the strength of correlation along the diagonal, resulting in underestimates of the lower left and top right corners. Despite these differences, the total correlation strength of the measured and estimated matrices is similar (Table 2). Running the SMM with the estimated matrix accurately predicts BTCs downstream (Figure 4) and outperforms an uncorrelated CTRW (Table 1). The earliest predicted arrival times are slightly faster than actual values due to the fact that transitions between fast arrival times (upper left corner of matrix) are overestimated.

Similar results were found in the Stokes flow system. Figure 5 shows the estimated and measured transition matrices for Pe$=1,000$, 200. Again, when Pe$=1,000$, there exists a strong diagonal correlation and low probabilities in the upper right and lower left corners on the measured transition matrix, which is captured in the estimated matrix. The diagonal elements as well as the upper right and lower left corners are slightly overestimated. However, the majority of estimated matrix elements, 79 out of 100, have less than a 2% error when compared to the actual measured values, which provides a comparable estimation of correlation strength (Table 2) for predicting downstream BTCs. Diffusive effects diminish the strong correlation in the Stokes system when Pe$=200$. This is clear by the decreased variance of values in the measured transition matrix, which also is found in the estimated matrix (Figure 5), as well as the decreased correlation strength (Table 2). Despite the decreased dominance of advection, the measured transition matrix still exhibits some correlation structure, notably with slightly higher transition matrix values in the upper left and lower right corners. In this low Pe number case, predicted BTCs are less sensitive to transition matrix values and so an uncorrelated CTRW predicts downstream BTCs with sufficient accuracy. This agrees with previous research that showed correlation effects can be neglected at Pe numbers of low order, Pe $< O(100)$ (Bolster et al., 2014).

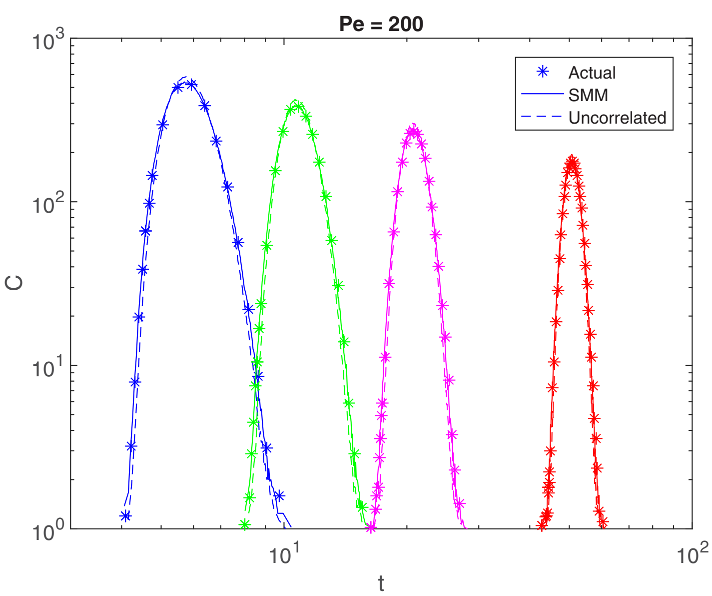

Figure 7. A comparison of predicted BTCs using an uncorrelated CTRW, predicted BTCs using the spatial Markov model with estimated correlation, and the actual BTCs at $x=5L$ (blue), $10L$ (green), $20L$ (magenta), and $50L$ (red) in the Stokes system with Pe = 200.

Figures 6 and 7 show predicted BTCs downstream using the estimated transition matrix in the Stokes flow system with Pe$=1,000$ and 200, respectively. At Pe$=1,000$, the SMM with the estimated transition matrix clearly outperforms the uncorrelated CTRW and accurately predicts tailing behavior (Table 1). The estimated SMM also correctly predicts inflection points at $x=20L, 50L$. The inflection points at $x=5L$, $10L$ are slightly underestimated, but they are more accurately predicted when compared with the uncorrelated model. As expected, the uncorrelated random walk method is able to accurately predict BTCs at Pe$=200$ (Figure 7). The SMM with the estimated transition matrix also accurately predicts BTCs. Thus, the proposed method to estimate the transition matrix captures BTC characteristics in both a strongly and weakly correlated system.

The proposed estimation method calculates a range of possible values for each element in the transition matrix because we calculate every value in the transition matrix for each discretized time in *BTC*2. We have demonstrated that in the systems under consideration, selecting the mean of this range enables us to capture the dominant matrix structure and accurately approximate values for the transition matrix. In addition to this mean matrix, we constructed matrices

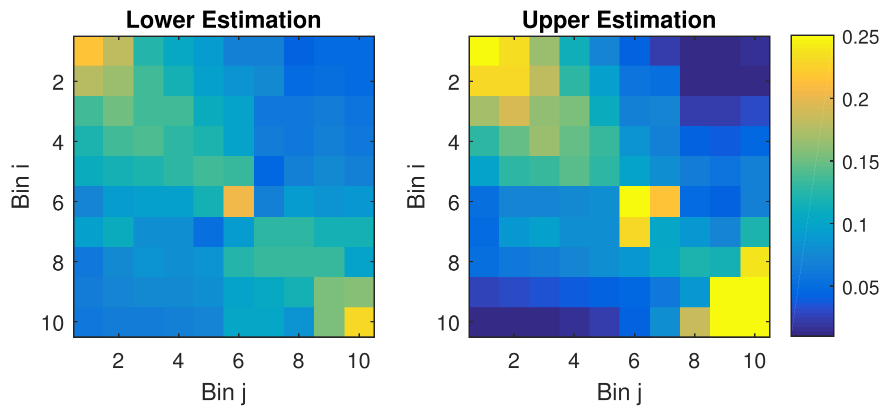

Figure 8. The (left) lower and (right) upper bound matrices for the Stokes system with Pe = 1,000.

using values one standard deviation above and below the mean (Figure 8). Note the lower transition matrix is less correlated, while the upper matrix contains stronger correlation than the mean (Figure 2). Thus, these matrices can be regarded as a lower and upper bound, and account for uncertainty.

In all our cases, the measured BTCs are contained within the predicted BTC range generated using the SMM with the upper and lower correlation matrices (Figure 9). Correlation strength is positively related to the range of arrival times for a given BTC. Consequently, the strongly correlated estimates will have the fastest and slowest arrival times for each predicted BTC. Interestingly, if the lower and upper uncertainty matrices are chosen such that they are truly bounds, then their respective predicted BTCs appear to intersect twice at points along the measured BTC. Note the uncorrelated matrix and identity matrix are the absolute extremum of possible lower and upper bounded matrices.

The lower and upper transition matrices in the SMM should represent the least and most correlated transition matrices that one might expect in a given system. This ensures that the actual BTC lies in the range of predicted BTCs. In this study, using 1 standard deviation to create upper and lower matrices created a

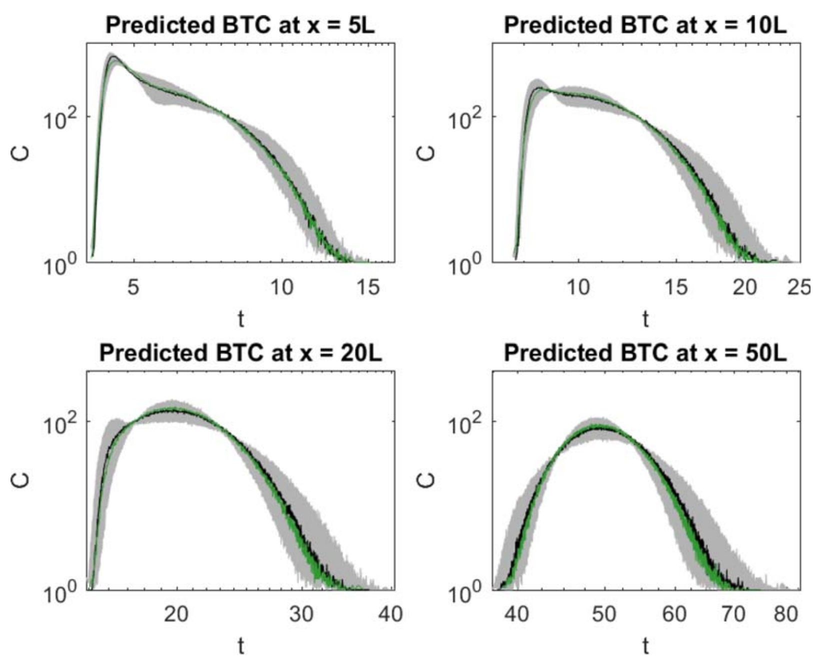

Figure 9. The predicted range of BTCs using the upper and lower bounded matrices in the Stokes system with Pe = 1,000 at (top left) $x=5L$, (top right) $x=10L$, (bottom left) $x=20L$, and (bottom right) $x=50L$. Gray is the predicted BTC range, black is the measured BTC, and green is the predicted BTC using the SMM with the mean correlation estimate.

predicted space that intersected the actual data. We envision that applying these upper and lower bounded transition matrices will be especially beneficial when predicting BTCs in laboratory and field systems, where errors in measurements will exist. Hence, in addition to enabling parameterization of the SMM using only BTC data, we have improved upon the current SMM framework by incorporating a scheme to classify uncertainty.

5. Conclusions

In this study, we answer our initial research question, "Does BTC data alone allow for parameterization of the SMM?" We demonstrate, through a novel theoretical mathematical framework, that it is in fact possible to calculate the transition matrix in stationary periodic flows using only data from two BTCs. Although in theory it is possible to recover the exact transition matrix, in practice, due to computational constraints and limited observational data, we achieve an estimation of the transition matrix by discretizing measured BTCs into finite sets of arrival times, and solving a system of equations that relates arrival time probabilities to transition matrix elements.

Our theoretical framework was put to the test using simulated experimental data. We considered two benchmark 2-D conceptual systems, a Poiseuille channel flow and a Stokes flow around an impermeable cylindrical obstacle, with $Pe=200, 1, 000$. Only BTC data measured at the outlet of the first two cells was used to estimate correlation statistics, parameterize the SMM, and predict BTCs downstream. Additionally, we estimated bounds with more highly and weakly correlated matrices, which were applied to predict a range of possible BTCs and account for uncertainty induced in the discretization process.

Results showed excellent agreement between predicted and actual BTCs, thereby validating the underlying approach for both weakly and strongly correlated systems. Specifically, the errors between predicted and measured BTCs were relatively small and in all cases actual BTCs fell within the range of the predicted range of uncertainty. Although we are only considering simple flows, the proposed framework can easily be applied to more complex systems that satisfy the SMM's underlying assumptions, thus improving our predictive power in modeling transport through field porous media systems and enhancing the applicability of the SMM beyond systems where Lagrangian statistics are calculated. A sample of the code used to implement our proposed method is openly shared and available for download via GitHub (see Acknowledgements section for link).

Acknowledgments
The code used to test our theoretical framework is made public at https://github:com/tjsherman24/SMM-Inverse-Model. The authors thank NSF for financial support for this work. This material is based upon work supported by, or in part by, NSF grants EAR-1351625, EAR-1417264, and EAR-1446236.

References

Aubeneau, A., Hanrahan, B., Bolster, D., & Tank, J. L. (2014). Substrate size and heterogeneity control anomalous transport in small streams. *Geophysical Research Letters*, *41*, 8335–8341. https://doi.org/10.1002/2014GL061838

Benson, D. A., & Wheatcraft, S. (2001). Fractional dispersion, Levy motion and the MADE tracer tests. In *Dispersion in heterogeneous geological formations* (Vol. 42, pp. 211–240). Dordrecht, the Netherlands: Kluwer Academic.

Benson, D. A., Wheatcraft, S. W., & Meerschaert, M. M. (2000). Application of a fractional advection-dispersion equation. *Water Resources Research*, *36*(6), 1403–1412.

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., Barahona, M., Dentz, M., Fernandez-Garcia, D., Sanchez-Vila, X., Trinchero, P., . . . Tartakovsky, D. M. (2009). Probabilistic risk analysis of groundwater remediation strategies. *Water Resources Research*, *45*, W06413. https://doi.org/10.1029/2008WR007551

Bolster, D., Méheust, Y., Borgne, T. L., Bouquain, J., & Davy, P. (2014). Modeling preasymptotic transport in flows with significant inertial and trapping effects—The importance of velocity correlations and a spatial Markov model. *Advances in Water Resources*, *70*, 89–103.

Cushman, J. H. (2013). *The physics of fluids in hierarchical porous media: Angstroms to miles* (Vol. 10). Berlin, Germany: Springer Science & Business Media.

de Anna, P., Le Borgne, T., Dentz, M., Tartakovsky, A. M., Bolster, D., & Davy, P. (2013). Flow intermittency, dispersion, and correlated continuous time random walks in porous media. *Physical Review Letters*, *110*(18), 184–502.

de Barros, F. P. J., Bolster, D., Sanchez-Vila, X., & Nowak, W. (2011). A divide and conquer approach to cope with uncertainty, human health risk, and decision making in contaminant hydrology. *Water Resources Research*, *47*, W05508. https://doi.org/10.1029/2010WR009954

Haggerty, R., Fleming, S. W., Meigs, L. C., & McKenna, S. A. (2001). Tracer tests in a fractured dolomite: 2. Analysis of mass transfer in single-well injection-withdrawal tests. *Water Resources Research*, *37*(5), 1129–1142.

Haggerty, R., & Gorelick, S. M. (1995). Multiple-rate mass transfer for modeling diffusion and surface reactions in media with pore-scale heterogeneity. *Water Resources Research*, *31*(10), 2383–2400.

He, X., & Luo, L.-S. (1997). A priori derivation of the lattice Boltzmann equation. *Physical Review E*, *55*(6), R6333.

Kang, P. K., Brown, S., & Juanes, R. (2016). Emergence of anomalous transport in stressed rough fractures. *Earth and Planetary Science Letters*, *454*, 46–54.

Kang, P. K., de Anna, P., Nunes, J. P., Bijeljic, B., Blunt, M. J., & Juanes, R. (2014). Pore-scale intermittent velocity structure underpinning anomalous transport through 3-D porous media. *Geophysical Research Letters*, *41*, 6184–6190. https://doi.org/10.1002/2014GL061475

Kang, P. K., Dentz, M., Le Borgne, T., & Juanes, R. (2011). Spatial Markov model of anomalous transport through random lattice networks. *Physical Review Letters*, *107*, 180–602. https://doi.org/10.1103/PhysRevLett.107.180602

Kang, P. K., Le Borgne, T., Dentz, M., Bour, O., & Juanes, R. (2015). Impact of velocity correlation and distribution on transport in fractured media: Field evidence and theoretical model. *Water Resources Research*, *51*, 940–959. https://doi.org/10.1002/2014WR015799

Le Borgne, T., Bolster, D., Dentz, M., de Anna, P., & Tartakovsky, A. M. (2011). Effective pore-scale dispersion upscaling with a correlated continuous time random walk approach. *Water Resources Research*, *47*, W12538. https://doi.org/10.1029/2011WR010457

Le Borgne, T., Dentz, M., & Carrera, J. (2008a). Lagrangian statistical model for transport in highly heterogeneous velocity fields. *Physical Review Letters*, *101*(9), 090601.

Le Borgne, T., Dentz, M., & Carrera, J. (2008b). Spatial Markov processes for modeling Lagrangian particle dynamics in heterogeneous porous media. *Physical Review E*, *78*, 026–308. https://doi.org/10.1103/PhysRevE.78.026308

Sund, N., Bolster, D., Mattis, S., & Dawson, C. (2015). Pre-asymptotic transport upscaling in inertial and unsteady flows through porous media. *Transport in Porous Media*, *109*(2), 411–432.

Sund, N., Porta, G., Bolster, D., & Parashar, R. (2017). A Lagrangian transport Eulerian reaction spatial (LATERS) Markov model for prediction of effective bimolecular reactive transport. *Water Resources Research*, *53*. https://doi.org/10.1002/2017WR020821

Sund, N. L., Bolster, D., & Benson, D. A. (2016). Testing the limits of the spatial Markov model for upscaling transport: The role of nonmonotonic effective velocity autocorrelations. *Physical Review E*, *94*(4), 43–107.

Sund, N. L., Bolster, D., & Dawson, C. (2015). Upscaling transport of a reacting solute through a peridocially converging–diverging channel at pre-asymptotic times. *Journal of Contaminant Hydrology*, *182*, 1–15.

Sund, N. L., Porta, G. M., & Bolster, D. (2017). Upscaling of dilution and mixing using a trajectory based spatial Markov random walk model in a periodic flow domain. *Advances in Water Resources*, *103*, 76–85.