AGU PUBLICATIONS

Water Resources Research

TECHNICAL REPORTS: METHODS

10.1002/2016WR019748

Key Points:
- `FracFit` is a parameter estimation tool based on four space-fractional and time-fractional models used by the hydrology community
- Parameter estimates are extracted from measured breakthrough curves using a weighted nonlinear least squares algorithm
- Future models may be implemented within the framework, allowing intercomparison of models

Correspondence to:
J. F. Kelly,
kellyja8@stt.msu.edu

Citation:
Kelly, J. F., D. Bolster, M. M. Meerschaert, J. D. Drummond, and A. I. Packman (2017), FracFit: A robust parameter estimation tool for fractional calculus models, *Water Resour. Res.*, *53*, 2559–2567, doi:10.1002/2016WR019748.

Received 3 SEP 2016
Accepted 1 MAR 2017
Accepted article online 7 MAR 2017
Published online 30 MAR 2017

FracFit: A robust parameter estimation tool for fractional calculus models

James F. Kelly[1], Diogo Bolster[2], Mark M. Meerschaert[1], Jennifer D. Drummond[3], and Aaron I. Packman[4]

[1]Department of Probability and Statistics, Michigan State University, East Lansing, Michigan, USA, [2]Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, South Bend, Indiana, USA, [3]Integrative Freshwater Ecology Group, Centre for Advanced Studies of Blanes (CEAB-CSIC), Blanes, Girona, Spain, [4]Department of Civil and Environmental Engineering, Northwestern University, Evanston, Illinois, USA

Abstract Anomalous transport cannot be adequately described with classical Fickian advection-dispersion equations (ADE) with constant coefficients. Rather, fractional calculus models may be used, which capture salient features of anomalous transport (e.g., skewness and power law tails). `FracFit` is a parameter estimation tool based on space-fractional and time-fractional models used by the hydrology community. Currently, four fractional models are supported: (1) space-fractional advection-dispersion equation (sFADE), (2) time-fractional dispersion equation with drift (TFDE), (3) fractional mobile-immobile (FMIM) equation, and (4) temporally tempered Lévy motion (TTLM). Model solutions using pulse initial conditions and continuous injections are evaluated using stable distributions or subordination integrals. Parameter estimates are extracted from measured breakthrough curves (BTCs) using a weighted nonlinear least squares (WNLS) algorithm. Optimal weights for BTCs for pulse initial conditions and continuous injections are presented. Two sample applications are analyzed: (1) pulse injection BTCs in the Selke River and (2) continuous injection laboratory experiments using natural organic matter. Model parameters are compared across models and goodness-of-fit metrics are presented, facilitating model evaluation.

1. Introduction

Anomalous transport cannot be adequately described with classical Fickian advection-dispersion equations (ADE) with constant coefficients [*Metzler and Klafter*, 2004; *Neuman and Tartakovsky*, 2009]. So-called anomalous transport is quite ubiquitous, spanning a multitude of scientific disciplines [*Klages et al.*, 2008], including the hydrologic sciences where it has been observed in both surface [*Deng et al.*, 2006; *Phanikumar et al.*, 2007; *Haggerty et al.*, 2002; *Aubeneau et al.*, 2014] and subsurface [*Benson et al.*, 2001; *Berkowitz and Scher*, 1997; *Cortis and Berkowitz*, 2004; *Wang and Cardenas*, 2014; *LeBorgne and Gouze*, 2008; *Becker and Shapiro*, 2000] water environments. Anomalous transport is characterized by subdiffusive or superdiffusive spreading of a plume, as inferred from the growth rate of its second centered moment, as well as heavy power law tails in concentration distributions and breakthrough curves (BTCs).

Several modeling approaches have been developed for anomalous diffusion, including continuous time random walks (CTRW) [*Berkowitz et al.*, 2006; *Boano et al.*, 2007], multirate mass transfer (MRMT) [*Haggerty and Gorelick*, 1995] and fractional advection-dispersion equations [*Benson et al.*, 2000]. All have enjoyed remarkable success in matching observations from experiments, spanning laboratory to field scales. For both CTRW [*Cortis and Berkowitz*, 2005] and MRMT [*Haggerty*, 2009], publicly available computational toolboxes for parameter estimation exist. Alternative modeling approaches include spatial and temporal Markov models [*LeBorgne et al.*, 2008; *Meyer and Tchelepi*, 2010] and the adjoint equation method [*Maryshev et al.*, 2016]. The goal of this paper is to describe a new toolbox for fractional advection-dispersion models [*Liu et al.*, 2003; *Schumer et al.*, 2003; *Meerschaert et al.*, 2008]. Given the historical success of fractional calculus in hydrology [e.g., *Benson et al.*, 2001; *Chakraborty et al.*, 2009; *Shen and Phanikumar*, 2009], such a general tool is desirable, allowing for improved intermodel comparison and rapid model validation, as well as enabling use by a broader fraction of the hydrologic community, not to mention countless other disciplines where fractional dispersion models are used.

Motivated by this need, we have developed `FracFit`, a parameter estimation tool based on common space-fractional and time-fractional models. `FracFit` is modular, allowing new models to be developed, implemented, verified for correctness, and tested in a rapid fashion. A current version is available on GitHub (https://github.com/jfk-inspire/FracFit-v-0.9). This technical report provides a summary of the models and numerics used in `FracFit`, which includes novel optimal weights used in the weighted nonlinear least squares (WNLS) algorithm for parameter estimation. We then apply `FracFit` to two data sets, which have not previously been interpreted with fractional models, illustrating the automated fitting of pulse and continuous injection BTCs. Space-fractional, time-fractional, and tempered-fractional models are discussed and compared.

2. Overview of Fractional Models

`FracFit` is a collection of MATLAB scripts that find the optimal parameter vector θ for a particular fractional model. At present, four representative models are implemented; the code is modular allowing additional models to be implemented with relative ease. In particular, all models use a common interface. Here we consider the following four forms of FADE commonly used in hydrology:

1. Space-fractional advection-dispersion equation (sFADE) [*Benson et al*., 2000].
2. Time-fractional dispersion equation with drift (TFDE) [*Liu et al*., 2003].
3. Fractional mobile-immobile (FMIM) equation [*Schumer et al*., 2003].
4. Temporally tempered Lévy motion (TTLM) [*Meerschaert et al*., 2008].

For each model, we consider two setups and solve for concentration *C*(*x*, *t*). These are (i) a pulse initial condition $C(x,t{=}0){=}K\delta(x)$ on $-\infty < x < \infty$ where *K* is initial mass and (ii) a continuous injection $C(x,t{=}0){=}0$ and $C(x{=}0,t){=}C_0$, where C_0 is a prescribed concentration, on $0 < x < \infty$. The governing equations and solutions for each of the four models are summarized in Table 1. The sFADE model involves positive and negative Riemann-Liouville derivatives on the real line. The TFDE involves a Caputo derivative on the half-axis. The FMIM model utilizes a Riemann-Liouville derivative on the half-axis. The TTLM model utilizes a tempered Riemann-Liouville derivative on the half-axis. For the FMIM and TTLM models, the governing equations are for the *mobile* phase. The solutions are tabulated in terms of stable probability density functions (PDFs), stable cumulative density functions (CDFs), and subordination integrals, which can be calculated

Table 1. Summary of Models Available in `FracFit`[a]

Model	Governing Equation	Pulse Initial Condition Solution	Continuous Injection Solution
sFADE	$\frac{\partial C}{\partial t}+v\frac{\partial C}{\partial x}=D\frac{1+\beta}{2}\frac{\partial^\alpha C}{\partial x^\alpha}+D\frac{1-\beta}{2}\frac{\partial^\alpha C}{\partial(-x)^\alpha}$	$\frac{H(t)}{(Dt)^{1/\alpha}}f_{\alpha,\beta}\left(\frac{x-vt}{(Dt)^{1/\alpha}}\right)$	$\bar{F}_{\alpha,\beta}\left(\frac{x-vt}{(Dt)^{1/\alpha}}\right)$
TFDE	$\left(\frac{\partial}{\partial t}\right)^\gamma C=-v\frac{\partial C}{\partial x}+D\frac{\partial^2 C}{\partial x^2}$	$\int_0^\infty h_\gamma(u,t)C_{ADE}(x,u)\,du$	$\int_0^\infty h_\gamma(u,t)C_{CBTC}(x,u)\,du$
FMIM	$\frac{\partial C}{\partial t}+\beta\frac{\partial^\gamma C}{\partial t^\gamma}+v\frac{\partial C}{\partial x}=D\frac{\partial^2 C}{\partial x^2}$	$\int_0^t g_\gamma(t-u,\beta u)C_{ADE}(x,u)\,du$	$\int_0^t g_\gamma(t-u,\beta u)C_{CBTC}(x,u)\,du$
TTLM	$\frac{\partial C}{\partial t}+\beta\frac{\partial^{\gamma,\lambda} C}{\partial t^{\gamma,\lambda}}+v\frac{\partial C}{\partial x}=D\frac{\partial^2 C}{\partial x^2}$	$\int_0^t g_{\gamma,\lambda}(t-u,\beta u)C_{ADE}(x,u)\,du$	$\int_0^t g_{\gamma,\lambda}(t-u,\beta u)C_{CBTC}(x,u)\,du$

Function	Equation
ADE solution	$C_{ADE}(x,u)=\frac{K}{\sqrt{4\pi Du}}\exp\left(-\frac{(x-vu)^2}{4Du}\right)$
CBTC solution	$C_{CBTC}(x,u)=\frac{C_0}{2}\operatorname{erfc}\left(\frac{x-vu}{2\sqrt{Du}}\right)$
Stable subordinator density	$g_\gamma(t,u)=u^{-1/\gamma}g_\gamma\left(tu^{-1/\gamma}\right)$
Tempered stable subordinator density	$g_{\gamma,\lambda}(t,u)=e^{-\lambda t+u\beta\lambda^\gamma}g_\gamma(t,u)$
Inverse stable subordinator density	$h_\gamma(u,t)=\frac{t}{\gamma}u^{-1-1/\gamma}g_\gamma\left(\frac{t}{u^{1/\gamma}}\right)$

[a]Pulse and continuous injection solutions are tabulated for each model in terms of stable distributions or subordination integrals. For sFADE, $f_{\alpha,\beta}(z)$ denotes the stable PDF and $\bar{F}_{\alpha,\beta}(z)$ denotes stable complementary CDF and *H*(*t*) is the Heaviside function. For TFDE, $h_\gamma(u, t)$ denotes the density of the inverse stable subordinator. For FMIM, $g_\gamma(t,u)$ denotes the density of the stable subordinator. Finally, for TTLM, $g_{\gamma,\lambda}(t,u)$ denotes the density of the tempered stable subordinator. The stable density $f_{\alpha,\beta}(z)$, complimentary CDF $\bar{F}_{\alpha,\beta}(z)$, and the stable subordinator density $g_\gamma(u)$ are computed using the STABLE toolbox [*Nolan*, 1997], freely available codes [*Veillette*, 2012], or MATLAB's Statistics and Machine Learning Toolbox.

Table 2. Summary of Parameters θ_i for Four Fractional Hydrology Models: (1) sFADE, (2) TFDE, (3) FMIM, and (4) TTLM[a]

Model	Parameters	Units	Lower and Upper Bounds θ_l and θ_u
sFADE	Stable index α	Unitless	$1 < \alpha \leq 2$
$\theta_1=(\alpha,\beta,v,D)$	Skewness β	Unitless	$-1 \leq \beta \leq 1$
	Average plume velocity v	$[L/T]$	$v>0$
	Fractional dispersivity D	$[L^\alpha/T]$	$D>0$
TFDE	Time-fractional exponent γ	Unitless	$0 < \gamma \leq 1$
$\theta_2=(\gamma,v,D)$	Fractional velocity v	$[L/T^\gamma]$	$v>0$
	Fractional dispersivity D	L^2/T^γ	$D>0$
FMIM	Time-fractional exponent γ	Unitless	$0 < \gamma \leq 1$
$\theta_3=(\gamma,v,\beta,D)$	Average plume velocity v	$[L/T]$	$v>0$
TTLM	Capacity coefficient β	$1/T^\gamma$	$\beta > 0$
$\theta_4=(\gamma,v,\beta,D,\lambda)$	Fractional dispersivity D	$[L^2/T]$	$D>0$
	Tempering rate λ	$[1/T]$	$\lambda > 0$

[a]Parameters, units, and default parameter lower bounds θ_l and upper bounds θ_u are given, where L denotes a unit of length and T denotes a unit of time. The user has the option to modify θ_l and θ_u for a particular data set. For pulse initial condition $C(x,0)=K\delta(x)$ problems, the initial mass $K>0$ is an additional parameter.

with widely available STABLE toolboxes [e.g., *Nolan*, 1997; *Veillette*, 2012] or MATLAB's Statistics and Machine Learning Toolbox (R2016a and later).

Details on each of these models are available in the noted references. The parameter vector θ_i associated with each model is listed in Table 2, along with a description of each parameter, parameter units, and bounds for each parameter.

3. Parameter Estimation

`FracFit`'s parameter estimation is based on the weighted nonlinear least squares (WNLS) approach developed in *Chakraborty et al.* [2009]. The original method is directly applicable for pulse initial condition cases as the solutions are either scalar multiples of PDFs or subordination integrals involving PDFs. For the continuous injection cases, the solutions involve CDFs or subordinated CDFs; for these functions, the specific techniques presented in *Chakraborty et al.* [2009] do not hold and the estimation method requires modification. Here we briefly review the WNLS method and propose an extension for the estimation of CDFs required for continuous injection cases.

Using a particle-tracking model, *Chakraborty et al.* [2009] showed that concentration variance is proportional to concentration, implying that data are heteroscedastic; therefore, a weighted nonlinear regression is used where the weights are proportional to the reciprocal of measured concentration. As a result, areas of lower concentration receive greater weight, which is important for capturing anomalous transport characteristics. Assuming we have N measurements of a BTC C_i at times $t_1,\ldots,t_N$, we wish to fit a candidate analytical model $C(x,t)$ to the observed data by minimizing the weighted mean square error (WMSE) function:

$$E(\theta)=\frac{1}{N}\sum_{i=1}^{N} w_i(C_i-C(x,t_i))^2, \tag{1}$$

where $C(x,t)$ is the appropriate PDF and the weights are given by $w_i=1/C_i$. These weights are applicable to any BTC that can be normalized into PDFs, including bimodal or multimodal BTCs. However, all the fractional calculus models considered in this report have solutions that are unimodal.

The continuous injection breakthrough curves (CBTCs) are fit in terms of a CDF instead of a PDF; hence, we expect a different set of weights w_i. In Appendix B, we construct an estimator for the CDF showing that the optimal weights for CBTCs are

$$w_i=\frac{1}{(1-C_i^*)C_i^*}, \tag{2}$$

where $C_i^*=C_i/C_0$. Hence, the weights are largest when C_i^* is near either one or zero; i.e., at early and late arrival times, similar to the lower concentrations in the pulse case at early and late times. Since the measured normalized CBTC contains some (relative) experimental error of order $\epsilon \ll 1$, we assign weights of

Table 3. Parameter Estimates for a Synthetic Breakthrough Curve Using (a) sFADE, (b) FMIM, and (c) TTLM

(a) sFADE

Parameter	α	β	v	D	K
Known θ_t	1.3	-1	0.02	0.002	25
Estimated θ	1.3	-0.99	0.02	0.002	24.9

(b) FMIM

Parameter	γ	v	β	D	K
Known θ_t	0.85	0.03	0.12	1.0×10^{-5}	25.0
Estimated θ	0.841	0.0297	0.111	1.02×10^{-5}	24.80

(c) TLLM

Parameter	γ	v	β	D	λ	K
Known θ_t	0.85	0.0300	0.12	1.00×10^{-5}	0.003	25.0
Estimated θ	0.855	0.0301	0.125	1.00×10^{-5}	0.00247	24.42

zero if $C_i < \epsilon$ or $C_i > (1-\epsilon)$. We note that this truncation is a modeling choice and may bias the fit. Alternatively, the variance of the CBTC may be modeled as $\sigma_i^2=\max\left(0,(1-C_i^*)C_i^*\right)+\epsilon$, thereby modifying equation (2). For pulse initial conditions, the variance may be modeled as $\sigma_i^2=\max\left(0,C_i^*\right)+\epsilon$. The curve fits in sections 4 and 5 use truncation, while the nontruncated weights are provided as an option in `FracFit`.

The WMSE function given by equation (1) is optimized with respect to θ using the local optimization `lsqnonlin` routine in MATLAB's Optimization Toolbox. Since `lsqnonlin` finds a local minimum to the objective function (1), `FracFit` requires a reasonable guess θ_0 to find a global minimum. For sFADE, we first fit the ADE to find (v, D) and then set $\alpha=1.5$ and $\beta=0$ as the initial guess. Similarly, for TFDE, we use (v, D) from the ADE fit and set $\gamma=0.9$. For the FMIM initial guess, we numerically compute the median and mode and estimate v and β assuming $\gamma=0.75$. Finally, for TLLM, we use the FMIM initial guess and set $\lambda=1/\max(t)$. We stress that these estimates are ad hoc and may not be appropriate for all data sets. Hence, we also allow the user to manually select both an initial guess θ_0 and a lower bound θ_l and upper bound θ_u of the search region.

Since local optimization may not converge for all data sets, we have also implemented a global optimization option using a genetic algorithm `(ga)` routine [*Conn et al.*, 1991] in MATLAB's Global Optimization Toolbox. The `ga` is much more expensive than `lsqnonlin`. Future generations of `FracFit` may utilize a two-step optimization scheme, where global optimization is used to find the initial guess for the local optimization scheme.

To evaluate the goodness of fit (GOF), we calculated the mean absolute residual (MAR) defined by

$$MAR=\frac{1}{N}\sum_{i=1}^{N}|C_i-C(x,t_i)/C_0|. \tag{3}$$

MAR quantifies the mean error between model and data and demonstrates the relative change in error reduction achieved by applying different models to the same data set. Alternative GOF measures, such as the (corrected) Akaike information criterion (AIC$_c$), are only valid for maximum likelihood estimation, which we have not implemented in `FracFit`.

As an initial test, we generated synthetic pulse injection data for the sFADE, FMIM, and TTLM models and present a representative subset here. The time-axis consisted of 400 samples logarithmically spaced on [40, 2000] with an observation point of $x=1.5$. A known parameter θ_t was chosen for each model to produce a synthetic BTC that resembled measured data. `FracFit` was then used to estimate θ. The results of this experiment are shown in Table 3 in dimensionless units. The MAR for sFADE, FMIM, and TTLM are 0.00324, 0.00950, and 0.00488, respectively. For this data set, `FracFit` is able to estimate the known parameters for all the models, although the estimate for the tempering rate λ in TTLM is off by about 20%. This is unsurprising and we note that the algorithm is sensitive to the number, duration, and sampling of the synthetic BTC. For the tempering parameter in TTLM, estimates will be poor if the duration of the BTC is limited relative to the tempering time scale [*Aubeneau et al.*, 2014].

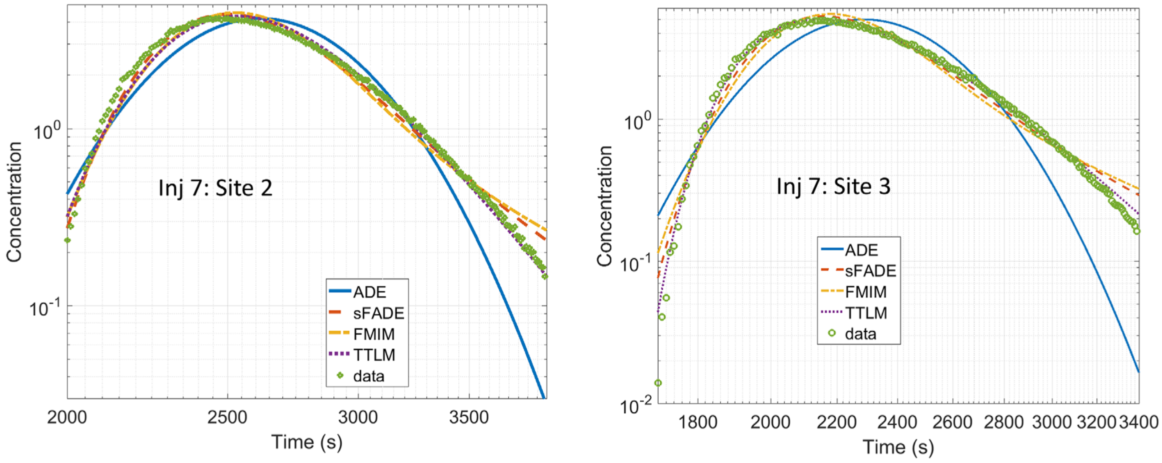

Figure 1. Model intercomparison using Selke River data for (left) Injection 7: Site 2 and (right) Injection 7: Site 3. The ADE, sFADE, FMIM, and TTLM models are fit to a pulse injection BTCs at the four sites.

Table 4. Parameter Estimates for the Selke River Breakthrough Curve Using (a) sFADE, (b) FMIM, and (c) TTLM Models for Injection 7: Sites 2 and 3

(a) sFADE

BTC	α	β	v (m/s)	D (m$^{\alpha}$/s)	K (ppm)
Inj 7: Site 2	1.59	-1	0.339	0.549	1306.8
Inj 7: Site 3	1.49	-1	0.338	0.376	1330.9

(b) FMIM

BTC	γ	v (m/s)	β (s$^{\gamma-1}$)	D (m^2/s)	K (ppm)
Inj 7: Site 2	0.78	0.421	0.0528	1.563	1693.2
Inj 7: Site 3	0.79	0.445	0.0717	1.048	1796.7

(c) TLLM

BTC	γ	v (m/s)	β (s$^{\gamma-1}$)	D (m^2/s)	λ (s^{-1})	K (ppm)
Inj 7: Site 2	0.67	0.498	0.0948	1.166	0.00219	106098
Inj 7: Site 3	0.64	0.568	0.110	0.108	0.00233	60173

4. Application 1: Pulse Initial Condition Breakthrough Curves From Transport Experiment in the Selke River

Our first example with observed data is a series of in-stream pulse injection experiments conducted in the Selke River [*Schmadel et al*., 2016]. In this experiment, there were seven in-stream monitoring sites that were sampled throughout each of the seven tracer injection experiments, leading to 49 BTCs. Three fractional models are evaluated: sFADE, FMIM, and TTLM, as well as the ADE. `FracFit` is useful for this study in terms of efficiency and consistency in BTC fitting, especially when considering multiple models.

Four representative BTCs were selected from the data set: two from the first injection, measured at site 6 ($x = 428$ m) and site 7 ($x = 294$ m), and two from the seventh injection, measured at site 2 ($x = 928$ m) and site 3 ($x = 819$ m). The BTC fits for the seventh injection and measured concentration data are shown on log-log scale in Figure 1. Parameter estimates for these BTCs are shown in Table 4. A GOF metric (MAR) evaluated for the three fractional models and the ADE are shown in Table 5 for all four BTCs.

Table 5. MAR for the Selke River BTCs for ADE, sFADE, FMIM, and TTLM Models

BTC	ADE	sFADE	FMIM	TTLM
Inj 1: Site 6	0.3509	0.0459	0.0556	0.0537
Inj 1: Site 7	0.4837	0.0566	0.0640	0.0586
Inj 7: Site 2	0.3248	0.1099	0.1704	0.1193
Inj 7: Site 3	0.4927	0.1630	0.2515	0.0631

Examining the fits in Figure 1, note that neither the main plume nor the heavy late-time tail was captured by ADE for any of the BTCs shown. For the sFADE model, all fits were negatively skewed

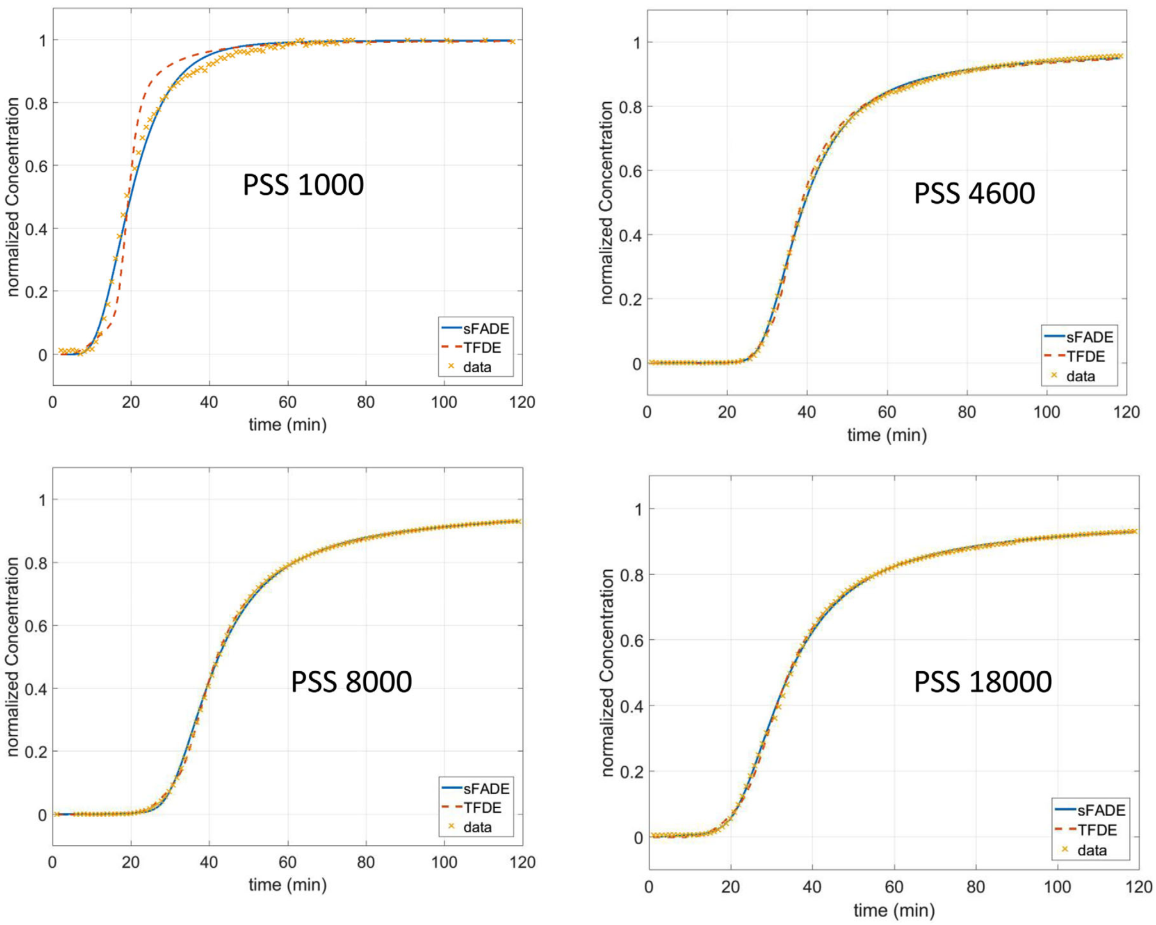

Figure 2. Parameter fits $\theta_1=(\alpha,\beta,v,D)$ for the sFADE model and $\theta_2=(\gamma,v,D)$ for the TFDE model for the four PSS continuous injection BTCs.

with $\beta=-1$, which agrees with earlier studies [*Chakraborty et al*., 2009; *Deng et al*., 2004]. This negative skewness has been attributed to retention and the existence of "dead zones." While sFADE provides acceptable fits for the BTCs under consideration, sFADE does admit nonphysical behavior (negative dispersion) that may manifest itself at other measurement locations/times. Space-time duality calculations in *Baeumer et al*. [2009] show an equivalence between space-fractional and time-fractional models, which may account for the good sFADE fit in Figure 1 and provide a more physical interpretation.

Examining the fits in the left and right sides of Figure 1, sFADE, FMIM, and TTLM yielded a better fit than ADE. However, sFADE and FMIM failed to capture the late-time truncation of the power law, while TTLM captured this feature. Recall that TTLM imposes an exponential cutoff to power law waiting times, allowing TTLM to transition from anomalous to Fickian transport [*Meerschaert et al*., 2008]. This transition is governed by the tempering rate λ. We note that simultaneous estimation of the capacity coefficient β and tempering rate λ is problematic with a single (mobile) BTC since the parameters act in a coupled fashion.

To address this problem, additional data, such as the BTC at another location, or measured mobile or immobile mass, may be utilized [e.g., *Briggs et al*., 2009]. As an example, we simultaneously fit the BTCs for Sites 2 and 3 using Injection 7. We allowed the velocities v_i and

Table 6. Parameter Fit $\theta_1=(\alpha,\beta,v,D)$ for the sFADE Model

Sample	α	β	v (cm/min)	D (cm$^\alpha$/min)	MAR
PSS1000	1.9565	-1	0.33108	0.17855	0.01639
PSS4600	1.4404	-0.93009	0.145	0.050669	0.00410
PSS8000	1.4095	-0.88437	0.12994	0.059243	0.00349
PSS18000	1.4475	-0.66565	0.1629	0.11041	0.00369
NOM	1.08956	0.16792	0.13126	0.27184	0.00543
HPOAs	1.04927	0.04555	0.10141	0.44257	0.00423
TPIAs	1.21822	0.04564	0.09207	0.24646	0.00304

Table 7. Parameter Fit $\theta_2=(\gamma, v, D)$ for the TFDE Model

Sample	γ	v (cm/min$^{\gamma}$)	D (cm^2/min)	MAR
PSS1000	0.98087	0.36339	0.78824	0.03611
PSS4600	0.91581	0.21482	0.020473	0.00769
PSS8000	0.90001	0.21157	0.03856	0.00228
PSS18000	0.8695	0.28709	0.13912	0.00453
NOM	0.95682	0.11083	1.96505	0.01423
HPOAs	0.84591	0.15961	0.51451	0.01550
TPIAs	0.75066	0.25419	1.15895	0.00980

dispersion coefficients D_i to vary between the sites but used the same exponent γ, capacity coefficient β, and tempering rate λ, yielding a parameter $\theta=(\gamma, \beta, \lambda, v_1, v_2, D_1, D_2)$ with seven degrees of freedom. This simultaneous fit yielded estimates of the exponent $\gamma=0.63$, capacity coefficient $\beta=$ 0.115 s$^{\gamma-1}$ and tempering rate $\lambda=$ 0.00239 s^{-1}. Parameters such as γ, β, and λ may also be allowed to vary with downstream distance. Analyzing multiple BTCs shows the variability of model parameters of a given stream and demonstrates local variations in transport and storage. Simultaneous fits for other models, such as sFADE, are also available in `FracFit`.

5. Application 2: Continuous Injection Breakthrough Curves From Natural Organic Matter (NOM) Transport

As a second example, we fit continuous injection breakthrough curves (CBTCs) from laboratory experiments. These experiments studied transport of organic matter through porous media columns and displayed strong anomalous transport characteristics [*Dietrich et al*., 2013; *McInnis et al*., 2014, 2015]. The data were originally fit with a CTRW model using the CTRW toolbox [*Cortis and Berkowitz*, 2005].

Two data sets are considered: (1) synthetic polystyrene sulfonates (PSSs) in columns packed with naturally Fe/Al-oxide-coated sands from Oyster, Virginia [*McInnis et al*., 2015] and (2) dissolved organic matter (DOM) from Nelson's Creek, MI, in a column of porous medium (oxide-coated quartz sand) [*McInnis et al*., 2014]. Both are continuously injected through sands via a gravity feed system with concentration measured at the outlet. Full details of the experiments are available in *McInnis et al*. [2014, 2015].

Figure 2 displays the sFADE and TFDE fits for the PSS samples. Comparable fits (not shown) were obtained for the DOM cases. The fitted parameter $\theta_1=(\alpha, \beta, v, D)$ for the sFADE and $\theta_2=(\gamma, v, D)$ for the TFDE are shown in Tables 6 and 7, respectively, along with the mean absolute residual (MAR), allowing comparison with the CTRW model fits from *McInnis et al*. [2015].

For both models, PSS1000 yields the poorest fit, with an MAR an order of magnitude larger than all others. For all cases, except PSS8000, the sFADE appears to yield slightly smaller MAR, although it benefits from having one additional free parameter. Generally, the MAR is comparable to those obtained by the CTRW in *McInnis et al*. [2015]. Our goal is not to compare CTRW and FADE model fits but rather demonstrates `FracFit`'s ability to interpret a continuous injection anomalous transport breakthrough curve, which is clearly shown here.

6. Conclusion

`FracFit` is a flexible tool that facilitates parameter estimation for a variety of models, such as sFADE, TFDE, FMIM, and TTLM. Future models may be implemented within this framework; since models are treated in a consistent manner, intercomparison of models may be performed seamlessly. The user may choose either a local, gradient-based optimization scheme or a global optimization scheme. One interesting application is studying the duality between space-fractional and time-fractional models [*Baeumer et al*., 2009]: under certain conditions, a time-fractional model can be equivalent to a space-fractional model.

Appendix A: Derivation of an Approximate sFADE CBTC Expression

The CBTC solution requires a fixed boundary condition at $x=0$; however, no closed form analytical solution exists at this time. The CBTC solution may be approximated by the "dam break" problem on the real line. We derive an analytical approximation following what is done for the classical ADE ($\alpha=2$) in *Danckwerts* [1953]. Consider the sFADE model (top row of Table 1) on $-\infty < x < \infty$ subject to initial condition $C_0(x, 0)=C_0$ if $x<0$ and $C_0(x, 0)=0$ if $x \geq 0$. Using the sFADE pulse initial condition solution, the CBTC solution is approximated by

10.1002/2016WR019748

$$C(x,t)=\int_{-\infty}^{\infty} C_{\mathrm{sFADE}}(x',0)G(x-x',t)\,dx', \tag{A1}$$

where $G(x, t)$ is the Green's function of sFADE. Evaluating the integral in equation (A1) yields

$$\frac{C(x,t)}{C_0}=1-F_{\alpha,\beta}\left(\frac{x-vt}{(Dt)^{1/\alpha}}\right)=\bar{F}_{\alpha,\beta}\left(\frac{x-vt}{(Dt)^{1/\alpha}}\right), \tag{A2}$$

where $\bar{F}_{\alpha,\beta}(z)$ denotes the complementary CDF function (survival function). To verify this approximation, we compared it to a complete numerical solution in *Zhang et al*. [2007]. Agreement between equation (A2) and the numerical solution is very good, indicating that equation (A2) is a good approximation for continuous injection BTCs.

Appendix B: Optimal Weights for CBTCs

Assume we have n statistically independent particles representing the tracer plume. The time-dependent location of the k-th particle is given by the random variable $X_t^{(k)}$, which is distributed according to the density $f_\theta(x,t)$. The vector θ specifies the model parameters, and the CDF, as in equation (A2), is $F_\theta(x,t)=\int_{-\infty}^{x} f_\theta(x',t)\,dx'$. We construct an estimator of $F_\theta(x,t)$ via the *empirical cumulative distribution function* [*van der Vaart*, 1998, chapter 19]:

$$\hat{F}_\theta(x,t)=\frac{1}{n}\sum_{k=1}^{n} I\left(X_t^{(k)}\leq x\right), \tag{B1}$$

where $I(X\leq x)$ is the indicator function defined such that $I(X\leq x)=1$ if $X\leq x$ and zero otherwise. Suppressing the time dependence, the expected value of equation (B1) is

$$\begin{aligned} E\left[\hat{F}_\theta(x)\right] &= \frac{1}{n}\sum_{k=1}^{n}\int_{-\infty}^{\infty} I(x'\leq x)f_\theta(x')\,dx' \\ &= \frac{1}{n}\sum_{k=1}^{n} F_\theta(x) \\ &= F_\theta(x), \end{aligned} \tag{B2}$$

indicating that the empirical CDF is an unbiased estimator. Calculating moments using the standard argument for Kolmogorov-Smirnov statistics [*van der Vaart*, 1998, chapter 19] yields

$$E\left[\hat{F}_\theta(x)\hat{F}_\theta(y)\right]=\frac{1}{n}F_\theta(\min(x,y))+\frac{n-1}{n}F_\theta(x)F_\theta(y). \tag{B3}$$

Use equation (B2) along with the identities $\mathrm{Var}[X]=\mathrm{E}[X^2]-(\mathrm{E}[X])^2$ and $\mathrm{Cov}[X,Y]=\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]$, yielding

$$\mathrm{Var}\left[\hat{F}_\theta(x)\right]=\frac{1}{n}F_\theta(x)(1-F_\theta(x)) \tag{B4a}$$

and

$$\mathrm{Cov}\left[\hat{F}_\theta(x),\hat{F}_\theta(y)\right]=\frac{1}{n}\left[F_\theta(\min(x,y))-F_\theta(x)F_\theta(y)\right]. \tag{B4b}$$

For n particles, we have

$$\mathrm{Var}\left[\sqrt{n}\hat{F}_\theta(x)\right]=F_\theta(x)(1-F_\theta(x)). \tag{B5}$$

Unlike the PDF estimator in *Chakraborty et al*. [2009], the covariance does not approach zero, implying that measurements of the CDF are correlated. Numerical evaluation of the covariance showed that the correlation was small, so weighted nonlinear least squares was chosen over generalized least squares, which minimizes the functional $Q(\theta)=[\mathbf{C}-F_\theta(\mathbf{x})]^T\Sigma_\theta^{-1}[\mathbf{C}-F_\theta(\mathbf{x})]$. Under this small correlation assumption, equation (B5) implies that the variance of the CDF is proportional to $C_i(1-C_i)$. Since the CBTC solutions for all models under consideration are either complementary CDFs or subordinated CDFs, we conclude that the CBTC has a variance proportional to $C_i(1-C_i)$, yielding equation (2).

Acknowledgments
Kelly was partially supported by ARO MURI grant W911NF-15-1-0562 and NSF grant EAR-1344280. Meerschaert was partially supported by ARO MURI grant W911NF-15-1-0562 and NSF grants DMS-1462156 and EAR-1344280. Bolster was partially supported by NSF grants EAR-1351625 and EAR-1417264. Packman was supported by NSF grant EAR-1344280 and ARO grant W911NF-15-1-0569. John Nolan (Department of Mathematics and Statistics, American University, Washington, DC) graciously provided the STABLE toolbox (www.RobustAnalysis.com). We acknowledge Noah Schmadel and Adam S. Ward (Department of Environmental Engineering, Indiana University) for providing the Selke River data. Financial support for the Selke experiment was provided by The Leverhulme Trust through the project "Where rivers, groundwater and disciplines meet: A hyporheic research network." Insightful comments by Yong Zhang, Department of Geological Sciences, University of Alabama, are also acknowledged. The Selke River data are available from Adam S. Ward (adamward@indiana.edu). The NOM data are available from Bolster (bolster@nd.edu).

References

Aubeneau, A., B. Hanrahan, D. Bolster, and J. Tank (2014), Substrate size and heterogeneity control anomalous transport in small streams, *Geophys. Res. Lett.*, *41*, 8335–8341, doi:10.1002/2014GL061838.

Baeumer, B., M. M. Meerschaert, and E. Nane (2009), Space–time duality for fractional diffusion, *J. Appl. Probab.*, *46*(4), 1100–1115.

Becker, M. W., and A. M. Shapiro (2000), Tracer transport in fractured crystalline rock: Evidence of nondiffusive breakthrough tailing, *Water Resour. Res.*, *36*(7), 1677–1686.

Benson, D. A., S. W. Wheatcraft, and M. M. Meerschaert (2000), The fractional-order governing equation of Lévy motion, *Water Resour. Res.*, *36*(6), 1413–1423.

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

Berkowitz, B., and H. Scher (1997), Anomalous transport in random fracture networks, *Phys. Rev. Lett.*, *79*, 4038.

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*(2), RG2003, doi:10.1029/2005RG000178.

Boano, F., A. 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.

Briggs, M. A., M. N. Gooseff, C. D. Arp, and M. A. Baker (2009), A method for estimating surface transient storage parameters for streams with concurrent hyporheic storage, *Water Resour. Res.*, *45*, W00D27, doi:10.1029/2008WR006959.

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.

Conn, A. R., N. I. M. Gould, and Ph. L. Toint (1991), A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds, *SIAM J. Numer. Anal.*, *28*(2), 545–572.

Cortis, A., and B. Berkowitz (2004), Anomalous transport in "classical" soil and sand columns, *Soil Sci. Soc. Am. J.*, *68*, 1539–1548.

Cortis, A., and B. Berkowitz (2005), Computing anomalous contaminant transport in porous media: The CTRW MATLAB toolbox, *Ground Water*, *43*(6), 947–950.

Danckwerts, P. (1953), Continuous flow systems: Distribution of residence times, *Chem. Eng. Sci.*, *2*(1), 1–13.

Deng, Z., L. Bengtsson, and V. P. Singh (2006), Parameter estimation for fractional dispersion model for rivers, *Environ. Fluid Mech.*, *6*(5), 451–475.

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

Dietrich, L., D. McInnis, D. Bolster, and P. Maurice (2013), Effect of polydispersity on natural organic matter transport, *Water Res.*, *47*, 2231–2240.

Haggerty, R. (2009), *STAMMT-L 3.0. A Solute Transport Code for Multirate Mass Transfer and Reaction Along Flowlines*, Sandia Natl. Lab., Albuquerque, N. M.

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

Haggerty, R., S. M. Wondzell, and M. A. Johnson (2002), Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream, *Geophys. Res. Lett.*, *29*(13), doi:10.1029/2002GL014743.

Klages, R., G. Radons, and I. M. Sokolov (2008), *Anomalous Transport: Foundations and Applications*, Wiley-VCH Verl., Weinheim, Germany.

LeBorgne, T., and P. Gouze (2008), Non-Fickian dispersion in porous media: 2. Model validation from measurements *at different scales, Water Resour. Res.*, *44*, W06427, doi:10.1029/2007WR006279.

LeBorgne, T., M. Dentz, and J. Carrera (2008), Lagrangian statistical model for transport in highly heterogeneous velocity fields, *Phys. Rev. Lett.*, *101*, 090601.

Liu, F., V. V. Anh, I. Turner, and P. Zhuang (2003), Time fractional advection-dispersion equation, *J. Appl. Math. Comput.*, *13*(1), 233–245.

Maryshev, B., A. Cartalade, C. Latrille, and M.-C. Néel (2016), Identifying space-dependent coefficients and the order of fractionality in fractional advection diffusion equation, *Transp. Porous Media*, *116*(1), 53–71.

McInnis, D. P., D. Bolster, and P. A. Maurice (2014), Natural organic matter transport modeling with a continuous time random walk approach, *Environ. Eng. Sci.*, *31*(2), 98–106.

McInnis, D. P., D. Bolster, and P. A. Maurice (2015), Mobility of dissolved organic matter from the Suwannee River (Georgia, USA) in sand-packed columns, *Environ. Eng. Sci.*, *32*(1), 4–13.

Meerschaert, M. M., Y. Zhang, and B. Baeumer (2008), Tempered anomalous diffusion in heterogeneous systems, *Geophys. Res. Lett.*, *35*, L17403, doi:10.1029/2008GL034899.

Metzler, R., and J. Klafter (2004), The restaurant at the end of the random walk: Recent developments in the description of anomalous transport by fractional dynamics, *J. Phys. A Math. Gen.*, *37*(31), R161.

Meyer, D. W., and H. A. Tchelepi (2010), Particle-based transport model with Markovian velocity processes for tracer dispersion in highly heterogeneous porous media, *Water Resour. Res.*, *46*, W11552, doi:10.1029/2009WR008925.

Neuman, S. P., and D. M. Tartakovsky (2009), Perspective on theories of non-Fickian transport in heterogeneous media, *Adv. Water Resour.*, *32*(5), 670–680.

Nolan, J. P. (1997), Numerical calculation of stable densities and distribution functions, *Commun. Stat. Stochastic Models*, *13*(4), 759–774.

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.

Schmadel, N. M., et al. (2016), Stream solute tracer timescales changing with discharge and reach length confound process interpretation, *Water Resour. Res.*, *52*, 3227–3245, doi:10.1002/2015WR018062.

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., and M. S. Phanikumar (2009), An efficient space-fractional dispersion approximation for stream solute transport modeling, *Adv. Water Resour.*, *32*(10), 1482–1494.

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

Veillette, M. (2012), STBL: Alpha stable distributions for MATLAB, *Matlab Cent. File Exch. 10*. [Available at https://www.mathworks.com/matlabcentral/fileexchange/37514-stbl–alpha-stable-distributions-for-matlab.]

Wang, L., and M. Cardenas (2014), Non-Fickian transport through two-dimensional rough fractures: Assessment and prediction, *Water Resour. Res.*, *50*, 871–884, doi:10.1002/2013WR014459.

Zhang, X., M. Lv, J. W. Crawford, and I. M. Young (2007), The impact of boundary on the fractional advection–dispersion equation for solute transport in soil: Defining the fractional dispersive flux with the Caputo derivatives, *Adv. Water Resour.*, *30*(5), 1205–1217.