Parameter estimation for fractional transport: A particle-tracking approach

Paramita Chakraborty,[1] Mark M. Meerschaert,[1] and Chae Young Lim[1]

Received 18 November 2008; revised 6 July 2009; accepted 13 July 2009; published 10 October 2009.

[1] Space-fractional advection-dispersion models provide attractive alternatives to the classical advection-dispersion equation for model applications that exhibit early arrivals and plume skewness. This paper develops a flexible method for estimating the parameters of the fractional transport model on the basis of spatial plume snapshots or temporal breakthrough curve data. A particle-tracking approach provides error bars for the parameter estimates and a general method for model fitting and comparison via optimal weighted least squares. A simple model of concentration variance, based on the particle-tracking approach, identifies the optimal weights.

Citation: Chakraborty, P., M. M. Meerschaert, and C. Y. Lim (2009), Parameter estimation for fractional transport: A particle-tracking approach, *Water Resour. Res.*, *45*, W10415, doi:10.1029/2008WR007577.

1. Introduction

[2] Non-Fickian transport of conservative solutes has been widely observed in laboratory and field data [*Benson et al.*, 2000a, 2000b, 2001; *Levy and Berkowitz*, 2003; *Bromly and Hinz*, 2004; *Klise et al.*, 2004]. The resulting anomalous dispersion is not well described by the classical second-order advection-dispersion equation (ADE) without extensive site characterization [*Zheng and Gorelick*, 2003]. A number of transient storage-based ADE models have been proposed for simulating the anomalous dispersion [*Bencala and Walters*, 1983; *Boano et al.*, 2007; *Deng and Jung*, 2009]. The space-fractional advection-dispersion equation (FADE) provides an attractive alternative that can represent plume skewness and early arrivals:

$$\frac{\partial C}{\partial t} = -v\frac{\partial C}{\partial x} + D\frac{1+\beta}{2}\frac{\partial^\alpha C}{\partial x^\alpha} + D\frac{1-\beta}{2}\frac{\partial^\alpha C}{\partial(-x)^\alpha}, \qquad (1)$$

where $C(x, t)$ (M/L^3) is tracer concentration, v (L/T) is the average plume velocity, D (L$^\alpha$/T) controls rate of spreading [*Benson et al.*, 2000b], β (dimensionless) is the skewness parameter ($-1 \le \beta \le 1$ with $\beta = 0$ for a symmetric plume), and the space-fractional index $1 < \alpha \le 2$ (dimensionless) codes the heterogeneity of the porous medium [*Clarke et al.*, 2005]. When $\alpha = 2$, (1) reduces to the classical ADE with constant parameters. The FADE (1) has been successful at modeling unsaturated transport [*Pachepsky et al.*, 2001; *Zhang et al.*, 2005], transport in saturated porous media [*Zhou and Selim*, 2003; *Chang et al.*, 2005; *Huang et al.*, 2006], and river flows [*Deng et al.*, 2004, 2006; *Zhang et al.*, 2005; *Kim and Kavvas*, 2006].

[3] This paper develops a general method of parameter estimation for the FADE parameters α, β, v, D from plume concentration data. Both spatial snapshots (observations of concentration $C(x, t)$ for t fixed and $x = x_1, \ldots, x_N$) and temporal breakthrough curves (measurements of $C(x, t)$ for x fixed and $t = t_1, \ldots, t_N$) are considered, since these are the data typically available. Naturally these data are contaminated by measurement error as well as model error (no model takes into account every source of variation). The FADE (and the ADE) relate to a specific stochastic process limit for a random walk model of particle movement [*Bhattacharya et al.*, 1976; *Zhang et al.*, 2006; *Chakraborty*, 2009]. Using a particle-tracking approach (Bhattacharya et al. use the term "pseudoparticle"), and equating the relative concentration of particles with the limit probability density, we can equate the measured concentration with a histogram consisting of the observed number of particles in each bin, where bin size is chosen to represent the volume sampled in a concentration measurement, and the number of particles is calibrated with plume roughness. Careful simulations (results not shown) were conducted to validate the fitting method for simulated plumes. Several illustrative applications were then fit to demonstrate the method: spatial snapshots from a tracer test at the MADE site in Mississippi; breakthrough curve data from tracer tests along the Grand River and the Red Cedar River in Michigan; and simulated ensemble snapshots from the Integrated Groundwater Modeling facility at Michigan State University. A recent review [*Neuman and Tartakovsky*, 2009] places the FADE in the context of modern stochastic theory, and highlights the relationships between alternative approaches. For example, the FADE may be considered as a type of nonlocal transport equation with convolution-Fickian flux [*Cushman and Ginn*, 2000] or a limit case of the continuous time random walk model [*Meerschaert and Scheffler*, 2004; *Berkowitz et al.*, 2006]. The model (1) is applicable under the ergodic hypothesis [*Gelhar*, 1993, *Dagan*, 1989] when particles sample enough aquifer heterogeneity so that the resulting plume resembles ensemble limit behavior, allowing the effective application of particle-tracking schemes. The general fitting methodology presented in this paper, based on an optimal weighted least squares approach, can also be adapted to alternative transport models in one or

[1]Department of Statistics and Probability, Michigan State University, East Lansing, Michigan, USA.

0043-1397/09/2008WR007577$09.00

several spatial dimensions. The optimal weights are determined via the particle-tracking approach, so the method is applicable to any model that admits a particle-tracking solution, where each particle moves independently of the other particles. Hence the model can include retention, a distributed source, and/or multiple interacting tracers (e.g., bacterial growth). Several example applications and model comparisons are included in section 4, where we use our general procedure to fit the FADE (1) to field and simulated data, and compare with alternative models.

2. Concentration Variance and Model Fitting

[4] This section describes a particle-tracking approach [*Bhattacharya et al.*, 1976] to estimate transport model parameters. The main technical issue is concentration variance. A stochastic process models the movement of particles, so that the relative concentration of particles approximates the probability density of the underlying stochastic process. The approach is quite general, and can be applied to any model that has a particle-tracking analogue. The main technical assumption is that each particle moves independently of the other particles. We highlight the FADE model (1) in this section, in order to focus the discussion. For the FADE with point source initial condition, the underlying stochastic process is a stable Lévy motion, a Markov process whose transition densities have no closed form in general, but can be efficiently computed by well established numerical methods. Extensions to more general initial/boundary conditions are possible, the only essential difference arises in the solution to the model equation. For example, any initial condition can be handled by convolution of the point source (Green's function) solution with the initial particle distribution.

[5] We begin with the problem of fitting to a spatial snapshot. Suppose that the measured concentration data at some fixed time $t > 0$ is given in the form (x_i, c_i) for $i = 1, \ldots, N$, where c_i is the observed concentration of particles at location x_i. Particles evolve independently according to the underlying stochastic process, and concentration is a histogram of particle location data, measured at locations $x_1, \ldots, x_N$. The underlying stochastic model implies that particle location is a random variable X_t whose probability density $f_{\boldsymbol{\theta}}(x, t)$ governs the chance of finding the particle in the histogram bin at location x. The notation explicitly acknowledges the unknown parameters $\boldsymbol{\theta}$ that determine the density function, since our goal is to estimate these parameters. The FADE parameter vector is $\boldsymbol{\theta} = (\alpha, \beta, \nu, \sigma)$ where the scale σ is related to the dispersion parameter D in (1) by $\sigma^\alpha = Dt|\cos(\pi\alpha/2)|$ [e.g., see *Benson et al.*, 2001; *Samorodnitsky and Taqqu*, 1994]. Model concentration is given by $C(x, t) = Kf_{\boldsymbol{\theta}}(x, t)$ where $K > 0$ is the total mass of the plume.

[6] Model parameter estimation is essentially a least squares curve fitting problem. We conceptualize the concentration data as a histogram, essentially a rescaled (by the total mass K) density estimate. For the FADE, the relevant $f_{\boldsymbol{\theta}}(x, t)$ is a stable density that has no closed form [*Samorodnitsky and Taqqu*, 1994]. The stable density represents the most general limit for sums of independent and statistically homogeneous particle movements, which explains its importance in stochastic process limits [*Feller*, 1971]. The simplest way to express the general stable density $f_{\boldsymbol{\theta}}(x, t)$ is in terms of its Fourier transform:

$$\int e^{ikx} f_{\boldsymbol{\theta}}(x,t)\,dx = \exp\bigl(i\mu k - \sigma^\alpha \omega_{\alpha,\beta}(k)\bigr), \tag{2}$$

where

$$\omega_{\alpha,\beta}(k) = \begin{cases} |k|^\alpha[1 - i\beta\,\mathrm{sign}(k)\tan(\pi\alpha/2)] & \text{for} \quad \alpha \neq 1, \\ |k|[1 + i\beta(2/\pi)\mathrm{sign}(k)\log|k|] & \text{for} \quad \alpha = 1, \end{cases}$$

where the tail index $0 < \alpha \leq 2$, skewness $-1 \leq \beta \leq 1$, center $-\infty < \mu < \infty$, and scale $\sigma \geq 0$. When $\alpha = 2$, the stable reduces to a normal density with mean μ and standard deviation $\sigma\sqrt{2}$. Codes for computing the stable density based on analytical inversion of the Fourier transform and numerical integration of the resulting formula are widely available [e.g., see *Nolan*, 1999].

[7] Model fitting is complicated by another factor as well. If we treat the fitting problem as a simple least squares optimization, we would seek the value of the parameter vector $\boldsymbol{\theta}$ and the total mass $K > 0$ that minimizes

$$\frac{1}{N}\sum_{i=1}^{N}\left(c_i - Kf_{\boldsymbol{\theta}}(x_i,t)\right)^2. \tag{3}$$

However, this assumes that the concentration variance is the same at each point x_i, which is far from the truth. Certainly the deviation between theoretical and measured concentration will not be the same at low and high concentrations. Statistically, this is a heteroscedastic nonlinear regression. The key to correctly fitting the parameters is to understand this error structure, and take it into account in a generalized least squares approach.

[8] A simple computation in Appendix A shows that the concentration variance in a particle-tracking setup is always proportional to the measured concentration. This result is universal, and does not depend on the choice of model. Hence any model that admits a particle-tracking solution, where each particle moves independently of the other particles, can be fit using this general procedure. In particular, our approach does not require the underlying stochastic process to be Markovian, nor does it require successive movements of any given particle to be uncorrelated. Since the variance of the measured concentration $c_i = \hat{C}(x_i, t)$ is proportional to model concentration $C(x_i,t)$, we expect measured and model concentration to agree more closely, in absolute terms, at lower concentrations. Hence a simple least squares fit is not appropriate. Instead, we will see that a weighted least squares is optimal. In Appendix A, we show that the optimal fitting procedure is to choose model parameters $\boldsymbol{\theta}$ and K to minimize the weighted mean square error function

$$e(\boldsymbol{\theta}, K) = \frac{1}{N}\sum_{i=1}^{N} w_i\left(c_i - Kf_{\boldsymbol{\theta}}(x_i,t)\right)^2, \tag{4}$$

where the weight $w_i = 1/(Kc_i)$ is essentially the reciprocal of the concentration variance. The procedure for fitting a breakthrough curve is essentially the same, except that the

data is measured concentration (c_i, t_i) for times $t_1, \ldots, t_N$ at one point x in space, and the space-varying model concentration $Kf_{\theta}(x_i, t)$ is replaced by its time-varying analogue $Kf_{\theta}(x, t_i)$. Appendix A also develops asymptotic expansions for the measured concentration and the estimated model parameters, leading to confidence bands that describe the quality of the approximations. Since the concentration variance is proportional to the concentration, low concentrations imply thin error bars. If this were not the case, the concentration error bars at low concentrations would allow negative observations, which makes no sense physically. Error bars for the model parameters depend on the first partial derivatives of the model concentration (for the FADE, this involves the stable density function), which can be computed numerically in practice. Details of the fitting procedure and error bar calculation are laid out in section 3. Finally, we point out that the procedure does not take measurement error into account. For concentrations well above the detection limit, this is a reasonable approximation. Concentrations below the detection limit should be set to zero. A zero concentration is ignored in the optimization, since we use the generalized inverse of the diagonal matrix Σ_{θ} in (A16).

[9] Concentration variance has been studied extensively in traditional stochastic hydrology. In that theory, concentration variance represents the variability of measured concentration at a fixed point in space, between different realizations of a random velocity field. Our approach represents a simplification of that model. *Fiori and Dagan* [2000] view particle motion as a Gaussian diffusion with drift, in a random velocity field. Our model differs in that the underlying diffusion is a non-Gaussian stable Lévy process, whose heavy tails also code the variations in the velocity field. This allows the effective use of a FADE (1) in which the constant average plume velocity v and fractal dispersivity D are constitutive parameters. Fiori and Dagan derive the explicit form of the observed concentration variance, which indicates that concentration variance is highest at the plume center [see also *Fiori and Dagan*, 2000, Figure 7]. *Hu and He* [2006] extend that result to account for kinetic sorption and local dispersion in a heterogeneous porous medium, and again they find that concentration variance is highest at high relative concentrations [*Hu and He*, 2006, Figures 1 and 2]. This is in agreement with our simplified model, in which the Gaussian pdf of concentration is approximated by a binomial, and a spatial snapshot is approximated by a histogram (particle-tracking approach).

[10] To further clarify the theory behind concentration variance, consider a typical contaminant transport simulation. A random hydraulic conductivity (K) field is generated, statistically consistent with the aquifer properties we wish to capture, and the corresponding velocity and dispersivity are computed based on relevant model assumptions, including appropriate initial and boundary conditions. If we track particles through a single aquifer realization, then each particle evolves independently of the other particles in that simulation, and the results of this paper are immediately applicable. If we take an ensemble average over a small number of realizations, this induces a correlation between particle movements, since two particles in the ensemble may undergo similar motion because they are following the same realization. This is the source of the two particle correlation in the classical theory of stochastic hydrology. If we take an ensemble average of a sufficiently large number of realizations, then an ergodic limit is achieved. If we run the simulation over a sufficiently long period of time, so that a typical particle samples all the heterogeneity of the aquifer, then again an ergodic limit is achieved. In both cases, the results of this paper are again applicable, since the simplifying assumption of particles that evolve independent of one another is close enough to reality to provide a simple, descriptive model of plume behavior. It is also important to point out that the subject of this paper is model fitting. It is not our intent to champion one model over another in this paper, but rather to develop statistically sound methods for fitting a given model to concentration data. Model error, a mismatch between the chosen model and the underlying physical reality, contributes another source of "noise" in measured concentration data, in that it contributes to the deviation of measured concentration values from a fitted curve derived from that model. We have not considered the affect of model error in our work.

[11] Following the work of *Cassiani et al.* [2005] on turbulent atmospheric flow, *Bellin and Tonina* [2007] model concentration itself as a Gaussian diffusion. They assume a quadratic dispersivity in their diffusion model, which leads to an analytical solution with concentration following a Beta pdf. They then fit a Beta pdf to point data and vertically averaged concentrations from the Cape Cod site, and obtain superior fits to a Gaussian or Lognormal. It is important to note that this Beta pdf represents the overall distribution of concentration, disregarding spatial location. Hence the Beta pdf fits a histogram with normalized concentration on the horizontal axis, and frequency on the vertical [*Bellin and Tonina*, 2007, Figures 2–6]. This is unlike the approach taken in this paper, where the FADE is fit to a histogram with spatial location on the horizontal axis, and concentration on the vertical. Hence the normal/binomial pdf of concentration at a fixed point in space, found in this paper as well as traditional stochastic hydrology, does not conflict with the findings of *Bellin and Tonina* [2007].

[12] An alternative method for fitting the FADE parameters (D. A. Benson, private communication, 2009) consists of minimizing mean square error on a log-log plot of the concentration data. This is equivalent to minimizing $N^{-1} \sum_{i=1}^{N} (\log c_i - \log C)^2$ where $C = Kf_{\theta}(x_i, t)$ is the model concentration. A Taylor expansion yields $\log c_i \approx \log C + (c_i - C)/C$ to first order. Then we are essentially minimizing $N^{-1} \sum_{i=1}^{N} W_i (c_i - C)^2$ where the weights $W_i = 1/C^2$ represent an improvement over unweighted least squares, but are not equivalent to the optimal weighting.

3. Parameter Estimation for the FADE Model

[13] This section describes a practical method for estimating coefficients in the space-fractional advection-dispersion model (1) based on laboratory or field data. We begin with the problem of fitting FADE parameters given a spatial snapshot of concentration data c_i collected at locations x_i for $i = 1, \cdots, N$ at a fixed time $t > 0$. The best fitting parameters are found by minimizing the weighted mean square error function (4) where $K > 0$ is the total mass and $f_{\theta}(x, t)$ is the stable density with parameters $\boldsymbol{\theta} = (\alpha, \beta, \nu, \sigma)$ where α, β are from the FADE (1), the plume center of mass

$\mu = vt$, and the scale σ is given by $\sigma^\alpha = Dt|\cos(\pi\alpha/2)|$. For a given value of $\boldsymbol{\theta}$, the optimal K can be computed from (A17). Hence we optimize (4) by an iterative two-step approach. Fix K and minimize (A17) as a function of the remaining parameters using standard nonlinear optimization software, and the method of *Chambers et al.* [1976] for simulating the stable density [see also *Nolan*, 1999]. Fix this value of $\boldsymbol{\theta}$ and compute K from (A17). Iterate until both estimates converge. The optimization problem is highly nonlinear with several local extrema, hence we advise a multistart approach where the two-step iterative method is applied to a grid of starting values representing a reasonable range of parameters, and the overall optimum is determined. Another reasonable approach is to fit the model by eye, and then use this as a starting value in the optimization. Typically the parameter range can be judged from the context of the problem. It would also be reasonable to employ more sophisticated nonlinear optimization methods such as simulated annealing or a genetic algorithm, but for the FADE fitting this was not attempted, as the multistart two-step nonlinear optimization method produced reasonable fits. Example fits to field data and numerical simulations will be shown in section 4.

[14] Next we consider the problem of parameter estimation for temporal breakthrough curve data, where plume concentration c_i is observed at one location x at different time points $t_1, \ldots, t_N$. As before, we conceptualize observed concentration as a histogram from an ensemble of particles, where the density estimate $c_i = \hat{C}(x, t_i)$ from (A1) approximates the model density $C(x, t_i) = Kf_{\boldsymbol{\theta}}(x, t_i)$. From this point of view, the observed concentration $\hat{C}(x, t)$ is a random variable whose mean is approximately equal to the model concentration $C(x, t)$ and whose variance is approximately equal to $KC(x, t_i)/(n\,dx)$. Detailed calculations in Appendix A show that the optimal fitting procedure is to minimize the weighted mean square error function

$$E(\boldsymbol{\theta}, K) = \frac{1}{N}\sum_{i=1}^{N} w_i (c_i - Kf_{\boldsymbol{\theta}}(x, t_i))^2 \qquad (5)$$

as a function of the FADE parameters $\boldsymbol{\theta} = (\alpha, \beta, v, \sigma)$ and the total mass K, where the weights $1/w_i = Kc_i$. This is completely analogous to the fitting procedure for spatial snapshots, the only difference being that the space-varying concentrations in (4) are replaced by their time-varying counterparts. Note this procedure also assumes no mass is lost over time, consistent with the FADE model for total concentration.

[15] The underlying statistical theory is complicated by correlations between c_i and c_j, which can be significant if the observation times t_i and t_j are closely spaced. This correlation between concentration measurements can occur due to correlation between movements of different particles, or between successive movements of an individual particle, but it can also occur in simple random walk models, since each particle moves only a limited distance in a very short period of time. The correlation becomes negligible if the time spacing is sufficiently wide so that the conditional probability of a single particle occupying the x bin at time t_j, given that the same particle occupies the same bin at time t_i, is approximately equal to the unconditional probability, essentially an ergodic assumption. This probability cannot be directly estimated from the concentration data, which contains no information about which particles are sampled. Now the computation of the parameter estimates is exactly the same as for spatial snapshot data, except that we substitute the objective function (5) in place of (4).

[16] The fitting procedure for other mass-preserving models, such as the CTRW or convolution flux model, is quite similar. All that is required is that: (1) the ergodic hypothesis is satisfied so that a particle-tracking model is applicable and (2) model concentrations $C(x, t)$ can be efficiently computed at each iteration of the optimization code. The same procedure could also be applied, in principle, for any partial differential equation model that admits a numerical solution as a function of its parameters, and a particle-tracking analogue. For example, in section 4.2 we fit a CTRW (time-fractional) model to breakthrough data for a river flow tracer test, using the same general procedure (5). The validity of this procedure for the CTRW model is supported by *Meerschaert and Scheffler* [2004] where the CTRW ergodic limit is proven to be a time-fractional ADE.

4. Application to Measured Plume Data

[17] This section demonstrates our parameter fitting methods in the case of the FADE model (1). The method was implemented using an R code (available from the authors) which takes advantage of a library function to compute the stable density.

4.1. Spatial Snapshots From a MADE Tracer Test

[18] Natural gradient tracer tests were conducted at the Macrodispersion Experimental (MADE) site at Columbus Air Force Base in northeastern Mississippi. The method of this paper was applied to the MADE-2 tritium plume data [*Boggs et al.*, 1993]. The data has four spatial snapshots at day 27, day 132, day 224, and day 328 days after injection, respectively. We consider the longitudinal distribution of total mass. The data points represent the maximum concentration measured in vertical slices perpendicular to the direction of plume travel, similar to Figure 4 of *Benson et al.* [2001].

[19] The resulting concentration curve (solid line) is shown against the measured concentration data in Figure 1. We attribute the lack of fit near the injection point to retention at the source, which is not captured by the FADE model (1) used in this study. See *Schumer et al.* [2003] for a modified FADE model that includes retention, which improves the fit. *Benson et al.* [2001] gave a range of values for α with the most favored value around $\alpha = 1.1$, in close agreement with our result of $\alpha = 1.0915$. The FADE presented by *Benson et al.* [2001] is limited to the case $\beta = 1$. Our estimate of the dimensionless skewness parameter $\beta = 0.99$ is sufficiently close to validate that assumption. The remaining parameter fits are $v = 0.196$ m/d, $D = 0.186$ m$^\alpha$/d, and $K = 56{,}778$ pCi/ml. Velocity v could also be estimated from the plume center of mass, which is located downstream of the peak concentration, since the plume is strongly skewed [*Benson et al.*, 2001]. The dashed lines in Figure 1 represent 95% confidence bands using equation (A15) where we calibrate $n\,dx_n = 300$ (number of particles times the length of the cube that defines the sampling volume) so that the confidence bands bracket all but a few of the concentration measurements. Note that the value for $n\,dx_n$ is purely

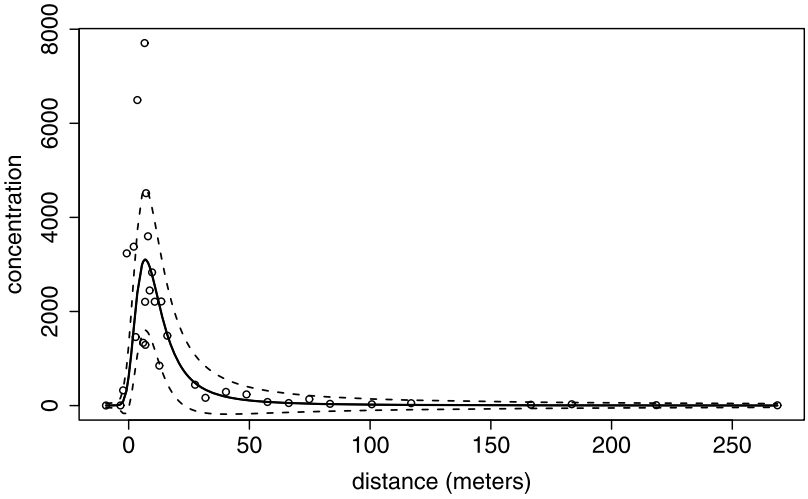

Figure 1. Concentration data and fitted FADE curve (solid line) for the MADE-2 tritium plume, day 224. The dashed lines represent 95% confidence bands for FADE concentration.

empirical, since there is no way to know the actual number of particles in the concentration measurements. Using the calibrated figure for $n\,dx_n$ we can compute confidence intervals for the FADE parameter estimates from equation (A14), numerically approximating the partial derivatives of the stable density. For example, the 95% interval for α is [1.08,1.11] which easily shows that α is statistically significantly less than two. The 95% interval for plume velocity v is [0.15,0.23] so that there is much more uncertainty in this parameter. Generally, the fractional parameters α and β can be estimated more precisely than v and D. Finally, we note that the parameters n and dx are known exactly in particle-tracking simulations. In that case, equation (A14) can be used directly to get confidence bands for the model parameters.

[20] The parameter fits for the MADE-2 tritium plume on day 224 shown in Figure 1 were then used to predict the later snapshot on day 328. Figure 2 shows that the upscaled FADE curve fits the day 328 data reasonably well, providing further evidence that the FADE is a useful model to predict plume evolution. See also the discussion by *Benson et al.* [2001].

[21] The physical derivation of the FADE in the work by *Schumer et al.* [2001] implies that the tail parameter $1 < \alpha \leq 2$, and we restricted to that range in our optimization. The solid line in Figure 3 shows the same FADE curve as in Figure 1, in log-log scale to accentuate the tail behavior. If we allow $0 < \alpha \leq 2$, which is the parameter range for the stable density, we find an alternative fit (dash-dotted line in Figure 3) with $\alpha = 0.70$ and $\beta = 1.0$. Certainly the tail fit seems superior. The development of *Schumer et al.* [2001] assumes a classical conservation of mass $\partial C/\partial t = -q$, along with a fractional flux $q = vC - D\partial^{\alpha-1} C/\partial x^{\alpha-1}$ that codes a power law distribution of particle jumps. Recent work by *Meerschaert et al.* [2006] outlines a more general approach that includes a fractional divergence (fractional conservation of mass). This theory extends the FADE to the full range of $0 < \alpha \leq 2$. Hence, while $\alpha = 0.70$ is outside the usual parameter range of the FADE (1), there is a well established physical basis that allows extending the model to $0 < \alpha \leq 2$. The weighted MSE (4) provides additional information. The fit in Figure 1 produced an optimal weighted MSE of 533 while the alternative fit in Figure 3 (dash-dotted line) gave 459. Thus

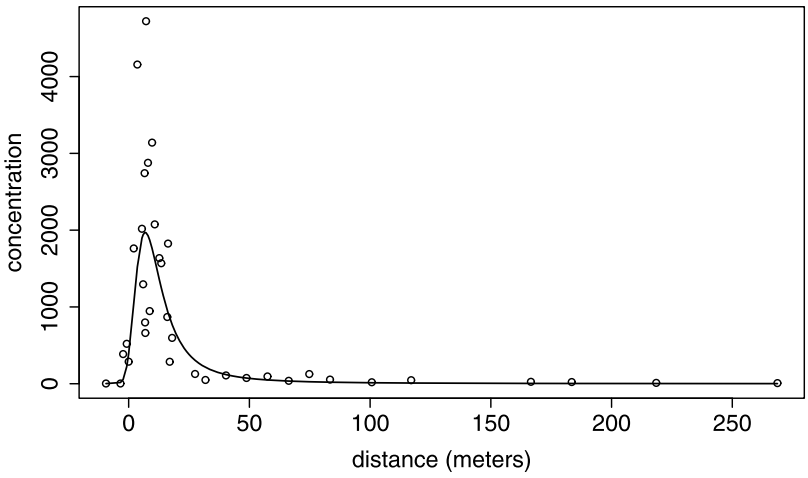

Figure 2. MADE-2 tritium plume, day 328, and FADE model prediction (solid line) using day 224 fitted parameter values.

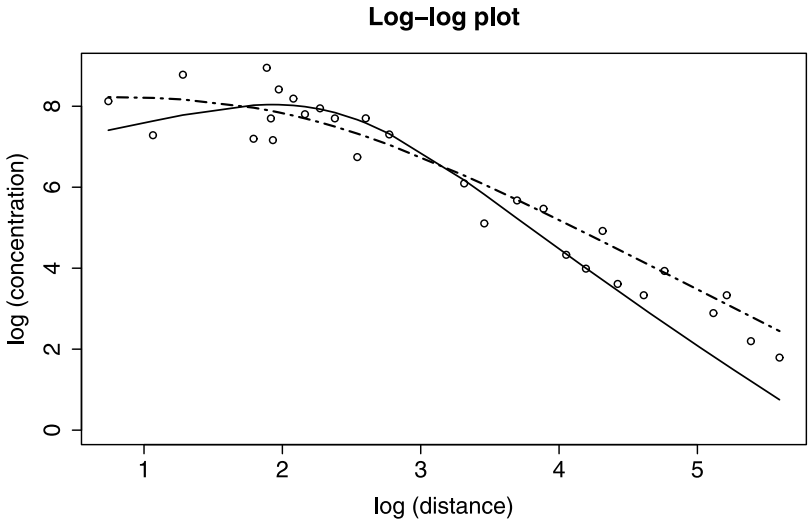

Figure 3. Log-log plot of concentration data and fitted FADE curve (solid line) for the MADE-2 tritium plume, day 224. The dash-dotted line shows an alternative stable fit with $\alpha < 1$.

the alternative FADE with $\alpha < 1$ gives a significantly better fit.

4.2. Breakthrough Curve for the Red Cedar River

[22] The methods of this paper were used to fit the FADE (1) to breakthrough curve data from a tracer test reported by *Phanikumar et al.* [2007]. The Red Cedar River is a fourth-order stream in south central Michigan, United States, that drains a landscape dominated by agriculture and urbanization. Four slug additions of fluorescein dye were released in the middle 75% of the channel to ensure conditions of instantaneous mixing. Sampling was done at the middle of the cross section for each dye release. The distances to the three sampling locations from the point of release were: 1.4 km, 3.1 km and 5.08 km from the injection site. Figure 4 illustrates the fit for the 3rd slug tracer sampled at all three sampling locations. The parameter fits for distance 1.4 km are $\alpha = 1.32$, $\beta = -0.99$, $v = 0.022$ km/min, $D = 0.00181$ km$^{\alpha}$/min, and $K = 22.64$ μg/L. For distance 3.1 km we get $\alpha = 1.56$, $\beta = -0.99$, $v = 0.026$ km/min, $D = 0.00131$ km$^{\alpha}$/min, and $K = 25.48$ μg/L. At distance 5.08 km we find $\alpha = 1.58$, $\beta = -0.96$, $v = 0.029$, $D = 0.00181$ km$^{\alpha}$/min, $K = 27.75$ μg/L. Note that the parameter values are fairly consistent between the three sampling sites, and the fitted values of $\alpha < 1.6$ shows that the FADE is appropriate as an alternative to the ADE (special case $\alpha = 2$) for this data. Note also that there is no evidence of mass loss downstream.

[23] *Deng et al.* [2004, 2006] report positive skewness $\beta = 1$ in their fits to similar river flow tracer tests. We believe that this is due to a small discrepancy between our parameterization and the one used by Deng et al. There are two competing parameterizations for the stable density, and a skewness of -1 in the work of *Zolotarev* [1986] corresponds to a skewness of +1 in the work by *Samorodnitsky and Taqqu* [1994]. The underlying random walk model [*Meerschaert et al.*, 1999] clarifies the meaning of the skewness parameter. Particles that jump downstream contribute to the positive skewness, and particle that jump upstream add to the negative skewness. Hence the positive skewness that predominates in the groundwater models of *Benson et al.* [2001], *Schumer et al.* [2001], and *Huang et al.* [2006] captures early arrivals caused by preferential flow paths. Negative skewness codes particle retention in dead zones or eddies. *Deng et al.* [2004] state that the fractional derivative in their river flow model captures a "wide spectrum of dead zones," and it is apparent from the fitted breakthrough curves in Figure 5 of *Deng et al.* [2004] that their model has a heavy tail at late time, consistent with $\beta = -1$ in the notation of our paper. The later work of *Deng et al.* [2006] adds retention to their model, but the late time tails in the breakthrough curves of Figures 4, 5, and 6 of *Deng et al.* [2006] indicate that heavy tails and negative skewness is also operating there. Related work by *Hunt* [2006, p. 88] gives a nice description of the fractional model as a way to capture "the effect of dead-zone regions in the river." Another application of the FADE appeared in work by *Xiong et al.* [2006], who fit the FADE to laboratory column data. Although the skewness was not reported there, we have learned (G. Huang, private communication, 2009) that a skewness of $\beta = -1$ was used, in the notation of our paper. It seems that the FADE with negative skewness is modeling retention. A physical explanation for the FADE model of river flows that allows negative skewness was recently accomplished by *Kim and Kavvas* [2006].

[24] *Xiong et al.* [2006] compare the FADE to an asymptotic CTRW model with retention. That CTRW model imposes power law waiting times between particle jumps, hence it coincides exactly with the time-fractional advection dispersion equation

$$\frac{\partial^{\gamma} C}{\partial t^{\gamma}} = -v\frac{\partial C}{\partial x} + D\frac{\partial^2 C}{\partial x^2}, \tag{6}$$

where the Caputo fractional derivative is used. The mathematical equivalence between the time-fractional ADE and the CTRW scaling limit is well known [*Meerschaert et al.*, 2002; *Berkowitz et al.*, 2006]. *Xiong et al.* [2006] find that the CTRW (time-fractional) model gives a better fit to their column data than the negatively skewed space-fractional ADE, or the traditional ADE, although the FADE

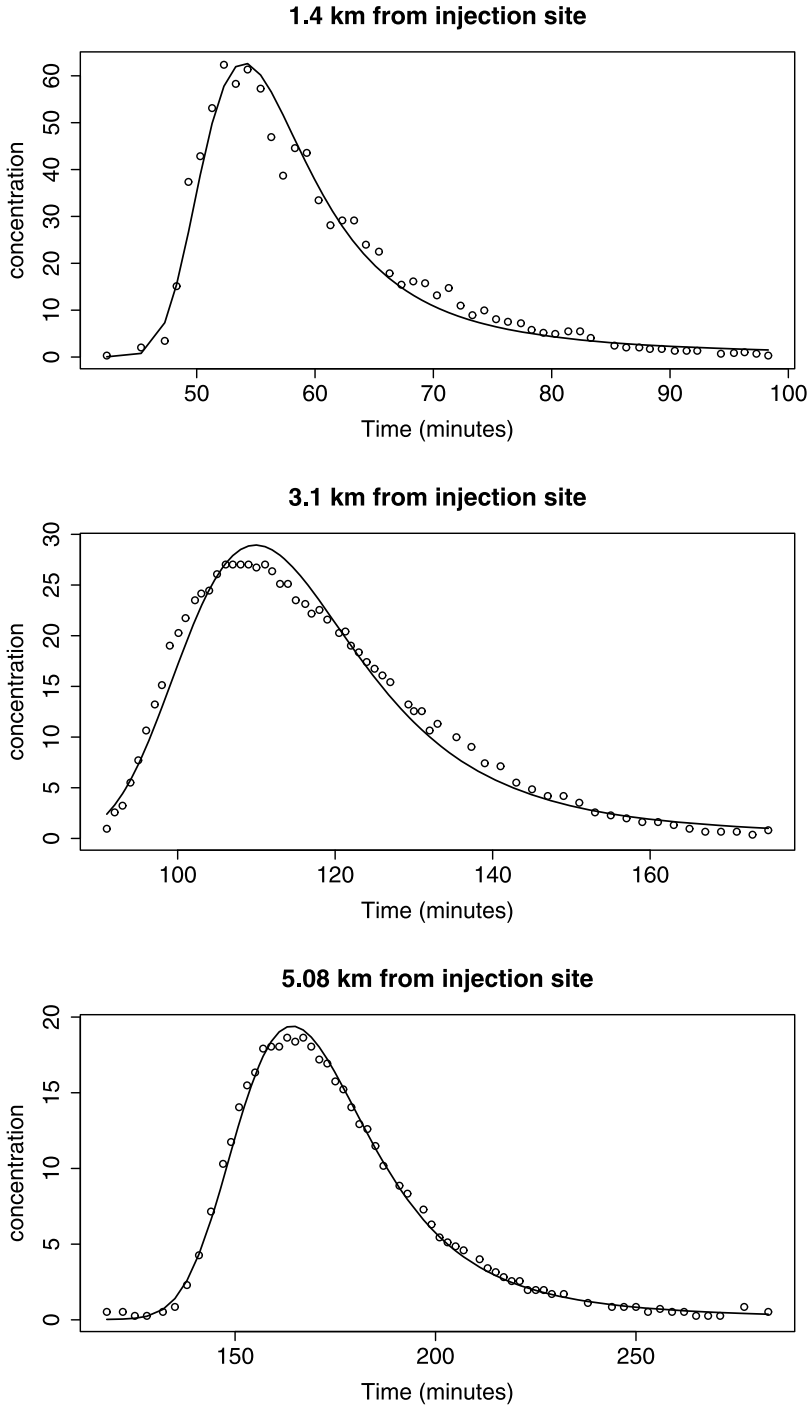

Figure 4. Breakthrough data and fitted FADE curves (solid lines) for the Red Cedar river fluorescein tracer test.

fits are also reasonable. Their fitting is based on unweighted least squares, and their measure of fit is the unweighted mean square error. It would be interesting to repeat their experiment with the optimal weighted mean square error procedure developed in our paper. In a similar vein, we compare the time-fractional model (6) to the space-fractional FADE (1). We fit the alternative model (6) to the Red Cedar breakthrough data measured 5.08 km downstream, using the same method of minimizing the weighted MSE function (5) as a function of the model parameters. The fitted parameter values for the time-fractional model are $\gamma = 0.978$ (dimensionless), $v = 0.029$ km/min$^\gamma$, $D = 0.000467$ km^2/min$^\gamma$ and $K = 27.85$ μg/L. Figure 5 shows the fitted curve against the data. The fit is reasonable, although visually not as good as the FADE, especially near the peak. The FADE fit gave an optimal weighted MSE of 0.058, while the time-fractional model gave 0.127, providing further evidence that the FADE fit is superior.

4.3. Breakthrough Curve for the Grand River

[25] The methods of this paper were used to fit the FADE (1) to breakthrough curve data from a tracer test

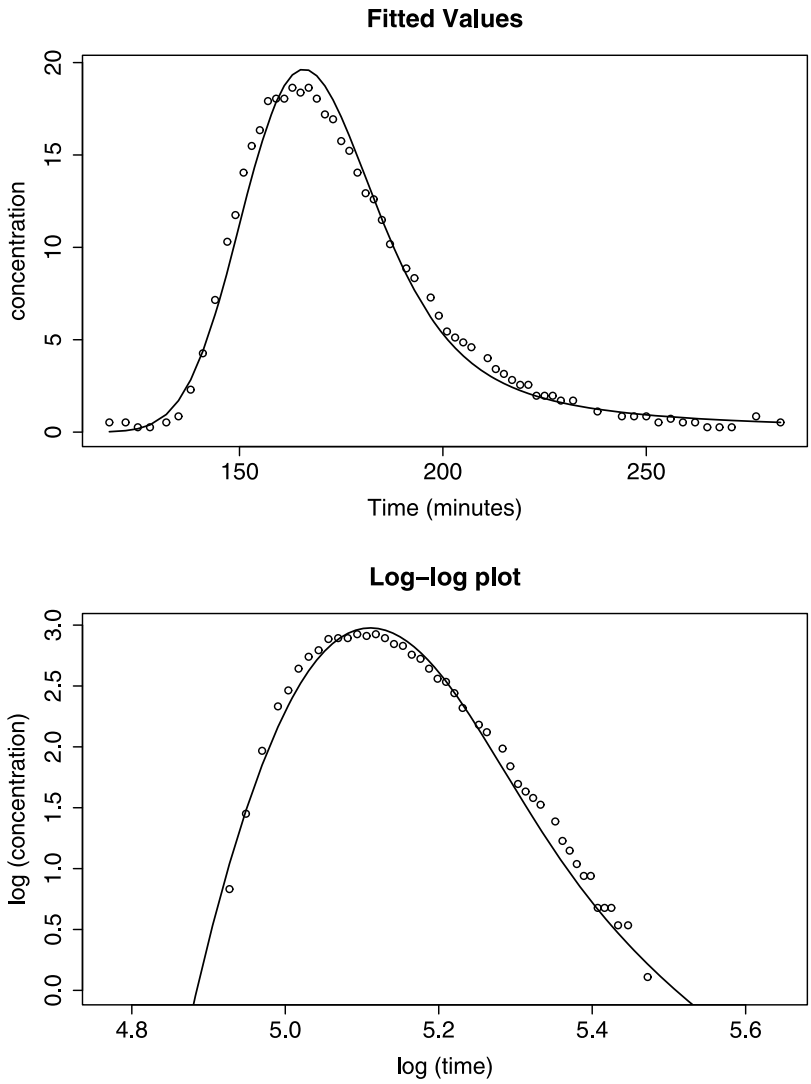

Figure 5. Alternative time fractional ADE model (6) fitted (solid line) versus breakthrough data for Red Cedar River fluorescein tracer test observed 5.08 km downstream from release. Compare with the FADE fit in Figure 4.

reported by *Shen et al.* [2008]. A tracer study was conducted on a 40 km stretch of the Grand River, a 420 km long tributary to Lake Michigan, traveling through the city of Grand Rapids and extending to Coopersville, Michigan, United States. Rhodamine WT 20% (weight) solution was used in the study. At each station, grab samples were collected from just below the surface using manual sampling. The distances to the four sampling locations from the point of release are: 4558 m (bridge 1), 13,687 m (bridge 2), 28,375 m (bridge 3) and 37,608 m (bridge 4). Figure 6 shows the fit for RWT breakthrough data collected at bridge 3. Fitted parameter values are $\alpha = 1.38$, $\beta = -1.0$, $v = 0.446$ m/s, $D = 0.887$ m$^{\alpha}$/s, and $K = 50{,}179$ ppb. The late time breakthrough curve shows significant retention. It appears that the FADE model is capturing the retention via the negative skewness parameter $\beta = -1$, which means that particles are falling behind the plume center of mass.

[26] While the negatively skewed stable provides a satisfying fit to the Grand river data, one can certainly entertain alternative models. A simple lognormal pdf captures some features of the data: It is strongly skewed and nonnegative. Figure 7 illustrates the best fit, using the same optimal weighted MSE procedure. The fitted parameter values are $\mu = 10.98$, $\sigma = 0.093$, and $K = 99945$. Lack of fit at the peak is evident, and the log-log plot shows that the lognormal model fails to capture the power law decline in the late-time breakthrough curve. The optimal weighted MSE for the lognormal model is 0.12 as compared to 0.018 for the FADE fit, providing an additional measure of how much better the FADE models this data.

4.4. Interactive Groundwater Simulation

[27] A sophisticated software environment, termed interactive groundwater (IGW), has recently been developed to provide unified deterministic, stochastic, and multiscale groundwater modeling [*Li and Liu*, 2006; *Li et al.*, 2006]. The FADE (1) was fit to an ensemble average plume simulated in IGW using a multiscale hydraulic conductivity field on a model domain of 500 m $\times$ 125 m. Three lognormal random fields with correlation lengths of 10 m, 100 m, and

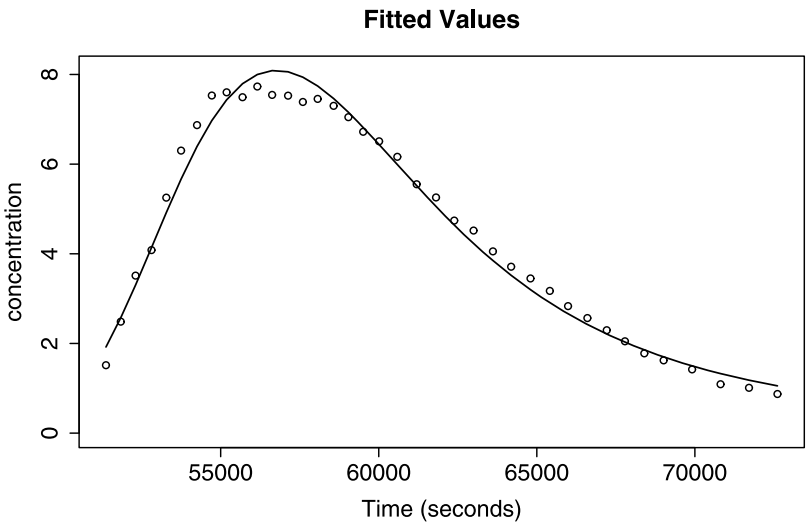

Figure 6. Breakthrough data and fitted FADE curve (solid line) for rhodamine tracer test in the Grand River 28.375 km downstream from release.

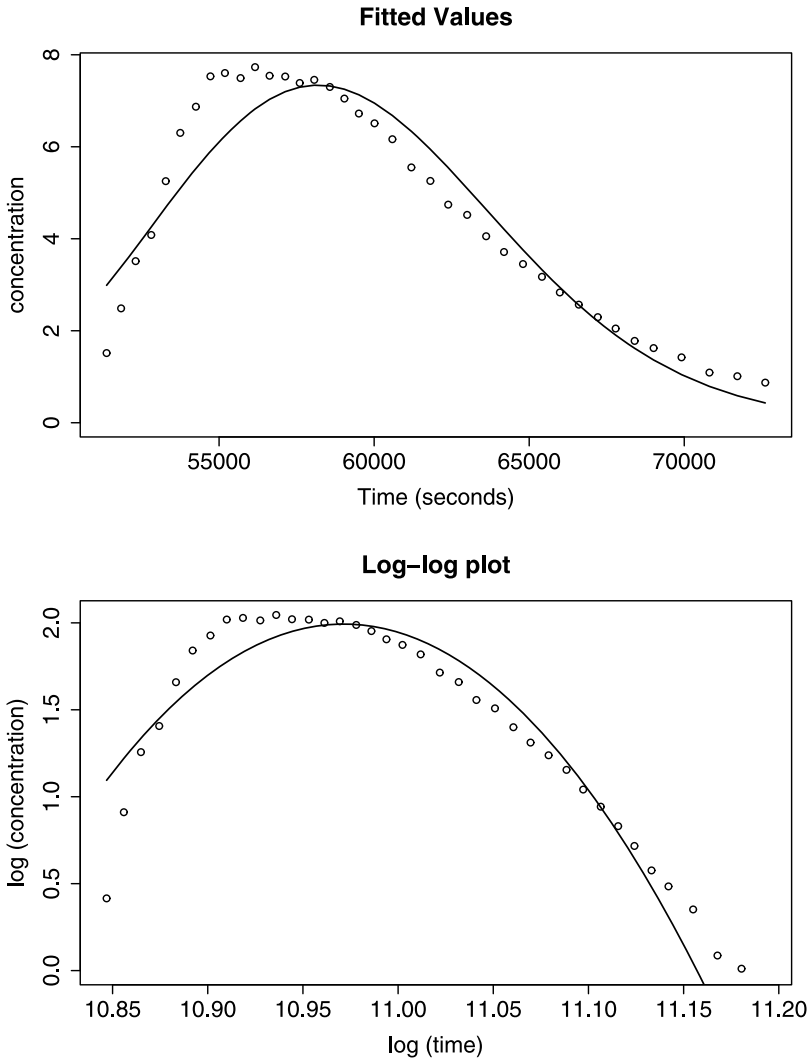

Figure 7. Alternative lognormal fit (solid line) and breakthrough data for Grand River rhodamine tracer test 28.375 km downstream from release. Compare with FADE fit in Figure 6.

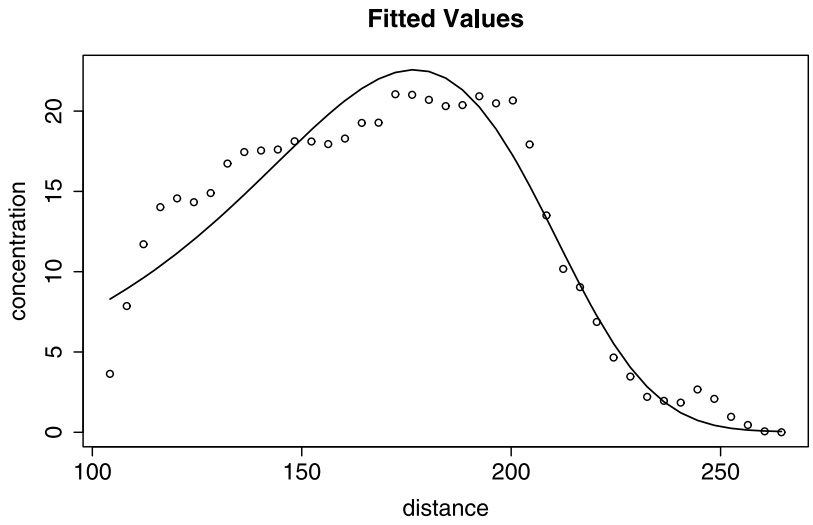

Figure 8. IGW simulated ensemble plume spatial snapshot and FADE model fit (solid line).

500 m and the same geometric mean of 10 m/d were superimposed to generate the conductivity field. A constant head difference of 1 m was imposed, porosity was set at 0.3, and a deterministic plume with a pulse initial condition was simulated using the IGW software on an 801 × 201 grid. An ensemble mean of 100 simulated plumes was averaged along the axis transverse to the flow to produce one dimensional concentration snapshots consistent with (1). Figure 8 shows the stable density fit [$\alpha = 1.25$, $\beta = -1.0$, $\mu = 93.4$, $\sigma = 31.8$, $K = 2578$]. The simulated concentration data is more variable than the river flow tracer data, but similar to the MADE snapshots. It seems that the stable density gives a reasonable fit, and since the tail parameter $\alpha = 1.25$ is much less than 2, this is evidence that the FADE provides a better descriptive model than the traditional ADE. The tail parameter indicates a high degree of heterogeneity, similar to the MADE plume, see the discussion by *Clarke et al.* [2005]. However, the skewness here is $\beta = -1$, quite different than the $\beta = 1$ value estimated for the MADE plume. One difference is that the MADE plume has a heavy leading edge, coded by the positive fractional derivative in (1). The simulated plume shows no early arrivals, and it seems likely that the FADE skewness is coding retention here, just as for the river flow data discussed earlier in this paper. We also note that the left (upstream) tail of the IGW plume falls off faster than the fitted concentration, so that the FADE model overestimates retention in this case. Note that the simulated conductivity field is multiscale lognormal, while the MADE conductivity field exhibits heavier tails [*Benson et al.*, 2001]. Although the multiscale simulated field successfully reproduces a high degree of heterogeneity, a different underlying distribution of conductivity may be required to fully capture the statistical properties of an aquifer similar to the MADE site, where a heavy tail conductivity field leads to significant early arrivals. The ensemble plume fitted here is quite different than the MADE plume, which is essentially one realization of a natural velocity field. The fitting procedure developed in this paper assumes a particle tracking Ansatz with independent particles. In the ensemble plume, there is a correlation between particles, imparted by the random velocity field. Under the usual ergodic hypothesis [*Zhan*, 1999] particles sample enough heterogeneity in the aquifer so that the resulting plume resembles the ensemble limit behavior, allowing the effective application of particle-tracking schemes. In that case, the methods of this paper remain valid. However, it may be that the IGW ensemble plume is preasymptotic. Further research to develop fitting schemes that include two particle correlations would be interesting, to facilitate modeling of pre-ergodic plumes.

4.5. Discussion

[28] The MADE site is a highly heterogeneous aquifer, and there is a large degree of scatter between the measured and fitted concentrations. Breakthrough concentrations at the Red Cedar River and Grand River are much smoother, with much less scatter between the measured and fitted concentrations. It is not common to fit the same model to both groundwater and surface water tracer tests in the same paper, and this provides an interesting opportunity for comparison. Transport models for both groundwater and surface water tracer tests recognize that hydrodynamic dispersion is a main driver of plume spreading. In groundwater, dispersion comes from intervention of the porous medium. At the MADE site, a highly heterogeneous medium leads to considerable dispersion, as compared to other groundwater sites. Model fitting to groundwater tracer tests typically leads to values of α between 1.1 and 1.7 in the FADE model (1). In general, the parameter α codes the heterogeneity of a porous medium. See *Clarke et al.* [2005] for further discussion. Tracer tests in 47 column experiments are fit to the symmetric FADE with $\beta = 0$ by *San José Martínez et al.* [2009] using an unweighted MSE criterion. The resulting α values range between 1.05 to 2.00. That study notes that α is affected by soil type, choice of tracer, whether soil is disturbed, and soil saturation. Saturated soil column data from *Huang et al.* [2006] fit by unweighted MSE produced values of α near 1.85 (with $\beta = 1$). They also found some evidence of scale dependence in the fractional dispersion parameter D in (1). *Xiong et al.* [2006] model the same soil column data, and compare the FADE and CTRW (time-fractional ADE) using the unweighted MSE criterion.

They find the CTRW (time-fractional) model gives a better fit, contrary to our findings for the Red Cedar. It would be interesting to repeat their study using our optimal weighted MSE criterion.

[29] The parameter fits for the Red Cedar and Grand River in this paper (α between 1.3 and 1.6) indicate less heterogeneity than the MADE site. *Hunt* [2006] fits the FADE (1) with α around 1.7 to three river flows. *Deng et al.* [2004] finds α between 1.6 and 1.8 for several reaches of the Missouri and Monocacy rivers. *Deng et al.* [2006] adds a retention term to the model, and fits to several additional rivers, with similar α values. In all these studies, the skewness $\beta = -1$ is used. Recall that the parameter α codes heterogeneity. In the case of river flows, heterogeneity refers to variability in the velocity field. An important difference between river flows and groundwater is the extent of mixing. The smoothness of the river flow concentrations is attributed to faster mixing in the flow. Faster mixing leads to less scatter of the measured concentrations away from the fitted concentration curve. This in turn implies tighter error bars for the parameter estimates. In river flow tracer tests, it is common for the fitted model parameters to vary from one measurement site to another. This is typically attributed to differences between reaches of the river, such as depth and meander.

[30] The heterogeneity parameter α seems to be similar for tracer tests in groundwater or river flows. In either case, a smaller value of α indicates a more heterogeneous velocity profile, causing more dispersion. The main difference between the two applications is the skewness. Typical groundwater plumes exhibit neutral to positive skewness, with a heavy leading tail. Many river flow plumes are negatively skewed, with a persistent late time tail. Hence the physical explanation of the fractional derivative used in groundwater theory [*Schumer et al.*, 2001] cannot be directly applied to river flows. A physical explanation for the FADE model of river flows that allows negative skewness was recently accomplished by *Kim and Kavvas* [2006].

[31] We also explored an alternative CTRW model that codes retention in terms of power law waiting times between particle movements. This model is equivalent to a time-fractional ADE. The CTRW (time-fractional) fit to the Grand River tracer test data was reasonably good, but not as good as the FADE fit. The idea of using space-fractional models to capture retention in river flows is fairly new, and controversial. *Zhang et al.* [2009] caution that the FADE with $\beta < 1$ may not be physically realistic, and recommend alternative time-fractional models. In previous applications to groundwater hydrology, space-fractional models capture early arrivals, and time-fractional or CTRW models code retention. One possible explanation for the use of space-fractional derivatives to model retention is given by B. Baeumer et al. (Space-time duality for fractional diffusion, submitted to *Journal of Applied Probability*, 2009) in terms of space-time duality: In short, falling behind the plume center of mass due to retention is mathematically equivalent to jumping upstream, in the ergodic limit.

5. Conclusions

[32] This paper develops a new method of parameter estimation for the space-fractional advection-dispersion equation (FADE) in equation (1). The method is based on a particle-tracking approach, where concentration measurements are interpreted as a random histogram. The method can also be used for any other transport model that admits a particle-tracking solution, where each particle moves independently of the other particles. Model fitting uses weighted least squares, with weights based on the concentration variance. The particle-tracking model implies that concentration variance is proportional to concentration, so a fitted curve should lie closer to the measured concentration data at lower concentrations. This weighted least squares approach is effective for both spatial snapshots and temporal breakthrough data. The method is used to fit the FADE to several data sets. The MADE tritium plume snapshot data was previously modeled using the FADE, and our parameter fits are similar. An alternative fit with an extended FADE, using a fractional derivative of order $\alpha < 1$, seems to provide a superior fit. We then discuss the underlying physical principles that support this model extension. Two different sets of breakthrough data for tracers tests in Michigan rivers are fit, with excellent results. The fit suggests significant deviation from the traditional ADE model, and it seems that the FADE (without any additional retention terms) is modeling retention through negative skewness. An alternative CTRW model is also examined, and the two competing model fits are compared using the metric of weighted MSE. It turns out that the FADE with negative skewness gives a somewhat better fit in this case. Simulated plume data using a multiscaling conductivity field also shows anomalous dispersion, similar to the MADE site, but again with negative skewness. In all cases, the fitting methods of this paper produce reasonable results, showing both the utility of our method, and the scope of the FADE as a model for transport.

Appendix A: Statistical Theory

[33] In the particle-tracking model, a tracer plume is represented by a large ensemble of statistically identical particles $[X_t^{(k)}: 1 \le k \le n]$. Each particle has the same probability density $f_{\boldsymbol{\theta}}(x, t)$, which depends on a vector of model parameters $\boldsymbol{\theta}$. We assume a fixed total mass $K > 0$, so that each particle carries mass K/n. Then a histogram of particle concentration at one fixed time $t > 0$ approximates the model concentration $C(x, t) = K f_{\boldsymbol{\theta}}(x, t)$ via the formula

$$\hat{C}(x,t) = \frac{K}{dx} \cdot \frac{N_x}{n}, \tag{A1}$$

where N_x counts the number of particles in the bin at location x, and dx is the bin width. The *concentration variance* can now be computed from the right-hand side of equation (A1). We sketch the ideas briefly here, and then proceed to a rigorous development.

[34] The count variable N_x is binomial with success probability $p \approx f_{\boldsymbol{\theta}}(x, t)\, dx$, the chance of a single particle occupying the bin at location x. The mean number of particles in the bin is $E[N_x] = np$, and the variance is $V[N_x] = np(1 - p)$, according to the standard formula for binomial random variables. Then the fraction of particles N_x/n in the bin at location x has mean $p = E[N_x/n]$ and variance $p(1 - p)/n =$

$V[N_x/n]$. Since N_x/n has mean $p \approx f_{\boldsymbol{\theta}}(x, t)\,dx$, the mean of the measured concentration $\hat{C}(x, t) = (K/dx) \cdot (N_x/n)$ is approximately equal to $(K/dx) \cdot f_{\boldsymbol{\theta}}(x, t)\,dx$, which reduces to the model concentration $C(x, t) = Kf_{\boldsymbol{\theta}}(x, t)$. In order to get an accurate estimate, we need a large number of particles (n) and a narrow bin width (dx), and hence the probability p is quite small. Then N_x/n has variance $p(1-p)/n \approx p/n \approx f_{\boldsymbol{\theta}}(x, t)dx/n$, so that the variance of the measured concentration is approximately $(K/dx)^2 \cdot f_{\boldsymbol{\theta}}(x, t)dx/n$ which reduces to $C(x, t) \cdot K/(n\,dx)$. In short, *concentration variance is proportional to concentration*.

[35] Estimation of the concentration via (A1) is complicated by the fact that the count variables N_x in (A1) are not independent for different locations x: If a particle $X_t^{(k)}$ inhabits bin x, it cannot inhabit bin $x' \neq x$ at that same time. To model this dependence, we introduce the cumulative distribution function $F_{\boldsymbol{\theta}}(x, t) = P(X_t \leq x)$. This distribution function is approximated by its empirical analogue

$$\hat{F}_{\boldsymbol{\theta}}(x,t) = \frac{1}{n}\sum_{k=1}^{n} I\left(X_t^{(k)} \leq x\right), \tag{A2}$$

where $I(X_t^{(k)} \leq x) = 1$ if $X_t^{(k)} \leq x$, and $I(X_t^{(k)} \leq x) = 0$ otherwise. Then the sum in (A2) is simply the number of particles located to the left of the point x. In the same notation, we write the density estimate as

$$\hat{f}(x,t)dx = \frac{1}{n}\sum_{k=1}^{n} I\left(x - dx < X_t^{(k)} \leq x\right), \tag{A3}$$

where again $I(\cdot) = 1$ if the statement in parentheses is true, and $I(\cdot) = 0$ otherwise. Note that $\hat{C}(x, t) = K\hat{f}_{\boldsymbol{\theta}}(x, t)$ where $\hat{f}_{\boldsymbol{\theta}}(x, t)\,dx = \hat{F}_{\boldsymbol{\theta}}(x, t) - \hat{F}_{\boldsymbol{\theta}}(x - dx, t)$ is the area of one histogram bar. Histogram density estimation is widely used, and asymptotic properties at a single point x can be obtained from the general theory of kernel density estimation [*Silverman*, 1986]. Since we need to consider the density estimation problem simultaneously at multiple points of x, a full treatment of the density estimation problem relies on the asymptotic theory of the empirical distribution function as a stochastic process in x. We consider the asymptotic approximation of the density function by the empirical density or histogram (A3) as the number of particles n tends to infinity and the bin size dx_n shrinks to zero. The notation $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ represents the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. Its probability density is $g(\boldsymbol{x}) = (2\pi)^{-d/2}|\Sigma|^{-1/2}\exp(-(\boldsymbol{x} - \mu)^T\Sigma^{-1}(\boldsymbol{x} - \mu)/2)$ in d dimensions. The notation $X_n \Rightarrow Y$ means that the probability distribution of the random variables X_n converges to the distribution of Y. We say that X_n is a consistent estimator of a if $P(|X_n - a| > \varepsilon) \to 0$ as $n \to \infty$ for any error tolerance $\varepsilon > 0$ (also called convergence in probability).

[36] Suppress t and $\boldsymbol{\theta}$ for ease of notation and write $p_\Delta(x) = F(x) - F(x - \Delta)$ where $\Delta = dx_n$. Then the density estimate has mean

$$E\left[\hat{f}(x)\right] = E\left[\frac{F_n(x) - F_n(x-\Delta)}{\Delta}\right] = \frac{p_\Delta(x)}{\Delta} \tag{A4}$$

and covariance $E[\hat{f}(x)\hat{f}(y)]$

$$\begin{aligned}
&= E\left[\left(\frac{1}{n\Delta}\sum_{k=1}^{n} I\left(x - \Delta < X^{(k)} \leq x\right)\right)\right.\\
&\qquad \left.\cdot\left(\frac{1}{n\Delta}\sum_{j=1}^{n} I\left(y - \Delta < X^{(j)} \leq y\right)\right)\right]\\
&= \frac{1}{(n\Delta)^2}E\left[\sum_{k=1}^{n} I\left(x - \Delta < X^{(k)} \leq x, y - \Delta < X^{(k)} \leq y\right)\right.\\
&\qquad \left.+\sum_{\substack{j,k=1\\ j\neq k}}^{n} I\left(x - \Delta < X^{(k)} \leq x\right) I\left(y - \Delta < X^{(j)} \leq y\right)\right]\\
&= \frac{1}{(n\Delta)^2}nP\left[\left(x - \Delta < X^{(1)} \leq x, y - \Delta < X^{(1)} \leq y\right)\right]\\
&\quad + \frac{n(n-1)}{(n\Delta)^2}\{F(x) - F(x-\Delta)\}\{F(y) - F(y-\Delta)\}.
\end{aligned}$$

Thus, for sufficiently small $\Delta = dx_n$, $E[\hat{f}(x)\hat{f}(y)]$

$$= \begin{cases} \dfrac{1}{n\Delta^2}(n-1)p_\Delta(x)p_\Delta(y) & \text{if } x \neq y \\ \dfrac{1}{n\Delta^2}\left[p_\Delta(x) + (n-1)p_\Delta^2(x)\right] & \text{if } x = y. \end{cases} \tag{A5}$$

Using (A4) and (A5)

$$\begin{aligned}
\mathrm{Var}\left[\hat{f}(x)\right] &= \frac{1}{n\Delta^2}p_\Delta(x)(1 - p_\Delta(x))\\
\mathrm{Cov}\left[\hat{f}(x), \hat{f}(y)\right] &= -\frac{1}{n\Delta^2}p_\Delta(x)p_\Delta(y).
\end{aligned}$$

Since f is the continuous density function of X,

$$\begin{aligned}
\lim_{\Delta\to 0}\frac{F(x) - F(x-\Delta)}{\Delta} &= \lim_{\Delta\to 0}\frac{1}{\Delta}p_\Delta(x) = f(x),\\
\lim_{\Delta\to 0} F(x) - F(x-\Delta) &= \lim_{\Delta\to 0} p_\Delta(x) = 0.
\end{aligned}$$

Hence if $\Delta \to 0$ and $n\Delta \to \infty$ we have

$$\begin{aligned}
E\left[\hat{f}(x)\right] &\approx f(x)\\
\mathrm{Var}\left[\sqrt{n\Delta}\hat{f}(x)\right] &= \frac{p_\Delta(x)}{\Delta}(1 - p_\Delta(x)) \approx f(x)
\end{aligned}$$

and $\mathrm{Cov}[\sqrt{n\Delta}\,\hat{f}(x), \sqrt{n\Delta}\,\hat{f}(y)]$

$$\begin{aligned}
&= -\frac{n\Delta}{n\Delta^2}p_\Delta(x)p_\Delta(y)\\
&= -\frac{p_\Delta(x)}{\Delta}p_\Delta(y) \approx -f(x) \times 0 = 0.
\end{aligned}$$

Since $\hat{f}$ is sum of independently and identically distributed random variables, using the central limit theorem and Slutsky's theorem [e.g., see *Ferguson*, 1996] we obtain

$$\sqrt{n\Delta}\begin{pmatrix}\hat{f}(x) - f(x)\\ \hat{f}(y) - f(y)\end{pmatrix} \to N\left(0, \begin{pmatrix} f(x) & 0\\ 0 & f(y)\end{pmatrix}\right). \tag{A6}$$

Since $n\Delta \to \infty$ it follows from (A6) that $\mathrm{E}[\hat{f}(x)] \to f(x)$ and $\mathrm{Var}[\hat{f}(x)] \to 0$, and hence $\hat{f} \to_p f$; that is, $\hat{f}$ is a consistent

estimator of f. Thus, in our original notation, we have shown that, for the density estimator $\hat{f}(x, t)$ as defined in (A3), if $n \to \infty$, $dx_n \to 0$ and $n\, dx_n \to \infty$, then

$$\sqrt{n\,dx_n}\begin{pmatrix}\hat{f}_\theta(x_i,t)-f_\theta(x_i,t)\\ \hat{f}_\theta(x_j,t)-f_\theta(x_j,t)\end{pmatrix} \Rightarrow \mathcal{N}\left(0,\begin{pmatrix}f_\theta(x_i,t) & 0\\ 0 & f_\theta(x_j,t)\end{pmatrix}\right) \tag{A7}$$

and $\hat{f}_{\boldsymbol{\theta}}(x, t)$ is a consistent estimator of $f_{\boldsymbol{\theta}}(x,t)$. Note that these results can be also derived from the asymptotics of empirical process [*van der Vaart*, 1998]. Note also that these results are true for any density estimation problem based on a histogram, and hence they apply to any transport model for which a corresponding particle-tracking method is available.

[37] From the above results, we can consider the observed concentrations after appropriate rescaling as a good approximation of density values so that we can estimate the parameters of the density. The remaining results are stated in terms of the stable density underlying the FADE. Extension to alternative transport models is similar, under suitable technical conditions. We assume that concentration data is observed at the vector of locations $\mathbf{x} = (x_1, \cdots, x_N)$, and we denote by diag[$\mathbf{x}$] the diagonal matrix with entries $x_1, \ldots, x_N$. Let Θ be an open subset of the parameter space on which the Fourier transform of the stable density (2) is continuous at all parameters. Suppose that the true value $\boldsymbol{\theta}_0$ is in interior of Θ. Suppose that $n \to \infty$, $dx_n \to 0$ in such a way that $\sqrt{n}dx_n \to \infty$. Let $\hat{\boldsymbol{\theta}}$ be the estimator obtained by minimizing

$$Q(\boldsymbol{\theta}) = \left(\hat{f}_\theta(\mathbf{x},t)-f_\theta(\mathbf{x},t)\right)^T \Sigma_\theta^{-1}\left(\hat{f}_\theta(\mathbf{x},t)-f_\theta(\mathbf{x},t)\right), \tag{A8}$$

where $\Sigma_{\boldsymbol{\theta}} = \text{diag}[\hat{f}_{\boldsymbol{\theta}}(\mathbf{x}, t)]$. We wish to show that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$.

[38] Consider a neighborhood of $\boldsymbol{\theta}_0$, N_δ, such that $N_\delta = \{\boldsymbol{\theta}: |\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta\}$ for $\delta > 0$. Here, $|v| = \max_i |v_i|$ for $v = (v_1, \cdots, v_k)$. Then it is enough to show $P(\hat{\boldsymbol{\theta}} \in N_\delta^c \cap \Theta) \to 0$ which is equivalent to $\hat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$. Write

$$\begin{aligned}P\left(\hat{\boldsymbol{\theta}} \in N_\delta^c \cap \Theta\right) &= P\left(\inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} Q(\boldsymbol{\theta}) \le \inf_{\boldsymbol{\theta}\in N_\delta\cap\Theta} Q(\boldsymbol{\theta})\right)\\ &\le P\left(\inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} (Q(\boldsymbol{\theta})-Q(\boldsymbol{\theta}_0)) \le 0\right)\end{aligned}$$

since $\inf_{\boldsymbol{\theta}\in N_\delta\cap\Theta} Q(\boldsymbol{\theta}) \le Q(\boldsymbol{\theta}_0)$. Note that

$$\begin{aligned}&Q(\boldsymbol{\theta}) - Q(\boldsymbol{\theta}_0)\\ &\quad= \sum_{i=1}^N \frac{\left(\hat{f}(x_i)-f_\theta(x_i)\right)^2}{\hat{f}(x_i)} - \sum_{i=1}^N \frac{\left(\hat{f}(x_i)-f_{\boldsymbol{\theta}_0}(x_i)\right)^2}{\hat{f}(x_i)}\\ &\quad= \sum_{i=1}^N \frac{(f_{\boldsymbol{\theta}_0}(x_i)-f_\theta(x_i))^2}{\hat{f}(x_i)}\\ &\qquad+ 2\sum_{i=1}^N \frac{\left(\hat{f}(x_i)-f_{\boldsymbol{\theta}_0}(x_i)\right)(f_{\boldsymbol{\theta}_0}(x_i)-f_\theta(x_i))}{\hat{f}(x_i)}\\ &\quad=: A(\boldsymbol{\theta}) + B(\boldsymbol{\theta}).\end{aligned}$$

Then, we have

$$\begin{aligned}&P\left(\inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} (Q(\boldsymbol{\theta})-Q(\boldsymbol{\theta}_0)) \le 0\right)\\ &\quad\le P\left(\inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} A(\boldsymbol{\theta}) + \inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} B(\boldsymbol{\theta}) \le 0\right).\end{aligned}$$

By the consistency of $\hat{f}(x_i)$ we have

$$A(\boldsymbol{\theta}) \to_p \sum_{i=1}^N \frac{(f_{\boldsymbol{\theta}_0}(x_i)-f_{\boldsymbol{\theta}}(x_i))^2}{f_{\boldsymbol{\theta}_0}(x_i)}$$

and thus there exists a C such that

$$\inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} A(\boldsymbol{\theta}) \ge C > 0$$

because $f_{\boldsymbol{\theta}}(x)$ is different from $f_{\boldsymbol{\theta}_0}(x)$ for $\boldsymbol{\theta} \in N_\delta^c \cap \Theta$.

[39] We can also choose an $\epsilon > 0$ such that $|B(\boldsymbol{\theta})| \le \boldsymbol{\epsilon} < C$ for large enough n. Thus,

$$\inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} A(\boldsymbol{\theta}) + \inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} B(\boldsymbol{\theta}) \ge C - \epsilon > 0$$

and it follows that

$$P\left(\inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} A(\boldsymbol{\theta}) + \inf_{\boldsymbol{\theta}\in N_\delta^c\cap\Theta} B(\boldsymbol{\theta}) \le 0\right) \to 0, \tag{A9}$$

which shows that $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$.

[40] Next we consider the asymptotics of the parameter estimates. Let $\boldsymbol{\Theta}$ be an open subset of the parameter space on which the density $f_{\boldsymbol{\theta}}(\mathbf{x}, t)$ is twice continuously differentiable. As before, suppose that $n \to \infty$, $dx_n \to 0$ in such a way that $\sqrt{n}dx_n \to \infty$. By consistency of $\hat{\boldsymbol{\theta}}$ and the mean value theorem,

$$\frac{\partial Q(\hat{\boldsymbol{\theta}})}{\partial\theta_j} - \frac{\partial Q(\boldsymbol{\theta}_0)}{\partial\theta_j} = \frac{\partial^2 Q(\overline{\boldsymbol{\theta}})}{\partial\theta_j^2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),$$

where $\overline{\boldsymbol{\theta}}$ is in a line segment between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ contained in $\boldsymbol{\Theta}$. Since $\partial Q(\hat{\boldsymbol{\theta}})/\partial\theta_i = 0$, we have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\left[\frac{\partial^2 Q(\overline{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}^2}\right]^{-1}\frac{\partial Q(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}},$$

where $\partial Q(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = (\partial Q(\boldsymbol{\theta})/\partial\theta_1, \cdots, \partial Q(\boldsymbol{\theta})/\partial\theta_4)^T$ with jth element defined as

$$\frac{\partial Q(\boldsymbol{\theta})}{\partial\theta_j} = -2\sum_{i=1}^N \frac{\left(\hat{f}(x_i)-f_\theta(x_i)\right)}{\hat{f}(x_i)}\frac{\partial f_\theta(x_i)}{\partial\theta_j} \tag{A10}$$

and $\partial^2 Q(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^2$ is a 4×4 matrix whose (j, k) entry is

$$\begin{aligned}\frac{\partial^2 Q(\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k} &= -2\sum_{i=1}^N \frac{\left(\hat{f}(x_i)-f_\theta(x_i)\right)}{\hat{f}(x_i)}\frac{\partial^2 f_\theta(x_i)}{\partial\theta_j\partial\theta_k}\\ &\quad+ 2\sum_{i=1}^N \frac{1}{\hat{f}(x_i)}\frac{\partial f_\theta(x_i)}{\partial\theta_j}\frac{\partial f_\theta(x_i)}{\partial\theta_k}.\end{aligned}$$

[41] Since $f_{\boldsymbol{\theta}}(x)$ is twice continuously differentiable with respect to $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}$, we have

$$\begin{aligned}&|\hat{f}(x_i) - f_{\bar{\boldsymbol{\theta}}}(x_i)| \\ &\quad \le |\hat{f}(x_i) - f_{\boldsymbol{\theta}_0}(x_i)| + |f_{\boldsymbol{\theta}_0}(x_i) - f_{\bar{\boldsymbol{\theta}}}(x_i)| \to_p 0 \\ &\quad \left.\frac{\partial f_{\boldsymbol{\theta}}(x_i)}{\partial \theta_j}\right|_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} \to \frac{\partial f_{\boldsymbol{\theta}_0}(x_i)}{\partial \theta_j}\end{aligned}$$

and thus we have

$$\frac{\partial^2 Q(\bar{\boldsymbol{\theta}})}{\partial \theta_j \partial \theta_k} \to_p 2\sum_{i=1}^{N} \frac{1}{f_{\boldsymbol{\theta}_0}(x_i)} \frac{\partial f_{\boldsymbol{\theta}_0}(x_i)}{\partial \theta_j} \frac{\partial f_{\boldsymbol{\theta}_0}(x_i)}{\partial \theta_k}. \tag{A11}$$

Consequently,

$$\frac{\partial^2 Q(\bar{\boldsymbol{\theta}})}{\partial \theta^2} \to_p 2 \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x})}{\partial \boldsymbol{\theta}}^T [\mathrm{diag}(f_{\boldsymbol{\theta}_0}(\mathbf{x}))]^{-1} \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x})}{\partial \boldsymbol{\theta}}, \tag{A12}$$

where $\partial f_{\boldsymbol{\theta}_0}(\mathbf{x}, t)/\partial \boldsymbol{\theta}$ is the $N \times 4$ matrix whose (j, k) entry is $\partial f_{\boldsymbol{\theta}}(x_j, t)/\partial \theta_k$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. From (A10), we can express

$$\frac{\partial Q(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = -2 \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x})}{\partial \boldsymbol{\theta}}^T \left[\mathrm{diag}\left(\hat{f}(\mathbf{x})\right)\right]^{-1} \left(\hat{f}(\mathbf{x}) - f_{\boldsymbol{\theta}_0}(\mathbf{x})\right).$$

Thus, It follows from (A7) and $\hat{f} \to_p f$ that

$$\begin{aligned}\sqrt{n\,dx_n} \frac{\partial Q(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} &= -2 \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x})}{\partial \boldsymbol{\theta}}^T \left[\mathrm{diag}\left(\hat{f}(\mathbf{x})\right)\right]^{-1} \\ &\quad \times \sqrt{n\,dx_n}\left(\hat{f}(\mathbf{x}) - f_{\boldsymbol{\theta}_0}(\mathbf{x})\right) \\ &\quad \Rightarrow 2 \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x})}{\partial \boldsymbol{\theta}}^T [\mathrm{diag}(f_{\boldsymbol{\theta}_0}(\mathbf{x}))]^{-1/2} \mathcal{N}(0, I).\end{aligned} \tag{A13}$$

where I is the 4×4 identity matrix. Now (A12) and (A13) combine to show that

$$\begin{aligned}\sqrt{n\,dx_n}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \Rightarrow A\mathcal{N}(0, I) \quad \text{where} \\ A = \left[\frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x}, t)}{\partial \boldsymbol{\theta}}^T \mathrm{diag}[f_{\boldsymbol{\theta}_0}(\mathbf{x}, t)]^{-1} \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x}, t)}{\partial \boldsymbol{\theta}}\right]^{-1} \\ \times \frac{\partial f_{\boldsymbol{\theta}_0}(\mathbf{x}, t)}{\partial \boldsymbol{\theta}}^T [\mathrm{diag}(f_{\boldsymbol{\theta}_0}(\mathbf{x}))]^{-1/2}.\end{aligned} \tag{A14}$$

This shows that the parameter estimate $\hat{\theta}_j$ is approximately normal with mean θ_j and standard deviation $\sigma_j/\sqrt{n\,dx_n}$ where σ_j is the jth entry along the diagonal of the matrix A in the formula (A14). Note that σ_j depends on the partial derivatives of the stable density, which are computed numerically. If $\hat{f}_{\boldsymbol{\theta}}(\mathbf{x}, t)$ follows a normal distribution exactly, $Q(\boldsymbol{\theta})$ in (A8) is similar to the negative log likelihood function. Then asymptotic properties of the estimates in this paper can be carried over from those of MLEs [e.g., see *Ferguson*, 1996]. However, the difference between parameter estimates in this paper and classical MLEs lies in the fact that $\hat{f}_{\boldsymbol{\theta}}(\mathbf{x}, t)$ is only asymptotically normal.

[42] Finally we discuss fitting the total mass. We assume that $c_i = K\hat{f}_{\theta}(x_i, t)$ where $K > 0$ is the (unknown) total plume mass, a nuisance parameter. We also write $c_i = K\hat{f}_{\boldsymbol{\theta}}(x_i, t)$ the observed concentration. Then it follows from (A7) by a simple rescaling that

$$\begin{aligned}&\sqrt{n\,dx_n}\begin{pmatrix} c_i - C(x_i, t) \\ c_j - C(x_j, t)\end{pmatrix} \\ &\quad \Rightarrow N\left(0, \begin{pmatrix} KC(x_i, t) & 0 \\ 0 & KC(x_j, t)\end{pmatrix}\right)\end{aligned} \tag{A15}$$

and c_i is a consistent estimator of $C(x_i, t)$. Similarly, it follows from (A9) that a consistent estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ comes from minimizing

$$e(\boldsymbol{\theta}, K) = (\mathbf{c} - C(\mathbf{x}, t))^T \Sigma_{\boldsymbol{\theta}}^{-1} (\mathbf{c} - C(\mathbf{x}, t)), \tag{A16}$$

where $\Sigma_{\boldsymbol{\theta}} = \mathrm{diag}[KC(\mathbf{x}, t)]$, $\mathbf{c} = (c_1, \cdots, c_N)$ is the vector of concentration measurements, and $C(\mathbf{x}, t) = (C(x_1, t), \ldots, C(x_N, t))$. In practice, we replace the unknown concentration $C(x_i, t)$ in the covariance matrix $\Sigma_{\boldsymbol{\theta}}$ by the observed concentration c_i, resulting in the weighted least squares objective function (4). Solve $\partial e/\partial K = 0$ by simple algebra to obtain

$$K = \sqrt{\frac{\sum_{i=1}^{n} c_i}{\sum_{i=1}^{n} \left[f_{\hat{\boldsymbol{\theta}}}^2(x_i, t)/c_i\right]}} \tag{A17}$$

so that K can be obtained easily once the remaining parameters $\boldsymbol{\theta}$ have been estimated. Note also that $\partial^2 e/\partial K^2 = 2K^{-3}\sum_{i=1}^{n} c_i > 0$ so that the solution to $\partial e/\partial K = 0$ yields a minimum.

[43] **Acknowledgments.** The authors wish to thank Rina Schumer, Desert Research Institute, for sharing the MADE concentration data and for helpful discussions. Thanks to M. S. Phanikumar, Civil and Environmental Engineering, Michigan State University, for providing the river flow tracer data. We also appreciate the cooperation of Shu-Guang Li and Dipa Dey, Civil and Environmental Engineering, Michigan State University, for assistance with the IGW modeling software suite. This research was partially supported by grants EAR-0823965 and DMS-0803360 from the U.S. National Science Foundation.

References

Bellin, A., and D. Tonina (2007), Probability density function of non-reactive solute concentration in heterogeneous porous formations, *J. Contam. Hydrol.*, *94*, 109–125.

Bencala, K. E., and R. A. Walters (1983), Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model, *Water Resour. Res.*, *19*(3), 718–724.

Benson, D. A., S. W. Wheatcraft, and M. M. Meerschaert (2000a), Application of a fractional advection–dispersion equation, *Water Resour. Res.*, *36*(6), 1403–1412.

Benson, D. A., S. W. Wheatcraft, and M. M. Meerschaert (2000b), The fractional-order governing equation of Lévy motion, *Water Resour. Res.*, *36*(6), 1413–1424.

Benson, D. A., R. Schumer, M. M. Meerschaert, and S. W. Wheatcraft (2001), Fractional dispersion, Lévy motion, and the MADE tracer tests, *Transp. Porous Media*, *42*(1–2), 211–240.

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

Bhattacharya, R. N., V. K. Gupta, and G. Sposito (1976), On the stochastic foundation of the theory of water flow through unsaturated soil, *Water Resour. Res.*, *12*(3), 503–512.

Boano, F., A. I. Packman, A. Cortis, R. Revelli, and L. Ridolfi (2007), A continuous time random walk approach to the stream transport of solutes, *Water Resour. Res.*, *43*, W10425, doi:10.1029/2007WR006062.

Boggs, J. M., L. M. Beard, S. E. Long, and M. P. McGee (1993), Database for the second Macro Dispersion Experiment (MADE-2), *Rep. TR-102072*, Electr. Power Res. Inst., Palo Alto, Calif.

Bromly, M., and C. Hinz (2004), Non-Fickian transport in homogeneous unsaturated repacked sand, *Water Resour. Res.*, *40*, W07402, doi:10.1029/2003WR002579.

Cassiani, M., P. Franzese, and U. Giostra (2005), A pdf micromixing model of dispersion for atmospheric flow. Part I: Development of the model, application to homogeneous turbulence and to neutral boundary layer, *Atmos. Environ.*, *39*, 1457–1469.

Chakraborty, P. (2009), A stochastic differential equation model with jumps for fractional advection and dispersion, *J. Stat. Phys.*, *136*(3), 527–551, doi:10.1007/s10955-009-9794-1.

Chambers, J. M., C. L. Mallows, and B. W. Stuck (1976), A method for simulating stable random variables, *J. Am. Stat. Assoc.*, *71*(354), 340–344.

Chang, F. X., J. Chen, and W. Huang (2005), Anomalous diffusion and fractional advection-diffusion equation, *Acta Phys. Sin.*, *54*(3), 1113–1117.

Clarke, D. D., M. M. Meerschaert, and S. W. Wheatcraft (2005), Fractal travel time estimates for dispersive contaminants, *Ground Water*, *43*(3), 401–407.

Cushman, J. H., and T. R. Ginn (2000), Fractional advection-dispersion equation: A classical mass balance with convolution-Fickian flux, *Water Resour. Res.*, *36*(12), 3763–3766.

Dagan, G. (1989), *Flow and Transport in Porous Formations*, Springer, Berlin.

Deng, Z.-Q., and H.-S. Jung (2009), Variable residence time-based model for solute transport in streams, *Water Resour. Res.*, *45*, W03415, doi:10.1029/2008WR007000.

Deng, Z.-Q., V. P. Singh, and L. Bengtsson (2004), Numerical solution of fractional advection-dispersion equation, *J. Hydraul. Eng.*, *130*(5), 422–431.

Deng, Z.-Q., L. Bengtsson, and V. P. Singh (2006), Parameter estimation for fractional dispersion model for rivers, *Environ. Fluid Mech.*, *6*, 451–475.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, vol. 2, 2nd ed., John Wiley, New York.

Ferguson, T. S. (1996), *A Course in Large Sample Theory*, Chapman and Hall, London.

Fiori, A., and G. Dagan (2000), Concentration fluctuations in aquifer transport: A rigorous first-order solution and applications, *J. Contam. Hydrol.*, *45*, 139–163.

Gelhar, L. W. (1993), *Stochastic Subsurface Hydrology*, Prentice-Hall, Englewood Cliffs, N. J.

Hu, B. X., and C. He (2006), Effects of local dispersion and kinetic sorption on evolution of concentration variance in a heterogeneous aquifer, *Math. Geol.*, *38*(3), 327–342.

Huang, G. H., Q. Z. Huang, and H. B. Zhan (2006), Evidence of one-dimensional scale-dependent fractional advection-dispersion, *J. Contam. Hydrol.*, *85*, 53–71.

Hunt, B. (2006), Asymptotic solutions for one-dimensional dispersion in rivers, *J. Hydraul. Eng.*, *132*(1), 87–93.

Kim, S., and M. L. Kavvas (2006), Generalized Fick's law and fractional ADE for pollutant transport in a river: Detailed derivation, *J. Hydrol. Eng.*, *11*(1), 80–83.

Klise, K. A., V. C. Tidwell, S. A. McKenna, and M. D. Chapin (2004), Analysis of permeability controls on transport throught laboratory-scale cross-bedded sandstone, *Geol. Soc. Am. Abstr. Programs*, *36*(5), 573.

Levy, M., and B. Berkowitz (2003), Measurement and analysis of non-Fickian dispersion in heterogeneous porous media, *J. Contam. Hydrol.*, *64*, 203–226.

Li, S. G., and Q. Liu (2006), A real-time, computational steering environment for integrated groundwater modeling, *J. Ground Water*, *44*(5), 758–763.

Li, S. G., Q. Liu, and S. Afshari (2006), An object-oriented hierarchical patch dynamics paradigm (HPDP) for modeling complex groundwater systems across multiple scales, *Environ. Modell. Software*, *21*(5), 744–749.

Meerschaert, M. M., and H. P. Scheffler (2004), Limit theorems for continuous time random walks with infinite mean waiting times, *J. Appl. Probab.*, *41*(3), 623–638.

Meerschaert, M. M., D. A. Benson, and B. Bäumer (1999), Multivariable advection and fractional dispersion, *Phys. Rev. E*, *59*, 5026–5028.

Meerschaert, M. M., D. A. Benson, H. P. Scheffler, and B. Baeumer (2002), Stochastic solution of space-time fractional diffusion equations, *Phys. Rev. E*, *65*, 1103–1106.

Meerschaert, M. M., J. Mortensen, and S. W. Wheatcraft (2006), Fractional vector calculus for fractional advection-dispersion, *Physica A*, *367*, 181–190.

Neuman, S. P., and D. M. Tartakovsky (2009), Perspective on theories of non-Fickian transport in heterogeneous media, *Adv. Water Resour.*, *32*, 670–680, doi:10.1016/j.advwatres.2008.08.005.

Nolan, J. P. (1999), An algorithm for evaluating stable densities in Zolotarev's (M) parameterization, *Math. Comput. Modell.*, *29*, 229–233.

Pachepsky, Y., D. Timlin, and D. A. Benson (2001), Transport of water and solutes in soils as in fractal porous media, *Soil Sci. Soc. Am. J.*, *56*, 51–75.

Phanikumar, M. S., I. Aslam, C. Shen, D. T. Long, and T. C. Voice (2007), Separating surface storage from hyporheic retention in natural streams using wavelet decomposition of acoustic Doppler current profiles, *Water Resour. Res.*, *43*, W05406, doi:10.1029/2006WR005104.

Samorodnitsky, G., and M. S. Taqqu (1994), *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*, Chapman and Hill, New York.

San José Martínez, F., Y. A. Pachepsky, and W. J. Rawls (2009), Advective-dispersive equation with spatial fractional derivatives evaluated with tracer transport data, *Vadose Zone J.*, *8*, 242–249.

Schumer, R., D. A. Benson, M. M. Meerschaert, and S. W. Wheatcraft (2001), Eulerian derivation of the fractional advection-dispersion equation, *J. Contam. Hydrol.*, *38*, 69–88.

Schumer, R., D. A. Benson, M. M. Meerschaert, and B. Baeumer (2003), Fractal mobile/immobile solute transport, *Water Resour. Res.*, *39*(10), 1296, doi:10.1029/2003WR002141.

Shen, C., M. S. Phanikumar, T. T. Fong, I. Aslam, S. L. Molloy, and J. B. Rose (2008), Evaluating bacteriophage P22 as a tracer in a complex surface water system: The Grand River, Michigan, *Environ. Sci. Technol.*, *42*(7), 2426–2431.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Univ. Press, Cambridge, U. K.

Xiong, Y., and G. Huang (2006), Modeling solute transport in one dimensional homogeneous and heterogeneous soil columns with continuous time random walk, *J. Comtam. Hydrol.*, *86*, 163–175.

Zhan, H. (1999), On the ergodicity hypothesis in heterogeneous formations, *Math. Geol.*, *31*(1), 113–134.

Zhang, X., J. W. Crawford, L. K. Deeks, M. I. Stutter, A. G. Bengough, and I. M. Young (2005), A mass balance based numerical method for the fractional advection-dispersion equation: Theory and application, *Water Resour. Res.*, *41*, W07029, doi:10.1029/2004WR003818.

Zhang, Y., D. A. Benson, M. M. Meerschaert, E. M. LaBolle, and H.-P. Scheffler (2006), Random walk approximation of fractional-order multiscaling anomalous diffusion, *Phys. Rev. E*, *74*, 026706.

Zhang, Y., D. A. Benson, and D. M. Reeves (2009), Time and space nonlocalities underlying fractional-derivative models: Distinction and literature review of field applications, *Adv. Water Resour.*, *32*, 561–581.

Zheng, C., and S. M. Gorelick (2003), Analysis of the effect of decimeter-scale preferential flow paths on solute transport, *Ground Water*, *41*(2), 142–155.

Zhou, L., and H. M. Selim (2003), Application of the fractional advection-dispersion equation in porous media, *Soil Sci. Soc. Am. J.*, *67*, 1079–1084.

Zolotarev, V. M. (1986), *One-Dimensional Stable Distributions*, Am. Math. Soc., Providence, R. I.

P. Chakraborty, C. Y. Lim, and M. M. Meerschaert, Department of Statistics and Probability, Michigan State University, Wells Hall, East Lansing, MI 48823, USA. (chakrab@stt.msu.edu; lim@stt.msu.edu; mcubed@stt.msu.edu)