Exploring Tracer Information and Model Framework Trade-offs to Improve Estimation of Stream Transient Storage Processes

Christa Kelleher[1,2], Adam Ward[3], J L A Knapp[4,5], P J Blaen[6,7], M J Kurz[8,9], J D Drummond[10], J P Zarnetske[11], D M Hannah[6], C Mendoza-Lera[12,13], N M Schmadel[3,14], T Datry[12], J Lewandowski[15], A M Milner[6] and S Krause[6,7]

1) Department of Earth Sciences, Syracuse University, Syracuse, NY 13244

2) Department of Civil Engineering, Syracuse University, Syracuse, NY 13244

3) School of Public and Environmental Affairs, Indiana University, 430 MSB-II, Bloomington, IN 47405, USA;

4) Department of Environmental Systems Science, ETH, Zurich, Switzerland;

5) Center for Applied Geoscience, University of Tübingen, Germany;

6) School of Geography, Earth and Environmental Sciences, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK;

7) Birmingham Institute of Forest Research (BIFoR), University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK;

8) The Academy of Natural Sciences of Drexel University, Philadelphia, USA;

9) Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany;

10) Integrative Freshwater Ecology Group, Centre for Advanced Studies of Blanes (CEAB-CSIC), C/Accés a la Cala St Francesc, 17, 17300, Blanes, Girona, Spain;

11) Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI 48824, USA;

12) Irstea, UR RIVERLY, Centre de Lyon-Villeurbanne, Villeurbanne Cedex, France;

13) Department of Freshwater Conservation, Brandenburg University of Technology Cottbus-Senftenberg, Seestraße 45, 10435 Bad Saarow, Germany;

14) Now U.S. Geological Survey

15) Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Müggelseedamm 310, 12587 Berlin, Germany

Corresponding author: Christa Kelleher (ckellehe@syr.edu)

Key Points:

- TSM interpretation improved with analysis of multiple tracers, but results in increased parameter uncertainty for a more complex model
- Nonconservative tracers enabled interpretation of parameters that were highly uncertain when estimated by conservative tracers alone
- Achieving reliable parameter estimates depends on choice of tracers and model framework, and should be coupled with uncertainty assessment

Abstract

Novel observation techniques (e.g., "smart" tracers) for characterizing coupled hydrological and biogeochemical processes are improving understanding of stream network transport and transformation dynamics. In turn, these observations are thought to enable increasingly sophisticated representations within transient storage models (TSM). However, TSM parameter estimation is prone to issues with insensitivity and equifinality, which grow as parameters are added to model formulations. Currently, it is unclear whether (or not) observations from different tracers may lead to greater process inference and reduced parameter uncertainty in the context of TSM. Herein, we aim to unravel the role of in-stream processes alongside metabolically active (MATS) and inactive storage zones (MITS) using variable TSM formulations. Models with one (1SZ) and two storage zones (2SZ) and with and without reactivity were applied to simulate conservative and "smart" tracer observations obtained experimentally for two reaches with differing morphologies. As we show, "smart" tracers are unsurprisingly superior to conservative tracers when it comes to partitioning MITS and MATS. However, when transient storage is lumped within a 1SZ formulation, little improvement in parameter uncertainty is gained by using a "smart" tracer, suggesting the addition of observations should scale with model complexity. Importantly, our work identifies several inconsistencies and open questions related to reconciling timescales of tracer observation with conceptual processes ("parameters") estimated within TSM. Approaching TSM with multiple models and tracer observations may be key to gaining improved insight into transient storage simulation as well as advancing feedback loops between models and observations within hydrologic science.

Plain Language Summary

Solute experiments and transport models, called commonly "tracer experiments," are used to understand the relative importance of different stream processes, especially those that influence water, solutes, and nutrients as they move through a stream network. Within these tracer experiments, there are processes that exchange mass beyond the main stream channel to other parts of the river valley bottom environment. Sometimes there are single or multiple types of tracers used and modeled to try to understand this exchange. There are also multiple models with different equations and structures to simulate these tracers. This study shows that what you can learn about these stream processes depends on experiment choices and which model you use. Hence, refining future multiple tracer experiments and models is needed to determine how we best obtain consistent measurements of key stream processes.

1 Introduction

The last decade has seen an explosion of novel techniques for collecting data used to characterize dynamic hydrologic systems. Tools and techniques that fall under this umbrella include the burgeoning field of hydrogeophysics (e.g., Ward et al., 2010; St Clair et al., 2015), the use of unmanned aerial vehicles (e.g., Vivoni et al. 2014; Brenner et al. 2017), high space-time resolution sensing systems (e.g. Blaen et al., 2016; Khamis et al., 2016) and the growing use of "smart" and conservative tracers in the environment (e.g., Haggerty et al., 2008; González-Pinzón et al., 2012; Runkel 2015; Knapp et al., 2017; Blaen et al., 2017). Observational data obtained from these techniques has been used to reveal new process dynamics and to refine current understanding of hydrological systems. These techniques have also advanced process-based mathematical representations within computational models as well as new approaches to assess whether model frameworks ensure model realism (e.g., Seibert and McDonnell, 2002; Li et al., 2015; Clark et al., 2017). Furthermore, research communities have developed approaches for testing multiple model frameworks that explore different mathematical representations of hydrological processes (e.g., Clark et al., 2015a, 2015b) as well as approaches for comparing the performance of different models applied to a variety of systems (e.g., Butts et al., 2004; Best et al., 2015). Our goal herein is to build on recent progress made by these communities to explore the relationship between empirical observations, model performance, and model complexity to inform the value of new information for addressing historic limitations. We use the example of stream solute transport, transient storage, and solute transformation as a study topic.

In the field of groundwater-surface water interactions, hyporheic exchange remains one of the most difficult processes to quantify (Orghidan, 1959; Triska et al., 1989; Gooseff, 2010; Boano et al., 2014). The hyporheic zone, although defined in a variety of contexts (Krause et al., 2011), is often described as a region both beneath and surrounding the stream channel containing sediments, microbes, and benthic organisms where water and nutrient exchange with the main stream channel occurs (Orghidan, 1959; Gooseff, 2010; Ward, 2016). Identifying this zone and characterizing the relative rates and spatial extent of hyporheic exchange with the nearby stream channel has been and continues to be an area of ongoing research (e.g., Triska et al. 1989; Storey et al., 2003; Boano et al., 2014; Caruso et al., 2016; Knapp et al. 2017; Schmadel et al. 2017). While quantifying the role of the hyporheic zone in relation to solute transformations and ecosystem processes has remained elusive, the use of multiple tracers, specifically the emerging “smart” tracer (i.e., resazurin; Raz) technique has shown promise for characterizing stream reactivity and functioning by enabling researchers to quantify the portion of the transient storage that is metabolically-active (Karakashev et al., 2003; Haggerty et al., 2009; Argerich et al., 2011; Knapp et al., 2018). Resazurin decays when in the presence of respiring cells typically found in the hyporheic zone (e.g., González-Pinzón et al., 2012), producing a new chemical, resorufin (Rru). Following this transformation, as water is exchanged across the streambed interface, Rru is exchanged back to the main channel and can be detected downstream. Thus, releasing Raz into a stream reach produces two time-series of concentration, referred to as breakthrough curves (BTC), that may be sensitive to different types of either metabolically active (MATS) or inactive storage (MITS). Beyond hyporheic exchange, decay of Raz to Rru is being widely used to characterize MATS, stream reactivity, and ecosystem respiration in many different transient storage zones (Knapp et al., 2018), including biofilms (Haggerty et al., 2014), the benthic zone (Knapp et al., 2017), vegetation beds (Kurz et al., 2017), and around woody debris (Blaen et al., 2018). In contrast to MATS, MITS may correspond to portions of a stream reach with a high volume of water and conversely low contact with sediment (e.g., in-stream dead zones).

While MATS and MITS are recognized as having two very different effects on stream nutrient exchange, there are few examples of TSM applications to reactive solutes (e.g., Gooseff et al., 2005; Knapp and Cirpka, 2017). Commonly, quantifying reach-scale transient storage has drawn upon parameter estimation with TSM representing the lumped effects of transient storage via MATS and MITS alongside advective in-channel processes such as advection and dispersion (Bencala and Walters, 1983; Valett et al., 1996). When combined with field observations of tracers in the form of a BTC, estimates of model parameters representing the temporal (i.e., rate of exchange) and spatial scales (i.e., size) of reach-averaged transient storage zone exchange can be obtained via inverse modeling (Runkel, 1998). This is done by employing one of many methods (e.g., Runkel, 1998; Wagener et al., 2001; Kelleher et al., 2013; Knapp and Cirpka, 2017) to search the parameter space for a parameter set that produces the simulation with lowest model error, assessed between a simulated and observed BTC for a given stream reach (Runkel, 1998; Ward et al., 2017). There are several recognized limitations to the timescales of transient storage zone exchange that may be assessed using TSM (e.g., Harvey et al., 1996) as well as accounting for spatial heterogeneity that exists at sub-reach scales (Harvey et al., 1996; Knapp et al., 2017). Despite these limitations, TSM remains a popular approach that can provide an assessment of the relative roles of different in- and near-stream processes.

The most commonly applied TSM (known as the One-Dimensional Transport with Inflow and Storage model, or "OTIS"; Runkel, 1998) uses four parameters to simulate transport of a conservative tracer (five for a non-conservative tracer, e.g., Raz transforming to Rru) with a single transient storage zone, and is available as open-source software from the United States Geological Survey. However, this single-storage zone representation is inconsistent with current understanding of transient storage (Briggs et al., 2009), as there are multiple dominant domains of transient storage that may alter the flow of water and nutrients in different ways. As a result, several other structural forms of the solute transport equations have been proposed (e.g., Lees et al., 2000; Bencala and Walters, 1983; Runkel, 1998; Lees et al., 2000; Marion et al., 2008; Briggs et al., 2009; Liao and Cirpka, 2011; Ward et al., 2015). One recent iteration of this model separates the effects of transient storage into two zones described by parameters in terms of size and exchange rate with the main channel (Briggs et al. 2009). Generally, we desire models with process representations that most closely match our understanding of streams (e.g., Briggs et al., 2009). This desire often results in the addition of model parameters, with the tradeoff of introducing additional uncertainty and equifinality due to parameter interactions (Beven 1993; Butts et al. 2004; Beven 2006). This must also be balanced against the addition of observations to vet simulations and constrain realistic parameter estimates. For instance, as shown by Briggs et al. (2009), the addition of model parameters to segment transient storage was accompanied by additional BTC observations from in-channel dead zones. Though numerous TSM studies exist, there is a broad need to better understand the tradeoffs between parameter uncertainty and choices that determine the number of estimated model parameters (i.e., model complexity, motivated by more realistic representation of dominant processes) alongside the addition of field observations for estimating parameter values (i.e., “smart” tracers).

An added challenge to quantifying stream-reach transient storage is the growing body of evidence that has shown that TSMs are susceptible to parameter equifinality (e.g., Choi et al., 2000; Kelleher et al., 2013), such that parameter determinations may be uncertain and therefore uninformative for assessing the role of transient storage in physical and ecological river processes (Wagner and Harvey, 1997; Wagener et al., 2002; Ward et al., 2017). Existing studies suffering from equifinality issues have typically assessed parameter estimates and uncertainty through inverse modeling of a single conservative tracer. When used, as shown in a small but growing number of studies, “smart” tracers provided different estimates for dispersion and transient storage parameter values (Lemke et al., 2013). Adding observations to constrain models is often viewed as an approach for reducing parameter uncertainty (e.g. Nearing and Gupta, 2015; Nearing et al., 2016). In practice, this requires that the observations in question contain non-redundant information. If new or more observations lead to the same parameter estimates, or similar levels of parameter uncertainty, the added information is not useful in reducing parameter uncertainty. As parameter estimates are often used to characterize systems, collecting datasets that can reduce this uncertainty is a common goal. Consequently, there is a need to explore how equifinality and process inference may vary with multiple tracer observations, across different types of stream reach morphologies and across model formulations of varying complexity.

Within the context of TSM, we explore whether conservative and non-conservative "smart" tracers may better constrain different TSM parameters, providing alternative but potentially complimentary information. Furthermore, we offer a unique comparison of constraints on parameter uncertainty arising from estimation of parameters by fitting to different tracer BTCs across model frameworks of varying complexity (e.g., conservative and non-conservative tracers, single versus multiple storage zone models). We aim to address the following two fundamental questions in the context of TSM parameter estimation: (1) when are multiple tracers useful?, and (2) when is increasing model complexity beneficial? Drawing from several growing efforts in the land-surface and watershed modeling communities, we take a model intercomparison-based approach (e.g., Best et al., 2015; Clark et al., 2015a, b; Clark et al., 2017), treating these TSM model formulations as different testable hypotheses, comparing the performance and parameter uncertainty associated with each unique model formulation. We test this approach using data from conservative (uranine; Ura) and reactive (Raz, Rru) solute tracer experiments performed in two lowland stream reaches with distinct morphological settings located in the Hammer Stream, in West Sussex, UK. To evaluate the addition of observations alongside changes to model complexity, we test four different formulations of the TSM, ranging in complexity from four to seven parameters. While we expect that inverse models constrained by Raz and Rru will reduce uncertainty and yield similar parameter estimates for active and inactive storage zone parameters, we speculate that uncertainty in main channel parameters may grow in the more complex model framework associated with two storage zones.

2 Study Area

Field experiments were conducted in the Hammer Stream (West Sussex, UK; 51°0' N 0°47' W; Figure S1). The 2,640 ha catchment drains mixed land use and is primarily underlain by sandstone and mudstone. We identified two study reaches located upstream and downstream of Hammer Lake. Upstream of the lake the streambed material is sandy (hereafter the sand reach), whereas the reach downstream of the lake is armored gravel (hereafter the gravel reach) as result of the sand and other fine sediment having been trapped in Hammer Lake. Both reaches include large woody debris in the stream channel (Blaen et al., 2018; Shelley et al., 2017). For the sand reach, we established a study reach along 760 m of channel (mean width 5.3 m, mean depth 0.42 m). For the gravel reach, we established a study reach of 683 m immediately downstream of Hammer Lake (mean width 6.35 m, mean depth 0.28 m). In each reach, a combination of Ura and Raz was injected as an instantaneous pulse about 150-m upstream of the start of the study reach to ensure complete mixing, even at the start of the study reach. All injections occurred in late afternoon/evening to minimize the effect of tracer mass photodegredation. In-situ field fluorometers (GGUN-FL30, Albillia Sàrl, Switzerland) were used to monitor the fluorescence signals of all three tracers at 10-s time intervals at each end of the study reach. Discharge was 73.2 L s^{-1} at the upstream end of the sand reach and 86.5 L s^{-1} at the upstream end of the gravel reach, and calculated using dilution gaging with Uranine at the upstream end of the study site. Additional details regarding the sand reach and injection are provided by Blaen et al. 2018; the gravel reach injection replicated the same experimental methods.

3 Transient Storage Modeling

3.1 Model Formulation

We derive models representing transport and transformation of solute tracers following closely after the TSM (Thackston and Schnelle 1970; Bencala and Walters 1983) and integrate several subsequent extensions such as multiple storage zones (e.g., Briggs et al. 2009; Kerr et al., 2013), reactivity (e.g., Haggerty et al. 2009; Lemke et al. 2013), and transport of multiple interacting solutes (Keefe et al., 2004; Ward et al., 2015). While many TSM formulations have partitioned transient storage using location (e.g., surface and sub-surface transient storage; Briggs et al. 2009; Kerr et al. 2013), MATS and MITS formulations separate transient storage based on the apparent presence or absence of metabolic activity (Argerich et al. 2011).

In this TSM formulation, we simulate concentration in the main channel, the MITS domain, and the MATS domain via three equations with flexibility to vary up to seven different parameters. Concentrations in the stream domain are described according to:

$$\frac{\partial C}{\partial t} = -\frac{Q}{A}\frac{\partial C}{\partial x} + D\frac{\partial^2 C}{\partial x^2} + \frac{q_{L,in}C_L}{A} - \frac{q_{L,out}C}{A} + \alpha_{MATS}(C_{MATS} - C) + \alpha_{MITS}(C_{MITS} - C) \quad (1a)$$

where C is solute concentration (g m^{-3}), t is time (s), Q is discharge (m^3 s^{-1}), A is cross-sectional area of the stream (m^2), D is the longitudinal dispersion coefficient (m^2 s^{-1}), $q_{L,in}$ is the lateral inflow per meter of stream (m^2 s^{-1}), C_L is the concentration of the solute in the lateral inflow (g m^{-3}), $q_{L,out}$ is the lateral outflow per meter of stream (m^2 s^{-1}), and α describes the exchange rate between the stream and transient storage zones (s^{-1}). Inflows and outflows to the system simulated as occurring in the stream instead of the hyporheic zone, consistent with TSM conceptualization (Bencala and Walters, 1983). For the purposes of this experiment, both $q_{L,in}$ and $q_{L,out}$ were set to zero on the basis of minimal changes in discharge and the absence of known surface outflows along the study reach and to minimize the number of free parameters. Furthermore, surface inflows do not contain the tracers Raz, Rru or Ura (i.e. C_L is also zero).

Within the MATS domain (denoted by subscript MATS), the mass balances for Raz, Rru, and Ura (denoted by subscripts) are calculated as:

$$\frac{\partial C_{MATS,Raz}}{\partial t} = \alpha_{MATS}\frac{A}{A_{MATS}}\left(C_{Raz} - C_{MATS,Raz}\right) - kC_{MATS,Raz} \quad (1b)$$

$$\frac{\partial C_{MATS,Rru}}{\partial t} = \alpha_{MATS}\frac{A}{A_{MATS}}\left(C_{Rru} - C_{MATS,Rru}\right) + kC_{MATS,Raz} \quad (1c)$$

$$\frac{\partial C_{MATS,Ura}}{\partial t} = \alpha_{MATS}\frac{A}{A_{MATS}}\left(C_{Ura} - C_{MATS,Ura}\right) \quad (1d)$$

where k (s^{-1}) is a reactive rate constant that describes the transformation of the parent (Raz in our study) to product (Rru in our study), and C_{Ura}, C_{Raz}, and C_{Rru} are the in-stream concentrations of Uranine, Resazurin, and Resorufin based on solving equation 1a for each solute. Similarly, in the MITS domain (denoted by subscript MITS) the concentrations of Raz, Rru, and Ura are calculated as:

$$\frac{\partial C_{MITS}}{\partial t} = \alpha_{MITS}\frac{A}{A_{MITS}}\left(C_{Raz} - C_{MITS,Raz}\right) \quad (1e)$$

$$\frac{\partial C_{MITS}}{\partial t} = \alpha_{MITS} \frac{A}{A_{MITS}} \left(C_{Rru} - C_{MITS,Rru} \right) \quad (1f)$$

$$\frac{\partial C_{MITS,Ura}}{\partial t} = \alpha_{MITS} \frac{A}{A_{MITS}} \left(C_{Ura} - C_{MITS,Ura} \right) \quad (1g)$$

Simulations are performed through forward modeling based on the stream (1a), MATS (1b-1d), and MITS (1e-1g) equations for Ura, Raz, and Rru. Critical to this study is that all three equations are solved using the same physical transport parameters *(A, D,* A_{MATS}*,* A_{MITS}*,* α_{MITS}*,* α_{MATS}*)*, allowing for simulation of dynamic parent-to-product transformations. This solution allows the simultaneous transport and interaction of both conservative and interacting reactive solutes (after Ward et al. 2015). Model equations for all solutes were solved simultaneously using a Crank-Nicolson solution scheme, common for TSM applications (e.g., Runkel 1998; Ward et al., 2015). Models were implemented using measured discharge at the upstream end of each study reach, with observed breakthrough curves used as upstream boundary conditions. Spatial and temporal steps for the simulations were fixed at 5-m and 10-s, respectively. Important and commonly used assumptions of the model include laterally and vertically well-mixed domains, exponential residence time distributions within transient storage zones, temporal constancy for transient storage zone model parameters, perfect conversion of Raz to Rru, no retardation (sorption), and no additional transformation pathways for any solutes.

As derived, the model is flexible in that it can represent both one storage zone (1SZ) and two storage zone (2SZ) realizations of the TSM (Figure 1a). Within this framework, we test the following model-tracer combinations (Figure 1b):

(1) A one storage zone model fit to a conservative tracer (Ura), where transient storage combines MATS and MITS (four parameters; *A, D,* A_s*,* α_S)

(2) A two storage zone model fit to a conservative tracer (Ura), where MITS and MATS do not distinguish active storage, but instead represent two different storage zones (six parameters; *A, D,* A_{MATS}*,* α_{MATS}*,* A_{MITS}*,* α_{MITS})

(3) A one storage zone model fit to "smart" tracer (Raz) and biproduct (Rru), where transient storage refers to MATS, and MITS is effectively incorporated into the dispersion term *(D)* (five parameters; *A, D,* A_{MATS}*,* α_{MATS}, *k*)

(4) A two storage zone model fit to "smart" tracer (Raz) and biproduct (Rru) (seven parameters; *A, D,* A_{MATS}*,* α_{MATS}*,* A_{MITS}*,* α_{MITS}*,* *k*).

While comparison (2) is included in this study to assist with interpretation, this combination of a two-storage zone model fit with a conservative tracer is not expected to yield useful storage zone parameter estimates. Each tracer was independently tested as a source of parameter information for each tracer-model combination listed above. Importantly, MATS and MITS, as visualized in Figure 1a, are integrations of transient storage along the channel that may either be reactive or inactive, respectively. MATS and MITS, in our formulation, do not represent physical zones such as the hyporheic zone, as many of these physical locations may include zones of both active and inactive storage. For 1SZ models, the exchange coefficient α_{MITS} is set to zero, which eliminates any exchange between the stream and MITS. For Ura, this represents a 1SZ model identical in formulation and implementation to the broadly used USGS OTIS model (Runkel 1998). For Raz and Rru, the MATS storage retains the ability to simulate transformations in the storage zone and assumes that all transient storage is metabolically active (i.e., with α_{MITS} =0, A_{MITS} cannot affect concentrations in the model). We do assume Ura to be a conservative tracer (see Supporting Information), though it may decay in direct sunlight. Notably, Rhodamine WT is also non-conservative (Runkel, 2015), highlighting that there is no single perfect conservative tracer.

Conceptually, the combinations of observations and model formulations listed in Figure 1b also represent four different scenarios for gaining insights regarding parameter importance. Though we assess parameter sensitivity and uncertainty per tracer and per model, we do not expect all parameter estimates to be sensitive to all tracers, and seek to test these potential relationships. In this same vein, certain tracers are likely to be more or less informative for different parameters. We expect that Ura, as a conservative tracer, will yield the most physically representative parameter distributions for A and D. Similarly, we do not expect that Ura is capable of separating the influence of *MATS* and *MITS*, and are uncertain as to whether A and D are sensitive to Raz or Rru, given these tracers are nonconservative.

Importantly, parameters estimated via different tracers represent different processes. Within the 1SZ formulation, storage zone parameters estimated via Ura assume the transient storage zone combines both MATS and MITS, while storage zone parameters estimated via Raz estimate transient storage zone size for MATS, assuming MITS is incorporated into the dispersion term. Thus, we may not necessarily expect distributions of D or storage zone parameters to be similar when fitting to Ura versus Raz with the 1D model.

3.2 Computational Experiments

We performed several computational experiments with inputs (parameters) to and outputs (errors and simulations) from models of varying complexity. Simulations and parameter sets were constrained to match different observations, including conservative (Ura) and non-conservative (Raz, Rru) BTCs. Model formulations used in these experiments are outlined in Figure 1b. To interrogate parameter uncertainty and equifinality, we used a Latin Hypercube approach to sample the model parameter space (N = 27,000 runs; e.g., Pianosi et al., 2015). All parameters and associated ranges are listed in Table 1.

Within the 2SZ model formulation, we sampled total area (A_{TOT}), representing the combined area of the advective channel and the area of MITS. For this formulation,

$$A = A_{TOT} \cdot (1 - F_{MITS}) \qquad \text{2a}$$

$$A_{MITS} = A_{TOT} \cdot F_{MITS} \qquad \text{2b}$$

where F_{MITS} describes the fraction of the stream channel that is metabolically inactive. To enable comparisons across model formulations, all results are presented in terms of A and A_{MITS}. This does result in slightly wider bounds for A for the 2SZ model and narrower bounds for A for the 1SZ model, but otherwise is purely a function of model formulation.

Model complexity, defined by the number of parameters, ranged from four to seven parameters. We tested a 1SZ model (Fig. 1b, 1) and 2SZ model (Fig. 1b, 2) constrained by observations from only Ura. We also tested a 1SZ (Fig. 1b, 3) and a 2SZ (Fig. 1b, 4) with an added parameter representing reactive decay constrained by BTCs for Raz and for Rru. All computational experiments were performed using the same structural model equations (Equations 1a, 1b, 1c, and 1d). For model formulations 1 and 3, we use a model formulation that has a single transient storage zone (i.e., $\alpha_{MITS} = 0$). To model this, we sampled the first five parameters, setting values for the fraction of stream area as MITS and the MITS exchange rate to small non-zero values (10^{-10}). For model formulations 2 and 4, we sampled all seven parameters across feasible ranges.

3.3 Model Performance

For each of the 27,000 runs, we calculated model fits in terms of a normalized Root Mean Squared Error *(nRMSE)* for each BTC (Ura, Raz, Rru) independently, according to:

$$nRMSE = \frac{1}{C_p}\left(\frac{\sum_{t=1}^{n}(O_t - S_t)^2}{n}\right)^{0.5} \quad (2)$$

where O_t and S_t correspond to observations and simulations at a given time step, n is the total number of observations, and C_p is peak concentration for each tracer (employed to normalize RMSE values across tracers; g m^{-3}). RMSE (and close variants) remains one of the key objective functions used to assess BTC errors (Runkel, 1998; Ward et al., 2017). We also calculated a log-transformed Root Mean Squared Error *(LRMSE;* similar to a weighted *RMSE)*, where the observed and simulated time-series were log-transformed prior to applying equation (2) above. Past work has shown log-transformed error metrics can be particularly useful for obtaining reliable TSM parameter estimates (e.g., Wörman and Wachniew, 2007; Ward et al., 2017).

Our analysis relies on the use of behavioral thresholds to segment populations of error and parameter estimates (e.g., Hornberger and Spear, 1980, 1981; Spear and Hornberger, 1980). We employ a behavioral threshold to identify a subset of simulations and parameter sets that closely match BTCs by achieving low errors. Instead of selecting a single best simulation and parameter set, the use of behavioral thresholds allows us to identify a distribution of these values. Behavioral thresholds may be implemented by identifying parameter sets and simulations below a certain error value, or by identifying those with errors below some percentage criterion (i.e., top 10% of errors). We use the latter (thresholds of 10% and 1%) to compare error, simulations, and parameter estimates across different tracers and models.

3.4 Parameter Sensitivity and Uncertainty

For an ideal TSM inverse modeling exercise there is a unique 'best' estimate for each parameter, such that behavioral parameter values occupy a defined and narrow area of the parameter space (Wagener et al., 2001; Kelleher et al., 2013). However, parameters are often insensitive or uncertain. This may be represented as wide distributions of behavioral parameter values spanning the entire parameter range, or by no difference between parameter distributions between the best and worst simulations. The former may also occur when a parameter is largely unimportant, or due to interactions with other parameters. To understand the influence of model complexity and different tracer observations on parameter estimates, we assessed parameter sensitivity, parameter uncertainty, and parameter interactions for all model-tracer combinations. While some studies have estimated parameters by first fitting parameters associated with conservative transport, and then fitting nonconservative parameters (e.g., Gooseff et al., 2005), we treat all BTCs as independent sources of information for assessing parameter sensitivity and uncertainty. Our goal is to avoid making assumptions about which BTCs may contain information regarding certain parameters, and instead to use the analysis presented here to more thoroughly assess how parameter estimates are impacted by fitting to different BTCs.

Approaches to obtain parameter estimates include use of optimization algorithms (e.g., Runkel, 1998; Briggs et al., 2009; Kerr et al., 2013), Markov Chain Monte Carlo approaches (e.g., Lemke et al., 2013; Knapp and Cirpka, 2017), and Monte Carlo approaches coupled with behavioral thresholds (e.g., Wagener et al., 2001; Kelleher et al., 2013), as well as a broad literature on approaches to parameter sensitivity (see Pianosi et al., 2016). In this study, we employed approaches based on Monte Carlo methods to enable a tiered assessment that draws on both the very best and worst simulations and corresponding parameter estimates. To provide a global assessment of parameter sensitivity, we generated regional sensitivity analysis (RSA) plots for each parameter based on errors associated with *nRMSE* calculated from simulations and observations of Ura, Raz, and Rru (Fig. 2a). RSA is a useful technique for mapping portions of the parameter space corresponding to either best or worst errors (e.g., Freer et al., 1996; Pianosi et al., 2016) and has been commonly applied to assess TSM parameter sensitivity (e.g., Wagener et al., 2002; Wlostowski et al., 2013). To apply RSA, we identified the top (best) 10% of errors and the bottom (worst) 10% of errors for Ura, Raz and Rru across all simulations. Parameter values corresponding to these best and worst 10% of simulations were transformed into marginal empirical cumulative distribution functions (CDF; Freer et al., 1996; Wagener et al., 2001; Pianosi et al., 2016). Sensitive parameters satisfied two criteria: parameter CDFs corresponding to the top 10% of all error values (1) deviated from the 1:1 line (representing a purely uniform distribution), assessed by visual inspection, and (2) deviated from parameter CDFs corresponding to the worst 10% of all errors (Fig. 2a, b).

We assessed parameter uncertainty and model performance comparing the top 1% of all simulations per error metric. To test whether the parameter values corresponding to the lowest model errors converged, we applied a visualization based on the widely used dotty plot (e.g., Wagener and Kollat, 2007). Dotty plots visualize model error plotted against model parameter values for all simulations meeting a given behavioral threshold (Fig. 2c, d). To summarize the distribution of optimal parameter values (those corresponding to the lowest error) across each dotty plot, we identified the single best parameter value (with lowest error) within a moving window ($1/20^{th}$ the width of parameter range) incremented across each parameter range ($1/40^{th}$ the width of the parameter range). This distribution of optimal errors was then normalized to a cumulative value of one (Fig. 2c, d). We report all dotty plots in Supporting Information (Figs. S1-S4). Optimal parameter values and 90% confidence intervals are also reported. Finally, we also investigated parameter interactions via scatter plots of parameter values to assess the dependency between parameters and how this changes for subsets of the very best simulations. Together, these assessments yield transferable approaches for assessing parameter sensitivity and uncertainty within environmental models, and for comparing these outcomes across 1SZ and 2SZ models and error metrics.

4 Results

4.1 Model errors and simulations

Tracer observations obtained from the two reaches are shown in Figure 3. While peak concentrations for Ura and Raz are coincident, peak concentrations for Rru occur at a later time, representing a temporal lag as Raz is converted to Rru in the presence of aerobic respiration. All tracers are capably simulated by one and two-storage zone models (Figure 4). As we show, a behavioral threshold of 1% yielded envelopes of simulations that bracketed observations for all tracers. Upper and lower bounds, representing the range of the 270 simulations with lowest error, were nearly identical for the 1SZ and 2SZ models. Information on mass recovery is included in

Supporting Information (Text S2) and demonstrated acceptable levels of mass recovery conforming to model assumptions.

Model errors corresponding to the top 1% of all model simulations are visualized for *nRMSE* in Figure 5. Comparing distributions of error across tracers, $nRMSE_{Raz}$ had the lowest overall error across both reaches, though medians and ranges of error between tracers were similar for all but $nRMSE_{Rru}$. Minimum and median errors for Rru were larger for the 2SZ as opposed to 1SZ model, and $nRMSE_{Rru}$ had larger errors than $nRMSE_{Raz}$ and $nRMSE_{Ura}$. Comparing errors across models, we found that median errors for 2SZ models were slightly higher than median errors for 1SZ models for nearly all reach-tracer combinations. Though the ranges of error were found to be wider for 2SZ as opposed to 1SZ models, 2SZ models still yielded the simulation with the single lowest error across all parameter sets for $nRMSE_{Raz}$ for both reaches and $nRMSE_{Ura}$ for the gravel reach (Figure 5a).

4.2 Parameter sensitivities

Interpretation of global parameter sensitivities assessed via RSA are shown in Figure 6. A select number of RSA plots are included for the 2SZ models, with all plots included in Supporting Information (see Figure S8). In general, *D, A,* and $\propto_{MATS}$ were globally sensitive across tracers. Distributions for *D* differed between tracers. For the gravel reach, lower errors for $nRMSE_{Ura}$ corresponded to larger values for *D*, but smaller values of *D* for $nRMSE_{Raz}$ and $nRMSE_{Rru}$. A_S (1SZ) and A_{MATS} (2SZ) were both sensitive, the latter to $nRMSE_{Ura}$ and $nRMSE_{Raz}$ and the former to $nRMSE_{Raz}$ and $nRMSE_{Rru.}$ Estimates for A_{MITS} and $\propto_{MITS}$ were difficult to interpret, in part, because CDFs corresponding to both best and worst performing parameter values were similar, likely indicating that these parameters are influenced by interactions with other parameters. Finally, *k* was globally insensitive across all models and performance metrics.

4.3 Parameter uncertainties

Parameter estimates corresponding to the top 1% of *nRMSE* values for each tracer are summarized as distributions in Figure 7 (parameter values and confidence intervals are reported in Table S2). In general, flatter distributions indicate that all values across the parameter range produce equal model errors, while the presence of peaks indicates certain areas of the parameter space produce higher or lower errors, suggesting that there are optimal values that better simulate observations. Comparisons of distributions were informative for testing whether conservative versus “smart” tracer errors yielded differences in parameter uncertainty as well as whether regions of the parameter space corresponding to the best simulations and therefore minimum error were similar across tracers.

Across models and reaches, parameter estimates were peaked for *A* and *D*, and narrower for the 1SZ (vs. 2SZ) errors. Within the 1SZ models, parameter estimates were uncertain for lumped transient storage size *(A_S)* for all tracers (note that parameter values with lowest error were distributed across the entire parameter range, spanning two orders of magnitude). In contrast, PDFs for $\propto_S$ were peaky for $nRMSE_{Ura}$ and $nRMSE_{Rru}$ (Fig. 7).

Separating transient storage into two storage zones (two indiscernible zones for Ura; MATS and MITS for Raz, Rru) introduced different patterns of parameter uncertainty. PDFs for A_{MATS} were wide for all reach-tracer combinations. Empirical PDFs of $\propto_{MATS}$ suggest better estimates for this parameter correspond to lower values when fitting to Ura and Raz, and better estimates correspond to higher values when fitting to Rru. MITS parameters ($A_{MITS}, \propto_{MITS}$) were best constrained by $nRMSE_{Rru}$ (Fig. 7).

Employing a non-conservative tracer (e.g., Raz) introduced an additional parameter k to describe the rate of transformation from Raz to Rru within the 1SZ and 2SZ models. This implies that fitting to one or both of these "smart" tracers should reduce uncertainty in this value. Global sensitivity analyses suggest that k is less sensitive than some storage zone parameters (e.g., Figure 6, Figure 7). However, dotty plots of k for both reaches (Figure S7) do suggest that this parameter is both sensitive (i.e., errors vary across the parameter range) and unique (such that a single best value exists within the parameter range). Within these dotty plots, values for k appeared insensitive to Raz, suggesting that Rru may contain more information for estimating this parameter.

Our results also underscore the importance of considering alternative objective functions. While not the primary focus of this manuscript, we include an additional assessment of parameter uncertainty with respect to log-transformed *nRMSE* (*LRMSE*) for both reaches. PDFs for the gravel reach were generally wide across parameters, suggesting that *nRMSE* was a more informative error metric in this reach (Figure S4). In contrast, empirical PDFs for the top 1% of errors by *LRMSE* were peaked for nearly all sand reach parameter values (Figure S6; Figure S7). These results demonstrate potential value added by considering alternative error formulations in assessments of parameter uncertainty.

4.4 Joint Distributions

Given past work suggesting that TSM parameters are influenced by interactions, we examined joint distributions of parameter values to explore how the interactive nature of TSM may impact parameter estimates and parameter uncertainty (Figs. 8, 9). While some parameters may be globally insensitive (Fig. 6) or exhibit flat parameter distributions (Fig. 7), visualizing joint distributions can reveal the presence of more complex relationships as well as the value of different tracers to discern these relationships. Figure 9 displays how parameter estimates and joint distributions varied with model complexity. Joint distributions of A and D were bimodal, and widened for the 2SZ model. Similar patterns were also observed between A_S (1SZ) and A_{MATS} (2SZ). In particular, these plots display that estimates for $\propto_S$ were best constrained by Ura and Rru for the 1SZ model, but Rru for the 2SZ model, shown by the shrinking 2D boundary of highest performing parameter combinations. While PDFs of parameter estimates for k did not reveal any strong patterns, joint distributions suggest lower errors are concentrated in a distinct portion of the parameter range for k (Raz, 1SZ; Rru, 2SZ). Last, we explored joint distributions between parameters only present in the 2SZ model, A_{MITS} and $\propto_{MITS}$ (Fig. 9). In particular, these joint distributions display the importance of Rru for refining estimates of MITS parameters. We note that Figures 9 and 10 display results for the gravel reach, with visualizations for the sand reach included in Supporting Information (Fig. S8 and S9), as the patterns of these joint distributions were similar between the reaches.

5 Discussion

5.1 Model complexity and conceptualization, simulations, and errors

Behavioral simulation bounds (Fig. 4; Fig. S2) and model errors (Fig. 5) indicate that, regardless of the tracer error or model framework used to constrain behavioral simulations, observations of all tracers were simulated to a reasonable degree of accuracy. Average errors for behavioral simulations were similar between 1SZ and 2SZ models. However, it is important to note that accurate simulation from an inverse model do not necessarily indicate meaningful information was gained from parameter values. Parameters that are not identifiable may provide a good inverse model fit without characterizing system processes and should not be over-interpreted (e.g., by comparing or interpreting values of insensitive or uncertain parameters). Thus, we echo calls for assessment of model parameter uncertainties, interactions, and identifiability as a requisite step prior to their interpretation (Wagener and Harvey, 1997; Wagener et al., 2002; Kelleher et al., 2013; Ward et al., 2017).

For the stream reaches analyzed in this study, we found that employing a more complex model did not necessarily yield simulations that better approximated tracer observations. Given the increased degrees of freedom in a 2SZ (as opposed to 1SZ) model, we expected 2SZ models to display smaller magnitude and range of errors than the 1SZ formulation errors calculated between measured and simulated BTCs. Instead, 1SZ versus 2SZ model errors were similar (and even notably larger for Rru), though the simulation with the lowest error was almost always generated with a 2SZ model (Fig. 5). We do not believe that these similarities in error indicate that the model is a poor representation of reality, as simulations well approximated observations (Fig. 4). Instead, we postulate that this shows that adding additional parameters introduces further uncertainty in addition to degrees of freedom, yielding similar model fits. It is also possible that a more complex or alternative model formulation could lead to improvements in error, and potentially a better representation of reality. Given the many iterations of TSM formulations (e.g., Gooseff et al., 2003; Gooseff et al., 2007; Kerr et al., 2013), we advise future work is needed to perform TSM model intercomparison with respect to both conservative and "smart" tracer BTCs.

While we sought to compare parameter inference through uncertainty assessment across multiple models and tracers, this introduces some challenges in interpretation. This is because transient storage parameters conceptually represent different processes when inverse modeling is performed with respect to different tracers. Transient storage zones cannot be partitioned into MATS or MITS through inverse modeling to simulate Uranine. Instead, transient storage zone parameters estimated via fitting a 2SZ model to Uranine assumed these parameters represent two independent storage zones with no association with MATS or MITS. Therefore, parameter distributions for storage zone parameters represent fundamentally different processes when fitting to Ura versus Raz and Rru, and as such, are not expected to be comparable. For this reason, we do not recommend estimating 2SZ MATS and MITS parameters by fitting to Ura, but include this comparison to emphasize that combining different model formulations and tracers can lead to fundamentally different conceptual representations of a system. Likewise, in the 1SZ formulation, storage parameters are assumed to represent MATS processes when fitting to Raz, with MITS lumped with dispersion. Therefore, we did not expect empirical PDFs for these parameter values to be similar. Indeed, these differences likely explain why fitting to Ura versus Raz yields such different empirical PDFs for $\propto_S$ (Figure 7). These differences also show that parameter estimates obtained by fitting a 1SZ model to Ura are not comparable to parameter estimates obtained by fitting a 1SZ model to Raz.

5.2 How do conservative versus nonconservative tracers affect parameter uncertainty?

In contrast to the expectation that conservative tracers may not always provide meaningful parameter estimates, our results show that conservative tracer BTCs do contain useful information for estimating TSM parameter values. In support of this, we found parameter uncertainty tended to be lower for parameters fit to conservative tracer BTCs (i.e., black distributions are narrower than blue or red distributions in Fig. 7). Encouragingly, we found that relatively narrow estimates for α could be achieved using a conservative tracer with either a 1SZ or a 2SZ model. This is in contrast to studies that have concluded α is typically highly uncertain (Wagener et al., 2002; Kelleher et al., 2013; Wlostowski et al., 2013). Thus, estimates of α with low uncertainty can be achieved, but this model result may be dependent upon the system and tracers. Overall, using multiple tracers allowed us to estimate and evaluate BTC parameters to a higher degree than could be achieved by using a single tracer, with consistency in findings across both reaches. We therefore recommend TSM parameter estimates and subsequent process-based interpretation should be based on the combination of conservative and non-conservative tracers.

The parameter we found most problematic to estimate was k, which describes the transformation of Raz to Rru and effectively determines mass balance. As shown in Figures 8 and 9, we found k to be highly interactive, which may explain apparent insensitivity and uncertainty for this parameter (Figures 6 and 7, Table S1). Furthermore, dotty plots (Fig. S7) between k and objective functions ($nRMSE_{Rru}$, $LRMSE_{Rru}$) show that k is indeed sensitive, it is just less sensitive than other model parameters. While only a few studies exist that employ formulations of MITS and MATS alongside "smart" tracer observations, some have concluded, similar to our study, that k may be highly uncertain (Yakerivich et al., 2017). Others have found low uncertainty for k through joint fitting of multiple tracers (Lemke et al., 2013). As this value is of particular interest to biogeochemists, future research with paired conservative and nonconservative tracer experiments will be needed to identify conditions that may lead to more (or less) uncertain k estimates.

In a similar vein, parameters A_S (1SZ) and A_{MATS} (2SZ) were also uncertain across study reaches (Fig. 7). Though we observed some organization between the structure of first order parameter interactions and model errors, our work suggests that these processes were difficult to estimate in this particular system. While not performed here, other analyses of parameter sensitivity and uncertainty (e.g., Kelleher et al., 2013) have shown that sometimes nested sampling schemes (narrowing bounds on certain parameters before completing additional analysis) can improve estimates of parameter values and associated uncertainty. This is because fitting to all BTCs is likely to be dominated by first finding best estimates for A and D. Fixing these values to narrow ranges, thereby reducing degrees of freedom, enables the importance of other parameters less sensitive than A and D to be identified, and may be an approach for obtaining more reliable estimates of problematically uncertain parameters.

Consistent with several recent studies using reactive tracer systems and TSM, we broadly found improved parameter constraints for some, but not all parameters associated with inclusion of reactive tracers (e.g., Lemke et al., 2013; Yakerivich et al., 2017) or additional experimental observations (e.g., Briggs et al., 2009; Neilson et al., 2010). Transient storage parameter uncertainties were minimized when a more complex model was used, most likely because this leads to greater degrees of freedom for fitting observations. For researchers who wish to separate the relative influences of transient storage between MITS and MATS, a 2SZ model simulating both conservative and "smart" tracer BTCs was capable of narrowing nearly all parameter estimates. We did find variations in parameter sensitivity and uncertainty across reaches. This is not surprising, given that the relative importance of different processes varies at the reach scale, and will determine parameter sensitivity and uncertainty within TSM applications. Though "smart" tracers are unsurprisingly superior to conservative tracers when it comes to partitioning MITS and MATS, little improvement in parameter uncertainty was gained for 1SZ model formulations by using a "smart" tracer.

5.3 Is information obtained from conservative and "smart" tracers complimentary or redundant?

For 1SZ models of conservative and "smart" tracers, a similar number of sensitive parameters were identified, illustrating that both tracer types contain valuable and potentially complimentary information. Furthermore, parameter estimates obtained with respect to all tracers were similar, but differed in some cases. On one hand, some tracers are likely to be more sensitive to main channel (e.g., Ura) versus storage zone (e.g., Rru) parameters and corresponding processes. This is a likely explanation for the difference in the empirical PDFs obtained for *A* and *D* fitting to Ura and Rru. As we would not expect fitting to Rru would contain information about main channel processes, this is unsurprising. A further explanation for the non-ideal estimation of *A* and *D* may be its sorption behavior in the subsurface (e.g., Lemke et al., 2014). Conversely, Raz and Ura may both provide similar information regarding *A* and *D*. Therefore, our work shows that even non-conservative tracers like Raz may still be useful for estimating parameters conceptualizing main channel processes.

In contrast, we also found differences in parameter estimates for transient storage exchange rate, α_s, when fitting to different tracers. This outcome was also mirrored within the 2SZ formulation for MATS exchange rates, and similar to findings from Lemke et al. (2013). These differences in estimates of transient storage parameters indicate that conservative and "smart" tracers may be sensitive to different timescales of transient storage. It is not clear why Raz and Rru would lead to different empirical PDFs and therefore different parameter estimates, but merits future work to explore why this may arise. As we only consider one objective function in this analysis, and we do not combine and propagate these parameter estimates back into the observation space, we can only speculate on how these findings may lead to improved calibration strategies. We do note that our findings challenge a common approach where some model parameters are constrained first using a conservative tracer, then fixed and others constrained in a second step using a reactive tracer (e.g., Keefe et al., 2004, Claessens et al., 2010, Yakirevich et al., 2017). Lemke et al. (2013) also found differences in optimized parameters for transport when a conservative tracer was fitted alone or jointly with Raz. Thus, our results demonstrate that improved interpretation of BTCs may be aided by fitting conservative and nonconservative tracers separately and comparing parameter estimates, instead of using conservative tracers to constrain parameters associated with nonconservative behavior.

Within our exercise, the tracer that provided the least redundant information was Rru, which contained unique information regarding MITS processes ($\propto_{MITS}$, A_{MITS}). While we anticipated differences between empirical PDFs fit to conservative versus non-conservative tracers, differences were especially pronounced between empirical PDFs for Raz versus Rru. This difference suggests that "smart" tracers may be more useful than conservative tracers for separating the hydrological and biogeochemical impacts of transient storage.

While our study suggests that Raz and Ura provide in part redundant information, we caution that this may not be the case for all systems. Making such a claim of redundancy based on a modeling exercise considering two stream reaches is unrealistic; more studies are needed to resolve questions of redundancy between tracers and parameter information content. Future work, especially experimental observations of transient storage processes (e.g., Knapp et al., 2017), is needed to clarify and investigate timescales of MATS and MITS, and whether these tracers are truly redundant when it comes to estimating parameter values. This ultimately relies on improved reconciliation by the TSM community of what is captured by a tracer versus what is represented within a given TSM formulation. At this stage, we do not have enough information to assess whether tracer observations may provide complementary or redundant information, as such an assessment should be based on numerous paired conservative and non-conservative tracer observations coupled with TSM.

5.4 How does model complexity impact parameter estimates and uncertainty?

Regardless of model complexity, the goals of tracer experiments are often to obtain reliable estimates with low uncertainty for parameters describing the influence of transient storage. Our results demonstrate that achieving this objective will ultimately be affected by the choice of tracer(s) (e.g., Abbot et al., 2016) and the choice of model framework, including the level of process representation. Increasing model complexity through the addition of model parameters may allow more realistic representation of in- and near-stream processes, but also has important implications for parameter uncertainty. In our analysis, we found that parameters typically well-estimated by TSM, *A* and *D*, saw wider uncertainty bounds moving from a 1SZ to 2SZ formulation (Fig. 6). This is likely due to increased degrees of freedom and interactions with added parameters in the 2SZ formulation (Fig. 8). As with our analysis, other studies have found *A* and *D* to be the most sensitive parameters with narrow ranges of uncertainty across many TSM applications (Wagener et al., 2002; Kelleher et al., 2013; Ward et al., 2017). These studies have also found strong interactions between *A* and *D*, likely the cause of the bimodal behavior observed in Figure 8. Our work adds to this existing body of literature by demonstrating how uncertainty in these well-estimated parameters changes alongside model complexity. When considering these uncertainty bounds in the context of uncertainty for other parameters, differences in these uncertainty bounds were still relatively small, leading us to conclude that only minor inference was lost with increased model complexity. Regardless, this outcome is a good reminder that as parameters are added to a model framework, uncertainty for some parameter estimates is likely to grow, even with additional information in the form of added tracer observations.

Our study offers cautious optimism regarding use of 2SZ models to infer process-based understanding of solute transport. As we show, 2SZ models, while more complex than 1SZ counterparts, produced narrow estimates of transient storage parameters and showed promise for separating the effects of MITS and MATS. Though parameters were highly interactive within the 2SZ model formulation (Figs. 8 and 9), we encouragingly found that we could obtain consistent and precise estimates of transient storage zone parameters (e.g., α_{MITS}, α_{MATS}) that are traditionally dominated by interactions and therefore have proved difficult to estimate in past studies (Wagner and Harvey, 1997; Wagener et al., 2002; Kelleher et al., 2013; Ward et al., 2017). However, our results also demonstrate that with increased complexity comes increased uncertainty with respect to other model parameters. Studies utilizing 2SZ models, or any TSM for that matter, should ultimately evaluate the uncertainty associated with parameter estimates (echoing past recommendations; Wagener et al., 2002; Kelleher et al., 2013; Ward et al., 2017). This need for uncertainty evaluation is especially clear in our analysis, in that we demonstrate that while this uncertainty may be reduced for 2SZ as compared to 1SZ models for some scenarios and parameters, uncertainty can still increase for other scenarios and parameters.

6 Conclusions

While researchers may wish to estimate size and exchange rates associated with transient storage in streams, and further to separate the effects of different transient storage zones, these goals rely on parameter estimation within a TSM framework. Within this context, we explored the tradeoffs between model complexity and utility of novel observations to estimate the effects of transient storage within stream reaches. Our results were consistent across two stream reaches with distinct morphologies; they suggest that model complexity and the necessity for new tracer observations are highly connected. For a 1D TSM, we found that parameter estimates were well-constrained by conservative tracer BTCs, but that fitting TSM simulations to a nonconservative tracer (Raz) yielded minimal additional gains in parameter inference. Thus, if using only a conservative tracer, a simpler model may yield more informative parameter estimates. In contrast, estimating parameters within a more complex 2SZ formulation from both conservative and "smart" tracer BTC error metrics produced complimentary insights, suggesting that (if the goal of a given study is to characterize both MITS and MATS) conservative and "smart" tracers should be used in tandem. Our findings suggest cautious optimism that nearly all parameters in 2SZ TSM formulations may be capably estimated by jointly fitting simulations to both conservative and "smart" tracer observations. Though our study represents a first step towards this goal, future work is needed to translate evaluations of parameter sensitivity and uncertainty into robust approaches to fitting multiple BTCs.

Though we show "smart" tracers have value for improving TSM approaches, we must ultimately reconcile how different process representations within TSM and tracer observations can be used to better quantify and understand specific stream transport processes. This is highlighted by the fact that experiments conducted with "smart" tracers, compared to conservative single tracer studies, require additional instrumentation, consumable costs, field time, and expertise. It remains to be seen whether "smart" tracers provide enough extra information to warrant their use within TSM, given our study solely demonstrates this outcome for two reaches with data collected at a single flow state. This detailed model assessment of multiple tracer types from two morphologically distinct stream reaches gives future stream investigators some insights, but, more importantly, quantitatively demonstrates that there are difficult tradeoffs each researcher will face (e.g., tradeoffs between tracer observations and model process representation efforts) when conducting stream tracer experiments. Furthermore, if unique information from tracers does not improve our current modeling tools, this may also suggest we need to interrogate and refine our perceptual models of these processes with the goal of improving numerical modeling tools.

The caution we offer, and are even prone to in this work, is that so many TSM analyses are treated as case studies, and there are few TSM synthesis efforts that have examined model frameworks, approaches, and outcomes across multiple sites, flow states, and physical representations of transient storage, let alone streams with different types of MATS and MITS. We note that our conclusions are specific to stream setting and flow state, and that there are likely other settings where these findings may differ. Continued discussion and evaluation of TSM formulations applied to conservative and nonconservative BTCs is therefore needed to refine the inference we can gain from tracer experiments across different environments, and to deliver a set of defensible recommendations regarding what can be achieved via TSM to the community of ecologists, hydrologists, and biogeochemists that apply these models.

Overall, our results validate that novel techniques for hydrologic data collection can help constrain parameter estimates within more complex and potentially more physically realistic models. This progress moves us toward improved process inference within hydrologic modeling of streams. More broadly, the approach we have taken of using gradients of both model complexity and observations is one that could be adapted and utilized for other hydrological model-based investigations. By continuing to interrogate the relationships between observations and model outcomes, we ultimately have great potential to improve our understanding of reactivity and transport within streams, especially when and where disconnects between modeled processes and observed processes occur.

Acknowledgments

Funding for this research was provided by the Leverhulme Trust (Where rivers, groundwater and disciplines meet: a hyporheic research network) and the UK Natural Environment Research Council (Large woody debris – A river restoration panacea for streambed nitrate attenuation? NERC NE/L003872/1). Data collection would not have been possible without the Leverhulme Hyporheic Zone Network Team, as well as funding from participating institutions. We also extend our thanks to Chithurst Buddhist Monastery for permitting access to their woodland. Additional support for Ward's time and computational infrastructure was provided by the Lilly Endowment, Inc., through its support for the Indiana University (IU) Pervasive Technology Institute, in part by the Indiana METACyt Initiative, and by University of Birmingham's Institute for Advanced Studies. Ward's time in development and implementation of Monte Carlo software and time-series analysis was supported by National Science Foundation (NSF) Grant Nos. EAR 1331906, EAR 1505309, and EAR 1652293. Several authors were also supported by the European Commission supported HiFreq: Smart high-frequency environmental sensor networks for quantifying nonlinear hydrological process dynamics across spatial scales (project ID 734317)

Field experiments were primarily led by Blaen, Kurz and Knapp, with input and assistance from most co-authors. Kelleher and Ward primarily conceived of the analyses presented here and led the modeling and data analysis efforts. Kelleher and Ward led the writing of this manuscript, with input from all co-authors. Views expressed in this manuscript do not necessarily reflect those of any funding agencies or institutions. Experimental data are accessible via Hydroshare (Blaen et al., 2018; https://www.hydroshare.org/resource/7c63809428cf495faceb67c0278ed036/). Monte Carlo and uncertainty analysis software are available as the Sensitivity and Uncertainty Analysis for Everyone (SAFE) package (http://www.bris.ac.uk/cabot/resources/safe-toolbox/).

The authors especially wish to thank three anonymous reviewers who provided productive feedback to revising this manuscript.

References

Abbott, B.W., Baranov, V., Mendoza-Lera, C., Nikolakopoulou, M., Harjung, A., Kolbe, T., Balasubramanian, M.N., Vaessen, T.N., Ciocca, F., Campeau, A. & Wallin, M.B. (2016). Using multi-tracer inference to move beyond single-catchment ecohydrology. *Earth-Science Reviews*, *160*, 19-42. https://doi.org/ 10.1016/j.earscirev.2016.06.014.

Argerich, A., Haggerty, R., Martí, E., Sabater, F., & Zarnetske, J. (2011). Quantification of metabolically active transient storage (MATS) in two reaches with contrasting transient storage and ecosystem respiration. *Journal of Geophysical Research: Biogeosciences, 116,* G03034. https://doi.org/10.1029/2010JG001379.

Bencala, K.E. & Walters, R.A. (1983). Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model. *Water Resources Research, 19,* 718–724. https://doi.org/10.1029/WR019i003p00718

Best, M.J., Abramowitz, G., Johnson, H.R., Pitman, A.J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P.A., Dong, J. & Ek, M. (2015). The plumbing of land surface models: benchmarking model performance. *Journal of Hydrometeorology, 16*(3), 1425-1442. https://doi.org/ 10.1175/JHM-D-14-0158.1.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology, 320,* 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K. (1993). Prophecy, reality and uncertainty in distributed hydrological modelling. *Advances in Water Resources, 16,* 41–51. https://doi.org/10.1016/0309-1708(93)90028-E

Blaen P.J., Khamis K., Lloyd C.E.M., Bradley C., Hannah D. & Krause S. (2016). Real-time monitoring of nutrients and dissolved organic matter in rivers: Capturing event dynamics, technological opportunities and future directions, *Science of the Total Environment*, *569-570*, 647-660, https://doi.org/ 10.1016/j.scitotenv.2016.06.116.

Blaen, P. J., Brekenfeld, N., Comer-Warner, S., & Krause, S. (2017). Multitracer Field Fluorometry: Accounting for Temperature and Turbidity Variability During Stream Tracer Tests. *Water Resources Research*, *53*, 9118–9126. https://doi.org/10.1002/2017WR020815.

Blaen, P.J., Kurz, M.J., Drummond, J.D., Knapp, J.L.A., Mendoza-Lera, C., Schmadel, N.M., Klaar, M.J., Jager, A., Folegot, S., Lee-Cullin, J., Ward, A.S., Zarnetske, J.P., Datry, T., Milner, A.M., Lewandowski, J., Hannah, D.M., & Krause, S. (2018). Woody debris is related to reach-scale hotspots of lowland stream ecosystem respiration under baseflow conditions, *Ecohydrology,* Accepted 13 February 2018. https://doi.org/ 10.1002/eco.1952.

Blaen, P., M. Kurz, J. Knapp, S. Krause, A. Ward, C. Kelleher (2018). 2015 Hammer Solute Tracer Injections, HydroShare, http://www.hydroshare.org/resource/7c63809428cf495faceb67c0278ed036.

Boano, F., Harvey, J.W., Marion, A., Packman, A.I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics, 52,* 2012RG000417. https://doi.org/10.1002/2012RG000417

Brenner, C., Thiem, C.E., Wizemann, H.-D., Bernhardt, M., & Schulz, K. (2017). Estimating spatially distributed turbulent heat fluxes from high-resolution thermal imagery acquired with a UAV system. *International Journal of Remote Sensing, 38,* 3003–3026. https://doi.org/10.1080/01431161.2017.1280202

Briggs, M.A., 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,* W00D27. https://doi.org/10.1029/2008WR006959

Butts, M.B., Payne, J.T., Kristensen, M., & Madsen, H. (2004). An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation. *Journal of Hydrology, 298,* 242–266. https://doi.org/10.1016/j.jhydrol.2004.03.042

Caruso, A., Ridolfi, L., & Boano, F. (2016). Impact of watershed topography on hyporheic exchange. *Advances in Water Resources, 94,* 400–411. https://doi.org/10.1016/j.advwatres.2016.06.005

Choi, J., Harvey, J.W., & Conklin, M.H. (2000). Characterizing multiple timescales of stream and storage zone interaction that affect solute fate and transport in streams. *Water Resources Research, 36,* 1511–1518. https://doi.org/10.1029/2000WR900051

Clark, M.P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D.E., Woods, R.A., Freer, J.E., Gutmann, E.D., Wood, A.W., Brekke, L.D. & Arnold, J.R. (2015a). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research, 51*(4), 2498-2514. https://doi.org/ 10.1002/2015WR017198.

Clark, M.P., Nijssen, B., Lundquist, J.D., Kavetski, D., Rupp, D.E., Woods, R.A., Freer, J.E., Gutmann, E.D., Wood, A.W., Gochis, D.J. & Rasmussen, R.M. (2015b). A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resources Research, 51*(4), 2515-2542. https://doi.org/ 10.1002/2015WR017198.

Clark, M.P., Bierkens, M.F., Samaniego, L., Woods, R.A., Uijlenhoet, R., Bennett, K.E., Pauwels, V.R., Cai, X., Wood, A.W. & Peters-Lidard, C.D. (2017). The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. *Hydrology and Earth Systems Sciences, 21*(7), 3427-3440, https://doi.org/ 10.5194/hess-21-3427-2017.

Claessens, L., Tague, C.L., Groffman, P.M. and Melack, J.M. (2010). Longitudinal assessment of the effect of concentration on stream N uptake rates in an urbanizing watershed. *Biogeochemistry, 98*(1-3), 63-74. https://doi.org/10.1002/2012RG000417.

Damköhler, G. (1936). Einflüsse der Strömung, Diffusion und des Wärmeüberganges auf die Leistung von Reaktionsöfen.: I. Allgemeine Gesichtspunkte für die Übertragung eines chemischen Prozesses aus dem Kleinen ins Große. *Zeitschrift für Elektrochemie und angewandte physikalische Chemie, 42,* 846–862. https://doi.org/10.1002/bbpc.19360421203.

Freer, J. Beven, K., & Ambroise, B. (1996). Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach, *Water Resources Research, 32,* 2161-2173, https://doi.org/ 10.1029/95WR03723.

González-Pinzón, R., Haggerty, R., & Myrold, D.D. (2012). Measuring aerobic respiration in stream ecosystems using the resazurin-resorufin system. *Journal of Geophysical Research: Biogeosciences, 117,* G00N06. https://doi.org/10.1029/2012JG001965.

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. Doi: 10.1016/S0309-1708(03)00105-2.

Gooseff, M. N., R. O. Hall Jr., and J. L. Tank (2007), Relating transient storage to channel complexity in streams of varying land use in Jackson Hole, Wyoming, *Water Resour. Res.*, 43, W01417, doi: 10.1029/2005WR004626.

Gooseff, M.N. (2010). Defining Hyporheic Zones – Advancing Our Conceptual and Operational Definitions of Where Stream Water and Groundwater Meet. *Geography Compass, 4,* 945–955. https://doi.org/10.1111/j.1749-8198.2010.00364.x

Haggerty, R., Argerich, A., & Martí, E. (2008). Development of a "smart" tracer for the assessment of microbiological activity and sediment-water interaction in natural waters: The resazurin-resorufin system. *Water Resources Research, 44,* W00D01. https://doi.org/10.1029/2007WR006670

Haggerty, R., Martí, E., Argerich, A., von Schiller, D., & Grimm, N.B. (2009). Resazurin as a "smart" tracer for quantifying metabolically active transient storage in stream ecosystems. *Journal of Geophysical Research Letters,114,* G03014. https://doi.org/10.1029/2008JG000942

Harvey, J.W., Wagner, B.J., & Bencala, K.E. (1996). Evaluating the Reliability of the Stream Tracer Approach to Characterize Stream-Subsurface Water Exchange. *Water Resources Research, 32,* 2441–2451. https://doi.org/10.1029/96WR01268

Hrachowitz, M. & Clark, M.P. (2017). HESS Opinions: The complementary merits of competing modelling philosophies in hydrology. *Hydrology and Earth Systems Sciences, 21,* 3953–3973. https://doi.org/10.5194/hess-21-3953-2017

Karakashev, D., Galabova, D., & Simeonov, I. (2003). A simple and rapid test for differentiation of aerobic from anaerobic bacteria. *World Journal of Microbiology and Biotechnology, 19,* 233–238. https://doi.org/10.1023/A:1023674315047

Keefe, S. H., Barber, L. B., Runkel, R. L., Ryan, J. N., McKnight, D. M., & Wass, R. D. (2004). Conservative and reactive solute transport in constructed wetlands. *Water Resources Research, 40*, W01201, https://doi.org/10.1029/2003WR002130.

Kelleher, C., Wagener, T., McGlynn, B., Ward, A.S., Gooseff, M.N., & Payn, R.A. (2013). Identifiability of transient storage model parameters along a mountain stream. *Water Resources Research, 49,* 5290–5306. https://doi.org/10.1002/wrcr.20413

Kerr, P.C., Gooseff, M.N., & Bolster, D. (2013). The significance of model structure in one-dimensional stream solute transport models with multiple transient storage zones – competing vs. nested arrangements. *Journal of Hydrology, 497,* 133–144. https://doi.org/10.1016/j.jhydrol.2013.05.013

Knapp, J.L.A., González-Pinzón, R., Drummond, J.D., Larsen, L.G., Cirpka, O.A., & Harvey, J.W. (2017). Tracer-based characterization of hyporheic exchange and benthic biolayers in streams. *Water Resources Research, 53,* 1575–1594. https://doi.org/10.1002/2016WR019393.

Knapp, J.L.A., & Cirpka, O.A. (2017). Determination of hyporheic travel time distributions and other parameters from concurrent conservative and reactive tracer tests by local-in-global optimization, *Water Resources Research, 53,* 4984-5001, https://doi.org/10.1002/2017WR020734.

Knapp, J. L. A., González-Pinzón, R., & Haggerty, R. (2018). The resazurin-resorufin system: Insights from a decade of "smart" tracer development for hydrologic applications. *Water Resources Research*, 54, 6877–6889. https://doi.org/10.1029/2018WR023103.

Khamis K., Bradley C., Stevens R., & Hannah D.M. (2016). Continuous field estimation of dissolved organic carbon concentration and biochemical oxygen demand using dual-wavelength fluorescence, *Hydrological Processes – Scientific Briefings*, *31,* 540-555. https:/10.1002/hyp.11040.

Krause S., Hannah D.M., Fleckenstein J.H., Heppell C.M., Pickup R., Pinay G., Robertson A.L. & Wood P.J. (2011). Interdisciplinary perspectives on processes in the hyporheic zone, *Ecohydrology*, *4,* 481–499. https://10.1002/eco.176

Lees, M.J., Camacho, L.A., & Chapra, S. (2000). On the relationship of transient storage and aggregated dead zone models of longitudinal solute transport in streams. *Water Resources Research, 36,* 213–224. https://doi.org/10.1029/1999WR900265

Lemke, D., Liao, Z., Wöhling, T., Osenbrück, K., & Cirpka, O.A. (2013). Concurrent conservative and reactive tracer tests in a stream undergoing hyporheic exchange. *Water Resources Research, 49,* 3024–3037. https://doi.org/10.1002/wrcr.20277

Lemke, D., González-Pinzón, R., Liao, Z., Wöhling, T., Osenbrück, K., Haggerty, R., and Cirpka, O. A. (2014). Sorption and transformation of the reactive tracers resazurin and resorufin in natural river sediments, *Hydrol. Earth Syst. Sci.*, 18, 3151-3163, https://doi.org/10.5194/hess-18-3151-2014.

Li, H., Xu, C.-Y., & Beldring, S. (2015). How much can we gain with increasing model complexity with the same model concepts? *Journal of Hydrology, 527,* 858–871. https://doi.org/10.1016/j.jhydrol.2015.05.044

Liao, Z., & Cirpka, O.A. (2011). Shape-free inference of hyporheic traveltime distributions from synthetic conservative and “smart” tracer tests in streams. *Water Resources Research, 47,* W07510. https://doi.org/10.1029/2010WR009927

Marion, A., Zaramella, M., & Bottacin-Busolin, A. (2008). Solute transport in rivers with multiple storage zones: The STIR model. *Water Resources Research, 44,* W10406. https://doi.org/10.1029/2008WR007037

Nearing, G. S., & Gupta, H. V. (2015). The quantity and quality of information in hydrologic models. *Water Resources Research 51*, 524–538. https://doi.org/10.1002/2014WR015895.

Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V., & Xia, Y. L. (2016). Benchmarking NLDAS-2 Soil Moisture and Evapotranspiration to Separate Uncertainty Contributions. *Journal of Hydrometeorology, 17*(3), 745-759. https://doi.org/10.1175/jhm-d-15-0063.1.

Orghidan, T., (1959). Ein neuer Lebensraum des unterirdischen Wasser: der hyporheische Biotop. *Archiv für Hydrobilogie, 55,* 392–414.

Pianosi, F., SarRazin, F., & Wagener, T. (2015). A Matlab toolbox for global sensitivity analysis. *Environmental Modelling & Software*, *70*, 80-85, https://doi.org/ 10.1016/j.envsoft.2015.04.009.

Runkel, R. L. (1998). One-Dimensional Transport with Inflow and Storage (OTIS): A Solute Transport Model for Streams and Rivers (No. Water-Resources Investigations Report 98-4018). United States Geological Survey.

Runkel, R.L. (2015). On the use of rhodamine WT for the characterization of stream hydrodynamics and transient storage. *Water Resources Research, 51,* 6125–6142. https://doi.org/10.1002/2015WR017201

Schmadel, N.M., Ward, A.S., & Wondzell, S.M. (2017). Hydrologic controls on hyporheic exchange in a headwater mountain stream. *Water Resources Research 53,* 6260–6278. https://doi.org/10.1002/2017WR020576

Schoups, G., van de Giesen, N.C., & Savenije, H.H.G. (2008). Model complexity control for hydrologic prediction. *Water Resources Research, 44,* W00B03. https://doi.org/10.1029/2008WR006836.

Seibert, J., & J. J. McDonnell (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research, 38*(11), 1241, https://doi.org/10.1029/2001WR000978.

Shelley, F., Klaar, M., Krause, S., & Trimmer, M. (2017). Enhanced hyporheic exchange flow around woody debris does not increase nitrate reduction in a sandy streambed. *Biogeochemistry, 136,* 353-372, https://doi.org/ 10.1007/s10533-017-0401-2.

St Clair, J., Moon, S., Holbrook, W.S., Perron, J.T., Riebe, C.S., Martel, S.J., Carr, B., Harman, C., Singha, K., & Richter, D. deB (2015). Geophysical imaging reveals topographic stress control of bedrock weathering. *Science, 350,* 534–538. https://doi.org/10.1126/science.aab2210

Storey, R.G., Howard, K.W.F., & Williams, D.D. (2003). Factors controlling riffle-scale hyporheic exchange flows and their seasonal changes in a gaining stream: A three-dimensional groundwater flow model. *Water Resources Research,* 39, 1034. https://doi.org/10.1029/2002WR001367

Thackston, E.L., & Schnelle, K.B. (1970). Predicting Effects of Dead Zones on Stream Mixing. *Journal of the Sanitary Engineering Division, 96,* 319–331.

Triska, F.J., Kennedy, V.C., Avanzino, R.J., Zellweger, G.W., & Bencala, K.E. (1989). Retention and Transport of Nutrients in a Third-Order Stream in Northwestern California: Hyporheic Processes. *Ecology, 70,* 1893–1905. https://doi.org/10.2307/1938120

Valett, H.M., Morrice, J.A., Dahm, C.N., & Campana, M.E. (1996). Parent lithology, surface–groundwater exchange, and nitrate retention in headwater streams. *Limnology and Oceanography, 41,* 333–345. https://doi.org/10.4319/lo.1996.41.2.0333

Vivoni, E.R., Rango, A., Anderson, C.A., Pierini, N.A., Schreiner-McGraw, A.P., Saripalli, S., & Laliberte, A.S. (2014). Ecohydrology with unmanned aerial vehicles. *Ecosphere 5,* 1–14. https://doi.org/10.1890/ES14-00217.1

Wagener, T., Camacho, L.A., & Wheater, H.S. (2002). Dynamic identifiability analysis of the transient storage model for solute transport in rivers. *Journal of Hydroinformatics 4,* 199–211.

Wagener, T., & Kollat, J. (2007). Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. *Environmental Modelling & Software, 22,* 1021–1033. https://doi.org/10.1016/j.envsoft.2006.06.017.

Wagner, B.J., & Harvey, J.W. (1997). Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies. *Water Resources Research 33,* 1731–1741. https://doi.org/10.1029/97WR01067.

Ward, A.S. (2016). The evolution and state of interdisciplinary hyporheic research. *WIREs Water, 3,* 83–103. https://doi.org/10.1002/wat2.1120.

Ward, A.S., Gooseff, M.N., & Singha, K. (2010). Imaging hyporheic zone solute transport using electrical resistivity. *Hydrologic Processes, 24,* 948–953. https://doi.org/10.1002/hyp.7672.

Ward, A.S., Cwiertny, D.M., Kolodziej, E.P., & Brehm, C.C. (2015). Coupled reversion and stream-hyporheic exchange processes increase environmental persistence of trenbolone metabolites. *Nature Communications, 6,* 7067. https://doi.org/10.1038/ncomms8067.

Ward, A.S., Kelleher, C.A., Mason, S.J.K., Wagener, T., McIntyre, N., McGlynn, B., Runkel, R.L., & Payn, R.A. (2017). A software tool to assess uncertainty in transient-storage model parameters using Monte Carlo simulations. *Freshwater Science, 36*, 195–217. https://doi.org/10.1086/690444.

Wlostowski, A.N., Gooseff, M.N., & Wagener, T. (2013). Influence of constant rate versus slug injection experiment type on parameter identifiability in a 1-D transient storage model for stream solute transport. *Water Resources Research*, *49*, doi: 10.1002/wrcr.20103.

Wörman, A., & Wachniew, P. (2007). Reach scale and evaluation methods as limitations for transient storage properties in streams and rivers. *Water Resources Research, 43*, W10405. https://doi.org/10.1029/2006WR005808.

Yakirevich, A., Shelton, D., Hill, R., Kiefer, L., Stocker, M., Blaustein, R., Kuznetsov, M., McCarty, G., & Pachepsky, Y. (2017). Transport of Conservative and “smart” Tracers in a First-Order Creek: Role of Transient Storage Type. *Water, 9*, 485. https://doi.org/10.3390/w9070485.

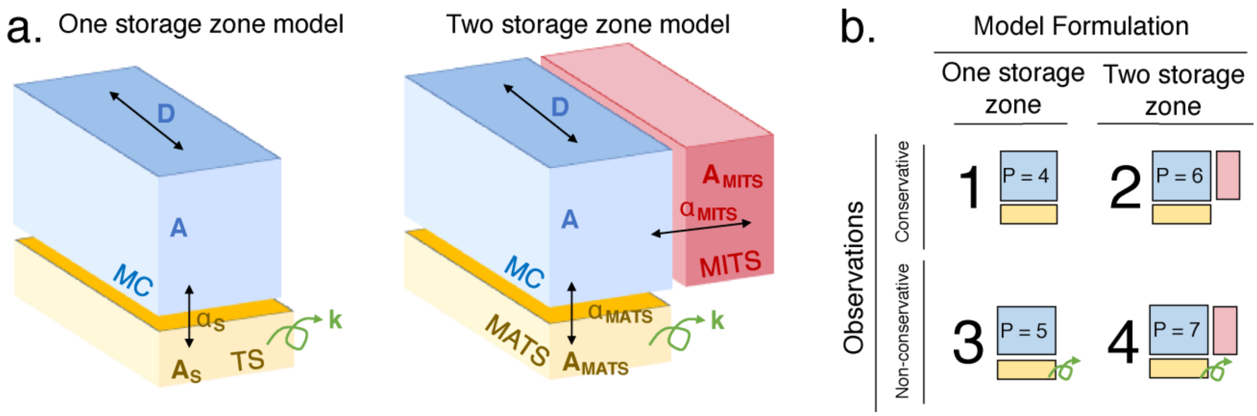


Figure 1. Model framework displaying (a) model parameters and the hypothetical compartments within the stream reach they are associated with (MC = main channel) for both one storage zone (1SZ) and two storage zone (2SZ) models and (b) the multiple model formulations utilized within this study (and corresponding numbers of parameters). In particular, we compare across the number of transient storage (TS) zones (one vs. two), as well as parameter estimates with respect to both conservative (Ura) or nonconservative (Raz, Rru) tracer dynamics. By combining these formulations and observations, we tested four different models ranging from four to seven model parameters (P). Additional figure abbreviations include: metabolically inactive storage (MITS), metabolically active storage (MATS), and parameters main-channel area (A), dispersive coefficient (D), transient storage zone exchange (α_S), transient storage zone size (A_s), conversion of Raz to Rru (k), MATS cross-sectional area (A_{MATS}), MATS exchange rate (α_{MATS}), MITS cross-sectional area (F_{MITS}), and MITS exchange rate (α_{MITS}).

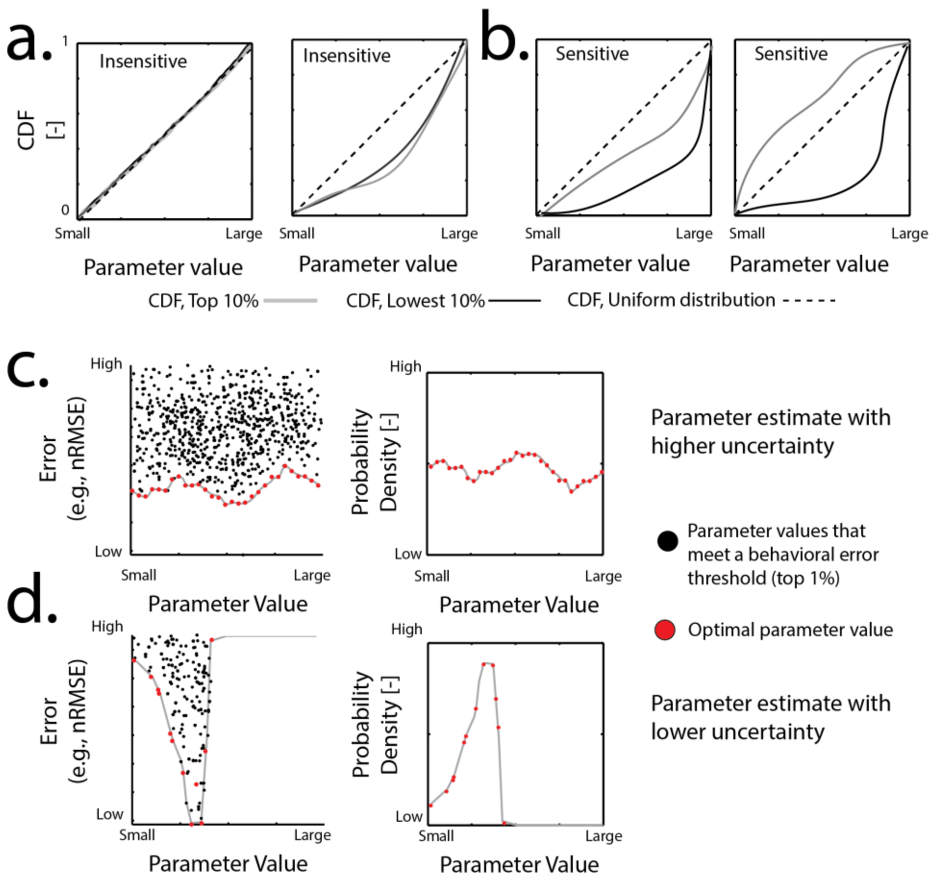


Figure 2. Interpretation and approaches for (a, b) sensitivity analysis and (c, d) uncertainty analysis. Regional sensitivity analysis is used to assess parameter sensitivity for parameter values with the best (top 10%) and worst (lowest 10%) errors, compared to a uniform distribution (1:1) line. Conceptual examples of cumulative distribution functions can be used to interpret whether parameters are insensitive (Fig. 2a), due to either falling along the 1:1 line or CDFs indistinguishable between parameter values corresponding to the best and worst error values, or sensitive (Fig. 2b), where the CDF of parameter values corresponding to the best errors are clearly distinguishable from the 1:1 line and the CDF corresponding to the worst errors. To compliment RSA, uncertainty is assessed by translating dotty plots to empirical probability density functions (PDFs) of optimal model errors across feasible parameter ranges. Optimal parameters (red) represent those with the lowest error for a narrow moving window along the parameter space. We display two hypothetical examples for a parameter with high uncertainty (Fig. 2c) and low uncertainty (Fig. 2d). Peaky distributions, found for a parameter with low uncertainty, indicate that certain regions of the parameter space yield better performance, while a flat distribution, corresponding to the parameter with greater uncertainty (Fig. 2d), suggests that all parameter values yield similar model performance.

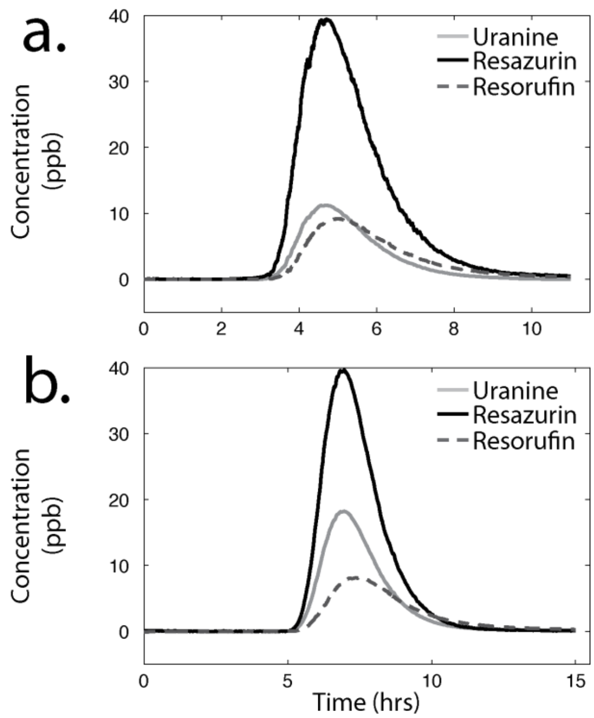


Figure 3. Observed breakthrough curves (concentration through time) for a conservative tracer (Ura) and nonconservative tracer Raz and biproduct Rru for (a) sand and (b) gravel reaches.

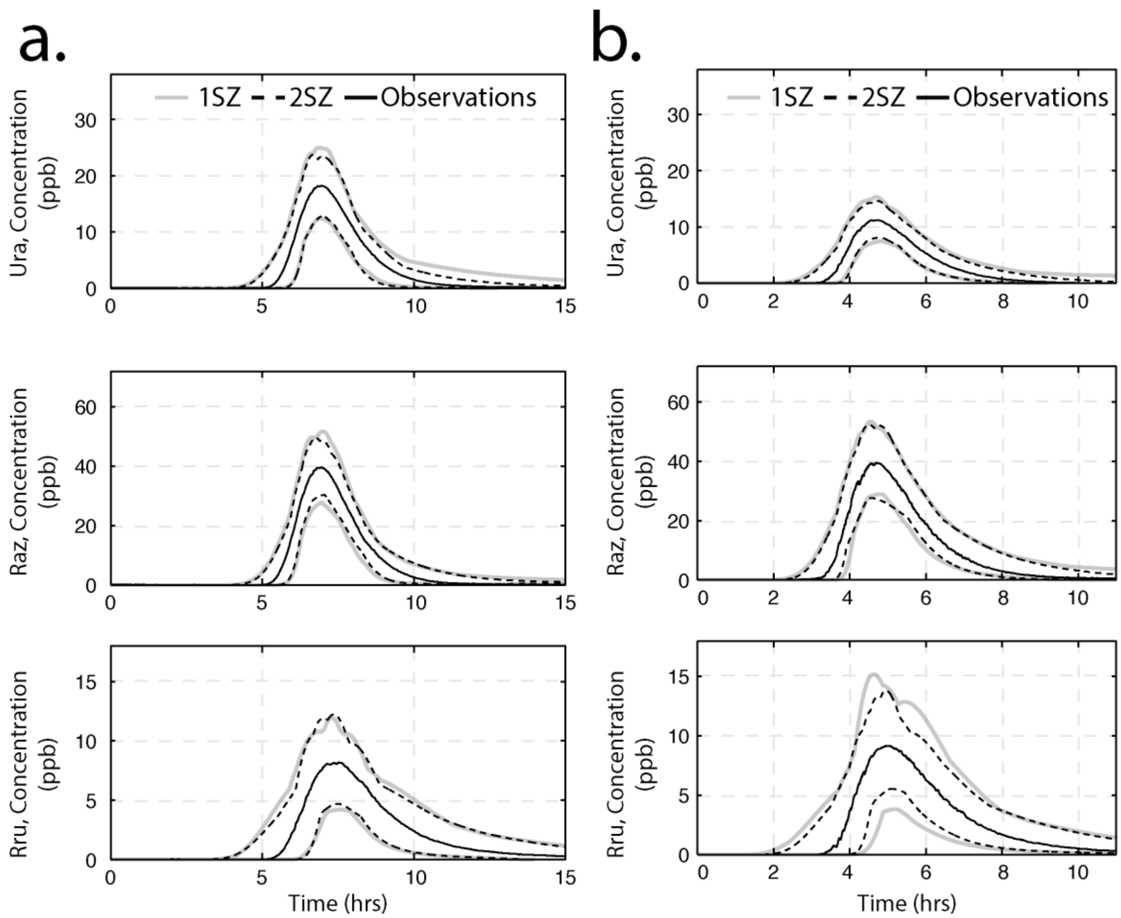


Figure 4. Upper and lower bounds for the 270 simulations corresponding to the minimum 1% of RMSE values for the (a) gravel and (b) sand reaches. These bounds represent the envelope encompassing the range of all simulations corresponding to the top 1% of values by *nRMSE*, per tracer and per model. Bounds are shown relative to observations. All simulations are included as ensemble averages in Figure S2.

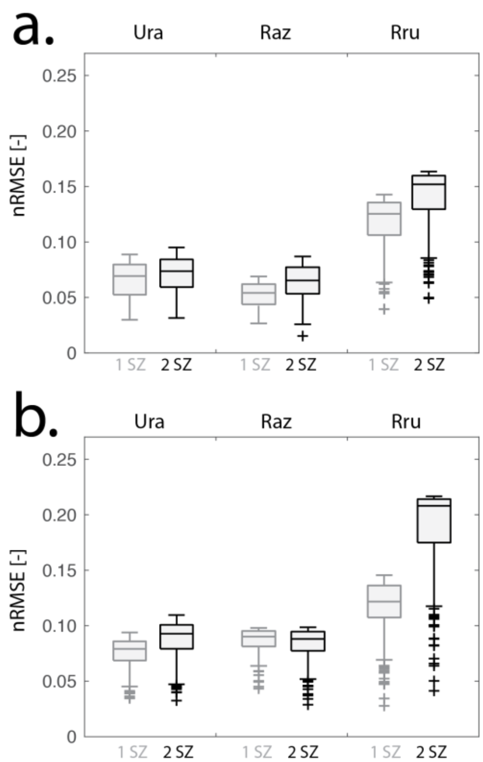


Figure 5. Distributions of model error shown for the top 1% of nRMSE values for all tracers and for a combined tracer metric for the (a) sand and (b) gravel reaches. Results are shown for the one (1SZ) and two storage zone models (2SZ).

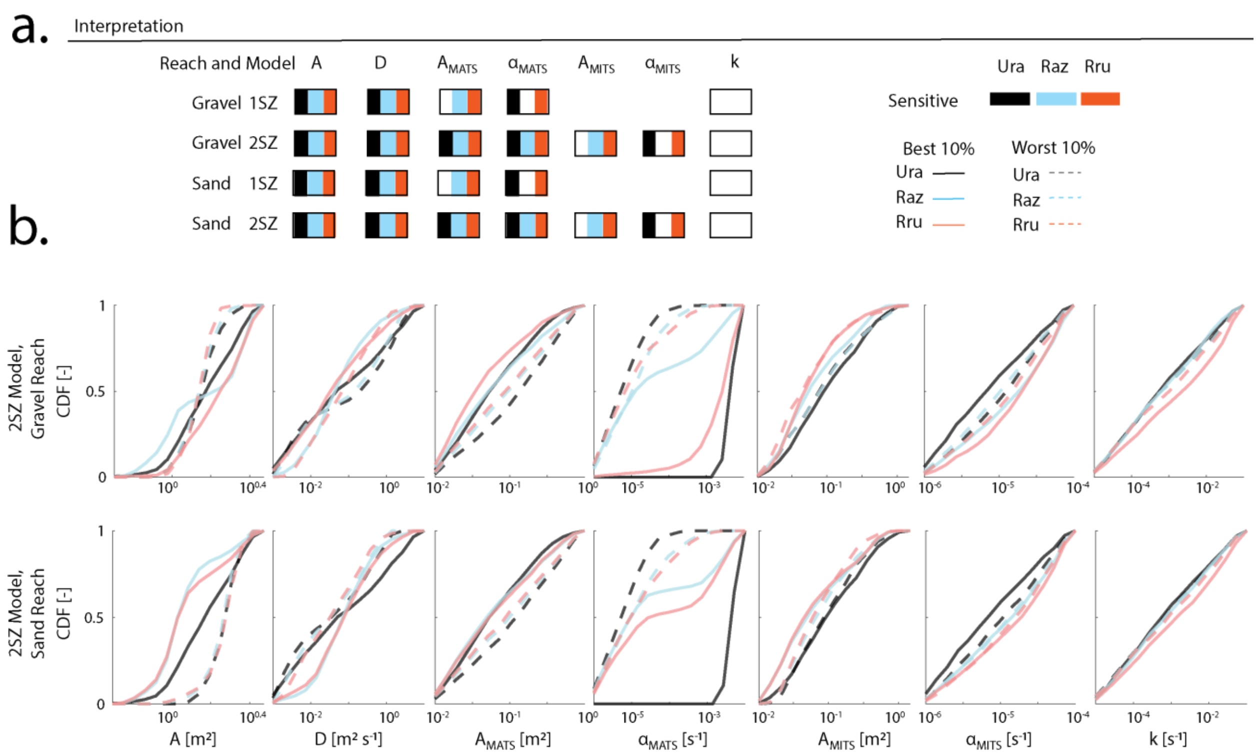


Figure 6. Analysis of parameter sensitivities including (a) interpretation of sensitivities across 1SZ and 2SZ models, reaches, and tracers, and (b) select RSA plots for 2SZ gravel and sand reaches for D, A_{MATS}, $\propto_{MATS}$, A_{MITS}, and $\propto_{MITS}$. Interpretation of (a) is based on Fig. 2a, with sensitive parameters deviating from a uniform CDF and from the CDF corresponding to the "worst 10%" of error values. A color shown in (a) indicates interpretation based on (b) that a parameter is sensitive. RSA plots compare empirical CDFs corresponding to the top 10% and worst 10% of all model simulations per tracer error metric.

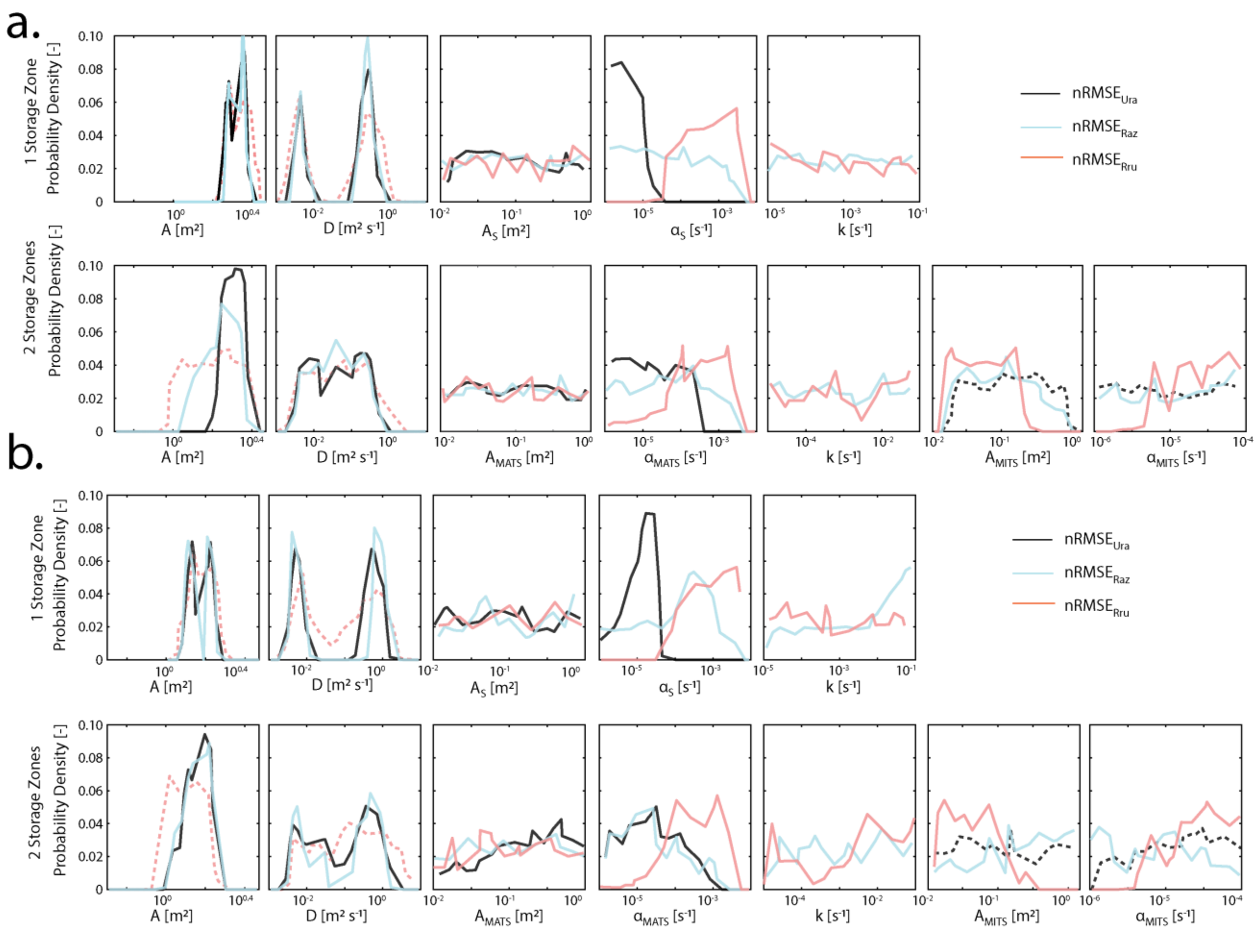


Figure 7. PDFs of the top 1% of RMSE values plotted across log-transformed parameter values for the 1SZ and 2SZ models of the (a) sand and (b) gravel reaches. Results were independently generated for each of three tracers (Fig. 2). Dotted lines indicate parameters that we do not expect to be physically related to or informed by a given tracer.

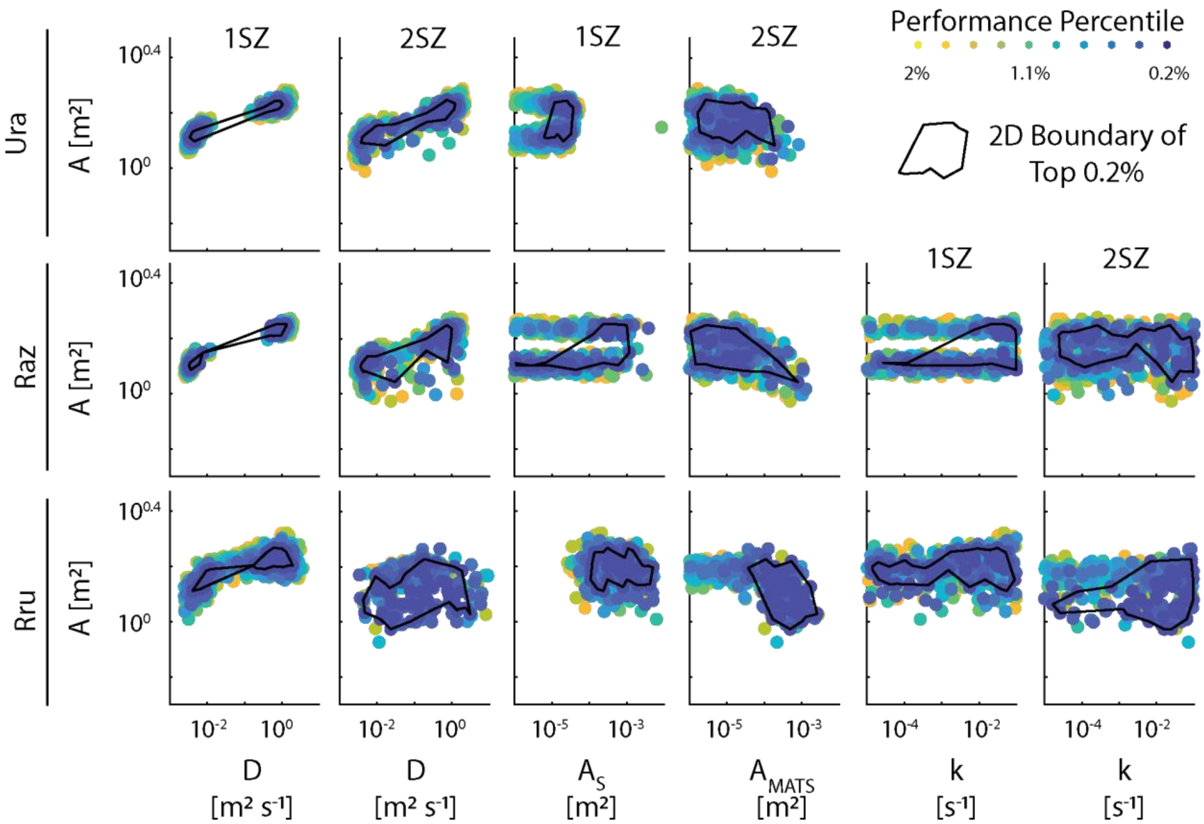


Figure 8. Joint distributions of 1SZ and 2SZ model parameters for the gravel reach. Black lines indicate the boundary of the top 0.2% of parameter sets (by *nRMSE* per tracer). Colors indicate different percentiles of performance corresponding to the top 2% of all parameter sets.

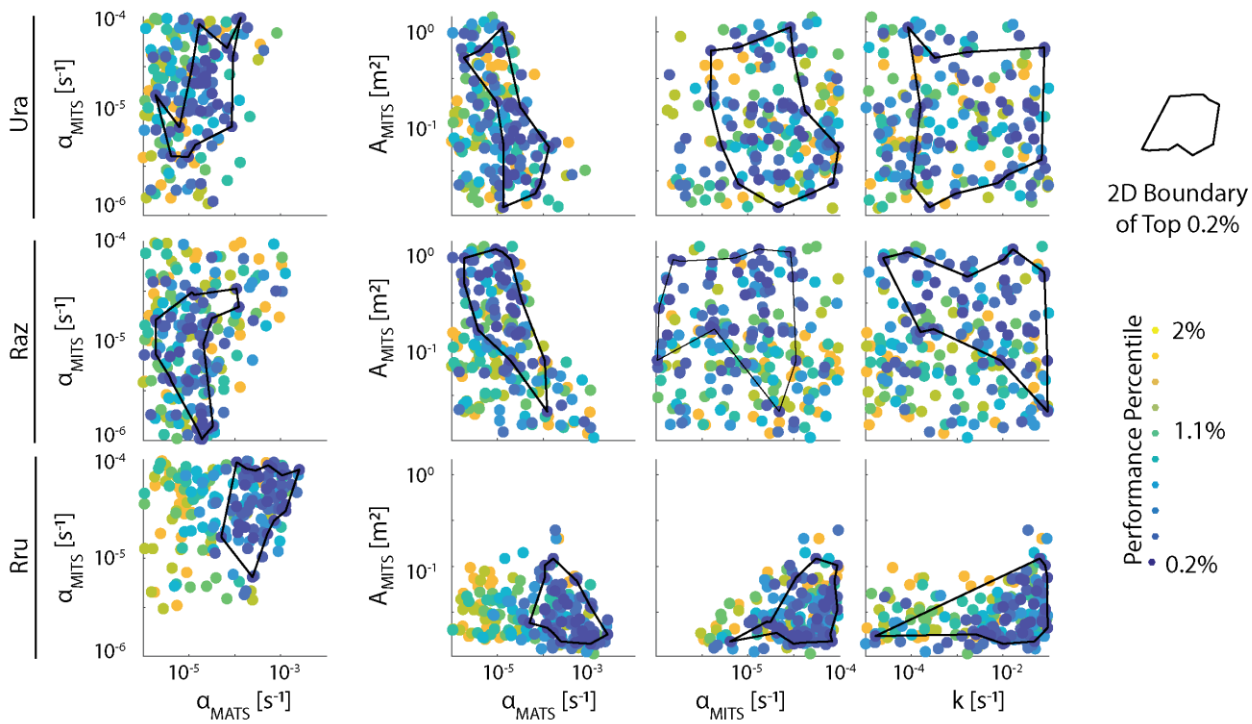
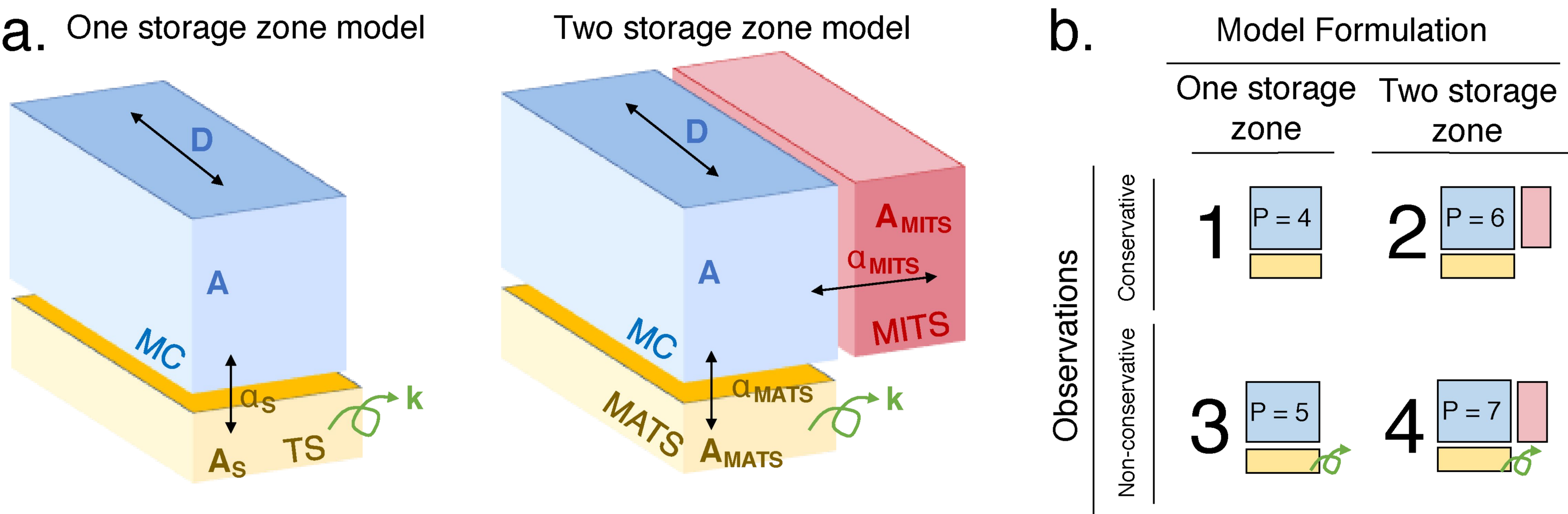


Figure 9. Joint distributions for the 2SZ gravel reach parameter sets. Black lines indicate the boundary of the top 0.2% of parameter sets (by nRMSE per tracer). Colors indicate different percentiles of performance corresponding to the top 2% of all parameter sets.

Table 1. Parameter names, abbreviations, and ranges for sensitivity and uncertainty analysis applied to variable TSM formulations (Figure 1).

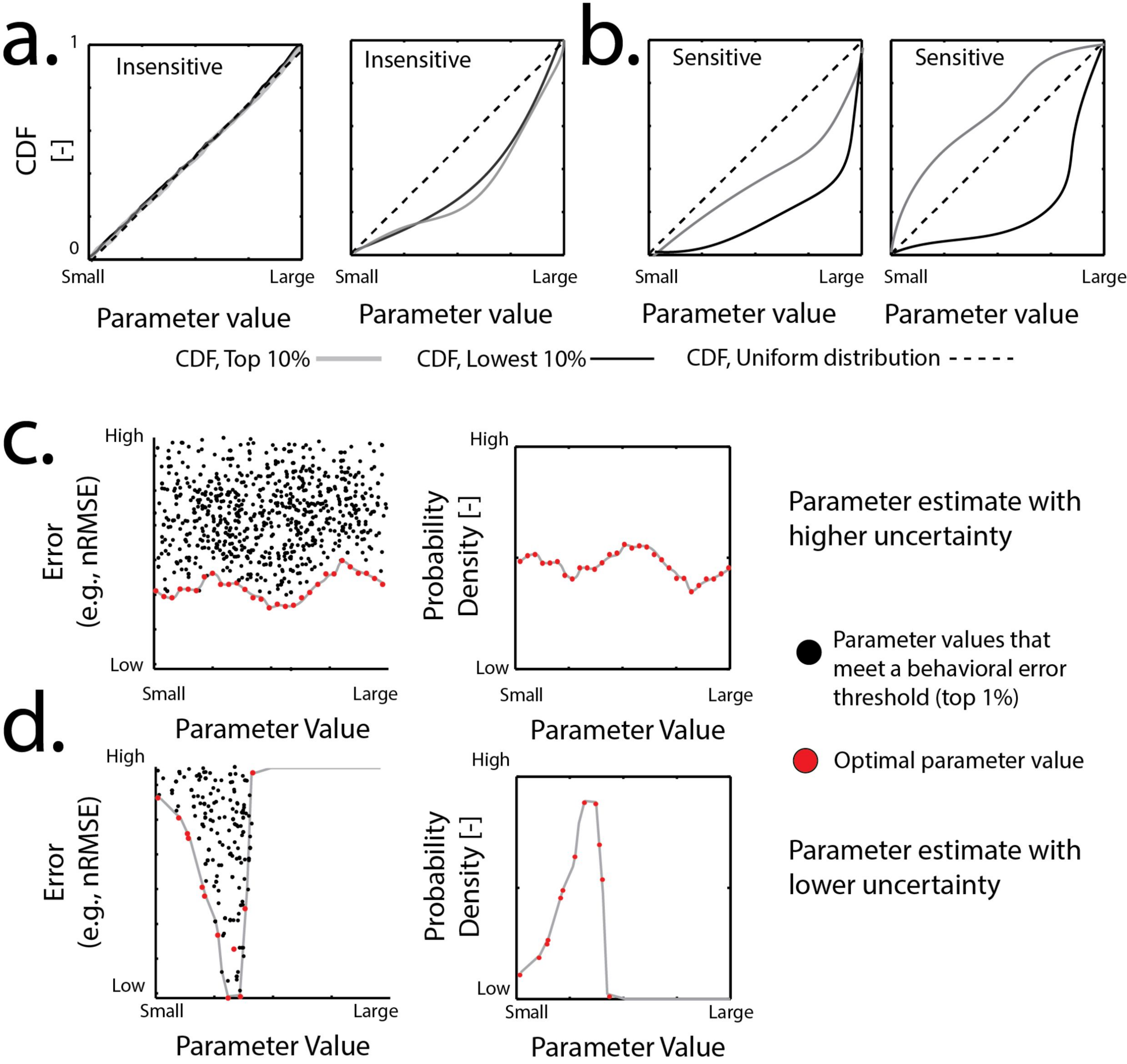
Abbrev.	Parameter	Model	Tracer	Units	Lower Bound	Upper Bound
D	Dispersion coefficient	1SZ, 2SZ	All	$m^2\ s^{-1}$	0.001	10
A	Advective channel cross-sectional area	1SZ	All	m^2	1	3
A_{TOT}	Total area	2SZ	All	m^2	1	3
A_S	Transient storage cross-sectional area	1SZ	All	m^2	0.01	1
α_S	Transient storage exchange rate	1SZ	All	s^{-1}	10^{-6}	10^{-2}
k	Conversion, Raz to Rru	2SZ	Raz, Rru	s^{-1}	10^{-5}	10^{-1}
A_{MATS}	MATS cross-sectional area	2SZ	All	m^2	0.01	1
α_{MATS}	MATS exchange rate	2SZ	All	s^{-1}	10^{-6}	10^{-2}
F_{MITS}	Fraction of stream area as MITS	2SZ	All	-	0.01	0.5
α_{MITS}	MITS exchange rate	2SZ	All	s^{-1}	10^{-5}	10^{-1}

Figure 1.



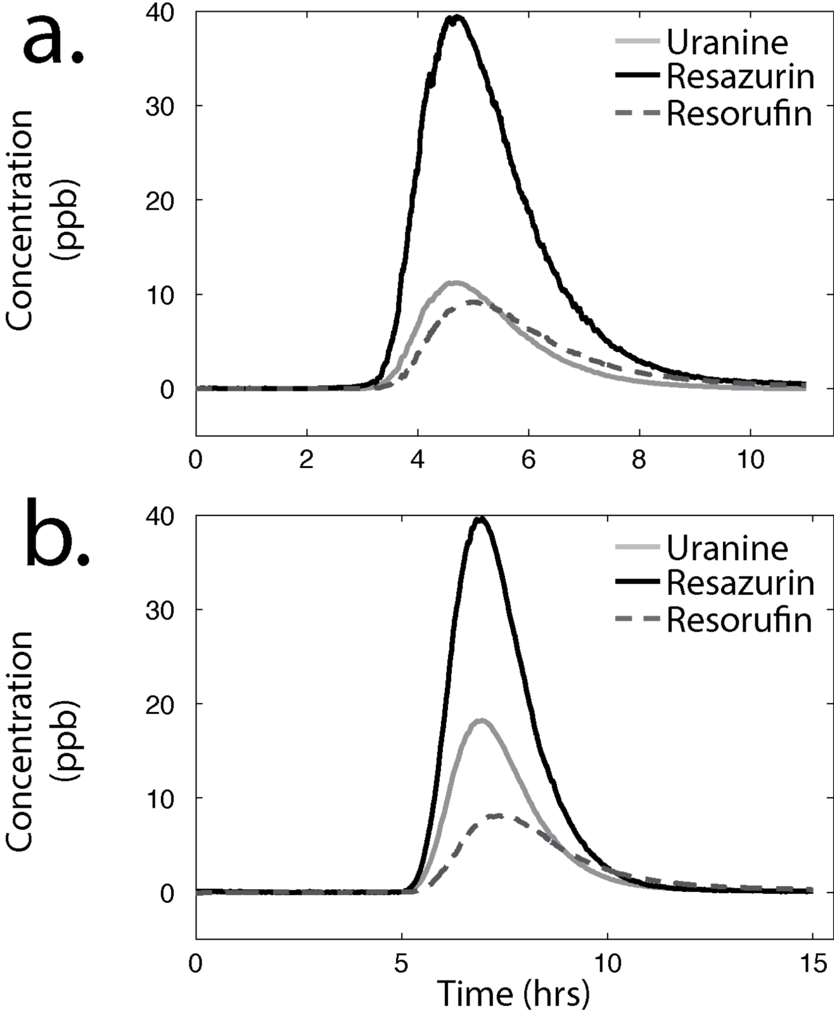
a.
One storage zone model
D
A
MC
α_S
k
A_S
TS
Two storage zone model
D
A_{MITS}
α_{MITS}
A
MITS
MC
α_{MATS}
k
MATS
A_{MATS}
b.
Model Formulation
One storage zone
Two storage zone
Observations
Conservative
Non-conservative
1
P = 4
2
P = 6
3
P = 5
4
P = 7

Figure 2.



a.
b.
1
0
CDF [-]
Insensitive
Insensitive
Sensitive
Sensitive
Small
Large
Parameter value
CDF, Top 10%
CDF, Lowest 10%
CDF, Uniform distribution
c.
d.
High
Low
Error (e.g., nRMSE)
Probability Density [-]
Parameter Value
Parameter estimate with higher uncertainty
Parameter values that meet a behavioral error threshold (top 1%)
Optimal parameter value
Parameter estimate with lower uncertainty

Figure 3.



a.
b.
Concentration (ppb)
Time (hrs)
Uranine
Resazurin
Resorufin
0
10
20
30
40
2
4
6
8
5
15

Figure 4.

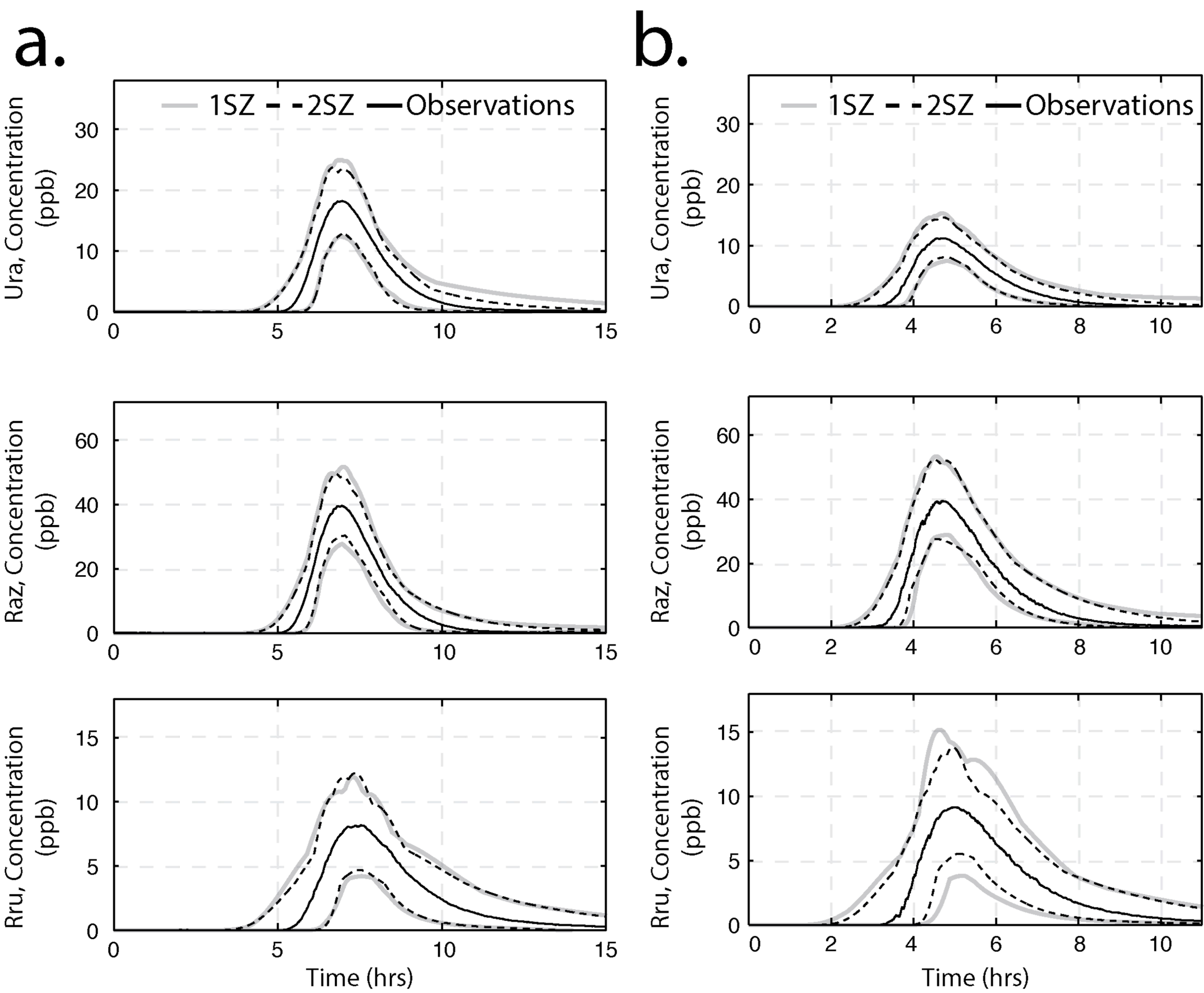
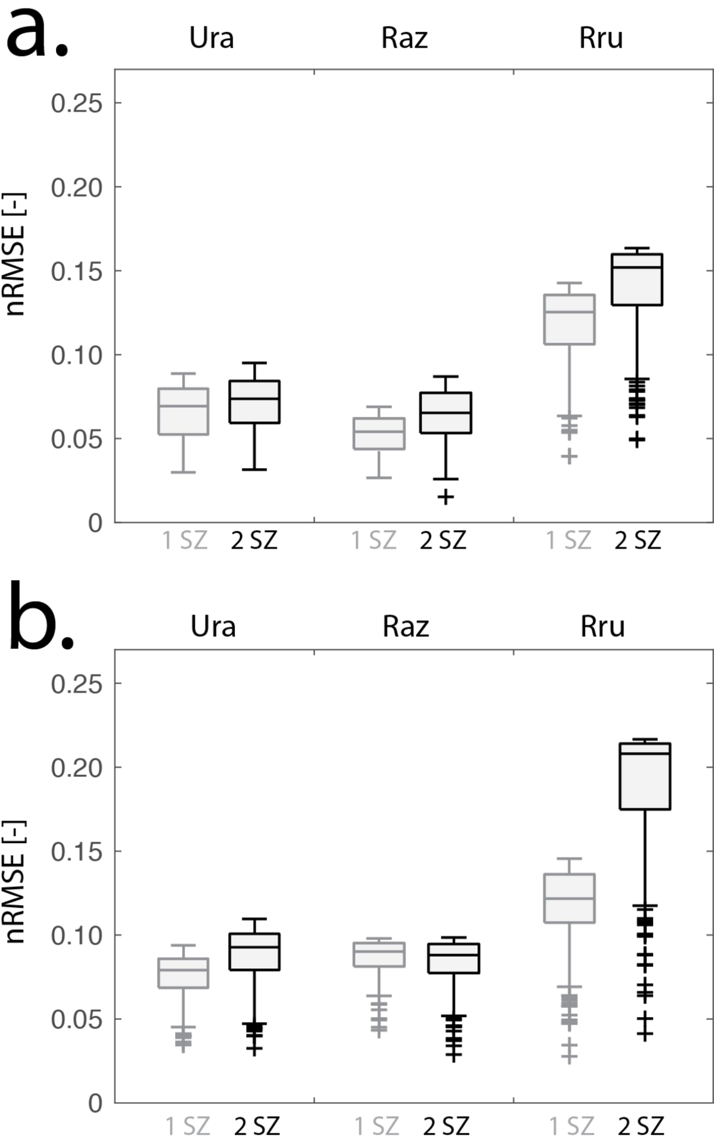
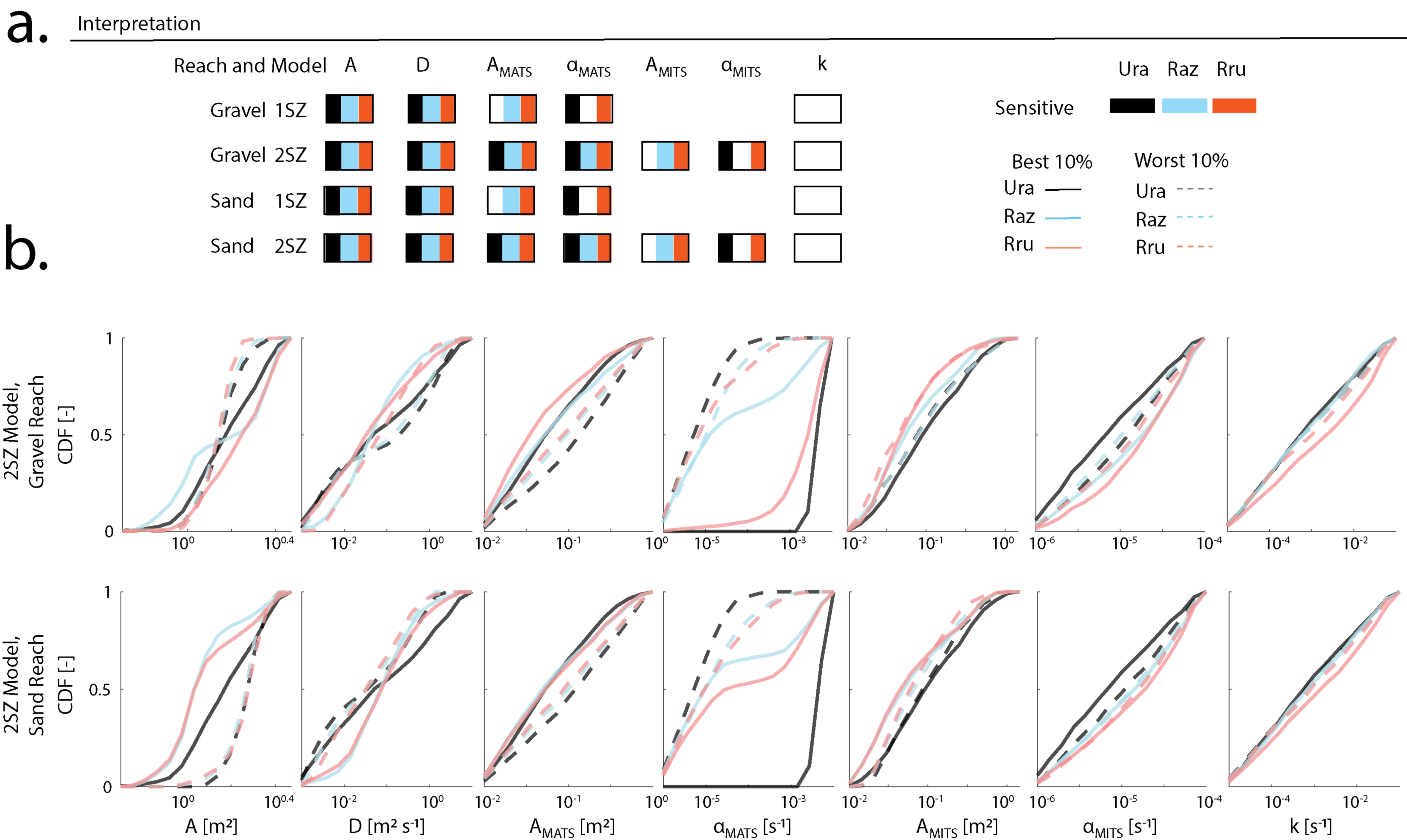

a.
b.
1SZ
2SZ
Observations
Ura, Concentration (ppb)
Raz, Concentration (ppb)
Rru, Concentration (ppb)
Time (hrs)

Figure 5.



a.
Ura
Raz
Rru
nRMSE [-]
0.25
0.20
0.15
0.10
0.05
0
1 SZ
2 SZ
1 SZ
2 SZ
1 SZ
2 SZ
b.
Ura
Raz
Rru
nRMSE [-]
0.25
0.20
0.15
0.10
0.05
0
1 SZ
2 SZ
1 SZ
2 SZ
1 SZ
2 SZ

Figure 6.



a.
Interpretation
Reach and Model
A
D
A_MATS
α_MATS
A_MITS
α_MITS
k
Gravel 1SZ
Gravel 2SZ
Sand 1SZ
Sand 2SZ
Ura
Raz
Rru
Sensitive
Best 10%
Worst 10%
Ura
Raz
Rru
b.
2SZ Model, Gravel Reach
CDF [-]
2SZ Model, Sand Reach
CDF [-]
A [m²]
D [m² s⁻¹]
A_MATS [m²]
α_MATS [s⁻¹]
A_MITS [m²]
α_MITS [s⁻¹]
k [s⁻¹]

Figure 7.

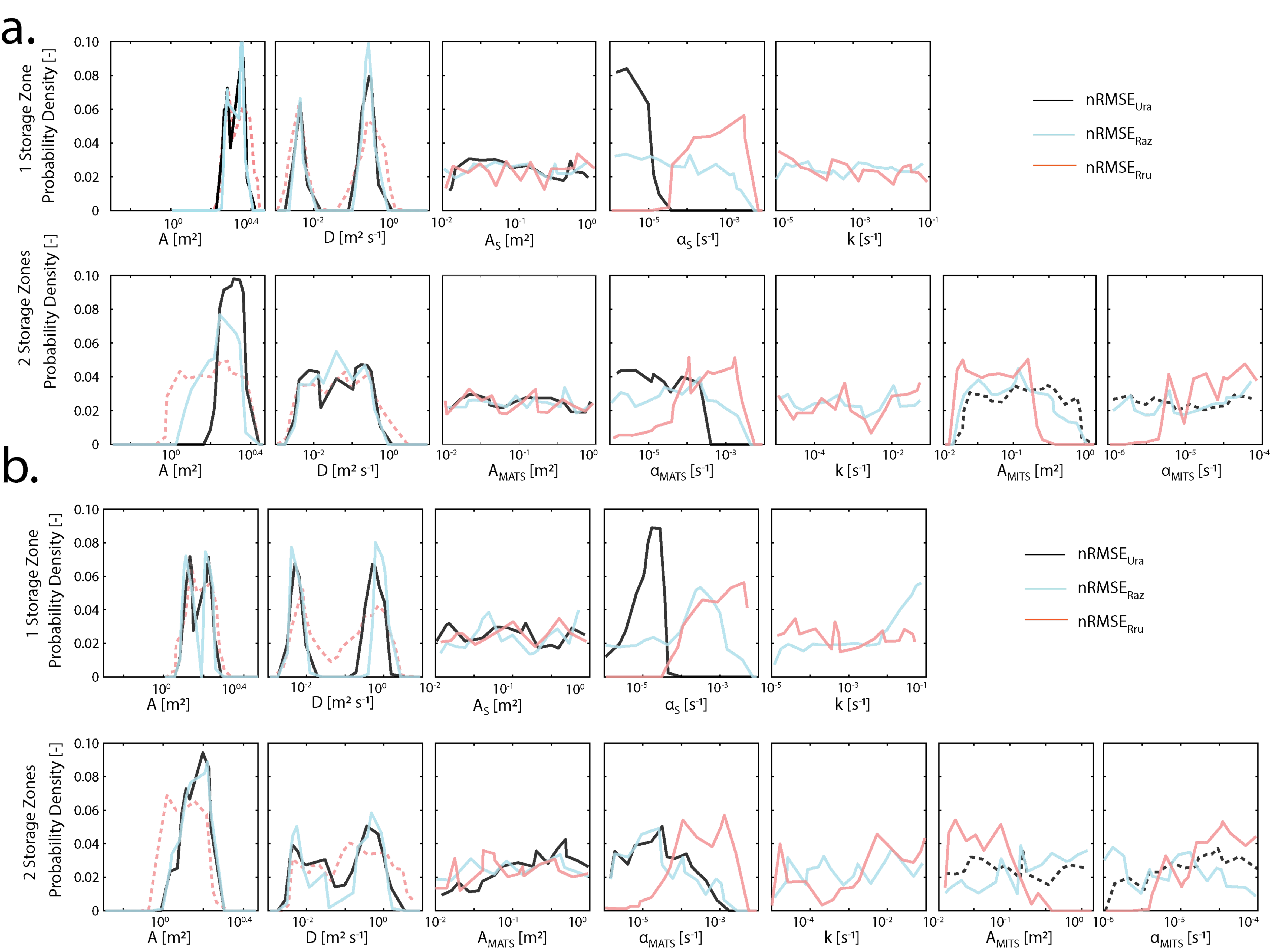
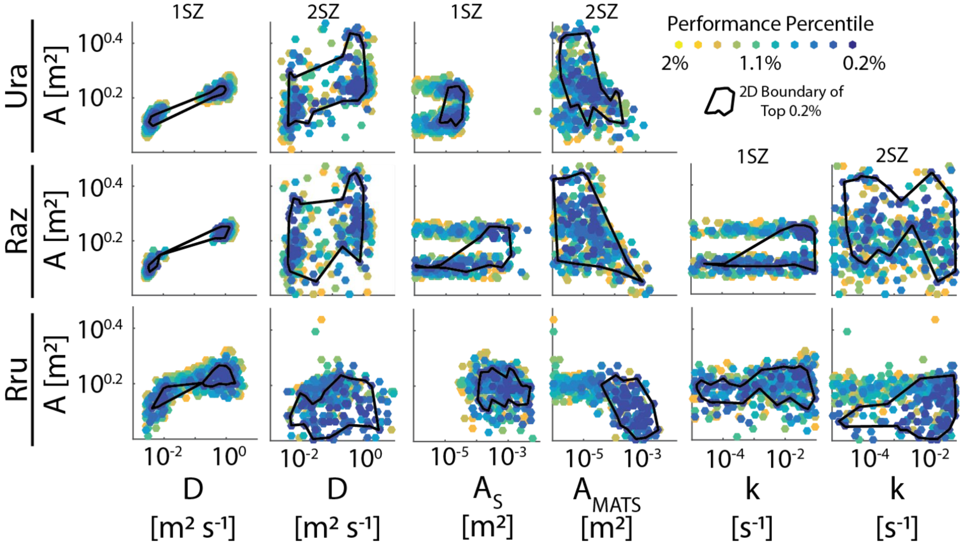
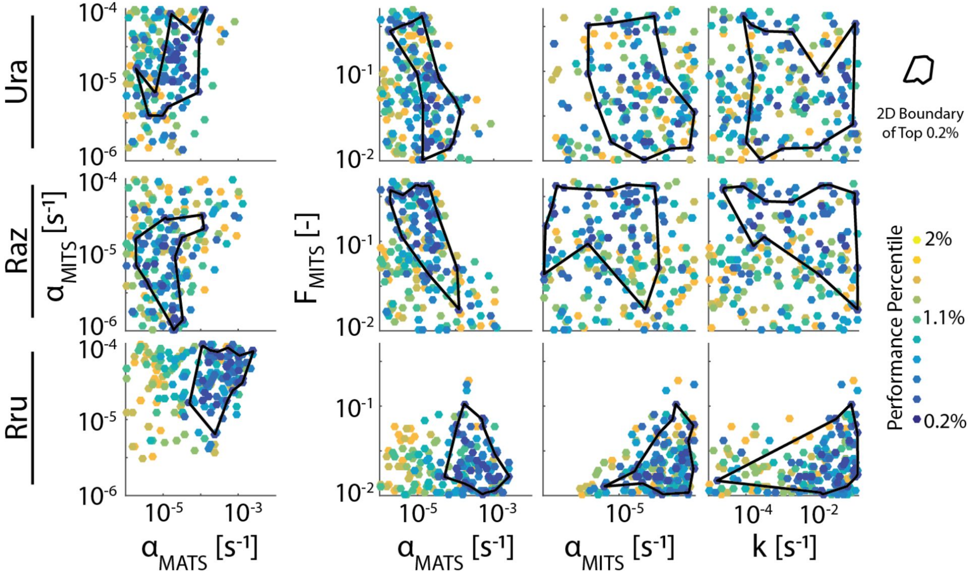

a.
b.
1 Storage Zone
2 Storage Zones
Probability Density [-]
0.10
0.08
0.06
0.04
0.02
0
A [m²]
D [m² s⁻¹]
A_S [m²]
α_S [s⁻¹]
k [s⁻¹]
A_MATS [m²]
α_MATS [s⁻¹]
A_MITS [m²]
α_MITS [s⁻¹]
nRMSE_Ura
nRMSE_Raz
nRMSE_Rru

Figure 8.



1SZ
2SZ
1SZ
2SZ
1SZ
2SZ
Performance Percentile
2%
1.1%
0.2%
2D Boundary of Top 0.2%
Ura
Raz
Rru
A [m²]
$10^{0.4}$
$10^{0.2}$
10^{-2}
10^{0}
10^{-5}
10^{-3}
10^{-4}
D [m² s⁻¹]
D [m² s⁻¹]
A_S [m²]
A_{MATS} [m²]
k [s⁻¹]
k [s⁻¹]

Figure 9.



Ura
Raz
Rru
10^{-4}
10^{-5}
10^{-6}
10^{-1}
10^{-2}
α_{MITS} [s^{-1}]
F_{MITS} [-]
10^{-5}
10^{-3}
10^{-4}
10^{-2}
α_{MATS} [s^{-1}]
α_{MATS} [s^{-1}]
α_{MITS} [s^{-1}]
k [s^{-1}]
2D Boundary of Top 0.2%
Performance Percentile
2%
1.1%
0.2%