Comparison of Negative Skewed Space Fractional Models with Time Nonlocal Approaches for Stream Solute Transport Modeling

Liwei Sun[1, 2], Han Qiu[3], Jie Niu[1, 2*], Bill X. Hu[1, 2], James F. Kelly[4], Diogo Bolster[5] and Mantha S. Phanikumar[6*]

[1] Institute of Groundwater and Earth Science, Jinan University, Guangzhou 510632, China

[2]Green Development Institute of Zhaoqing, Zhaoqing 526000, China

[3]Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD 20740, USA; email address: han.qiu@pnnl.gov

[4]U.S. Naval Research Lab, Washington, DC 20375, USA

[5]Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN 46556, USA

[6]Department of Civil & Environmental Engineering, Michigan State University, East Lansing, MI 48824, USA

Correspondence to: Jie Niu, jniu@jnu.edu.cn; Mantha S. Phanikumar phani@egr.msu.edu

Abstract

Continuous time random walks (CTRW), multi-rate mass transfer (MRMT), and fractional advection-dispersion equations (FADEs) are three promising models of anomalous transport as commonly found in natural streams. Although these paradigms are mathematically related, understanding their advantages and limitations poses a challenge for model selection. In this paper, we quantitatively evaluate the advection-dispersion equation (ADE), fractional-mobile-immobile (FMIM), fractional-in-space (sFADE), fractional in space transient storage (FSTS), truncated time-fractional model (TTFM), and CTRW models with truncated power-law waiting time distribution (CTRW-TPL) by fitting them first to synthetic data. We then applied these models to observations from tracer experiments conducted in several rivers. Based on the extensive analysis, we conclude that

the FSTS model $\beta = -1$) is comparable, if not superior, to the other nonlocal models evaluated in the paper; therefore, the model represents an alternative to existing models for simulating stream solute transport for spatially-homogeneous flows.

Keywords:

transient storage; fractional derivative; hyporheic zone; continuous time random walk

1. Introduction

Anomalous or non-Fickian transport of solutes is often found in streams at all scales (*Burnell et al.*, 2017; *Edery et al.*, 2010; *Liu et al.*, 2017; *Shen and Phanikumar*, 2009; *Vishal and Leung*, 2015). Exchange of water and solutes with the hyporheic zone produces delays in transport relative to the mainstream flow, often leading to long (or heavy) tails in concentration breakthrough curves (BTCs) (*Boano et al.*, 2007). After decades of effort, modeling of non-Fickian transport of tracers continues to be a challenging problem. Beyond being a purely academic problem, observed heavy tails play a critical role in the transport of toxic chemicals where underestimation poses more risk, while overestimation can increase cleanup costs (*de Barros et al.*, 2013).

Among many others, three anomalous nonlocal models are most widely used in hydrology; these are continuous time random walks (CTRW), multi-rate mass transfer (MRMT), and fractional advection dispersion equations (FADE). These models are mathematically interrelated as reviewed in the supplementary material (SI). For example, an analytical relationship between MRMT and CTRW is established in *Dentz and*

Berkowitz (2003); the fractional-in-time and fractional-in-space ADEs are limit forms of CTRWs (*Metzler and Klafter*, 2000; *Schumer et al.*, 2009). The fractional-in-time ADE can be shown to be mathematically equivalent to the fractional-in-space ADE via space-time duality (*Kelly and Meerschaert*, 2017; 2019). In addition, the fractional mobile/immobile model (FMIM) is derived from the MRMT (with a special memory function (*Schumer and Benson*, 2003)). While the overarching frameworks are related, choices of special memory functions or residence time distributions (RTDs) have resulted in particularly popular sub-models, such as the CTRW with a truncated power law waiting time distribution function (CTRW-TPL), the single rate MRMT model (MRMT-1 hereafter), the lognormal diffusion rate MRMT (MRMT-2 hereafter), and the power-law distribution of mass exchange rates MRMT model (MRMT-3 hereafter). In addition, new forms of FADEs were introduced in hydrology such as the fractional mobile/immobile (FMIM) model, the truncated time fractional model (TTFM), and the fractional in space transient storage model (FSTS). The governing equations and non-local characteristics of the above models are summarized in Table 1.

The models in Table 1 are based on different underlying physical assumptions. Most obviously, they have different parameters each of which has a different physical interpretation and ranges of values. Despite significant progress in application of these models to describe solute transport in streams, the wide range of available sub-models and the complex relationships among them can be sources of confusion for selecting the best model for a given situation (e.g., solute transport in a river with significant hyporheic effect (SHE) that follows either a power-law or an exponential RTD). In particular, models with a large number of parameters also present the problem that they are difficult to optimize and could suffer from equifinality. Therefore, one of the objectives of this work was to address the following question: taking the number of model parameters into account, which model best captures BTCs with heavy tails? To address this question, we briefly review the interrelations between the different models first (see SI). Then, we generate synthetic data corresponding to different residence time distributions using the MRMT-1, MRMT-2, and MRMT-3 models and test the ability of the other non-local models to reproduce the synthetic data; finally we further evaluate all the models using tracer data from rivers with and without SHE.

This paper further explores the connection between space-fractional equations and other (CTRW and MRMT) approaches using numerical simulations and fits to observed tracer data to provide a clear physical and stochastic interpretation for space-fractional models in river flow hydrology. The FSTS model considered in our work is essentially the well-known transient storage (TS) model (*Bencala and Walters*, 1983; *Runkel*, 1993) with the second-order dispersion term in the TS model replaced with a more general fractional derivative term that includes positive and negative skewness terms. The TS model was applied extensively to address questions involving conservative and reactive transport in streams and rivers in the past. A major advantage of the model is that its parameters can be directly measured in the field using detailed velocity measurements or tracer studies or both (*Carr and Rehmann*, 2007, *Shen et al*., 2010; *Phanikumar et al*., 2007). The second-order dispersion term has been a source of some confusion in the TS modeling literature as some researchers found that a dispersion term was not needed to describe solute transport in some stream reaches (*Gupta and Cvetkovic*, 2000l; *Worman*, 1998). The analytical solution of the TS model (*De Smedt et al*., 2005) follows an exponential residence time distribution representing solute retention in either surface storage zones or shallow hyporheic regions but not in deep hyporheic zones. This separation of surface and deep hyporheic storage contributions is important for biogeochemical processes within streams (e.g., denitrification) and has been the focus of previous research (e.g., Briggs et al., 2009). One of the objectives of the paper is to further evaluate the FSTS model for its ability to represent both surface and hyporheic storages within a stream reach and to compare the performance of this model with other nonlocal approaches.

In section 2, we describe the different models considered in this paper. To understand the physical meaning of different parameters in the models, we evaluate the models using synthetic data generated using the three synthetic MRMT cases. In addition, we also compare parameter estimates based on the sFADE, FSTS, FMIM, TTFM, CTRW-TPL, MRMT-1, and MRMT-3 models with rhodamine WT (RWT) and sodium chloride tracer data collected in a total of 17 reaches in 4 different rivers. We also examine the physical interpretation of the backward dispersion term in the FSTS and sFADE models, given there is a debate related to previous studies as to whether this reflects a true physical process or an unphysical mathematical representation (*Zhang et al.*, 2009). In addition, we compare the sFADE and FSTS models to understand if there is any advantage of FSTS over sFADE and other time non-local models. Finally, we summarize our results and findings in the Conclusion section.

2. Methods

2.1 The FSTS and sFADE Model

The FSTS model (*Deng et al.*, 2006; *Shen and Phanikumar*, 2009) assumes a first-order mass exchange between the main channel and a storage zones (Eqs. (t1) and (t2) in Table 1). C is the solute concentration in the main channel (ML^{-3}) (L is the unit of distance), C_s is the concentration in the storage zone (ML^{-3}), v is the average water velocity (LT^{-1}), D is the coefficient of longitudinal dispersion $(L^{\alpha}T^{-1})$, x is the space coordinate in the flow direction (L), t is time (T), ε is a first-order exchange coefficient (T^{-1}), A is the main channel cross-sectional area (L^2), A_s is the size of the storage zones (L^2); C_L is the concentration associated with lateral inflow, and q_L $(L^3T^{-1}L^{-1})$ is the lateral inflow rate. The order of the fractional Riemann-Liouville (RL) derivative in (1) is $\alpha \in (1,2]$, while the parameter $\beta \in [-1,1]$ controls the skewness. When $\beta = 0$ dispersion in the main channel is symmetric. When $\beta < 0$, the solution of the FSTS is skewed backward, while when $\beta > 0$ the solution is skewed forward (*Zhang et al.*, 2005). When $\alpha = 2$ the FSTS reduces to the classical TS model for any choice of skewness β.

The fractional-in-space advection dispersion equation (sFADE) is a special case of the FSTS (with $\varepsilon = q_L = 0$), which only includes the mobile channel concentration (Eq. (t3) in Table 1). The analytical solution for sFADE with a pulse injection in an infinite domain (i.e. no boundary condition) (*Benson et al.*, 2000) is used in this study. Since the RL derivative of a constant is not zero, specifying boundary conditions for the continuous slug release and pulse injection of a tracer of a known constant concentration in a stream is problematic (*Baeumer et al.* 2018; *Zhang et al*, 2019). To mitigate this, in this work, the FSTS model is formulated by using the Caputo definition. The FSTS models were implemented using a mass conserving control volume method based on the Caputo fractional derivative as described in *Zhang et al*. (2005) and *Zhang et al*. (2007b). Boundary conditions for the FSTS model at the inlet and outlet correspond to a specified concentration and free drainage/zero-flux conditions, respectively (see Eq. (18) in *Zhang et al.*, 2007b). Computational domain lengths of each case are summarized in Table S1 (supplementary material).

2.2 FMIM Model

To explicitly model the mobile (main channel) and immobile (storage) zones using fractional calculus, *Schumer et al.* (2003) developed the fractional mobile immobile (FMIM) model (Eq. (t4) in Table 1). C_m is the main channel (mobile zone) concentration; γ is the fractional derivative order; and β_s (T^{γ}) is the fractional capacity coefficient. The term $\frac{\partial^{\gamma}}{\partial t^{\gamma}}$ is the fractional RL derivative on the half-axis, v is velocity (LT^{-1}), and D (L^2T^{-1}) is the dispersion coefficient. When $\gamma = 1$, the FMIM reduces to the classic ADE with a retardation factor $1+\beta_s$. An analytical solution to (5) with a pulse initial condition on an unbounded domain can be computed using a stable subordinator density (*Schumer et al.,* 2003) and is implemented in `FracFit` (*Kelly et al.,* 2017). The FMIM is a different model from the fractional-in-time ADE (FTADE, $\frac{\partial^{\gamma} p}{\partial t^{\gamma}} = -v\frac{\partial p}{\partial x} + D\frac{\partial^2 p}{\partial x^2}$), which is a limit form the CTRW (*Metzler and Klafter*, 2000). The stochastic process of the FMIM can be viewed as a power-law resident time (immobile state) among the mobile process (mobile state) (*Benson and Meerschaert*, 2009).

2.3 The TTFM model

Exponentially tempering heavy tailed power-law distributions in FMIM produces a waiting time distribution with a finite mean yielding a truncated time fractional ADE, i.e. the TTFM (*Meerschaert et al.*, 2008); see Eq. (t5) in Table 1. In TTFM, γ is functionally equivalent to that in FMIM; $\lambda \geq 0$ is the truncation parameter that controls the RTD transition from a power-law to an exponential. When $t \ll \frac{1}{\lambda}$, the mobile zone concentration decays as power-law as in FMIM; at later times $t \gg \frac{1}{\lambda}$ the tail of the mobile-phase BTC decays exponentially.

2.4 MRMT Model

The basic-form of MRMT can be expressed as (*Haggerty and Gorelick*, 1995):

$$\frac{\partial C_m}{\partial t} + \sum_{i=1}^{n} \beta_i \frac{\partial C_{im,i}}{\partial t} = -v \frac{\partial C_m}{\partial x} + D \frac{\partial^2 C_m}{\partial x^2} \tag{1}$$

$$\frac{\partial C_{im,i}}{\partial t} = \varepsilon_i (C_m - C_{im,i}) \tag{2}$$

where C_m is the concentration of the mobile zone; $C_{im,i}$ is the concentration of the i-th immobile zone. ε_i is the first-order rate coefficient of between the i-th immobile zone and mobile zone. When $i = 1$ the MRMT reduces to the single rate mass exchange model (Eqs.

(t6) and (t7) in Table 1). For the case of a continuous distribution of rate coefficients, Eq. (1) can be written as Eq. (t8) in Table1. In this form, $b(\varepsilon)$ is the PDF of the first order exchange rate coefficients and $\beta_{tot} = \int_0^\infty b(\varepsilon)d\varepsilon$ denotes the capacity coefficient. STAMMT-L is a code for the MRMT model and offers user-specified mass exchange rate coefficients. In this study, we choose the single rate mass transfer model, the lognormal distribution diffusion rate model (*Haggerty and Gorelick* (1998)), and the power-law distribution of first-order mass-transfer rates model to generate the synthetic data sets. In Appendix B, we show their specific functional forms.

2.5 CTRW Model

Continuous time random walk (CTRW) formulations have been widely used to quantify non-Fickian transport (*Berkowitz et al.*, 2006; *Burnell et al.*, 2017; *Muljadi et al.*, 2017; *Russian et al.*, 2016; *Scher et al.*, 2010). In the CTRW framework, transport processes are conceptualized as a series of temporal transitions om space of particles. In one dimension, the Laplace transformed concentration $\tilde{C}(x,u)$ can be expressed as Eq. (t10) in Table 1. The memory function $\tilde{M}(s) \equiv \bar{t}s\frac{\tilde{\phi}(s)}{1-\tilde{\phi}(s)}$ accounts for delays; here the notation $\tilde{}$ denotes that the term is Laplace transformed; s (T^{-1}) is the Laplace variable; C_0 is the initial condition. $\bar{t}$ is the characteristic time; v_ϕ and D_ϕ are the transport velocity and generalized dispersion coefficient respectively. The general CTRW model can show to be equivalent to the general MRMT model (*Dentz and Berkowitz* (2003)), yielding a one-to-one relationship between the waiting time distribution $\phi(t)$ and the memory function $g(t)$

(SI). The PDF $\phi(t)$ is the waiting time density, and can be regarded as the “heart” of the CTRW formulation. While CTRW and MRMT are mathematically equivalent, practical differences exist in typically applied formulations; e.g. the CTRW-TPL and the MRMT-3, which will both be used in this study, each have specific forms of $\phi(t)$ and $g(t)$, so they are not the same model. For the CTRW-TPL, $\tilde{\phi}(s)$ has the form:

$$\tilde{\phi}(s) = (1+\tau_2 s t_1)^{\delta} \exp(t_1 s)\Gamma(-\delta, \tau_2^{-1} + t_1 s) / \Gamma(-\delta, \tau_2^{-1}),\ 0 < \delta < 2 \qquad (3)$$

Where $\Gamma(a,z)$ is the incomplete gamma function; δ is the power law constant that denotes the proxy for the degree of velocity field heterogeneity; t_1 is a characteristic transition time that governs the onset of power law region and t_2 is a ‘‘cut-off” time that governs the crossover from power law to a decreasing exponential function ($\tau_2 = t_2 / t_1$). For $t_1 \ll t \ll t_2$, $\phi(t) \propto (t/t_1)^{-1-\delta}$. The memory function is determined by substitution of the expression above into Eq. (t11) in Table 1 with $\overline{t} = t_1$.

2.6 Model Parameter Estimation

The log-based root mean squared error (RMSE) was computed for each model simulation run as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (\log_{10}(C_{sim}(x,t_i)) - \log_{10}(C_{obs}(x,t_i)))^2}{n}} \qquad (4)$$

where n is the number of time samples in each BTC to provide a measure of goodness of fit (GOF). As a result, areas of lower concentration in BTCs receive greater weight, than in the absence of log transformation, which is important for assessing anomalous transport characteristics where heavy tails occur at lower concentrations. Smaller RMSE values indicate better agreement between simulated and observed datasets. Because of the characteristics of the logarithmic function, the absolute values of $\log_{10}(C_{sim})$ and $\log_{10}(C_{obs})$ become too large when C_{obs} and C_{sim} get close to zero, thus we eliminate data points where C_{sim} (and corresponding C_{obs}) are less than 10^{-6} when calculating RMSE. Based on the shuffled complex evolution (SCE) algorithm (*Duan et al.*, 1993; *Muttil et al.*, 2007), we developed the parallel version of SCE for parameter estimation in all presented cases. The parameters optimized with different models are shown in Table 1. The ADE model was only applied to the synthetic data due to its poor performance in simulating the late-time behavior of BTCs.

The small-sample-corrected Akaike information metric (AICc) that takes both GOF and number of parameters into account is an effective parameter for model comparison and evaluation for models with varying parameter numbers and is given by (*Akpa and Unuabonah*, 2011; *Anderson and Phanikumar*, 2011; *Saffron et al.* 2006; *Xia et al.* 2018):

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2M(M+1)}{n-M-1} \tag{5}$$

where AIC is the Akaike information criterion given by:

$$\mathrm{AIC} = n\ln\left(\frac{S}{n}\right) + 2M \tag{6}$$

n is the number of data points; M is the number of model parameters. S is the error sum of squares, which is log-transformed (similar to log based RMSE) to give the same weight to the tails. Smaller AICc values (may be negative) suggests the model is more justified by the data.

2.7 Sites Description

The models examined in the present work were evaluated against synthetic breakthrough data generated using the STAMMT-L code for different models (MRMT-1, MRMT-2, and MRMT-3). In the synthetic data, breakthrough curves were generated at 360 m downstream from the injection location and a value of 0.3 m^3s^{-1} was used for the discharge Q. In addition, field tracer data collected from natural streams were also used to test the models. Data from both large and small rivers were also used, including the Red Cedar River (RCR), Michigan, USA; the Grand River (GR), Michigan, USA; Uvas Creek (UC), California, USA; and the Ohio River (OR), Ohio, USA.

The tracer study of RCR was reported in *Phanikumar et al.* (2007). RCR, a fourth-order stream in south central Michigan, originates as an outflow from Cedar Lake, Michigan, and flows through East Lansing. The study reach is between Hagadorn Bridge (on the east) and the Kalamazoo Street Bridge (on the west). The RCR meanders through the Michigan State University (MSU) campus over a stretch of approximately 5 km. Tracers were released at Hagadorn Bridge and samples were collected at three downstream sites (Farm Lane, Kellogg and Kalamazoo Bridges) whose distances from the injection point are 1.4 km, 3.1 km, and 5.08 km, respectively.

GR is a 420 km long tributary to Lake Michigan. It originates from the city of Grand Rapids and extends to Coopersville. The tracer study was conducted on a 40 km stretch of the main stem. The Ann Street Bridge near downtown Grand Rapids was selected as the injection point. Sampling was carried out at four downstream sites; the distances from the injection site are 4.558 km, 13.678 km, 28.357 km, and 37.608 km respectively (more details are given in *Shen et al.*, 2008).

UC is a small cobble-bed stream located on the eastern slopes of the Santa Cruz mountains in California. The experiment was conducted near the headwaters of UC. The experimental reach includes a background monitoring station (15 m above the injection point) and five observation stations that are 38 m, 105 m, 281m, 433 m, 619 m downstream from the injection point, respectively (details are available in *Avanzino and Bencala*, 1972).

OR originates at the confluence of the Allegheny and Monongahela Rivers, and flows westward to the border of Pennsylvania, Ohio, and West Virginia, and then flows southwest-ward along the Ohio and West Virginia border. The observation sites were at 21.405 km, 51.017 km, 64.697 km, 87.549 km, and 135.508 km from the injection point respectively (see *Wilely*, 1997 for details).

3. Results

3.1. Comparisons with Synthetic Data

Parameter values used to generate the synthetic data sets with STAMMT-L are shown in Table S2, and the corresponding BTCs in Figure 1. Concentration peaks and peak times of all BTCs are approximately equal. The BTC from MRMT-3 model has the heaviest tail, which characterizes the long-term mass exchange between the mainstream and hyporheic zones. The AICc values are summarized in Table 2 and the calibrated parameters are presented in Tables S3 - S7. Since the results from the sFADE simulation show negative skewness with β values very close to -1 (see Tables S7, S11, S14 and S19), we present results mainly for FSTS $\beta = -1$, but also those for FSTS $\beta = 1$ for comparison.

Figure 2 shows the comparison between simulated BTCs and synthetic data generated by MRMT-1. As we can see, all models can reproduce the synthetic BTC accurately. Both the RMSE and AICc values indicate that the FSTS model with $\beta = -1$ shows the best agreement and the FSTS $\beta = 1$ model also produces a good agreement. The optimized α values of FSTS and sFADE are very close to 2 (Tables S5 and S7), meaning that they both reduce to the traditional TS and ADE model for this case (parameter β is canceled out when $\alpha = 2$ in FSTS and sFADE). In general, although the RMSE values are slightly different, all the models can fit the exponential case well. For the data set generated by the MRMT-2 (Figure 3), both the RMSE and AICc values indicate that the FSTS $\beta = -1$ performs best (AICc = -2.1721E3). The FSTS $\beta = 1$, however, shows the worst performance. The FMIM overestimates the tail of the BTC. Compared to FMIM, the TTFM yields a better simulation but still overestimates the tail. The estimated parameter values for FMIM are very similar to those of the TTFM (Table S3) and the better fitting of the tail for TTFM is due to the truncating effect by λ. Significantly, the CTRW-TPL and ADE cannot fit the BTC tail well in this case. For the data set generated by MRMT-3 (Figure 4), whose BTC has the most flattened tail, the simulation results of TTFM fit the data best (AICc = -1.8586E3). Next is the FMIM. The better simulation result of TTFM over the FMIM can be attributed to the truncation effect. But what should be pointed out is that the FSTS $\beta = -1$ also gives comparable accuracy. The sFADE overestimates the late time concentration. On the other hand, the ADE and CTRW-TPL underestimate the late time part of the BTC.

In general, FSTS $\beta = -1$ captures BTCs with different type of heavy tails well. In the MRMT-2 case FSTS $\beta = -1$ performs best and the TTFM gives better simulation than the

FMIM due to the truncating effect. In the power-law case, even though the TTFM and the FMIM give better results, the FSTS with $\beta = -1$ also performs well and results in comparable accuracy. The FSTS with $\beta = 1$, ADE and CTRW-TPL, however, cannot effectively capture the heavy tailing cases.

3.2. Comparisons with Tracer Studies in Natural Stream

3.2.1 Red Cedar River

Optimal parameter values of sFADE, FMIM, TTFM, CTRW-TPL, MRMT-1, MRMT-3 and FSTS models estimated for tracer experiments conducted on the RCR are presented in Table S8 - S12. Figure 5 - 9 show the comparisons between the simulation and the observed data. The AICc values are summarized in Table S13. *Phanikumar et al.* (2007) combined tracer data with wavelet decomposition of acoustic Doppler current profiler data to separate surface storage from hyporheic retention and indicated that reach 1 was dominated by surface storage. In contrast, hyporheic exchange mainly contributed to transient storage in reach 3 of the RCR. Meanwhile, reach 2 has comparable contributions of both surface storage and hyporheic exchange. Consistent with this, MRMT-1 is superior to MRMT-3 in reach 1 (Figure 7). For reach 3, the MRMT-3 fits the tail of BTC well but overestimates the leading edge which is better simulated by MRMT-1. However, MRMT-1 underestimates the late time concentrations due to the limitation of an exponential RTD. Similar results were also found by *Gooseff et al.* (2003). Both the FSTS $\beta = 1$ and the FSTS $\beta = -1$ fit the observed tracer concentration in reaches 1 and 2 of the RCR well (Figure 8). For reach 3, the positive skewness ($\beta = -1$) of FSTS fits the observed data well for early times from the start of tracer arrival through the passage of the advection peak, but it fits the data poorly at late-time ($t > 4$ hours). However, the FSTS model with negative skewness ($\beta = -1$) fits better over the late-time portion of the BTC of RCR reach 3.

The sFADE fits the leading edge well for all the three reaches but predicts longer residence times than the observed data in reach 1 and reach 2. For reach 3, the sFADE fits both the early and late time concentration well. Similarly, the FMIM overestimates the tail in reach 1 and reach 2 but fits the tailing of BTC in reach 3 well. Compared to the FMIM, the TTFM yields a better simulation especially for the late time concentration, but still overestimates the tails in reach 1 and reach 2. The CTRW-TPL (Figure 9) overestimates late time concentration at the level lower than ~10^0 but slightly underestimates the concentration at level of about 10^0 - 10^1 in reach 1. The truncation time t_2 is very large, indicating that transition to an exponential tail has not yet occurred. In reach 2 and reach 3, the CTRW-TPL fits the tail well but overestimates the leading edge. The RMSE values show that the FSTS $\beta = -1$ gives the best accuracy over all three reaches. One the other hand, the AICc values indicate that the FSTS $\beta = -1$ only performs best in reach 1 and reach 2. For reach 3, however, the FMIM is the best model. The reason is that the FMIM has less parameters than the FSTS but has comparable accuracy.

3.2.2. Grand River

The GR is a relatively large river and the tracer study conducted in all 4 reaches of it does not show significant hyporheic zone (storage) effects. Correspondingly, the BTCs of the observed data do not show heavy tailing. Figure S1 - S5 shows comparisons between the simulation results and observed data. Optimal parameter values for these simulations are listed in Table S14 - S18 and the AICc values are shown in Table S19.

As we can see, all models fit reach 3 and reach 4 well except the CTRW-TPL (Figure S5) and MRMT-3 (Figure S3), which overestimate the early time concentration. For reach 1 and reach 2, however, the MRMT-1 underestimates the late-time concentration, while the MRMT-3 model overestimates it. Similarly, the sFADE (Figure S1) and the FMIM (Figure S2) models overestimate the late time part BTC in reach 1 and reach 2. Meanwhile, the TTFM gives a better match than FMIM since it can better capture the tail of BTC. Similarly, the CTRW-TPL (Figure S5) fits the tail well but shows a deviation at early times. The AICc values suggest that the FSTS model (both $\beta = 1$ and $\beta = -1$) perform better than others. Given that the α values are very close to 2 for the FSTS, the better performance may be mainly attributed to the TS term, which is consistent with the MRMT-1 outperforming the MRMT-3. The limited resolution of the observations, especially at lower concentrations, can lead to underestimation of hyporheic exchange (*Drummond et al.*, 2012); this prevents us from drawing a strong conclusion on which FSTS model is best (between $\beta = 1$ and $\beta = -1$), but for the GR, the FSTS $\beta = -1$ is still a promising model.

3.2.3 Ohio River

Figures S6 - S8 show the simulation of OR tracer data. Best fitting parameters for these simulations are presented in Table S20 - S22. The OR tracer data are best simulated by the FSTS model. The optimized velocity values in sFADE and FMIM (3.0 - 5.0 ms^{-1}) are significantly larger than observations (0.044 - 0.065 ms^{-1}) (*Wiley*, 1997), suggesting perhaps an issue of equifinality. The FSTS simulations with positive and negative skewness fit the data well except for reach 1 (Figure S8), where the case with positive skewness underestimates the late-time concentration, while FSTS $\beta = -1$ gives a better fit.

Meanwhile, the α values for both models are very close to 2 (except for reach 1 for FSTS $\beta = -1$) indicating that the tails are mainly explained by the TS term. For reach 1, part of the tailing phenomenon of the BTC is explained by the space fractional term in FSTS $\beta = -1$. Both the sFADE and the FMIM tend to overestimate the tail of BTC.

3.2.4 Uvas Creek

For further comparison between the FSTS $\beta = 1$ and $\beta = -1$, we only use these two models to fit the UC experiment data. As we can see from Figure S9, FSTS with both positive and negative skewness fit observed data of UC well. In reaches 3, 4, and 5, however, FSTS with negative skewness reproduces the experiment data better than that with positive FSTS, especially for the peak and late time portion of BTCs. The calibrated parameters are listed in Table S23. This may indicate that, as the center of mass flows downstream with water, more particles experience retention in the storage zones (seems to be jumped to the upstream direction relative to the plume mass center) and this begins to affect the shape of BTCs which can be better described by FSTS $\beta = -1$.

3.3 Model Properties

To further explore the heavy tail characteristics exhibited by each model, the sensitivity of the BTC tails was tested. Figure S10 shows the BTCs generated by the sFADE with different α values (the other parameter values are fixed as: $\beta = -1$, $v = 1$, and $D = 5$). Obviously, as well known, smaller values of α corresponds to a heavier tail. Figure S11 is the BTCs generated by the FSTS with different ε values (the other parameter values are fixed as: $\beta = 1$, $D = 5$, $v = 0.8$, $\alpha = 1.6$, and $A_s = 3$). This model assumes the Lévy jumps (backward relative to mass center) with the exponential RTD for the tracer's transport. Compared to the purely Lévy jumps ($\varepsilon = 0$), this mixed transport makes more mass concentrated in the middle of the tail rather than the very late time. This characteristic can be very useful or reasonable, for example, the CTRW-TPL and sFADE underestimate the middle of the tail of BTC in GR reach 2 which can be well simulated by the FSTS. Thus, the FSTS $\beta = -1$ is more flexible when simulating the late time concentration. The non-Fickian nature of FMIM is governed by two independent parameters (β_s and γ). When set different β_s values (Figure S12, the other parameters are set as: $D = 2$, $v = 0.4$, and $\gamma = 0.63$), the slope of the tail doesn't change but the power-law tail appears at a lower concentration magnitude as β_s decreases. It is conceivable that when set $\beta_s = 0$, the BTC will have no heavy tail, i.e., the FMIM reduces to the ADE. On the other hand, when set different γ values (Figure S13, other parameters are fixed as: $D = 2$, $v = 0.4$, and $\beta_s = 1$) the slope of the tail change significantly, which is similar to the function of α in sFADE. In the MRMT-3, k is the main parameter that governs the non-Fickian transport feature. Figure S14 shows the BTCs from MRMT-3 with different values of k (other parameters are fixed as: $v = 0.8$, $D = 1$, $\beta_{tot} = 1$, $\varepsilon_{\max} = 0.5$, and $\varepsilon_{\min} = 1\mathrm{E}-6$). As we can see, the tail is heavier as k decreases. The heavy tail also appears at significant different concentration magnitudes with different k (i.e., k can influence both the slope and the concentration magnitude of the tail). In the CTRW-TPL, the δ is main factor that governs the non-local feature. However, as we can see form Figure S15 (other parameters are set as: $v = 0.8$, $D = 1$, $\beta = 1.6$, $\log 10(t_1) = -5$ and $\log 10(t_2) = 8$), the slope of the tail is not very sensitive to the value of δ, which, however, can significant influence both the peak concentration and the peak time. In general, the λ in TTFM, the $\varepsilon_{\min}$ in MRMT-3, and the t_2 in CTRW-TPL all have similar functions (i.e., the truncated power-law). Take the result from CTRW-TPL fitting the synthetic data generated by MRMT-2 for example, when set different truncating time t_2 (the other parameter values are the same as those estimated for MRMT-2 synthetic data) the tails of BTCs decrease sharply at different times (Figure S16).

4. Discussion

4.1 The Physical Interpretation of β in FSTS and sFADE

The stochastic model underlying the FSTS and the sFADE equations are solute particles undergoing deterministic drift with random heavy tailed jumps superimposed when solute particles are mobile. To date, a controversy associated with such heavy tailed random jumps is that long-distance and long-term backward dispersion (i.e., negative skewness in FSTS and sFADE) are unphysical (*Zhang et al.*, 2009); that is, the physical interpretation of this backward skewness, when applied to streamflow, remains controversial (*Deng et al.*, 2004; *Zhang et al.*, 2005; *Zhang et al.*, 2009) as it does not make sense for particles to make such large upstream jumps traveling against the mean flow. Our results, however, show that for our best models the optimized values of β are very close to -1, indicating that models with negative skewness may be suitable for solute transport modeling, particularly for cases with SHE. In order to reconcile this, we must understand the physical meaning of the parameter β. To this end, we present BTCs generated by the FSTS model for different values of β in Appendix B. In an open channel system, many factors including channel morphology can affect the transport of solutes. The existence of hyporheic zones (solutes moving into and out of near-bed sediments) or surface transient storage (*Ensign et al.*, 2005), such as side pools and other in-channel features, can both retain solutes. Other immobile objects can also hinder the migration of solutes in the mobile phase, such as fallen trees, vegetation, large stones in the river bottom, and organic debris (*Briggs, et al*, 2009). *Kelly and Meerschaert* (2017; 2019) argued that when the bulk of the plume travels downstream, since the center of mass moves downstream while the solute particles remains at rest for a long time due to retention, the particle seems to be displaced in the upstream direction relative to the plume center of mass. More importantly, Kelly and Meerschaert (2017; 2019) mathematically proved that long upstream particle jumps (described by negative skewed space-fractional models) and long resting times (described by time-fractional models) are two sides of the same coin by using the space-time duality (Appendix A). Our results further reinforce this message via fitting of real-world data. Thus, as long as one is willing to accept this physical and mathematical interpretation. the FSTS model with $\beta = -1$ is a valid representation of the underlying physics of such situation.

4.2 Analysis of Model Performance

4.2.1 The sFADE model

In our results, the sFADE tends to overestimate the late time concentration in the natural reaches without SHE. In the heavy tailing synthetic cases (i.e. data from MRMT-2 and MRMT-3), the sFADE shows a poorer performance than the FMIM. As is known, a smaller α correspond to a higher probability long distance jumps (i.e. a more flatten tail).

Compared to the FMIM, the major limitation of sFADE is that it does not explicitly distinguish mobile and immobile phases (*Boano et al.*, 2007, *Zhang et al.*, 2005). *Zhang et al*. (2009) pointed out that the FMIM can explain mass decline as it describes the dynamic partitioning of solute mass into the immobile phase, and the rate coefficient between the mobile and immobile phases was described by the fractional derivative in time. While, the pure sFADE (Eq. (t3) in Table 1) cannot account for the loss of mobile mass, it can be described by FSTS (Eqs. (t1) and (t2) in Table 1), which uses a first-order mass transfer between mobile and immobile regions. But one of the advantages of sFADE over the time nonlocal methods is that it can simulate the early arrivals by setting $\beta > 0$ (*Zhang et al.*, 2009).

4.2.2 The FMIM and TTFM models

As presented in section 3.3, the FMIM has two independent parameters (β_s and γ) that characterize nonlocal behavior., A larger value of β_s and a smaller γ correspond to a flatter tail. The parameter β_s in FMIM has a physical interpretation and it can be defined as the ratio of volume of the immobile to mobile zones (*Schumer et al.*, 2003), or as the ratio of expected time in the immobile versus mobile zones (*Benson and Meerschaert*, 2009). γ denotes the power-law decline rate of concentration with time. We find that the FMIM fits the cases with SHE (reach 3 in RCR and the synthetic case by MRMT-3) very well but tends to overestimate tailing in cases without SHE. The TTFM provides additional flexibility by exponentially truncating the tail via an additional parameter λ (*Meerschaert et al.*, 2008) which is functionally equivalent to $\varepsilon_{\min}$. Note, however, that the gain may be minimal: although the RMSE values show a slightly better performance of TTFM, the AICc indicates that the FMIM is better due to lesser number of parameters.

4.2.3 The CTRW-TPL model

For the synthetic cases, the CTRW-TPL performs poorly, particularly in capturing heavy tailing. When fitting the experimental data, it tends to overestimate the leading edge of BTC. Three additional parameters over an ADE are of interest (t_1, t_2 and δ). t_1 roughly sets a starting point for power law tailing that is cut off at t_2, but as pointed out by *Haggerty et al*. (2000), t_1 is of minor importance when capturing the late-time tailing. In our tests, the slope of the BTC tail does not change significantly (Figure S15), this may be due to the fact that the power law regime (from t_1 to t_2) is not long enough. However, a tempered power law can have an observed or inferred slope due to the interaction of the tempering time. When truncated at a specific time (t_2) the tail of BTC declines rapidly (exponentially) when $t \gg t_2$ (see Figure S16). But this decrease is too quick to effectively capture the tail from the synthetic BTC. Thus, in this context, it appears that the CTRW-TPL has limited ability to effectively capture different types of heavy tails, although it has had much success in other applications (e.g. *Burnell et al*. (2017)).

4.2.4 The MRMT-1 and MRMT-3 models

In the cases without SHE, the MRMT-1 performs well. But it underestimates the tail of the BTC significantly. In the MRMT-1, RTD of the solute in storage is assumed to be exponential (*Thorsten Wagener*, 2002). This is a significant limitation since the exponential RTD is not appropriate for characterizing late-time behaviors, which are better characterized by power-law tailing, as solutes are retained in the storage zones for prolonged periods of time (*Aubeneau et al*., 2014; *Gooseff et al*., 2003). The MRMT-3 can give a better match for the late time data but presents higher concentrations for early arrivals, which are still better simulated by MRMT-1. In MRMT-3, the parameter k governs the power-law distribution of mass exchange rate at the range of $\varepsilon_{\min} < \varepsilon < \varepsilon_{\max}$. $\varepsilon_{\min}$ and $\varepsilon_{\max}$ are functionally equivalent to the inverse of the t_2 and t_1 (*Lu et al*. 2018). Thus, k can control the slope of the late-time BTC in a log-log plot (a larger k leads to a faster decline of the late-time BTC). However, k also impacts the concentration magnitudes at which the heavy tail appears in an irregular way. In our tests, the concentration magnitude (where the heavy tail occurs) increase as k decreases in the range from 3.5 to 2.0 (Figure S14). When $k < 2.0$, the tails of BTCs are flatter as k decreases; however, more solute mass is concentrated in the peaks and heavy tails appear at significantly lower concentrations. With the enhancement of trapping effects, more solute trends to be distributed in the tail rather than the peak of the BTC. So, heavier trailing is more likely to begin at higher concentrations similar to the effect of the α in sFADE (Figure S10) and γ in MFIM (Figure S13). Thus, the changes of k in the MRMT-3 cause the irregular variation of the BTC tail and may result in inappropriate simulations.

4.2.5 The FSTS model with $\beta = -1$

In our study, both the RMSE and the AICc values indicate that the FSTS $\beta = -1$ model performs well in all cases. The FSTS is an extension of the TS model with a fractional dispersion term in the mobile zone. In the FSTS, the tracer particles experience Lévy jumps while mobile, and exponential RTDs when trapped. When $\beta = -1$ the space fractional term predicts large distance of upstream migrations. But this upstream migration is relative to the center of mass (that moves downstream with flowing water) rather than a stationary river bed. In this context, it can be regarded as backward migration relative to the center of mass of the migrating tracer plume when part of the tracer mass has a lower velocity relative to the center of mass (see Figure 10). In rivers, turbulent eddies, meander bends and pools, side pockets, local scale river-bed topography variations, fallen trees within a channel, could all lead to lower transport velocity of the trapped solutes relative to the center of mass of the tracer plume. Additionally, Zhang et al. (2019) showed that the negatively skewed sFADE is also physical meaningful when describing the non-Fickan transport in a river and could capture the heavy tail in a BTC similar to the time fractional ADE. Thus, when trapping in the physical system is short-term retention (exponential RTD), it can be explicitly modeled in the FSTS using a first order exchange with an immobile region as in the classic TS model (e.g., reach 1 and reach 2 in RCR). For the long-term trapping (e.g., larger hyporheic zones or deep hyporheic flow paths) heavy tails are captured by the space fractional term with $\beta = -1$ (e.g., reach 1 in Ohio and reach 3 in RCR). Again, here we invoke the mathematical equivalence and interpretation of *Kelly and Meershaert* (2017; 2019) that the fractional in space jumps are equivalent to heavy tailing waiting due to space-time duality. The FSTS seems to hit the sweet spot between other models. Only having surface transient storage (TS) typically underestimates late time concentrations in cases with SHE, while the sFADE tends to overestimate late time concentrations there. In the FSTS, these two seem to better balanced. Whether the model is truly physical or purely mathematically is likely still open to debate and comes down to whether one accepts the interpretation of the backward skewed space fractional process as representing lower velocities relative to the center of mass of the tracer plume rather than upstream jumps of tracer particles relative to the river bed. Regardless, the FSTS was able to describe different types of observed tailing behaviors. Although the parameters in a model may have interactions with one another, the above sensitivity analysis of the BTC tail for the isolated parameter can also explain the characteristics of the models used in this study.

4.3 Comparation of the FSTS $\beta = -1$ and the TTFM

Although the TTFM truncates the power-law tail, the resulting model can be used to represent transport in streams. However, as we can see from Figure S16, the truncation can lead to a decay of tracer mass that is too fast compared to the tail of the synthetic data generated using MRMT-2. The rationale behind truncation of tails in the TTFM is that natural rivers are finite systems and transport of tracers must eventually converge to a finite mean and variance (*Aubeneau et al*., 2014; 2016), even at very late times. But when tracers are trapped in immobile zones and never come out (i.e. infinite residence time), those assumptions are violated. Exchanges of water and solutes with the immobile zone will produce a delay in solute transport relative to the mainstream flow and lead to long tails in concentration BTCs (*Boano et al*., 2007). Different storage zones (e.g., surface storage zone and the deep hyporheic zone) can also lead to different types of RTD. The power-law tail (mathematically represented by the space fractional term in the FSTS) balanced by the exponential RTD term in the FSTS perform better in capturing the BTC.

5. Conclusions

Although a rigorous, but complex mathematical interrelation exists between CTRW, MRMT and FADE models, in their most commonly applied formats differences in memory or RTD functions lead to different classes of non-Fickian sub-models. In general, these models have different numbers of parameters with varying physical interpretations. The highlight of this study is a comparison and evaluation of seven anomalous transport models summarized in Table 1. Parameter estimation for these seven models was performed with both synthetic and field data, and the fits were quantitatively compared with a log-based RMSE and corrected AIC (AICc) metrics.

Compared to sFADE and CTRW-TPL, and MRMT-3, the FMIM performs better. While the FMIM can overestimate late time tailing in concentrations, improvements with the truncating scheme in TTFM appears limited, as the truncating time may lead to the decline of concentration that is too fast. In contrast, the FSTS ($\beta = -1$) assuming long-term upstream jumps (mathematically equivalent to power-law waiting times due to space-time duality; see Appendix A) that well balanced by an exponential RTD could better estimates both early and late-time portions of the BTCs. Both the RMSE and AICc indicate that the FSTS with negative skewness estimates the BTCs with greater fidelity both with and without SHE. In addition, the FSTS model could also serve as a useful and efficient diagnostic tool to assess the nature of storage in a stream reach (surface versus deep hyporheic). Thus, by fitting the tracer data to the FSTS model with positive ($\beta = 1$) and negative ($\beta = -1$) skewness values, users can quickly assess the relative importance of hyporheic storage in a given reach so that more appropriate models can be selected for further analysis. However, whether the negatively skewed space fractional term is truly physically meaningful is still open to debate. And it depends on whether one accepts the interpretation of this term as representing a slower transport velocity (or waiting) relative to the mass center rather than upstream jumps relative to the river bed. Additionally, our current conclusions should only be applied in the context of conservative transport as non-conservative (reactive) transport in systems displaying anomalous transport can be very different from that predicting by just naively adding chemical reaction terms to the governing equations (*Bolster et al.,* 2010, *Bolster et al.,* 2017).

Acknowledgments

We thank Roy Haggerty, Oregon State University for sharing the STAMMT-L model used in this work to generate the synthetic data corresponding to different residence time distributions. Sun was partially supported by NSF project of Guangdong, China 2018A030313165. This research was also funded by a grant from NOAA/CILER. Kelly was partially supported by ARO MURI grant W911NF-15-1-0562.

Appendix A. Space-Time Duality

The negatively-skewed derivative in the sFADE and FSTS models with $\beta = -1$ models long (power-law) upstream jumps. As noted by *Zhang et al*. (2009), these upstream jumps (or negative dispersion) appear unphysical. However, due to the advection term in the sFADE, these upstream jumps are relative to the center of mass in the underlying random walk. Hence, in the sFADE model, a particle moves downstream due to advection and then may jump back upstream. In the fractional-in-time ADE model, the particle remains upstream while the bulk of the plume moves downstream. In both models, the particle ends up behind the plume center of mass, resulting in an effective delay, or retention. Hence, there are two separative and equivalent descriptions for the same underlying hydrological mechanism (retention), thus resolving the controversy between the sFADE and the fractional-time ADE for river flows.

This simple observation may be understood mathematically using space-time duality (*Baeumer* et al., 2009; *Kelly and Meerschaert*, 2017; 2019). For simplicity, consider the sFADE (t3) with $\beta = -1$, $v = 0$, and $D = 1$:

$$\frac{\partial}{\partial t} C(x,t) = \frac{\partial^{\alpha}}{\partial(-x)^{\alpha}} C(x,t) \tag{A.1}$$

Applying a Fourier transform (FT) with respect to both *x* and *t* yields:

$$\left[(i\omega) - (-ik)^{\alpha}\right]\hat{C}(k,\omega) = 0 \tag{A.2}$$

where k is the wave number and ω is the angular frequency. The bracketed *dispersion relationship* is equivalent to $(i\omega)^{\gamma} = (-ik)$ where $\gamma = 1/\alpha$. Substituting back into (A.1) and inverting the Fourier transform leads to the *dual equation*:

$$\frac{\partial^{\gamma}}{\partial t^{\gamma}} C(x,t) = -\frac{\partial}{\partial x} C(x,t) \quad \text{(A.3)}$$

which is a special case of the fractional-in-time ADE with $v = 1$ and $D = 0$, where the left-hand side uses a Caputo derivative. Note that the order of the time-fractional derivative γ ranges from 0.5 to 1. This may be made rigorous using a complex plane argument (*Kelly and Meerschaert*, 2017). If the advection term in the sFADE is retained, it is shown in *Kelly and Meerschaert*, 2017) that the dual time-fractional PDE involves the fractional material derivative (*Sokolov and Metzler*, 2003). The resulting coupled space-time fractional PDE (see Equation (28) in *Kelly and Meerschaert* (2017)) governs power-law waiting times in a moving reference frame, or retention relative to plume center of mass. Using this duality analysis, negative derivative term in the sFADE and the FSTS can effectively model power-law retention. We note that this analysis is restricted to fractional models with constant coefficients in either an unbounded domain or a reflecting boundary at $x = 0$.

Appendix B: BTCs Generated by the FSTS Model with Different Values of β

The BTCs generated by the FSTS model with different values of β are presented in Figure D1 (other parameters were fixed as: $D = 0.7$, $\alpha = 1.9$, $v = 1.2$, $A_s = 5$, and $\varepsilon = 3\mathrm{E}^{-3}$). The total mass under each of the BTCs is the same for all values of β. As we can see from Figure Appendix B1. When $\beta = -1$ the BTC shows a flattened tail at late times and a sharp leading edge. When $\beta = 1$, however, the resulting BTC has a flattened leading edge and the tail is steep instead (Figure S4 also shows similar phenomenon). Thus, a larger β value results in a BTC with a more flattened leading edge. When $\beta = 0$ the BTC is approximate symmetric with heavy leading edge and tailing on both sides. In addition, we note that the peaks of all BTCs are almost synchronous regardless of the value of β, as these are dictated mostly by the deterministic drift.

References

Akpa, O. M., and Unuabonah, E. I., 2011. Small-Sample Corrected Akaike Information Criterion: An appropriate statistical tool for ranking of adsorption isotherm models, Desalination, 272(1-3), 20-26.

Anderson, E.J., M.S. Phanikumar, 2011. Surface Storage Dynamics in Large Rivers: Comparing Three-Dimensional Particle Transport, 1D Fractional Derivative and Multi - Rate Transient Storage Models, Water Resources Research, Vol. 47(9), W09511.

Aubeneau, A.F., Hanrahan, B., Bolster, D. Tank, J. L., 2014. Substrate size and heterogeneity control anomalous transport in small streams. Geophysical Research Letters, 41(23), 8335-8341.

Aubeneau, A. F., Hanrahan, B., Bolster, D., Tank, J., 2016. Biofilm growth in gravel bed streams controls solute residence time distributions. Journal of Geophysical Research: Biogeosciences, 121(7), 1840-1850.

Baeumer, B., Kovács, M., Meerschaert, M. M., Sankaranarayanan, H., 2018. Boundary conditions for fractional diffusion. Journal of Computational and Applied Mathematics, 336, 408-424.

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

Benson, D. A., Meerschaert, M. M., 2009. A simple and efficient random walk solution of multi-rate mobile/immobile mass transport equations, Advances in Water Resources, 32(4), 532-539.

Berkowitz, B., Andrea, C., Denz, M., Scher, H., 2006. Modeling non-Fickian transport in geological formations as a continuous time random walk, Review of Geophysics, 44(RG20032).

Boano, F., Packman, A. I., Cortis, A., Revelli, R. and Ridolfi, L., 2007. A continuous time random walk approach to the stream transport of solutes, Water Resources Research, 43(10).

Bolster, D., Benson, D. A., Le Borgne, T., Dentz, M., 2010. Anomalous mixing and reaction induced by superdiffusive nonlocal transport. Physical Review E, 82(02111922).

Bolster, D., Benson D.A., Singha, K., 2017. Upscaling chemical reactions in multicontinuum systems: When might time fractional equations work? Chaos Solitons & Fractals, 102, 414-425.

Briggs, M.A., et al., Gooseff, M. N., Arp, C. D., Baker, M. A., 2009. A method for estimating surface transient storage parameters for streams with concurrent hyporheic storage. Water Resources Research, 45(4).

Burnell, D. K., Scott K. H., Jie, X., 2017. Transient modeling of non-Fickian transport and first-order reaction using continuous time random walk, Advances in Water Resource, 107, 370-392.

Carr, M.L., Rehmann, C.R., 2007. Measuring the dispersion coefficient with acoustic Doppler current profilers. Journal of Hydraulic Engineering (ASCE), 133(8), 977-82.

de Barros, F.P.J., Fernàndez-Garcia, D., Bolster, D., Sanchez-Vila, X., 2013. A risk-based probabilistic framework to estimate the endpoint of remediation: Concentration rebound by rate-limited mass transfer. Water Resources Research, 2013. 49(4), 1929-1942.

de Hoog, F. R., J.H. Knight and A.N. Stokeset al. 1982. An Improved Method for Numerical Inversion of Laplace Transforms. *SIAM Journal on Scientific and Statistical Computing*, 3(3), 357-366.

De Smedt, F., Brevis, W., and Debels, P. 2005. Analytical solution for solute transport resulting from instantaneous injection in streams with transient storage, Journal of Hydrology, 315(1-4), 25-39.

Deng, Z. Q., Singh, V. P., Bengtsson, L., 2004. Numerical solution of fractional advection-dispersion equation, Journal of Hydraulic Engineering-ASCE, 130(5), 422-431.

Deng Z. Q., Bengtsson, L., Singh V. P., 2006. Parameter estimation for fractional dispersion model for rivers. Environmental Fluid Mechanics, 6(5), 451-75.

Dentz, M. and Berkowitz, B., 2003. Transport behavior of a passive solute in continuous time random walks and multirate mass transfer. Water Resources Research, 39(5).

Duan, Q. Y., Gupta, V. K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization, Journal of Optimization Theory and Applications, 76(3), 501-521.

Drummond, J. D., Covino, T. P., Aubeneau, A. F., Leong, D., Patil, S., Schumer, R., and Packman, A. I., 2012. Effects of solute breakthrough curve tail truncation on residence time estimates: A synthesis of solute tracer injection studies. Journal of Geophysical Research: Biogeosciences, 117(G3).

Edery, Y., Scher, H., Berkowitz, B., 2010. Particle tracking model of bimolecular reactive transport in porous media, Water Resources Research, 46(W07524).

Ensign, S.H., Doyle M.W., 2005. In-Channel Transient Storage and Associated Nutrient Retention: Evidence from Experimental Manipulations. Limnology and

Oceanography, 50(6),1740-1751.

Gooseff, M. N., Wondzell, S. M., Haggerty, R., Anderson, J, 2003. Comparing transient storage modeling and residence time distribution (RTD) analysis in geomorphically varied reaches in the Lookout Creek basin, Oregon, USA, Advances in Water Resources, 26(9), 925-937.

Haggerty, R. and Gorelick, S.M., 1995. Multiple-Rate Mass Transfer for Modeling Diffusion and Surface Reactions in Media with Pore-Scale Heterogeneity. Water Resources Research, 31(10), 2383-2400.

Haggerty, R. and Gorelick, S.M., 1998. Modeling Mass Transfer Processes in Soil Columns with Pore-Scale Heterogeneity. Soil Science Society of America Journal, 62(1), 62-74.

Haggerty, R., S.A. McKenna and L.C. Meigs, 2000. On the late-time behavior of tracer test breakthrough curves. *Water Resources Research*, 36(12),3467-3479.

Haggerty, R., Wondzell, S. M., Matthew A. J. 2002. Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream, Geophysical Research Letters, 29(164013).

Kelly, J.F., Meerschaert, M.M., 2017. Space-time duality for the fractional advection dispersion equation. Water Resources Research, 53(4): 3464–3475.

Kelly, J.F. and Meerschaert, M.M., 2019. Space-time duality and high-order fractional diffusion. Physical Review E, 99(2-1), 022122.

Kelly, J. F., Bolster, D., Meerschaert, M.M., Drummond, J. D., Packman, A. I., 2017. FracFit: A robust parameter estimation tool for fractional calculus models, Water Resources Research, 53(3), 2559-2567.

Liu, D. X., Jivkov, A. P., Wang, L. C., Si, G. H., Yu, J., 2017. Non-Fickian dispersive transport of strontium in laboratory-scale columns: Modelling and evaluation, Journal of Hydrology, 549, 1-11.

Lu, B., Zhang, Y., Zheng, C., Green, C.T., O'Neill, C., Sun, H.-G., and Qian, J., 2018. Comparison of Time Nonlocal Transport Models for Characterizing Non-Fickian Transport: From Mathematical Interpretation to Laboratory Application, Water, 10(778).

Meerschaert, M.M., Zhang, Y., Baeumer, B., 2008. Tempered anomalous diffusion in heterogeneous systems, Geophysical Research Letters, 35 (17).

Metzler, R. and Klafter, J.， 2000. The random walk's guide to anomalous diffusion: a fractional dynamics approach. Physics Reports-review Section of Physicsletters, 339(1), 1-77.

Muljadi, B. P., Bijeljic, B., Blunt, M. J., Colbourne, A., Sederman, A. J., Mantle, M. D., Gladden, L. F., 2017. Modelling and upscaling of transport in carbonates during dissolution: Validation and calibration with NMR experiments, Journal of Contaminant Hydrology, 212, 85-95.

Muttil, N., Liong, S. Y., Nesterov, O., 2007. A Parallel Shuffled Complex Evolution Model Calibrating Algorithm to Reduce Computational Time, MODSIM 2007: International Congress on Modelling and Simulation: Land, Water and Environmental Management: Integrated Systems for Sustainability, 1940-1946.

Phanikumar, M. S., Aslam, I., Shen, C. P., Long, D. T., Voice, T. C., 2007. Separating surface storage from hyporheic retention in natural streams using wavelet decomposition of acoustic Doppler current profiles, Water Resources Research, 43(5).

Ronald J. A., Gary, W. Z., Vance. C. K., S., Zand, S. M., Kenneth, E. B., 1972. Uvas Creek Tracer Study 1972, U. S. Geological Survey Open-File Report 84-236.

Russian, A., Dentz, M., Gouze, P., 2016. Time domain random walks for hydrodynamic transport in heterogeneous media, Water Resources Research, 52(5), 3309-3323.

Saffron, C. M., Park, J. H., Dale, B. E., Voice, T. C., 2006. Kinetics of contaminant desorption from soil: Comparison of model formulations using the Akaike Information Criterion, Environmental Science & Technology, 40(24), 7662-7667.

Scher, H., Willbrand, K., Berkowitz, B., 2010. Transport Equation Evaluation of Coupled Continuous Time Random Walks, Journal of Statistical Physics, 141(6), 1093-1103.

Schumer, R., Benson, D. A., Meerschaert, M. M., Baeumer, B., 2003. Fractal mobile/immobile solute transport, Water Resources Research, 39(10).

Schumer, R., Meerschaert, M. M., Baeumer, B., 2009. Fractional advection-dispersion equations for modeling transport at the Earth surface, Journal of Geophysical Research, 114.

Shen, C. P., Phanikumar, M. S., Fong, T. T., Aslam, I., McElmurry, S. P., Molloy, S. L., Rose, J. B., 2008. Evaluating Bacteriophage P22 as a Tracer in a Complex Surface Water System: The Grand River, Michigan, Environmental Science & Technology, 42(7), 2426-2431.

Shen, C. P., and Phanikumar, M. S., 2009. An efficient space-fractional dispersion approximation for stream solute transport modeling, Advances in Water Resources, 32(10), 1482-1494.

Shen, C.P., Niu, J., Anderson, E.J. and Phanikumar, M.S., 2010. Estimating longitudinal dispersion in rivers using acoustic Doppler current profilers, Advances in Water

Resources, 33(6), 615-623.

Sokolov, I.M. and Metzler R., 2003. Towards deterministic equations for Lévy walks: The fractional material derivative. Physical Review E, 67(1):010101.

Thorsten W., Luis, A. C., Howard S. W., 2002. Dynamic identifiability analysis of the transient storage model for solute transport in rivers, Journal of Hydroinformatics, 04(3), 199-210.

Vishal, V., Leung, J. Y., 2015. Modeling impacts of subscale heterogeneities on dispersive solute transport in subsurface systems, Journal of Contaminant Hydrology, 182, 63-77.

Wiley, J. B., 1997. USGS Geoglogical Survey Open-File Report, 97-562.

Xia, C. A., Tong, J. x., Hu, Bill X. Wu, X. j., Guadagnini, 2018. Assessment of alternative adsorption models and global sensitivity analysis to characterize hexavalent chromium loss from soil to surface runoff. Hydrological Processes, 32(20): p. 3140-3157.

Zhang, X. X., Crawford, J. W., Deeks, L. K., Stutter, M. I., Bengough, A. G., Young, I. M., 2005. A mass balance based numerical method for the fractional advection-dispersion equation: Theory and application, Water Resources Research, 41(7).

Zhang, Y., Benson, D. A., Baeumer, B. 2007a. Predicting the tails of breakthrough curves in regional-scale alluvial systems, Ground Water, 45(4), 473-484.

Zhang Y, Benson, D. A., Meerschaert, M. M., LaBolle, E. M., 2007b. Space-fractional advection–dispersion equations with variable parameters: diverse formulas, numerical solutions, and application to the macrodispersion experiment site data. Water Resources Research, 43(5): W05439.

Zhang, Y., Benson, D. A., Reeves, D. M., 2009. Time and space nonlocalities underlying fractional-derivative models: Distinction and literature review of field applications, Advances in Water Resources, 32(4), 561-581.

Zhang, Y., Yu, X. N., Li, X. C., Kelly, J. F., Sun, H. G., Zheng, C. M., 2019. Impact of absorbing and reflective boundaries on fractional derivative models: Quantification, evaluation and application. Advances in Water Resources, 128, 129-144.

Table

Table 1. Summary of the nonlocal models used in this study.

Model	Equation(s)	Solving scheme	Parameters estimated in each model	Characteristics
Fractional in space transient storage model (FSTS)	$\frac{\partial C}{\partial t}+v\frac{\partial C}{\partial x}=D\left[\frac{1+\beta}{2}\frac{\partial^\alpha C}{\partial x^\alpha}+\frac{1-\beta}{2}\frac{\partial^\alpha C}{\partial(-x)^\alpha}\right]+\frac{q_L}{A}(C_L-C)+\varepsilon(C_s-C)$ (t1) $\frac{\partial C_S}{\partial t}=\varepsilon\frac{A}{A_S}(C-C_S)$ (t2)	Finite difference scheme presented in *Shen and Phanikumar* (2009).	v, D, α, A_s, and ε.	Lévy jumps and exponential RTD.
Fractional in space advection dispersion equation (sFADE)	$\frac{\partial C}{\partial t}+v\frac{\partial C}{\partial x}=D\left[\frac{1+\beta}{2}\frac{\partial^\alpha C}{\partial x^\alpha}+\frac{1-\beta}{2}\frac{\partial^\alpha C}{\partial(-x)^\alpha}\right]$ (t3)	Analytical solution in *Benson et al.* (2000).	v, D, α, and β.	Solutes experience Lévy jumps.
Fractional mobile/immobile model (FMIM)	$\frac{\partial C_m}{\partial t}+\beta_s\frac{\partial^\gamma C_m}{\partial t^\gamma}=-v\frac{\partial C_m}{\partial x}+D\frac{\partial^2 C_m}{\partial x^2}-\beta_s C_m(x,0)\frac{t^{-\gamma}}{\Gamma(1-\gamma)}$ (t4)	Analytical solution in *Schumer et al.* (2003).	v, D, γ, and β_s.	The concentration in the mobile phase declines as a power law: $C_m(x,t)\propto t^{-1-\gamma}$. The memory function has the form: $g(t)=\frac{t^{-\gamma}}{\Gamma(1-\gamma)}$.
Truncated time fractional model (TTFM)	$\frac{\partial C_m}{\partial t}+\beta_s e^{-\lambda t}\frac{\partial^\gamma}{\partial t^\gamma}(e^{\lambda t}C_m)-\beta\lambda^\gamma C_m=$ $-v\frac{\partial C_m}{\partial x}+D\frac{\partial^2 C_m}{\partial x^2}-\beta_s C_m(x,0)\int_t^\infty e^{-\lambda\tau}\frac{\tau^{-\gamma-1}}{\Gamma(1-\gamma)}d\tau$ (t5)	Analytical solution in *Meerschaert et al.* (2008).	v, D, γ, β_s, and λ.	When t □ $1/\lambda$, the mobile zone concentration still declines as power-law like FMIM; while at a much later time t □ λ, the tail of the mobile-phase BTC declines exponentially. The memory function has the form: $g(t)=\int_t^\infty e^{-\lambda\tau}\frac{\gamma\tau^{-\gamma-1}}{\Gamma(1-\gamma)}d\tau$.
Single rate Multi-rate mass transfer model (MRMT-1)	$\frac{\partial C_m}{\partial t}+\beta_{tot}\frac{\partial C_{im}}{\partial t}=-v\frac{\partial C_m}{\partial x}+D\frac{\partial^2 C_m}{\partial x^2}$ (t6) $\frac{\partial C_{im}}{\partial t}=\varepsilon(C_m-C_{im})$ (t7)	Analytical solution presented in *Haggerty and Gorelick* (1995) implemented in STAMMT-L (Ver. 3.0).	v, D, β_{tot}, and ε.	The tail declines exponentially: $C_m\propto e^{-t}$. The memory function has the form: $g(t)=\varepsilon e^{-\varepsilon t}$.
Power-law mass exchange rate Multi-rate mass transfer model (MRMT-3)	$\frac{\partial C_m}{\partial t}+\int_{\varepsilon_{\min}}^{\varepsilon_{\max}}p(\varepsilon)\frac{\partial C_{im}}{\partial t}d\varepsilon=-v\frac{\partial C_m}{\partial x}+D\frac{\partial^2 C_m}{\partial x^2}$ (t8) $p(\varepsilon)=\beta_{tot}\frac{(k-2)\varepsilon^{k-3}}{\varepsilon_{\max}^{k-2}-\varepsilon_{\min}^{k-2}}$, $k\neq 2$; $p(\varepsilon)=\beta_{tot}\frac{1}{\ln(\varepsilon_{\max}/\varepsilon_{\min})\varepsilon}$, $k=2$ (t9)	Inverse Laplace transform implemented in STAMMT-L (Ver. 3.0).	v, D, β_{tot}, k, $\varepsilon_{\min}$, and $\varepsilon_{\max}$.	The first order mass exchange rate coefficients have a power-law distribution: for $\varepsilon_{\min}\le\varepsilon\le\varepsilon_{\max}$. The memory function has the form: $g(t)=\int_0^\infty\varepsilon p(\varepsilon)\exp(-\varepsilon t)d\varepsilon$.
Continuous time random walk with truncated power-law waiting time distribution (CTRW-TPL)	$s\tilde{C}(x,s)-C_0(x)=-\tilde{M}(s)\left[v_\phi\tilde{C}(x,s)-D_\phi\frac{\partial^2}{\partial x^2}\tilde{C}(x,s)\right]$ (t10) $\tilde{M}(s)\equiv\bar{t}s\frac{\tilde{\phi}(s)}{1-\tilde{\phi}(s)}$ (t11)	Inverse Laplace transform from *de Hoog et al.* (1982) implemented in the CTRW MATLAB toolbox (Ver. 4.0).	v_ϕ, D_ϕ, δ, t_1, and t_2.	For $t_1 = t = t_2$, the waiting time distribution flows a power-law: $\phi(t)\propto(t/t_1)^{-1-\delta}$. For t ? t_2, waiting time distribution is exponential. The memory function(in the Laplace domain) has the form: $\tilde{\phi}(s)=(1+\tau_2 st_1)^\delta\exp(t_1 s)\Gamma(-\delta,\tau_2^{-1}+t_1 s)/\Gamma(-\delta,\tau_2^{-1})$.

Table 2. AICc (×1e3) values for the parameter fits using synthetic data.

model	MRMT-1	MRMT-2	MRMT-3
FSTS $\beta = 1$	-0.4361	0.2180	0.7049
FSTS $\beta = -1$	-0.5221	-0.6413	-1.0738
FMIM	-0.3282	-0.5278	-1.8552
TTFM	-0.3212	-0.5925	-1.8586
sFADE	-0.3072	-0.5639	-0.6090
CTRW-TPL	-0.3117	-0.2172	-0.0740
ADE	-0.4938	0.1474	0.3726

Figures

Figure 1. Synthetic data generated by MRMT-1 (blue), MRMT-2 (red), and MRMT-3 (yellow).

Figure 2. Comparisons of the simulations for synthetic data (generated by MRMT-1).

Figure 3. Comparisons of the simulations for synthetic data (generated by MRMT-2).

Figure 4. Comparisons of the simulations for synthetic data (generated by MRMT-3).

Figure 5. sFADE simulated BTCs and observations for Red Cedar River.

Figure 6. FMIM and TTFM simulated BTCs and observations for Red Cedar River.

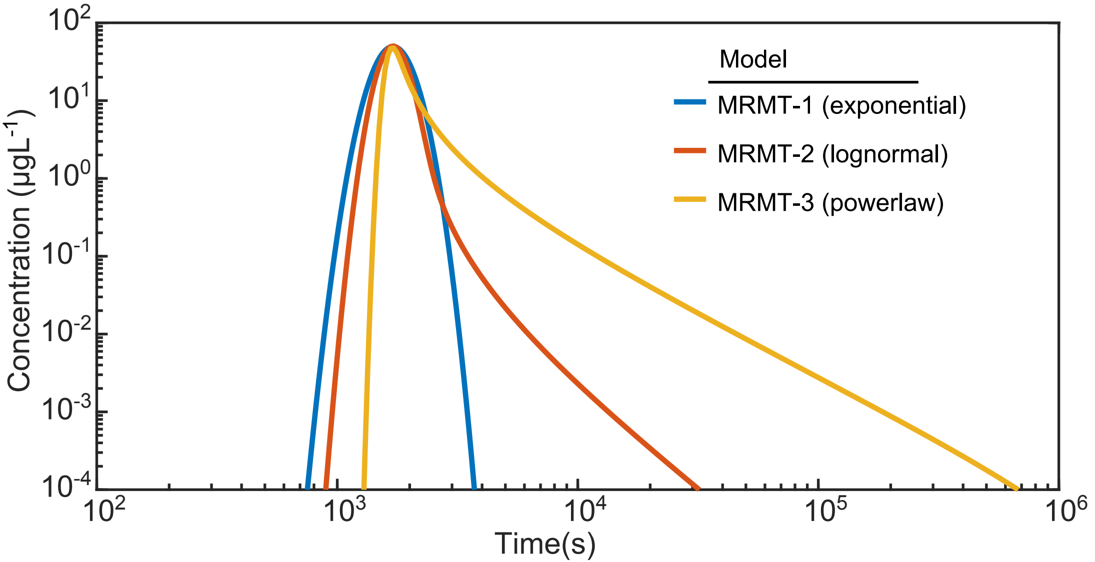
Figure 7. MRMT-1 and MRMT-2 simulated BTCs and observations for Red Cedar River.

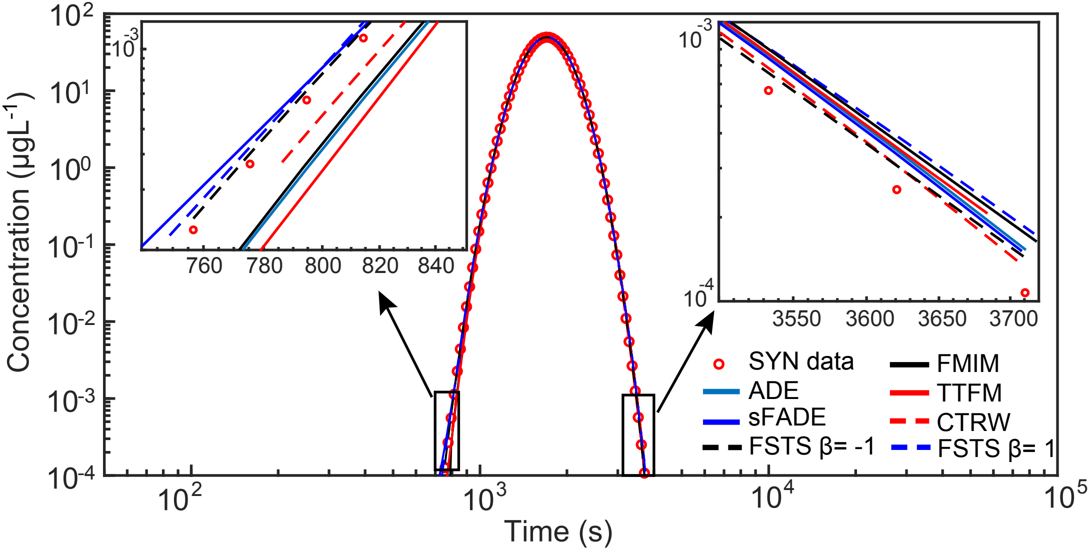
Figure 8. FSTS $\beta = 1$ and FSTS $\beta = -1$ simulated BTCs and observations for Red Cedar River.

Figure 9. CTRW-TPL simulated BTCs and observations for Red Cedar River.

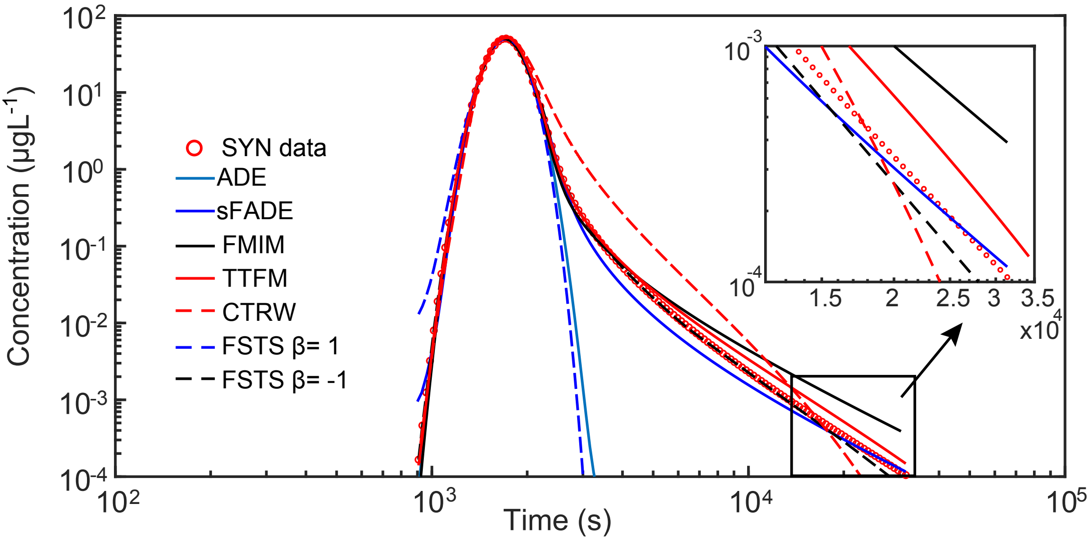
Figure 10. Schematic illustrating solute transport processes within a stream reach with exchange between surface storage zones and the main channel.

Figure Appendix B. The BTCs based on different β values of FSTS model. The other parameters are fixed as: $D = 0.7$, $\alpha = 1.9$, $v = 1.2$, $\varepsilon = 0.003$, and $A_s = 5$.

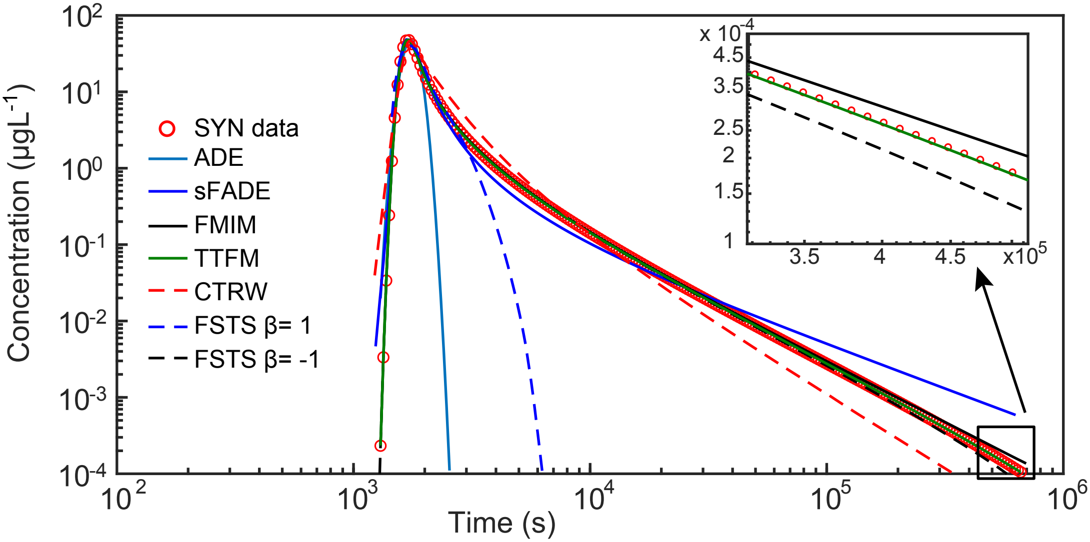

Model
MRMT-1 (exponential)
MRMT-2 (lognormal)
MRMT-3 (powerlaw)
Concentration (μgL-1)
Time(s)
10^2
10^1
10^0
10^-1
10^-2
10^-3
10^-4
10^2
10^3
10^4
10^5
10^6



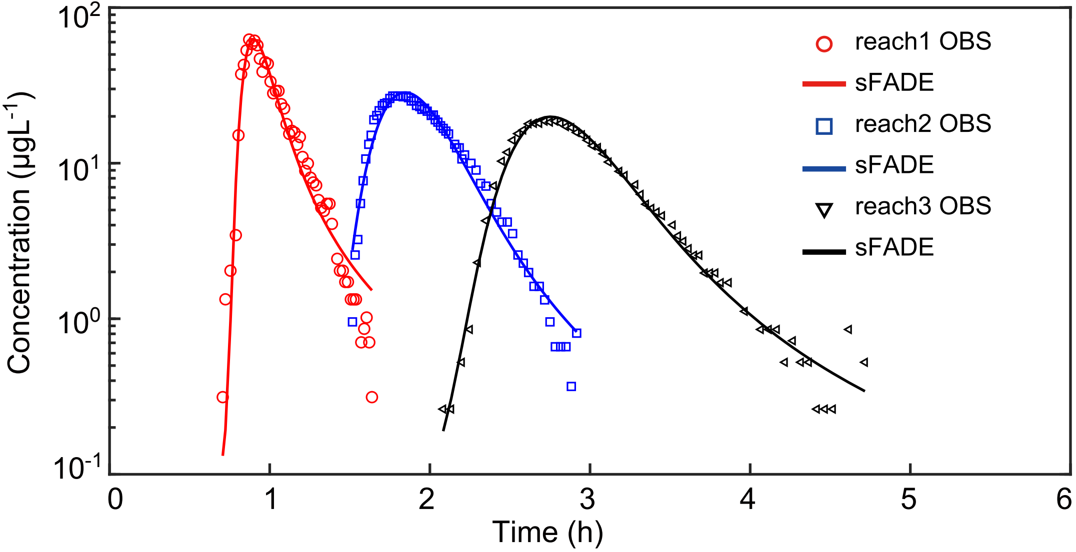
Concentration (μgL-1)
Time (s)
SYN data
ADE
sFADE
FSTS β= -1
FMIM
TTFM
CTRW
FSTS β= 1



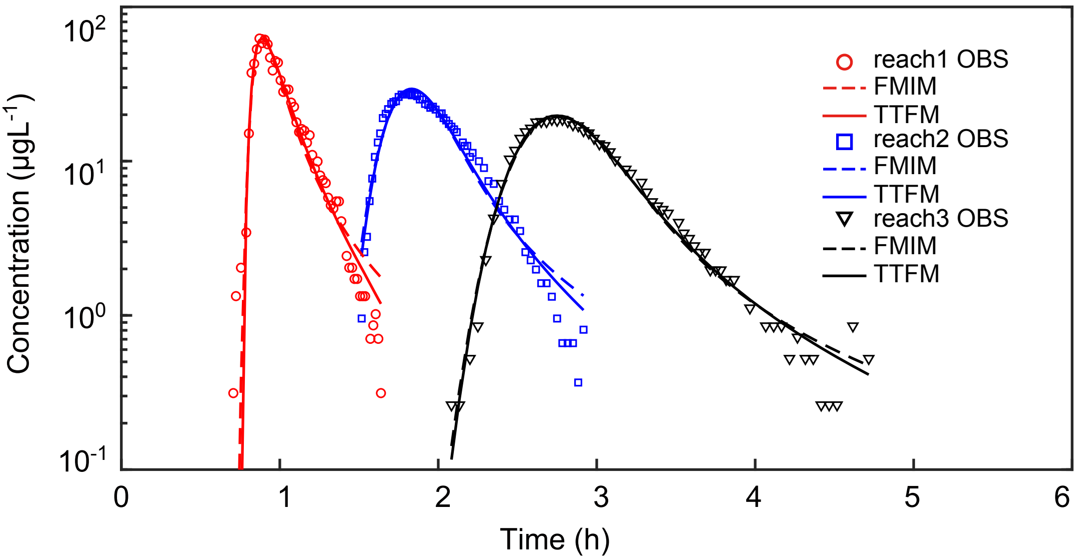
Concentration (μgL^-1)
Time (s)
SYN data
ADE
sFADE
FMIM
TTFM
CTRW
FSTS β= 1
FSTS β= -1
10^2
10^1
10^0
10^-1
10^-2
10^-3
10^-4
10^2
10^3
10^4
10^5
10^-3
10^-4
1.5
2
2.5
3
3.5
x10^4



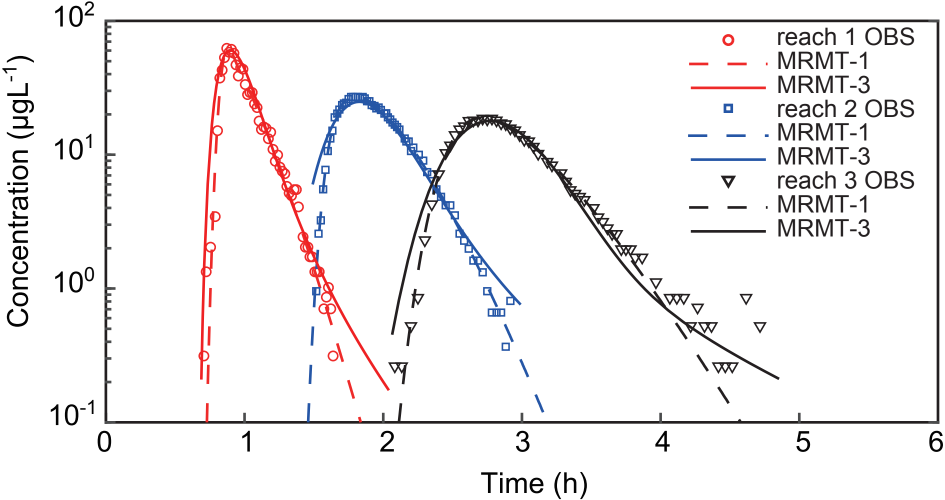
Concentration (μgL-1)
Time (s)
SYN data
ADE
sFADE
FMIM
TTFM
CTRW
FSTS β= 1
FSTS β= -1
x 10-4
x105



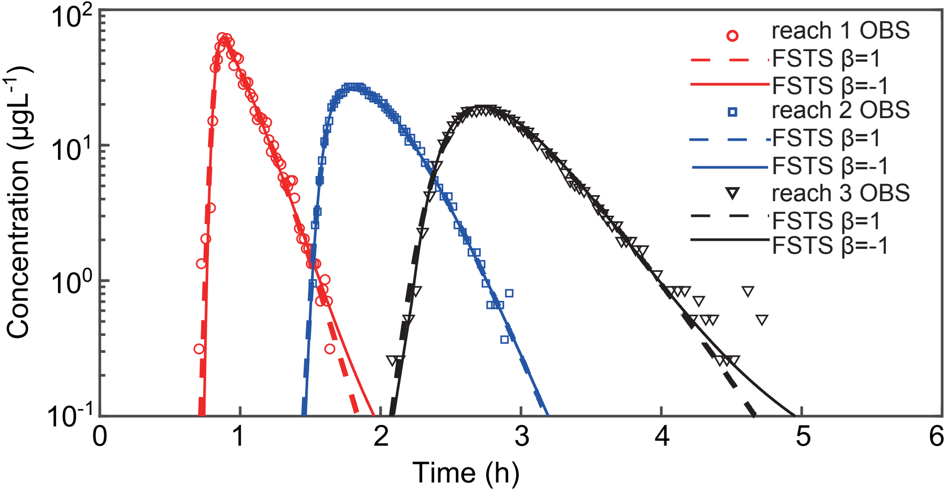
Concentration (μgL⁻¹)
10^2
10^1
10^0
10^-1
0
1
2
3
4
5
6
Time (h)
reach1 OBS
sFADE
reach2 OBS
sFADE
reach3 OBS
sFADE



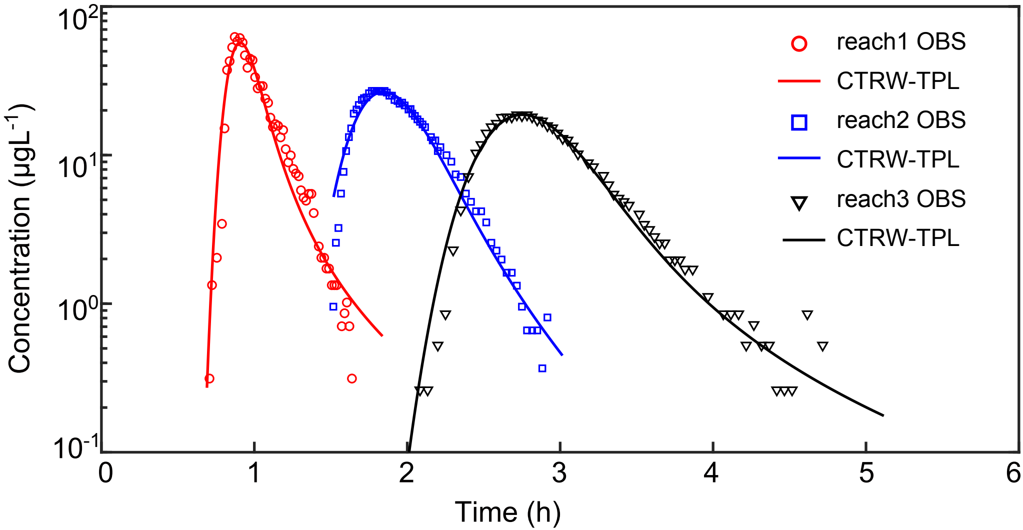
reach1 OBS
FMIM
TTFM
reach2 OBS
FMIM
TTFM
reach3 OBS
FMIM
TTFM
Concentration (μgL⁻¹)
Time (h)



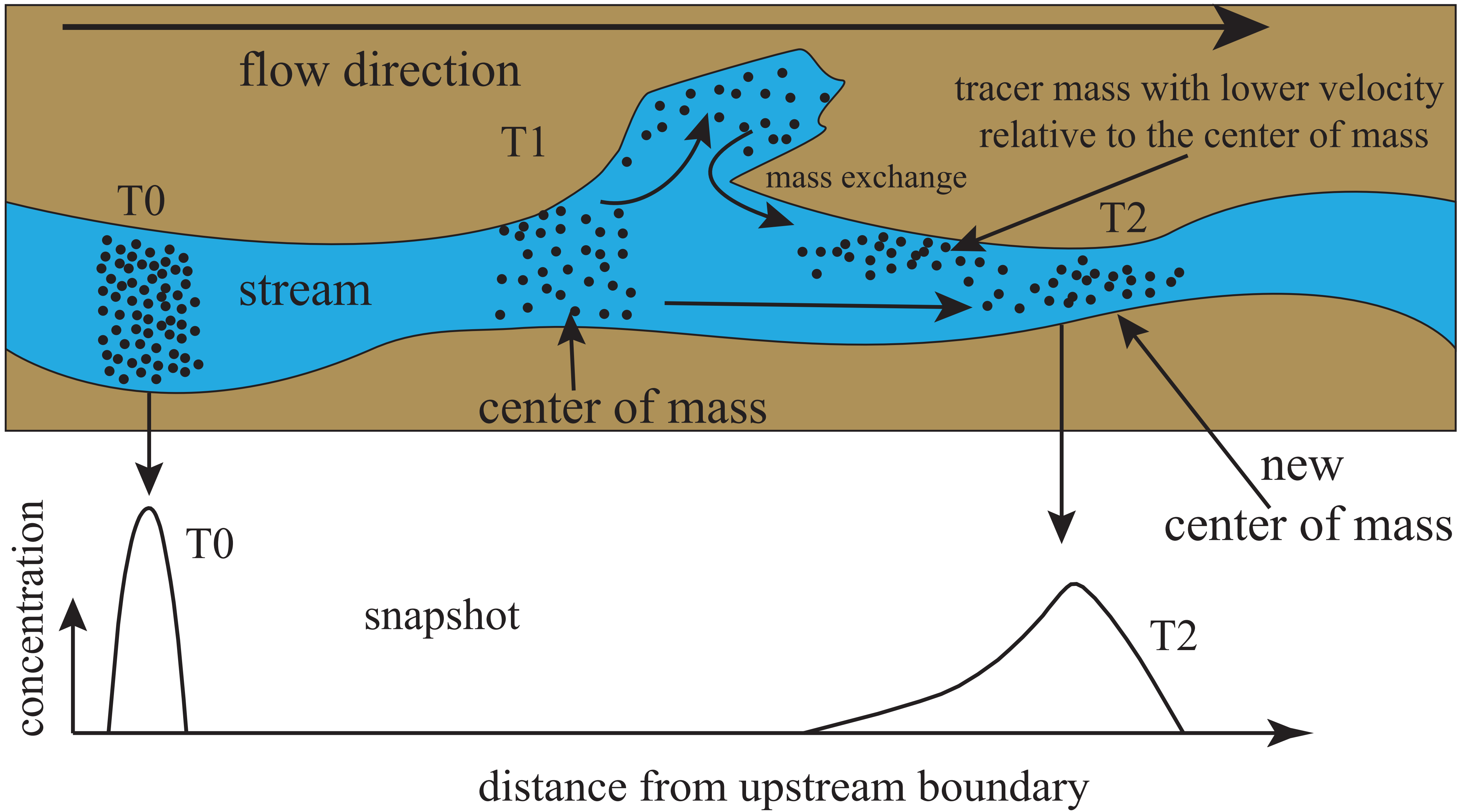
Concentration (μgL⁻¹)
10²
10¹
10⁰
10⁻¹
0
1
2
3
4
5
6
Time (h)
reach 1 OBS
MRMT-1
MRMT-3
reach 2 OBS
MRMT-1
MRMT-3
reach 3 OBS
MRMT-1
MRMT-3



reach 1 OBS
FSTS β=1
FSTS β=-1
reach 2 OBS
FSTS β=1
FSTS β=-1
reach 3 OBS
FSTS β=1
FSTS β=-1
Concentration (μgL-1)
Time (h)



Concentration (μgL⁻¹)
10²
10¹
10⁰
10⁻¹
0
1
2
3
4
5
6
Time (h)
reach1 OBS
CTRW-TPL
reach2 OBS
CTRW-TPL
reach3 OBS
CTRW-TPL


flow direction
T1
tracer mass with lower velocity
relative to the center of mass
mass exchange
T0
T2
stream
center of mass
new
center of mass
concentration
T0
snapshot
T2
distance from upstream boundary