A software tool to assess uncertainty in transient-storage model parameters using Monte Carlo simulations

Adam S. Ward[1,11], Christa A. Kelleher[2,3,12], Seth J. K. Mason[4,13], Thorsten Wagener[5,6,14], Neil McIntyre[7,15], Brian McGlynn[2,16], Robert L. Runkel[8,17], and Robert A. Payn[9,10,18]

[1]School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana 47405 USA
[2]Division of Earth and Ocean Sciences, The Nicholas School of the Environment, Duke University, Durham, North Carolina 27708 USA
[3]Earth Science and Civil Engineering Departments, Syracuse University, Syracuse, New York 13244 USA
[4]Lotic Hydrological LLC, P.O. Box 1524, Carbondale, Colorado 81623 USA
[5]Department of Civil Engineering, University of Bristol, Bristol, BS8 1TR UK
[6]Cabot Institute, University of Bristol, Bristol, BS8 1UJ UK
[7]Centre for Water in the Minerals Industry, Sustainable Minerals Institute, University of Queensland, St Lucia, Queensland 4072 Australia
[8]US Geological Survey, 3215 Marine Street, Building 6, Boulder, Colorado 80309 USA
[9]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana 59717 USA
[10]Montana Institute on Ecosystems, Montana State University, Bozeman, Montana 59717 USA

Abstract: Researchers and practitioners alike often need to understand and characterize how water and solutes move through a stream in terms of the relative importance of in-stream and near-stream storage and transport processes. In-channel and subsurface storage processes are highly variable in space and time and difficult to measure. Storage estimates are commonly obtained using transient-storage models (TSMs) of the experimentally obtained solute-tracer test data. The TSM equations represent key transport and storage processes with a suite of numerical parameters. Parameter values are estimated via inverse modeling, in which parameter values are iteratively changed until model simulations closely match observed solute-tracer data. Several investigators have shown that TSM parameter estimates can be highly uncertain. When this is the case, parameter values cannot be used reliably to interpret stream-reach functioning. However, authors of most TSM studies do not evaluate or report parameter certainty. Here, we present a software tool linked to the One-dimensional Transport with Inflow and Storage (OTIS) model that enables researchers to conduct uncertainty analyses via Monte-Carlo parameter sampling and to visualize uncertainty and sensitivity results. We demonstrate application of our tool to 2 case studies and compare our results to output obtained from more traditional implementation of the OTIS model. We conclude by suggesting best practices for transient-storage modeling and recommend that future applications of TSMs include assessments of parameter certainty to support comparisons and more reliable interpretations of transport processes.
Key words: solute transport, inverse modeling, tracer, transient storage, hyporheic zone, parameter estimation, OTIS, MCAT, OTIS-MCAT

Ecologists, hydrologists, and practitioners who study or manage stream and river networks need to quantify the relative importance of in- and near-stream processes for the purposes of understanding the fate and transport of solutes (e.g., nutrient spiraling and contaminant transport). Numerous experimental methods have been used to characterize these processes (Kalbus et al. 2006, Rosenberry and LaBaugh 2008, González-Pinzón et al. 2015, Harvey and Gooseff 2015), but a common approach is to assess characteristics of the temporary storage of water and solutes in the stream (i.e., transient storage) via comparison of solute transport model predictions with solute data from tracer tests conducted along stream reaches. Transient-storage modeling has long been applied to understand the fate and transport of solutes in streams and to characterize the relative magnitude of short-term water storage in the stream channel (e.g., eddies or pools) or in the subsurface near the stream (e.g., the hyporheic zone) (Thackston and Schnelle 1970, Bencala and Walters 1983). Values for parameters defined by the transient-storage model (TSM) formulation

E-mail addresses: [11]adamward@indiana.edu; [12]ckellehe@syr.edu; [13]seth@lotichydrological.com; [14]thorsten.wagener@bristol.ac.uk; [15]n.mcintyre@uq.edu.au; [16]brian.mcglynn@duke.edu; [17]runkel@usgs.gov; [18]rpayn@montana.edu

DOI: 10.1086/690444. Received 23 March 2016; Accepted 16 September 2016; Published online 12 December 2016.
 © 2017 by The Society for Freshwater Science.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

provide one tool for comparing fundamental solute transport characteristics within or among different streams. These parameter values usually are estimated via inverse modeling, with the goal to find a set of parameters that produces the 'best fit' between model predictions and experimentally obtained tracer concentration breakthrough curves (i.e., a time series of tracer concentrations at the downstream end of a study reach). These inferred parameter values are then applied to predictive modeling of stream solute transport in engineering pursuits or as comparative indices of stream hydraulics in scientific pursuits.

Several implementations of the TSM exist (e.g., Beer et al. 1983, Young and Wallis 1993, Camacho and González 2008), but the One Dimensional Transport with Inflow and Storage (OTIS) model (Runkel 1998) is a commonly applied and open-source computer program that has been used prolifically in studies of streams around the world. The OTIS model features a built-in nonlinear regression algorithm known as OTIS-P that uses inverse modeling to estimate parameter values and corresponding confidence intervals (CIs) based on typical statistical assumptions regarding the distribution and variance of model residuals. Parameter estimates are obtained via an iterative search of the parameter space that converges when there is either a minimal change in parameter values or model performance. Uncertainty estimates are calculated in the neighborhood of the solution (Runkel et al. 1998). Obtaining field experimental data for this approach is inexpensive for smaller streams, and completing a successful solute tracer study requires minimal expertise (Stream Solute Workshop 1990, Runkel 2015). Furthermore, OTIS-P software enables parameter estimation with limited specialized knowledge in numerical modeling or inferential statistics. As such, this approach is particularly useful for obtaining a 1st-order understanding of stream hydraulics and transport in the context of transient storage.

Despite the relative ease of the combined tracer test–OTIS-P approach, many investigators have concluded that particular data sets and stream conditions may yield substantial uncertainty in the inferred values of transient-storage parameters (Wagner and Gorelick 1986, Wagner and Harvey 1997, Wagener et al. 2002, Kelleher et al. 2013, Mueller Price et al. 2015). The limited power of transient-storage parameter estimates can be problematic because even the most basic statistical methods would emphasize that useful results from inverse modeling require an objective and rigorous consideration of uncertainty in parameter estimates. Optimization algorithms are powerful tools for arriving at a single set of parameter values for a given experiment, but we argue that reporting their uncertainty responsibly is paramount to meaningful interpretation and comparisons. Here, we present the OTIS Monte-Carlo Analysis Toolbox (OTIS-MCAT), a modeling toolbox that will allow users to gain a more robust understanding of uncertainty in parameter estimates for the OTIS model, which may contribute to a more robust representation of uncertainty in values of OTIS parameters reported in hydrologic literature.

Numerous investigators have provided evaluations of the uncertainty associated with characterizations of stream hydraulics based on inference from TSMs. Researchers now recognize that some parameters generally can be inferred with relatively high confidence and have some idea of specific conditions or experiment types that are likely to provide more or less confidence in estimates of more dubious parameters. For example, values for the cross-sectional area of the channel (A, [L^2]) can nearly always be inferred with high confidence because this parameter is the primary control on the average advective transport time that is readily and uniquely identifiable in the breakthrough curve (e.g., Stream Solute Workshop 1990, Wagener et al. 2002). Estimates of other hydraulic parameters are less reliable for a given experiment or stream condition. In particular, estimating values for parameters describing the extent of transient storage (characterized as cross-sectional area A_S, [L^2]) and the fractional rate of exchange between the channel and transient-storage zone (α, [1/T]) can be problematic (Wagener et al. 2002). This lack of confidence is unfortunate because these parameters are of particular interest for characterizing the potential for streams to alter the fate of solutes. This characterization, in turn, informs the role of stream hydraulics in controlling stream water or habitat quality. Furthermore, certainty in parameter estimation may fluctuate for multiple reasons across different stream reaches (e.g., Kelleher et al. 2013). This fluctuation implies that flow state, geomorphology, and reach length (e.g., Wagner and Harvey 1997) may influence our ability to extract parameter estimates with meaningful confidence. Here, we provide example applications of the OTIS-MCAT tools to previously published data sets to illustrate how multiple approaches to uncertainty estimation can lead to a more robust interpretation of transient-storage parameters.

Our 1st objective is to present OTIS-MCAT to the users of TSMs as a means to quantify parameter certainty for any transient-storage experiment. The toolbox was created by integrating OTIS with the MCAT (Wagener and Kollat 2007), which allows users to evaluate parameter uncertainty and sensitivity graphically. The analysis framework from OTIS-MCAT is similar to generalized likelihood uncertainty analysis (Beven and Binley 1992). Output from OTIS-MCAT aids users in deciding whether parameter estimates allow for meaningful conclusions about a given stream reach or experiment. The software developed for our study is available for download from https://github.com/WardHydroLab/OTIS-MCAT.git, and runs in the Matlab (Mathworks, Natick, Massachusetts) computing language. Our 2nd objective is to use 2 well-studied data sets to illustrate how the new toolbox complements information provided by OTIS-P and to demonstrate the value of using OTIS-MCAT with OTIS-P.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

METHODS

OTIS

All solute-transport simulations for our study were done with the OTIS and OTIS-P models (Runkel 1998). The OTIS model predicts a time series of solute concentrations (breakthrough curve) for a given parameterization at distances downstream from user-supplied upstream boundary conditions for solute and flow. The OTIS-P implementation contains an inverse modeling feature that estimates parameter values producing a predicted downstream breakthrough curve that best matches an observed breakthrough curve (e.g., from a tracer-release experiment). OTIS uses a Crank–Nicolson numerical scheme to solve the following equations for a conservative solute:

$$\frac{\partial C}{\partial t} = -\frac{Q}{A}\frac{\partial C}{\partial x} + \frac{1}{A}\frac{\partial}{\partial x}\left(AD\frac{\partial C}{\partial x}\right) + \frac{q_{Lin}}{A}(C_{Lin} - C) + \alpha(C_S - C) \quad \text{(Eq. 1)}$$

$$\frac{\partial C_S}{\partial t} = -\alpha\frac{A}{A_S}(C_S - C), \quad \text{(Eq. 2)}$$

where t is time (T), x is the longitudinal distance along the stream (L), Q is the stream discharge (L^3/T), A is the cross-sectional area of the stream channel (L^2), A_s is the cross-sectional area of the transient-storage volume (L^2), D is the longitudinal dispersion coefficient (L^2/T), α is the transient-storage exchange rate (1/T), q_{Lin} is the rate of lateral inflow to the channel per unit length along the stream (L^2/T), C is the stream concentration (M/L^3), C_s is the transient-storage zone concentration (M/L^3), and C_{Lin} is the solute concentration in the lateral inflow (M/L^3). Model parameters A, D, A_S, and α are estimated to compare the advective (A), dispersive (D), or storage-related (A_S, α) hydraulic characteristics of study reaches.

The model also parameterizes lateral outflows (q_{Lout}; L^2/T), which are used to calculate the mass balance of water embedded in the derivation of Eq. 1 (Runkel and Chapra 1994). The model equations and OTIS simulations may be extended for solutes that are not conservative, such as those that are reactive in the system or sorptive to streambed sediments. Runkel (1998) provided a full derivation and details on the numerical solution scheme, and Bencala and Walters (1983) described key assumptions of model applications to stream–hyporheic systems.

Parameter certainty

The goal of most transient-storage modeling exercises is to infer parametric information about stream hydraulics. However, appropriately characterizing or comparing stream reaches based on estimated parameter values requires an acceptable level of confidence in the parameter values themselves. Potential for uncertainty in parameter estimates obtained from OTIS model inversions has been demonstrated thoroughly in many examples (Wagner and Harvey 1997). Nevertheless, the certainty in inferred OTIS parameters is reported infrequently.

We explored parameter uncertainty with the OTIS-MCAT tool, which enables a Monte-Carlo-based evaluation of the parameter space. For nonlinear models, such as OTIS, assessing certainty in parameter estimates requires running a model with multiple parameter combinations perhaps as many as thousands of times. Within OTIS-MCAT, this exploration is completed via Monte-Carlo analysis, in which all parameters are varied randomly across user-defined ranges to produce many parameter sets. After the model is run for all parameter sets, objective functions, which compare how well (or poorly) model simulations match the observed breakthrough curve, are calculated for each parameter set generated via Monte Carlo sampling. When all parameters are varied together, many simulations approximate observations well and many simulations approximate observations poorly. Our analysis focuses on a 'best' set of runs, which we call 'behavioral', because the goal of such an analysis is to explore parameter certainty in the context of fitting the observed breakthrough curve. Behavioral runs are those regarded as consistent with observed behavior and represent the model simulations that most closely match the observed breakthrough curve. Within this framework, one can explore and assess parameter certainty vs uncertainty by visually evaluating relationships between parameter values and model performance for behavioral model runs. A simple approach to delineating behavioral simulations is to select a threshold in the objective function based on a percentile of the ensemble of simulated breakthrough curves that are considered viable descriptions of the field observations (e.g., the top 10% of objective function values; red or gray points in Fig. 1A, B). Iterative Monte-Carlo approaches are a feasible method for exploring potential sources of uncertainty in OTIS parameter estimates because the OTIS model has short run times and relatively few parameters.

We present 2 general examples in Fig. 1A, B to demonstrate how parameter certainty vs uncertainty may manifest in relationships between model performance and parameter values across behavioral model runs. One likely source of parameter uncertainty in model inversions is parameter insensitivity. An assessment of parameter sensitivity can be equated colloquially to answering the question "how much does model performance change if I change the value of a parameter?". The impact of sensitivity on model performance typically is quantified by changes in objective function values when the model is subjected to perturbations in parameter values. In OTIS-MCAT, parameter sensitivity can be assessed by using dotty plots (Fig. 1A, B) that display the relationship between objective functions and corre-

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

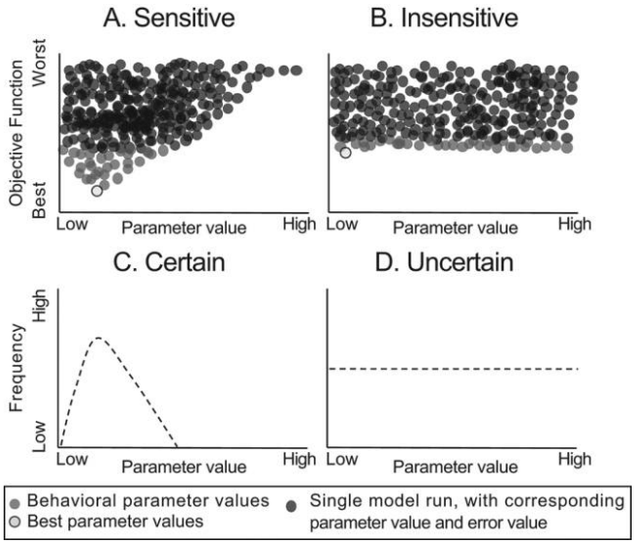

Figure 1. Conceptual illustrations of sensitive and insensitive (A, B) and certain and uncertain (C, D) parameters. Sensitivity is visualized as a relationship between parameter values and corresponding objective functions, where each point corresponds to a different value from a parameter set. The 'best' set of runs are termed 'behavioral'. See text for an explanation of sensitivity and certainty.

sponding parameter values across the entire Monte-Carlo ensemble. If model performance differs substantially across the ensemble of parameter values, one would say a parameter is sensitive within that defined range of values (Fig. 1A). An insensitive parameter has little effect on model performance over a given range of values (Fig. 1B). In general, estimates are more likely to be certain for sensitive parameters in a model inversion analysis.

We expect that parameter values that accurately simulate the observed breakthrough curves should be similar and concentrated in a narrow portion of the parameter space. Thus, certainty also may be inferred from the distribution of parameter values that produce behavioral model simulations, with confidence in estimates based on the behavioral parameter ensemble visualized as a probability density function over the sampling range of each parameter. If the range of behavioral values is peaked and narrow when compared to the sampling range, the estimated parameter has higher certainty (Fig. 1C) than if the range of behavioral values is flat and nearly as wide as the sampling range (Fig. 1D). These values can be translated into CIs based on descriptive statistics of the behavioral ensemble (e.g., percentiles or standard deviations) that should be reported with the best parameter values.

OTIS-P is an alternative but complementary approach to a Monte-Carlo analysis that produces a best set of parameter estimates, associated standard deviations of parameter estimates, and ~95% confidence limits. To calculate these outputs, OTIS-P uses an iterative nonlinear regression scheme to search the parameter space for a set of parameters that minimizes the residual sum of squared errors (RSS) between the observed and simulated downstream breakthrough curves. This parameter estimation scheme was implemented in the OTIS-P software by iterating OTIS model runs within a nonlinear least-squares algorithm described by Dennis et al. (1981) and executed within the Standards Time Series and Regression Package (STARPAC) (Donaldson and Tryon 1990). Updates to parameter estimates within the iterative framework are based on numerically approximated partial derivatives of parameter values from the previous iteration and the 2nd-order terms (Hessian) for the RSS. Changes to parameter estimates from one iteration to the next must not exceed an adaptively selected trust region diameter, which defines the scale for reliability for local model approximation (Donaldson and Tryon 1990). The algorithm converges on a final set of parameter estimates based on user-defined criteria that either specifies a minimal change in parameter value or objective function value relative to output from the previous iteration. Uncertainty estimates for OTIS-P are calculated based on a linear numerical approximation of the variance–covariance matrix within a local neighborhood of the solution, a key difference between OTIS-P and OTIS-MCAT. Whether or not this linear approximation within a neighborhood of the solution is a reasonable assumption for estimating parameter certainty depends on the nonlinearity of the model (Donaldson and Tryon 1990). Users may also opt to estimate all model parameters (A, D, A_S, and α), or to fix parameter values that are considered to be reliable estimates. A complete description of OTIS-P features was published by Runkel (1998).

Like similar single-point optimization schemes, this approach may be sensitive to where it is initialized in the parameter space (Hill and Tiedeman 2007). For robust parameter estimates, OTIS-P optimization should be executed from different initial parameter estimates to ensure that the algorithm converges to the same final set of parameter values. Beyond initializing OTIS-P from different parts of the parameter space, Runkel (1998) also recommended executing OTIS-P multiple times and using final estimates from the previous execution as initial estimates for the next. For some scenarios, often when data quality is too low to estimate all parameters or when parameter effects cannot be separated, OTIS-P may not converge on a final set of parameter estimates. However, completion without error generally is considered indicative of reliable parameter estimates.

Results from OTIS-P provide a best-fit parameter value and a 95% CI, but optimizations like OTIS-P provide more-localized information about parameter uncertainty than do Monte-Carlo-based uncertainty analyses. The uncertainty ranges reported by OTIS-P provide valuable information about uncertainty in the local vicinity of the optimized parameter values but may not always provide generalized in-

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

formation about parameter sensitivity and certainty across the orders of magnitude that OTIS model parameter values can span. Thus, researchers need to place OTIS-P estimates and their associated certainty into a context that enables a broader Monte-Carlo-based assessment of parameter sensitivity and certainty by using OTIS-MCAT and OTIS-P in tandem. Addressing this need and paring the output from these 2 tools should ultimately enable more-thoughtful use of the OTIS model to characterize stream transport characteristics by taking into account both uncertainty bounds reported by OTIS-P and broader investigation of parameter sensitivity and certainty for tracer-test experiments. The general considerations about parameter uncertainty that we present are not unique to OTIS and tracer experiments, but are applicable in any effort to infer system properties from nonlinear model inversions based on observed data. As discussed in the following sections, our toolbox introduces several visualizations that allow researchers to assess parameter certainty and sensitivity. These plots inform users about which values are appropriate for characterizing a given stream and inform comparisons between solute-tracer experiments.

OTIS-MCAT

The workflow for modeling and analysis software detailed in the following material is summarized in a Quick Start manual provided with the model code on the software GitHub site. Subsequent sections detail the implementation of the Monte-Carlo simulation framework and analysis of results using the MCAT (Wagener and Kollat 2007). The entire analysis is designed to be self-contained in its execution, insomuch as users construct the input files and execute the program in Matlab without any interactive input until results have been saved. This design allows remote execution on shared computing resources rather than requiring user interaction on a desktop system. This feature is useful for potentially long runtimes when a large number of simulations is necessary to explore a complex parameter space. We refer to the coupled OTIS model and MCAT collectively as OTIS-MCAT. We outline the workflow and details described below in a conceptual figure that summarizes steps from the generation of model input data, to model execution, to interpretation of OTIS-MCAT plots, to arriving at conclusions regarding parameter certainty (Fig. 2).

Overview of computational steps The OTIS model is executed as a forward-model run for a single stream reach to perform multiple runs with different combinations of parameters. Execution requires an upstream concentration time series as a boundary condition (observed from field data) and a set of randomly generated parameter values within user-defined ranges. The model equations in OTIS are formulated to include additional processes, including 1^{st}-order decay in the stream, 1^{st}-order decay in the transient-storage zone, sorption to streambed sediments from stream water, and sorption to hyporheic sediment from the hyporheic water (Runkel 1998). Input files include parameters for all processes and flags that are used to enable these mechanisms (if desired) and for routing of a single dissolved solute. Our implementation is applicable for only steady-state discharge conditions, although the OTIS model is capable of other discharge schemes. OTIS-MCAT could be modified to allow for such implementations with modest effort.

Monte-Carlo analysis The OTIS-MCAT software package uses the forward-modeling implementation of the OTIS model to conduct an unstructured Monte-Carlo analysis, which means that the parameter space is randomly sampled. This strategy is based on the work presented by Ward et al. (2013b) and Kelleher et al. (2013). In our implementation, the user specifies parameter ranges for all parameters controlling the transport and fate of tracers in the simulation.

Parameters can span orders of magnitude for typical values, so sampling is done with a uniform distribution in a base-10 logarithmically transformed space to allow selection of values to be distributed evenly across each order of magnitude (after Kelleher et al. 2013 and Ward et al. 2013b). The large number of parameter sets tested is designed to overcome any nonuniform distribution in the underlying parameters. We recommend that users sample parameters across $\geq$2 orders of magnitude for A_s, D, and α. Users must also specify the total number of parameter sets to be tested. Limited recommendations exist for how many parameter sets a user should sample when using Monte-Carlo analysis. We recommend sampling between 10,000 and 100,000 parameter sets as an initial search, consistent with recent guidance (Pianosi et al. 2016). We also suggest that users perform the analysis for multiple Monte-Carlo sample sizes to verify that parameter certainty outcomes do not change substantially for a larger number of parameter samples. Simulations are run for the user-specified reach length plus 200 finite-difference elements downstream from the observation point. The additional simulated stream length isolates the downstream boundary condition from the simulation point of interest to minimize the chance that results are a computational artifact of the simplified conditions assumed at the boundary.

The choice of spatial and temporal resolution for the numerical solution scheme in OTIS can negatively affect the accuracy of differential equation solutions in solute transport simulations. The Crank–Nicolson solution scheme implemented in OTIS is unconditionally stable, but oscillations in the resultant values can exist when time steps are too large or spatial elements are too small relative to

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

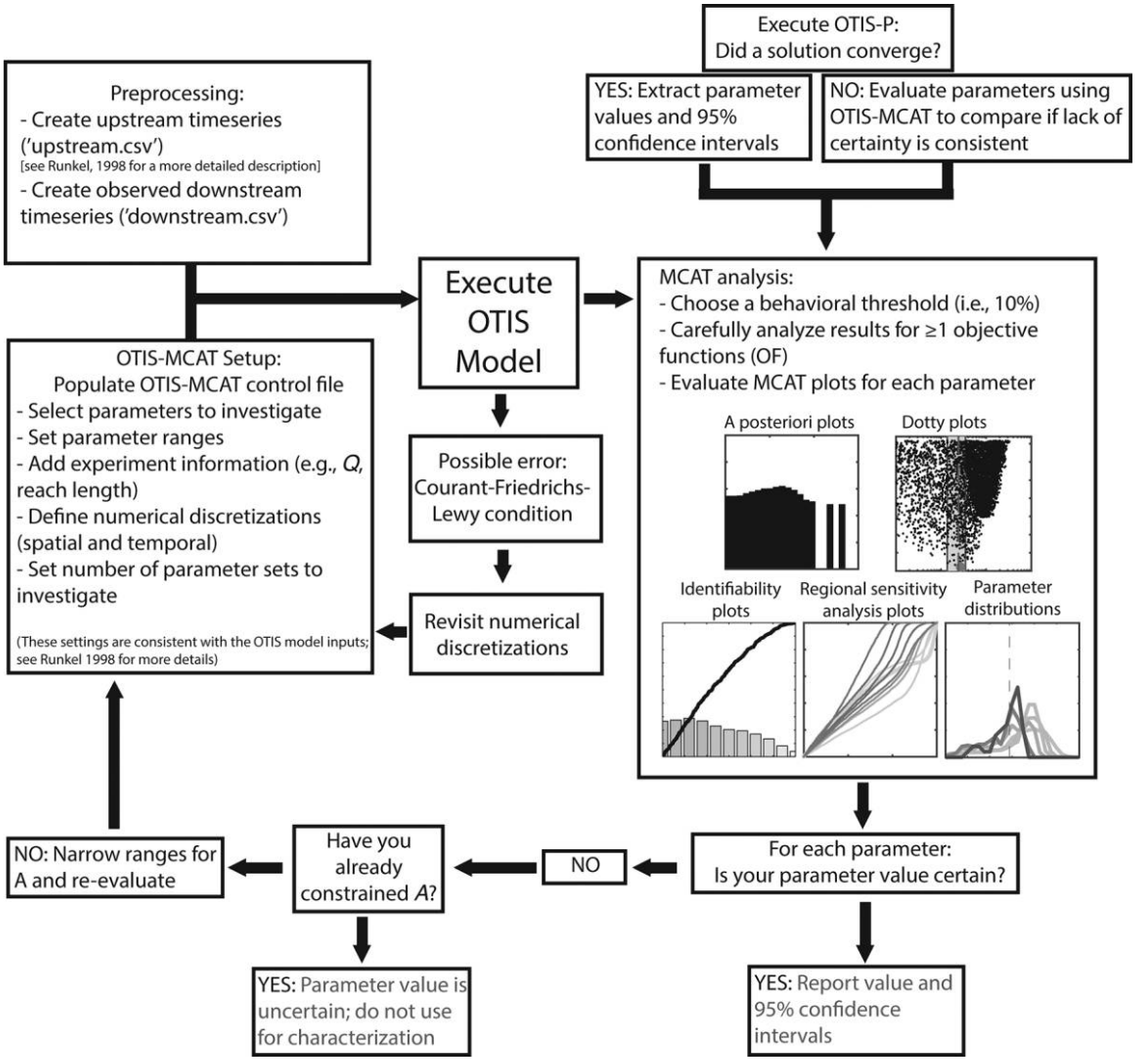

Figure 2. Conceptual workflow of One Dimensional Transport with Inflow and Storage–Monte-Carlo Analysis Toolbox (OTIS-MCAT) implementation. Q = discharge, A = cross-sectional area.

the advective transport rate of the solute. To minimize numerical error, the code calculates the Courant–Friedrichs–Lewy (*CFL*) condition:

$$CFL = \frac{u\Delta t}{\Delta x}, \qquad \text{(Eq. 3)}$$

where u is estimated velocity based on modal transit time, Δx is the user-specified spatial step, and Δt is the user-specified temporal step for simulations (Courant et al. 1928). OTIS-MCAT includes warnings for questionable stability ($0.5 < CFL < 1$) and breaks to the command line when solutions are unstable ($CFL > 1$). This step acts as a basic filter to help users select spatial and temporal steps that will yield results with minimal numerical error and will remove nonphysical computational artifacts from simulations. Adjusting spatial and temporal steps to yield smaller values of the *CFL* can reduce numerical error in the Crank–Nicholson solution scheme, but adjustments must be balanced against computation time, which increases for smaller time steps.

Model objective functions Numerous objective functions have been used to evaluate how well simulated breakthrough curves derived with OTIS-based analyses match the observed downstream breakthrough curve. For flexible comparisons with these studies, OTIS-MCAT calculates a suite of commonly used objective functions (Gupta and Cvetkovic 2000, Schmid 2003, Mason et al. 2012, Ward et al. 2013b), including RSS, the singular objective function implemented within OTIS-P.

All objective functions, unless otherwise noted, are calculated for both observed and normalized time series of

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

concentration. Time series were normalized by their 0th temporal moments as:

$$C_{norm}(t) = \frac{C(t)}{\int_0^{t_{99}} C(t)dt}, \quad \text{(Eq. 4)}$$

where t_{99} corresponds to the time at which 99% of the recovered tracer signal above background noise has passed the monitoring location. Several investigators have identified t_{99} as a meaningful time for truncation (e.g., Mason et al. 2012, Ward et al. 2013b). We use the subscript *norm* throughout to denote the normalized concentration time series or metric.

The root mean squared error (*RMSE*) is an objective function that effectively emphasizes fitting to the higher concentrations in the breakthrough curve. In notation below, C_{obs} is the downstream breakthrough curve (M/L^3), C_{sim} is the downstream solute tracer [M/L^3], and n is the number of observations in C_{obs}. The *RMSE* is calculated as:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{t=n} (C_{obs}(t) - C_{sim}(t))^2}{n}}. \quad \text{(Eq. 5)}$$

The *RMSE* for log-transformed values of the observed and simulated concentrations ($RMSE_{log}$) effectively adds more weight to errors in the smaller concentrations when calculating the objective function. The $RMSE_{log}$ is calculated as:

$$RMSE_{log} = \sqrt{\frac{\sum_{t=1}^{t=n} (\log_{10}(C_{obs}(t)) - \log_{10}(C_{sim}(t)))^2}{n}}. \quad \text{(Eq. 6)}$$

The *RMSE* and $RMSE_{log}$ metrics are not normalized to the magnitude of observed or simulated concentrations. Thus, their values are not necessarily comparable among model fits from different data sets. OTIS-MCAT also calculates the square of the Pearson product–moment correlation coefficient for both the observed (r^2) and log-transformed ($r_{log}{}^2$) values using the Matlab function *corrcoef*. We included these values in the software because they are widely reported model results, but we emphasize that users should apply these objective functions carefully because high values for r^2 can be achieved when breakthrough curve magnitudes are not matched (indicating good correlation, but a slope between observed and simulated values is very different from 1). To maintain consistency across metrics such that smaller values indicate a better fit, values of r^2 *and* $r_{log}{}^2$ are also saved as $1 - r^2$ and $1 - r_{log}{}^2$ for ease of use. The *RMSE*, r^2, and $1 - r^2$ metrics, in both normal and log forms, also are calculated for the normalized concentration time series.

Several additional metrics are calculated for the observed and simulated downstream breakthrough curves to provide flexibility in the selection of an objective function. These metrics enable users to estimate parameter sets based on specific fit characteristics of interest. For example, users interested in correctly representing behavior on the rising limb of the breakthrough curve can select a different objective function than those for whom accurately representing the maximum concentration is the most important objective.

OTIS-MCAT includes objective functions aimed specifically at matching characteristics of the peak concentration, including the difference in predicted and observed peak concentrations (ΔC_{peak}) and in predicted and observed peak arrival times (Δt_{peak}). Temporal moments of breakthrough curves can contain meaningful information regarding solute transport (e.g., Gupta and Cvetkovic 2000, Schmid 2003), so OTIS-MCAT includes objective functions that quantify differences in predicted and observed temporal moments. These differences are calculated for raw and normalized time series, where the 1st-order temporal moment (M_1) is calculated as:

$$M_1 = \int_0^{t_{99}} C(t)tdt, \quad \text{(Eq. 7)}$$

where C represents the concentration breakthrough curve of interest. Here, we follow the analyses by Ward et al. (2013a, b), where integral methods are based on the first 99% of the signal recovered (t_{99}), although other truncation timescales also have been suggested (Koestel et al. 2011). We calculate differences in the 1st temporal moment for the observed (M_1) and normalized ($M_{1,norm}$) time series. Higher-order central temporal moments (μ_n) were calculated about the mean arrival time (M_1) as:

$$\mu_n = \int_0^{t_{99}} C(t)(t - M_1)^n dt. \quad \text{(Eq. 8)}$$

Central temporal moments were calculated for both observed and normalized time series. We specifically compared the 2nd- and 3rd-order temporal variance (μ_2 and μ_3, respectively) and skewness (γ):

$$\gamma = \mu_3 / \left(\mu_2{}^{3/2}\right), \quad \text{(Eq. 9)}$$

a measure of asymmetry of the tracer mass around the mean, for the observed and normalized time series. These temporal moments can be used to calculate the apparent dispersivity (λ_{app}) and apparent dispersion (D_{app}) as (after Koestel et al. 2011):

$$\lambda_{app} = \frac{\mu_{2,norm}{}^{L}}{2}, \text{and} \quad \text{(Eq. 10)}$$

$$D_{app} = \lambda_{app} u = \frac{\mu_{2,norm}{}^{L^2}}{2M_{1,norm}}. \quad \text{(Eq. 11)}$$

OTIS-MCAT also calculates the holdback function (H), an integral measure that describes the deviation from perfect piston flow (Danckwerts 1953), which is a dimensionless

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

value describing how much tracer mass has moved through the reach over the average transport time. H ranges from 0 (perfect piston flow) to 1 (no movement) calculated as:

$$H = \frac{1}{M_{1,norm}} \int_{t=0}^{M_{1,norm}} F(t)dt \qquad \text{(Eq. 12)}$$

$$F(t) = \int_{\tau=0}^{t} C_{norm}(\tau)d\tau \qquad \text{(Eq. 13)}$$

where τ is the time variable of integration. The value of H is strongly affected by tailing in the breakthrough curve because long residence times in transient-storage result in higher mean residence times relative to the rest of the breakthrough curve. Thus, higher values of H indicate more 'holdback', which can be interpreted as greater influence of transient storage in the system.

OTIS-MCAT also calculates objective functions that compare the times of arrival for specific fractions of the solute mass in the predicted and observed breakthrough curves. Koestel et al. (2011) quantified the time at which specific quantiles of recovered solute tracer mass passed the downstream observation point. For example, the time to 5% recovery (t_5) is calculated by solving:

$$0.05 = \frac{\int_0^{t_5} Cdt}{\int_0^{t_{99}} Cdt}. \qquad \text{(Eq. 14)}$$

A series of quantile arrival times are essentially a compressed version of the information contained in a relatively high-resolution breakthrough curve. Last, OTIS-MCAT calculates the transient-storage index (TSI) as the time elapsed from the peak to t_{99} (i.e., TSI = $t_{99} - t_{peak}$) (after Mason et al. 2012). The TSI is another metric that is particularly sensitive to the effects of tailing behavior in the breakthrough curve.

Altogether, OTIS-MCAT calculates a total of 33 metrics that can all be used in an objective function for analysis of parameter sensitivity and uncertainty. Customized metrics could be defined in the MCAT code with modest effort. As with all model inversions, the choice of the objective function has the potential to bias parameter estimates, depending on whether the associated assumptions about the structure of residual error between predictions and observations are met. In practice, this statistical detail has been largely ignored in parameterizations with OTIS, and it is outside the scope of our paper. However, from the statistical perspective, each objective function is likely to bias parameter estimates in a different way, and characterizing which of these parameter sets are closest to the 'true' nature of channel hydraulics is problematic. For a given observational data set, the differences in parameter estimates obtained via different objective functions are a form of statistical bias (i.e., a computational artifact), and thus, these differences should not be subjected to physical interpretation. Furthermore, the existence of a single truly unbiased objective function is rare, because nonrandom structure in residual error is quite common in most model fits (e.g., autocorrelation). We leave formal analysis of statistical bias as a source of parameter uncertainty to future research, but we remind the reader that the choice of an objective function is frequently (and perhaps ironically) an injection of some degree of subjectivity into a model inversion analysis.

MCAT Results from the Monte-Carlo simulations using OTIS are stored and passed into the MCAT, a Matlab library of tools created to analyze parameter sensitivity and uncertainty in inversions of environmental models (Wagener and Kollat 2007). MCAT contains tools to evaluate model performance, parameter sensitivity, and predictive uncertainty. The MCAT library (version 5) is bundled with the OTIS-MCAT software package. Additional details on the interpretation of MCAT results can be found in papers by Wagener et al. (2004) and Wagener and Kollat (2007).

Acknowledging other sources of error and uncertainty The inference of parameter values from inverse modeling of field data is subject to several limitations beyond those associated with parameter uncertainty and insensitivity (the primary subjects of our study). First, all field data are subject to uncertainty. In solute-tracer studies, this uncertainty is most notably associated with the truncation of late-time tailing caused by limited sensitivity of the instrument, commonly the 'window of detection' problem (Harvey et al. 1996). One strategy to minimize this error is artificial extension of late-time behavior by fitting a distribution to the observed information (e.g., Drummond et al. 2012). Second, the structure of the TSM itself may be a source of error. Investigators have demonstrated that inclusion of multiple storage zones (e.g., Choi et al. 2000, Briggs et al. 2010) and their configuration (Kerr et al. 2013) may improve model fits, although none of these authors evaluated these improvements within an uncertainty framework. Third, the formulation of the TSM assumes an exponential residence-time distribution for the storage zone, although other models may relax this limitation (e.g., Haggerty et al. 2000, 2002, Wörman et al. 2002). Fourth, many models include assumptions of spatially homogeneous and temporally static parameter values. Last, we limit our analyses to fitting a single solute in our study. Fitting multiple species, such as temperature and solute concentration, may lead to improvements in parameter certainty (e.g., Neilson et al. 2010).

RESULTS

Case study 1: benchmarking the OTIS-MCAT approach at Uvas Creek

Introduction and field experiment We applied OTIS-P and OTIS-MCAT software to model inversions using the

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

well-studied Uvas Creek data set. We used this data set to illustrate the ability of the coupled approach to characterize a broader region of parameter space and to compare results obtained with existing tools. This case study also serves to demonstrate the steps required to apply the OTIS-MCAT tools responsibly and to interpret parameter certainties. This case study highlights the extension of the OTIS-P approach toward a better understanding of the parameter space around the best-fit values that can be obtained via OTIS-MCAT.

The Uvas Creek experiment (Zand et al. 1976, Avanzino et al. 1984) has become a benchmark data set for investigators who collect and analyze stream solute-tracer data, and it has been analyzed by numerous investigators (Bencala 1983, Bencala and Walters 1983, Jackman et al. 1984, Kennedy et al. 1984, Wagner and Gorelick 1986, Runkel 1998, Wörman 1998, 2000, Runkel et al. 1999, Schmid 2003, Scott et al. 2003, Gooseff et al. 2005, 2013, Kazezyilmaz-Alhan and Medina 2006, Kumar and Dalal 2010). Briefly, Uvas Creek is a small, headwater stream in Santa Clara County, California, USA, with a down-valley gradient of ~3% through the study reach. The stream drains a forested tributary area of ~9 km^2 and has primarily a pool and riffle morphology (Bencala and Walters 1983, Avanzino et al. 1984). NaCl was injected into the stream channel above the study reach at a steady rate for 3 h. At the injection location, discharge was 12.5 L/s and background Cl^- concentration was 3.7 mg/L. Grab samples were collected 38, 105, 281, 433, and 619 m downstream from the injection location to characterize resulting breakthrough curves in the main channel. Samples were collected regularly before, during, and after passage of the solute tracer. We used the data from the 38- and 619-m sampling locations as the upstream boundary condition and downstream observed breakthrough curve, respectively, for our experimental reach.

Model application Input files for the case study are bundled with the software download (upstream.csv, downstream .csv, and control.txt). For demonstration, we conducted 2 suites of model runs. In suite 1, we ran the OTIS model for 50,000 parameter sets using a broad range of potential values for A (0.1–1 m^2). For the TSM, A is a highly sensitive parameter because of its dominant control on the bulk advection through the system (Wagner and Harvey 1997, Wagener et al. 2002, Scott et al. 2003, Kelleher et al. 2013). In suite 2, we created a denser coverage of parameter space by narrowing the range of potential values for A (0.3–0.5 m^2) and ran 50,000 simulations. This range for A was based on the optimal value near 0.4 m^2 identified from analysis of suite 1. A also could be constrained based on observations in the field or estimated from $A = Q/u$ where the peak transit time is used to estimate u. This estimation would eliminate the need for the first suite of model runs to narrow the potential range for A. We included suite 1 in our results to demonstrate that narrowing the possible parameter value range for A may be needed to evaluate other model parameters meaningfully (Kelleher et al. 2013). For other parameters, ranges of $0.01 \le D \le 10$, $0.01 \le A_S \le 1$, and $1 \times 10^{-5} \le \alpha \le 1 \times 10^{-1}$ were used for both suites of simulations. Ranges for A, A_S, and D were set based on those used by Ward et al. (2013b). The range for α is narrower than that used by Ward et al. (2013b) based on our experience that the lowest values are functionally indistinguishable for this parameter.

We designed the 2 suites of model runs to show typical parameter-set behavior and to enable users to explore all functions in the MCAT. Simulation of 50,000 parameter combinations took <24 h to complete on a single-core processor of Indiana University's BigRedII supercomputer. Analyses presented by Ward et al. (2013b) and Kelleher at el. (2013) simulated 100,000 and 42,000 combinations, respectively, of the 4 parameters estimated in their studies (A, A_s, α, D). Users running any analysis should verify that conclusions regarding parameter certainty do not change for an increasing number of parameter sets.

We evaluated only 4 hydraulic parameters in this example, but other OTIS parameters, such as those characterizing reactive transport, could be added to an MCAT analysis. As the number of free variables increases (e.g., including reactive tracer terms, lateral inflows and outflows, sorption–desorption dynamics) the number of model runs also should increase to sample the parameter space with adequate density. Furthermore, parameter sensitivity and certainty are likely to be reduced as more estimated parameters are added to the model inversion. In the following sections we use examples from the Uvas Creek simulations to describe how a user would evaluate the sensitivity and uncertainty of a given parameter.

Interpreting parameter certainty from OTIS-MCAT We analyzed parameter certainty with respect to 2 questions: 1) is the parameter estimate sensitive to the observations?, and 2) is confidence in the best-fit parameter estimate sufficient (i.e., is the CI narrow relative to the range of possible parameter values)? We applied these questions to assess certainty with respect to the 4 OTIS model parameters and in the context of several different objective functions. We performed OTIS-MCAT analyses for 2 different ranges for A (suites 1 and 2), but we focused most of our discussion on parameter sensitivity and certainty examples from suite 2 where A is drawn from a narrower range of values.

Are parameters sensitive? A parameter is considered sensitive if variation in model performance can be clearly explained by variation in that parameter to whatever extent is considered meaningful by the modeler. Sensitivity can be assessed based on plots that display the relationship between performance (i.e., different objective function values) and parameter values. Within MCAT, 4 plots can be

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

used to assess parameter sensitivity: 1) dotty plots, 2) identifiability plots, 3) posteriori distribution plots, and 4) regional sensitivity analysis plots. Dotty plots map objective functions against parameter values, isolating how model performance (objective function value) varies with parameter values for a single parameter (Fig. 3A–O). In the dotty plot display, each point represents an objective function calculation associated with simulation from a candidate parameter set. Sensitivity of each parameter is assessed by the nature of variation in objective function values across the sampling range for that parameter. In the context of a dotty plot, a parameter is considered sensitive if there exists a region of samples that 'dips' toward a particular parameter value (Fig. 1A). This phenomenon can be observed in Fig. 3, particularly for smaller percentages of behavioral thresholds based on *RMSE* (panels K–N). The width and 'sharpness' of the dip quantify the degree of sensitivity of the parameter, where wider and flatter distributions in the dip indicate less sensitivity. A completely insensitive parameter is indicated by an uninterrupted horizontal 'front' of dots across the minimal values of the objective function (Fig. 1B). This front demonstrates a lack of any particular parameter samples that produce a better fit (lower objective function value) than other samples across the front.

To explore how parameter sensitivity varies for different definitions of behavioral runs, the MCAT allows users to interactively set the maximum threshold value of the objective function, below which parameter values will be considered a behavioral subset of the Monte-Carlo results. In some cases, reducing the parameter space examined to only behavioral runs may display a very similar picture of parameter sensitivity. More commonly, including all Monte-Carlo parameter sets as opposed to just the behavioral subset may make evaluating whether a parameter is sensitive within the behavioral region difficult. Thus, sensitivity may become more apparent when dotty plots are reduced to only behavioral runs (e.g., the increasingly clear sensitivity for *D* when decreasing the number of data points in the behavioral set in Fig. 3A, F, K). When evaluating parameter sensitivity with dotty plots, we recommend that users evaluate sensitivity for a range of behavioral thresholds. In addition, behavioral parameter sets can be used to generate the corresponding behavioral breakthrough curve simulations and examined alongside the dotty plots (Fig. 3E, J, O).

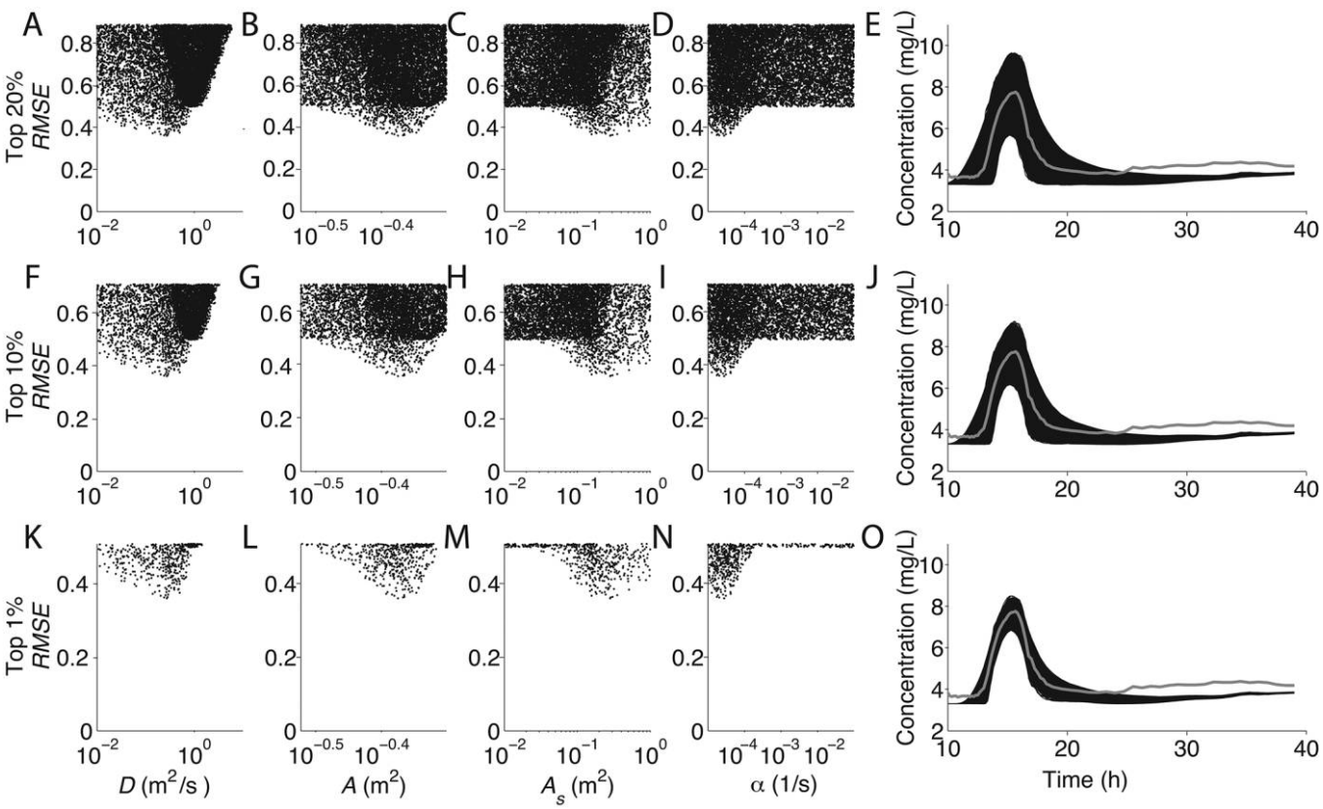

Figure 3. Dotty plots for parameter values vs Root Mean Square Error (*RMSE*) for behavioral thresholds including the top 20 (A–E), 10 (F–J), and 1% (K–O) of sets of simulations for the longitudinal dispersion coefficient (*D*) (A, F, K), cross-sectional area (*A*) (B, G, L), cross-sectional area of the transient-storage volume (A_s) (C, H, M), the transient-storage exchange rate (α) (D, I, N), and simulations (black) of tracer-solute concentrations that correspond to the behavioral parameter sets compared to the observed breakthrough curve (red) (E, J, O).

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Comparing the parameter sets to the behavioral predictions highlights how choice of the behavioral threshold corresponds to the range of simulations included in the analysis with respect to the observed breakthrough curve.

Identifiability plots display cumulative distributions and gradients along that cumulative distribution for a user-defined behavioral threshold (Fig. 4A–L). Gradients can be extracted in 2 ways from this figure: based on the slope of the line for the cumulative distribution function (a plot of the cumulative gradient) and based on the height and tint of the bar (a measure of the local gradient at a given point in the parameter space). In Fig. 4A–D, these results are displayed for a threshold of 10%. We would conclude from Fig. 4A, C, and D that parameter estimates for A_S, α, and D all have portions of the parameter space with greater likelihood based on steeper gradients (high line slope and tall bars). Behavioral parameter estimates for A (Fig. 4B), in comparison, have similar gradients across much of the parameter range (equal height and tint of bars depicting constant gradient in cumulative distribution). The parameter A appears insensitive, and therefore uncertain, from this analysis. This apparent insensitivity is in part because these results reflect sampling of a narrow range of values for A, identified as the most likely values in the analysis of a 1st suite of model runs, and wide ranges for all other parameters. Thus, Fig. 4B suggests that parameter estimates for A between 0.3 and 0.5 m^2 yield similar model performance.

A posteriori parameter distribution plots condense the data shown in dotty plots into histograms of parameter values and performance (Fig. 4E–H). Parameters from the behavioral runs are binned into 20 groups of equal size, with bar height visualized as the sum of likelihoods (calculated as 1 – [the objective function normalized across its range]). In contrast to identifiability plots, which display the gradient associated with the cumulative distribution of behavioral parameter sets, a posteriori parameter distribution plots display the likelihoods of behavioral parameter values conditioned on a given objective function (Wagener and

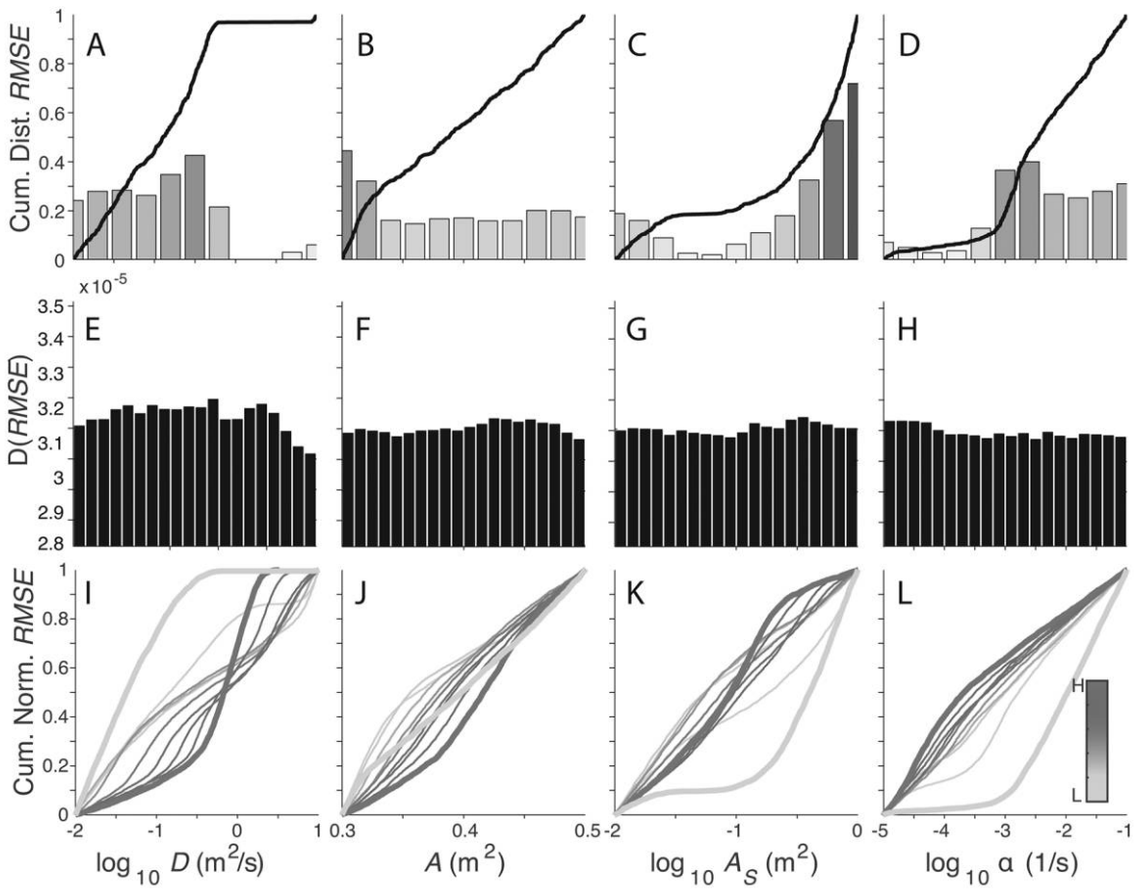

Figure 4. Parameter identifiability (the black line represents the normalized cumulative distribution (Cum. dist.) of *RMSE* for the top 10% of model solutions and bar height and shading represent the gradient of the likelihoods across behavioral parameter values approximated for 11 equally sized bins across the parameter space (A–D), *RSME* likelihood based on the top 10% of all model runs (E–H), and normalized cumulative (Cum. norm.) distributions of *RMSE* (lines represent the best 10% of model runs binned into 1% increments; each line represents 1% of all model simulations) (I–L) for D (A, E, I), A (B, F, J), A_s (C, G, K), and α (D, H, L). Steeper gradients indicate increased sensitivity at a given parameter value. Abbreviations are as in Fig. 3.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Kollat 2007). Taller bars indicate higher likelihood of a parameter value. Sensitive parameters should exhibit higher variation in likelihood across different parameter values, and insensitive parameters will have similar likelihoods across different parameter values.

A posteriori parameter distribution plots may capture similar behavior to those shown in dotty plots (Fig. 3A–D, F–I, K–N). For instance, higher likelihoods are concentrated in the middle of the parameter range for D (Fig. 4E), on the upper end of the parameter ranges for A and A_s (Fig. 4F, G), and the lower end of the parameter range for α (Fig. 4H), consistent with areas associated with better fit (i.e., lower *RMSE*; Fig. 3A–D, F–I, K–N). This visualization treats all parameter values that meet an objective function threshold as behavioral, and therefore, as having an equal weight regardless of objective function value. Thus, cases may exist where a higher density of points in an area with worse performance could still produce similar levels of likelihood as areas of the parameter space with high performance, leading to misdiagnosis of a parameter as insensitive. This problem can be seen for parameter D (cf. Figs 3A, F, K, 4E). The areas of highest likelihood are concentrated around $10^{-0.5}$ (Fig. 4E), corresponding to the portion of the parameter space with the best objective function fits (Fig. 3A, F, K), and 10^{0}, where there is a high concentration of behavioral values with higher (worse) *RMSE* values. Thus, we recommend that users carefully select and evaluate multiple behavioral thresholds, especially when using these plots, and use all plots to judge parameter certainty.

Parameter sensitivity can further assessed based on the MCAT regional sensitivity analysis plots (Fig. 4I–L). These figures are produced by dividing the population of parameters for behavioral runs into 10 bins of equal size sorted by objective function values, which are converted into likelihoods. The cumulative distribution of each bin is plotted as the parameter value vs the cumulative likelihood. The steepest portions of the cumulative distributions represent the highest sensitivities, where large changes in the objective function value are associated with small changes in the parameter value (as is the case for the highest-likelihood simulations of D, Fig. 4I). The spread of the lines is another indicator of parameter sensitivity. Tightly clustered lines indicate low sensitivity of the objective function to parameter values (e.g., α, highest likelihood distributions; Fig. 4L), whereas increased spread in the lines indicates a large difference in likelihood corresponding to large differences in the parameter value, consistent with the definition of a sensitive parameter (e.g., D and A_s; Fig. 4I, K).

One advantage of MCAT is the automated display of the data in a variety of forms, which allows users to make robust interpretations based on all visualizations. Figure 4A–L provides an example of how multiple plots can be used in concert to interpret parameter sensitivity. Interpreting across this suite, we would conclude that parameter D is very sensitive, parameters A_s and α are less sensitive, and parameter A is least sensitive among the 4, but still displays some sensitivity across parts of the parameter space. We would conclude that no parameters are insensitive. Overall, no clear-cut quantitative definition exists when using these plots to define whether a parameter is sensitive or insensitive. We recommend, at minimum, considering how parameter values vary for all model runs and for some behavioral threshold (selected interactively by the MCAT user), to verify that a parameter is both globally and locally sensitive.

A lack of parameter sensitivity could be the result of significant parameter interactions (Wagener et al. 2002, Gooseff et al. 2005, Kelleher et al. 2013). When 2 different parameters are capable of producing similar behavior in the model, they are said to interact, which ultimately reduces the ability to differentiate their individual effects on model predictions. In the TSM, both longitudinal dispersion and transient-storage produce spreading and tailing in the in-stream breakthrough curve. As such, the parameters interact, challenging the ability of model users to find an area of the parameter space with a single parameter value that produces a single best objective function value because this interaction probably will highlight multiple portions of the parameter space for different reasons (i.e., fitting the tail with a high value of D and low values of A_s and α, and the converse of these values). This interaction does not mean that the observations contain no useful information about the processes, but that the processes cannot be differentiated given the model structure and observed data.

Is a parameter value certain? Parameter certainty can be assessed by comparing a single best-fit value to the range or distribution of values across a suite of behavioral runs for a given objective function. We cannot assume that any number of random samples will explore the entire parameter space, but the goal of a Monte-Carlo exercise is to characterize the distribution of possibilities. We include 1 additional plot in the MCAT that displays the distribution of parameter values meeting a range of behavioral thresholds beside the 'best' parameter value (smallest objective function) across the MCAT ensemble for a single objective function.

Interpretation of parameter certainty will vary for different objective functions. For the Uvas Creek Experiment, we denoted behavioral runs as those corresponding to the top 10% of all values. Figure 5A–P displays interpretation of parameter certainty as a function of multiple objective functions, with frequency distributions of behavioral parameter values (Fig. 5E–H, M–P) shown below dotty plots (*RMSE*, absolute error in t_{99}; Fig. 5A–D, I–L). Deciding which parameters are certain will require use of dotty plots in the case when ranges from OTIS-P are available and parameter value probability distribution functions, which display the

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

distribution of parameter values corresponding to a range of behavioral thresholds. All parameters appear certain for *RMSE* based on both of these plots (Fig. 5A–H). We reached this conclusion because, for each parameter, probability distributions corresponding to smaller subsets of the parameter space produce a narrower range of parameter values and display peakedness around the best value indicated by the vertical gray line. In addition, the best value on each dotty plot and the CIs obtained from OTIS-P are similar and converge on a narrow area of the parameter space. We evaluated results against OTIS-P estimates for *RMSE*, but this comparison should be made carefully for other objective functions because only *RMSE* is similar to the objective function used by OTIS-P.

We also included an example, fitting the model to t_{99}, when parameters are highly uncertain. In contrast to *RMSE*, dotty plots for t_{99} indicate a general lack of sensitivity (Fig. 5I–L). Each parameter probability distribution, even at the lowest behavioral threshold (0.1%), contains parameter values that span a large part of the parameter range. This pattern is especially noticeable when comparing ranges across behavioral thresholds for α (cf. Fig. 5H,

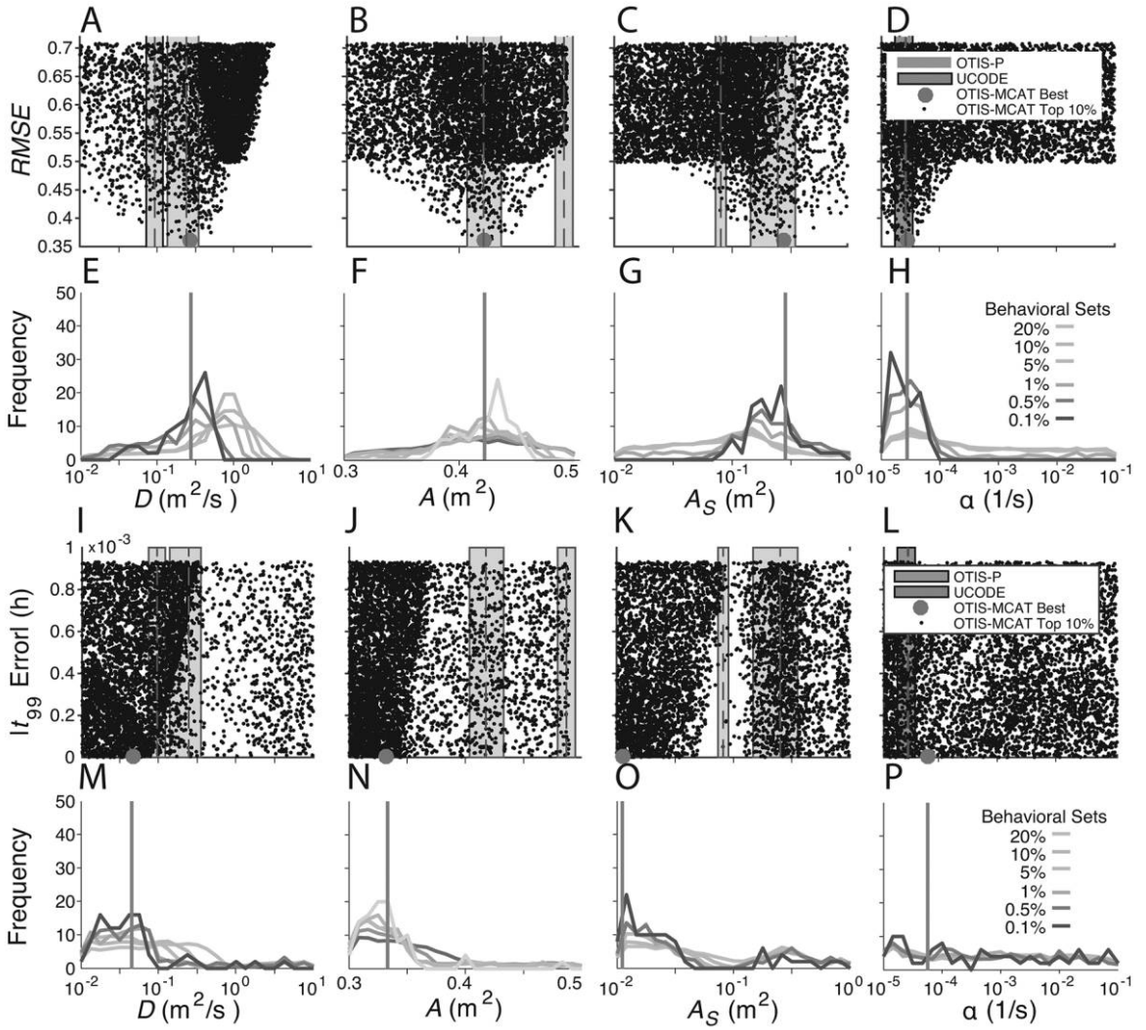

Figure 5. Dotty plots (A–D, I–L) displaying parameter values for *D* (A, E, I, M), *A* (B, F, J, N), A_s (C, G, K,)), and α (D, H, L, P) vs objective function values for the best 10% of model runs judged by *RMSE* (A–D) and absolute error in the time at which 99% of the recovered tracer signal above background noise has passed the monitoring location (t_{99}) (I–L). The mean (vertical dashed line) and 95% confidence intervals (shaded regions) for One Dimensional Transport with Inflow and Storage nonlinear regression algorithm (OTIS-P) optimized parameters (our study) and UCODE-optimized parameters (Gooseff et al. 2013) are shown in red and blue, respectively. OTIS–Monte-Carlo Analysis Toolbox (OTIS-MCAT) best-fit parameters for each objective function are shown in the large magenta circle on each subplot. We also show probability density functions for each parameter (E–H, M–P) corresponding to the top 20, 10, 5, 1, and 0.1% of values, as selected by each objective function. The parameter value corresponding to the best objective function is shown as a gray solid line. Abbreviations are as in Fig. 3.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

P). Behavioral values for t_{99} error are spread nearly evenly across the parameter sample range (Fig. 5P), but ranges for behavioral values of α by *RMSE* are concentrated $<10^{-4}$ (Fig. 5H). As demonstrated across these examples, the information contained in different objective functions will vary, and not all objective functions will be useful for all applications.

As Fig. 5A–P illustrates, users will benefit from considering parameter certainty across multiple objective functions because some may contain information about all parameters (e.g., *RMSE*), whereas others may contain little information (e.g., t_{99}). Objective functions also can be selected based on the goals of a given study (e.g., using t_{peak} if advection is a key interest, using skewness if breakthrough curve asymmetry is a key interest). We expect interpretation of sensitivity and uncertainty to change for every tracer test because the conditions of a given stream (e.g., reach length, flow rate, and other characteristics) will influence the shape of the observed breakthrough curve. As such, some objective functions should contain information about parameter values across most applications (e.g., *RMSE* and its analog, *RSS*, used by OTIS-P), but the usefulness of other objective functions probably will change across tracer tests and stream conditions. The OTIS-MCAT user bears the responsibility of selecting an appropriate objective function for a given application and evaluating this selection and its potential bias across the suite of information contained in dotty plots and behavioral simulations (Fig. 3A–O), sensitivity displays (Fig. 4A–L), and comparisons to OTIS-P (Fig. 5A–P).

Which parameters are certain? On the basis of global sensitivity analyses conducted with respect to *RMSE*, shown in the context of best values determined via UCODE (Poeter and Hill 1998) and OTIS-P from previous studies (Scott et al. 2003, Gooseff et al. 2013; Fig. 4A–L), we would regard all parameters as both sensitive and certain. The steps to arriving at these conclusions are outlined conceptually in Fig. 2. This procedure can be broadly implemented to investigate parameter sensitivity and certainty with respect to a given objective function for any tracer test. Results of our study agree with past analyses of the TSM, wherein A is the most sensitive and identifiable variable and widespread parameter uncertainty is present in combinations of other parameters (Harvey et al. 1996, Wagner and Harvey 1997, Scott et al. 2003, Kelleher et al. 2013). As such, our main contribution is to provide a tool to transient-storage modelers that enables quantification of parameter certainty and sensitivity and allows the interpretation of model results in the context of parameter identifiability.

User-defined parameter ranges may negatively affect the interpretation of OTIS-MCAT results An important choice within the MCAT analysis is selecting user-defined parameter ranges. The width of these ranges will significantly change interpretation of model performance and parameter sensitivities. Here, we describe a 2-step approach that illustrates how interpreting parameter certainty based on information obtained via OTIS-MCAT may change as parameter ranges narrow, specifically A. By constraining A, more realizations for other parameters are likely to fine-tune better model fits because advective velocity, which is mostly controlled by A, is more accurately matched. Conversely, the best values for α, D, and A_s will not produce high-quality model fits if the peak does not arrive at the observation location at the right time. We display this influence for dotty and identifiability plots in Fig. 6A–D. The range of behavioral *RMSE* values for suite 1 (Fig. 6A–D left columns) is much wider than the range for suite 2 (Fig. 6A–D right columns) (comparing all parameters), indicating that the behavioral runs for suite 2 include many better-performing parameter sets. Dotty and identifiability plots displaying parameter values for α (Fig. 6D), D (Fig. 6A), and A_S (Fig. 6C) after A was narrowed have more peaked, defined surfaces (dotty plots) and steeper gradients (identifiability plots). Thus, we conclude that analyses may contain more information regarding parameter sensitivities when the parameter ranges for A are constrained.

The role of alternate or multiple objective functions OTIS-MCAT calculates a suite of 33 alternative objective functions to quantify model outcomes. General recommendations about which objective functions may contain information for a given application are difficult to make because the information contained in a given breakthrough curve is likely to vary based on the type of stream, flow conditions, length of the tracer test, or other environmental aspects related to the site and the tracer data, all of which contribute to the truncation of observed signals (Drummond et al. 2012). As a first step, we recommend that users evaluate parameter certainty with respect to *RMSE* because, among the many objective functions calculated by OTIS-MCAT, it most closely emulates the form of RSS and, therefore, is most readily compatible with outputs from OTIS-P.

To demonstrate how different objective functions may be used to inform parameter certainty (and how different objective functions may be highly uncertain), we present box plots of parameter values for the top 10% of model outcomes for each objective function (Fig. 7A–D). Results are shown only for the suite 2 model runs (narrowed range for A), but are similar to those calculated for suite 1. In Fig. 7A–D, some combinations of parameters and objective functions produce very narrow ranges of optimal parameter values. For example, α values were constrained to narrow ranges for t_{90}, t_{95}, t_{90norm}, and t_{95norm} (see Fig. 7 for definitions), results suggesting these objective functions (Fig. 7D), for this particular case, may contain more information regarding exchange between the advective channel and

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

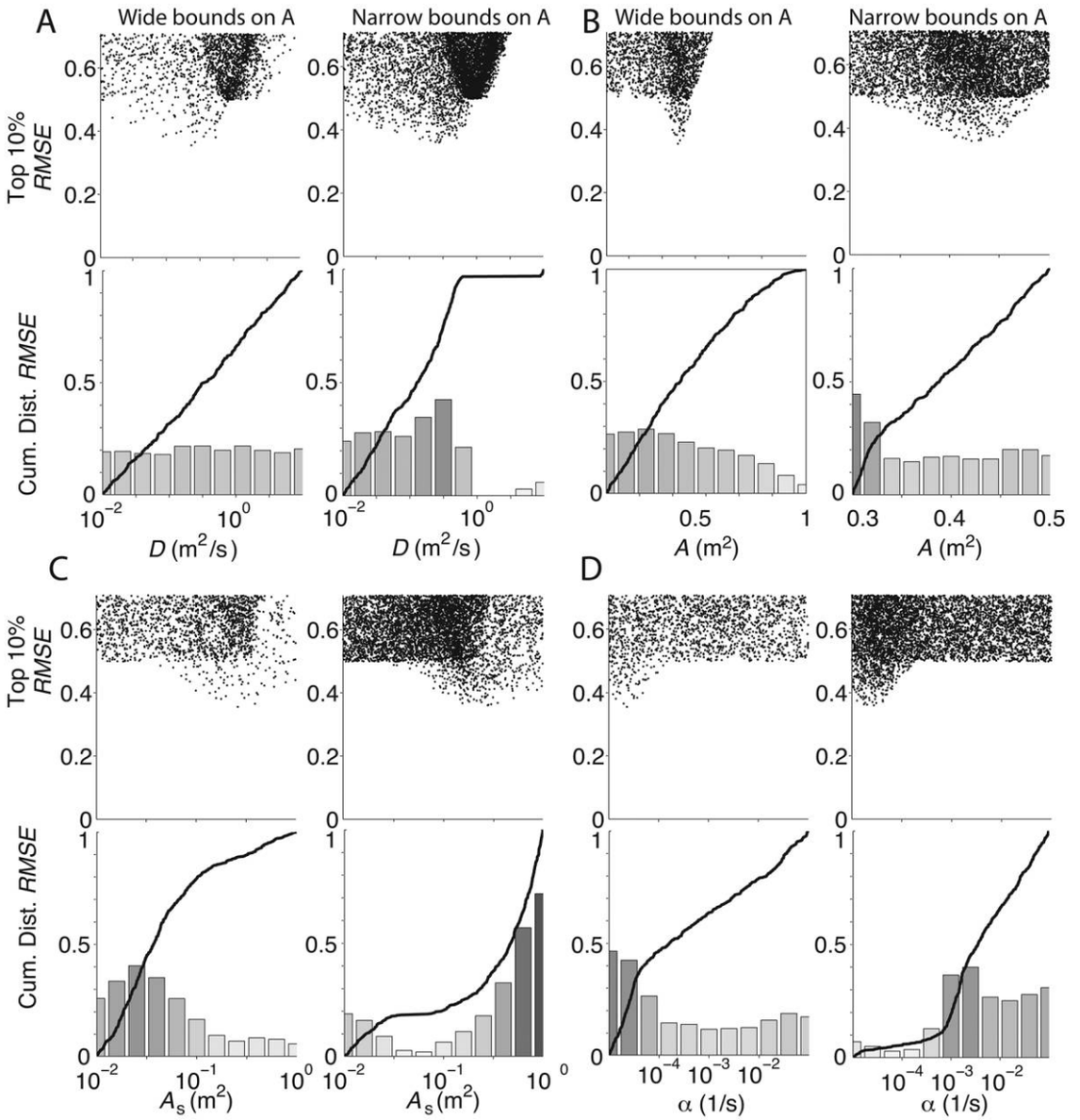

Figure 6. Dotty plots and identifiability plots (A–D) for a simulation set with wide bounds on cross-sectional area (A) (suite 1; A range = 0.1–1 m^2) (left columns in A–D) and one with narrow bounds on A (suite 2; presented in all other figures, range = 0.3–0.5 m^2) (right column in A–D) for parameter values D (A), A (B), A_s (C), and α (D). Both figures display results for the top 10% of model runs for Root Mean Square Error (*RMSE*) for a suite of 50,000 parameter sets. Where A is sampled across wide bounds, uncertainty of all other parameters is high because of overwhelming sensitivity of results to A (i.e., if A is not correct, no values of other parameters can help the model fit). Narrowing the sampling range for A based on results from suite 1, improved certainty of other parameters. The narrower range for A in suite 2 is denoted on the dotty plot for A for suite 1. Cum. dist. = cumulative distribution. Abbreviations are as in Fig. 3.

transient-storage zone. This interpretation was confirmed by the high information content for α observed at late times in analyses here and elsewhere (Wagener et al. 2002, Wlostowski et al. 2013). Wider box plots demonstrate a larger range of parameter values that produce behavioral results for a given objective function. Parameter estimates are likely to be uncertain for objective functions with wide interquartile ranges.

Given that the ultimate goal is to characterize a stream reach based on sensitive and certain parameters, OTIS-MCAT enables evaluation of these criteria across a range of different objective functions. This approach extends

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

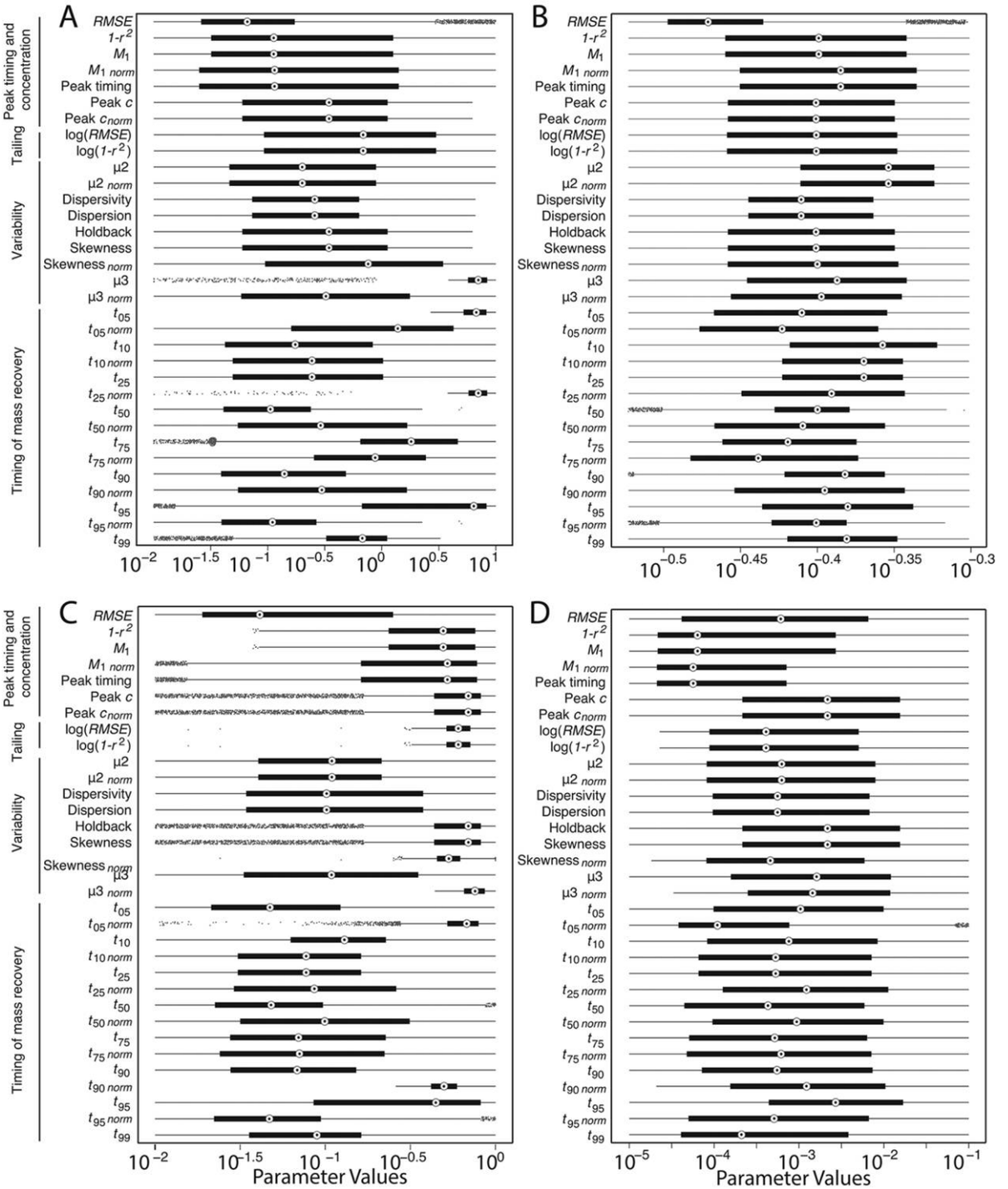

Figure 7. Box plots for the distribution of D (A), A (B), A_s (C), and α (D) for the top 10% of model runs for each objective function, enabling a preliminary assessment of how robust a parameter value may be across a range of possible objective functions. More tightly grouped parameter values indicate a more certain parameter. Objective functions are broadly organized by the objective they typically address. Black boxes represent the interquartile range, whiskers designate the 10th and 90th percentiles, small dots are outliers, and the circle within the interquartile range is the median value. *RMSE* = root mean square error, r^2 = coefficient of determination, $M_n = n^{th}$ temporal moment, $\mu_n = n^{th}$ central temporal moment, t_X = time to X% of signal passed, and the subscript *norm* indicates the value was calculated for the concentration time series normalized with Eq. 4).

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

beyond the standard objective functions used by OTIS and OTIS-P. However, this extension limits our ability to compare results to OTIS-P estimates. If a parameter is well identified for an objective function not included within OTIS-P, obtaining its unique value for stream-reach characterization can be done in 1 of 2 ways. For the case where a large number of model runs already have been performed, the unique value can be approximated from the OTIS-MCAT analysis. Otherwise, researchers should consider implementing an optimization algorithm that can search according to the given objective function. For example, the Shuffled Complex Evolution algorithm is one optimization tool that readily interfaces with Matlab (Duan et al. 1993, 1994). Each objective function has inherent assumptions and possible biases. Users should use a single objective function to assess all parameters. For example, assessing A on the basis of *RMSE* and D on the basis of r^2 might be inappropriate because these objective functions emphasize different aspects of the breakthrough curve.

Comparison to OTIS-P and UCODE analyses A premise of our paper is that the OTIS-MCAT outputs complement the OTIS-P estimates of parameter uncertainty. To demonstrate the paired use of OTIS-MCAT and OTIS-P, we also completed parameter optimization based on OTIS-P following procedures outlined by Runkel (1998). We executed multiple sets of optimization runs, starting from an initial estimate of the best-fit values based on minimum *RMSE* from the Monte-Carlo simulations. We stopped iterating with OTIS-P when parameter values changed by <0.1% between subsequent runs with parameter results summarized in Table 1. Gooseff et al. (2013) presented results from parameter optimization for the same data set that were conducted with UCODE based on minimizing the sum of weighted squared errors (Table 1). We constructed dotty plots superimposed by the OTIS-P and UCODE optimized parameter values and 95% CIs to compare OTIS-P and UCODE results to the OTIS-MCAT outputs (Fig. 5A–D, I–L).

The best-fit OTIS-MCAT value (conditioned on *RMSE*) and OTIS-P values (conditioned on RSS) were nearly identical for all parameter estimates obtained via OTIS-MCAT and OTIS-P, and the OTIS-MCAT best-fit values always were within the OTIS-P 95% CI (Table 1, Fig. 5A–D). OTIS-MCAT, UCODE, and OTIS-P all suggest similarly narrow (and therefore high) levels of confidence in parameter estimates, with at least order-of-magnitude accuracy in model predictions based on existing tools (Table 1). However, interpreting parameter sensitivity and certainty, while supported by the information in Table 1, is further justified by the suite of analyses contained in OTIS-MCAT. Thus, OTIS-MCAT, especially when used alongside OTIS-P, can provide a clear picture of parameter certainty and sensitivity for several combinations of parameters and objective functions, thereby allowing users to exploit fully all of the information available across the breakthrough curve.

In contrast, when the objective function is changed, parameter estimates obtained OTIS-MCAT and OTIS-P may not identify best estimates in similar areas of the parameter space (Fig. 5I–L). This situation is especially apparent when comparing best estimates for OTIS-MCAT and OTIS-P to UCODE estimates, which are based on a slightly different objective function. The 95% CI obtained from UCODE (Gooseff et al. 2013) did not include the OTIS-MCAT best fit for A, D, or A_S (Fig. 5I–K). Results for the objective function of maximizing r^2 also demonstrate how use of a differ-

Table 1. Comparison of best-fit transient-storage model parameters and 95% confidence intervals (CIs; in parentheses) based on multiple fitting techniques for Uvas Creek. OTIS = One Dimensional Transport with Inflow and Storage, OTIS-P = OTIS nonlinear regression algorithm, OTIS-MCAT = OTIS-Monte Carlo Analysis Toolpack, *RSME* = Root Mean Square Error, D = the longitudinal dispersion coefficient, A = cross-sectional area, A_s = cross-sectional area of the transient-storage volume, α = the transient-storage exchange rate.

			OTIS-MCAT		
Parameter	Gooseff et al. (2013)[a]	OTIS-P[b]	*RMSE*[c]	r^{2c}	t_{99}[c]
D (m^2/s)	0.094 (0.073–0.120)	0.242 (0.137–0.348)	0.261 (0.237– 0.285)	0.147 (0.139–0.157)	0.465 (0.286–0.644)
A (m^2)	0.497 (0.489–0.505)	0.424 (0.409–0.440)	0.416 (0.413–0.420)	0.395 (0.391–0.398)	0.344 (0.339–0.350)
A_S (m^2)	0.082 (0.074–0.091)	0.253 (0.148–0.357)	0.279 (0.254–0.303)	0.126 (0.101–0.151)	0.138 (0.112–0.165)
α (10^{-5}/s)	2.63 (2.07–3.34)	2.56 (1.70–3.43)	3.42 (3.15–3.68)	631 (424–839)	747 (537–957)

[a] Reach "2-5" used by Gooseff et al. (2013). Optimization with UCODE (Poeter and Hill 1998). Objective function is minimization of weighted least squares objective function (Scott et al. 2003, Gooseff et al. 2013).

[b] Optimization with OTIS-P (Runkel 1998) using STARPAC (Donaldson and Tryon 1990). Objective function is minimization of residual sum of squares.

[c] Monte-Carlo simulation with OTIS-MCAT. Mean and 95% CI for the mean reported for the top 0.5% of model runs based on minimizing *RMSE*, maximizing r^2, and minimizing the difference in t_{99} between observed and predicted breakthrough curves.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

ent objective function can lead to disagreement in best-fit parameter estimates when compared to OTIS-P results (Table 1).

Case study 2: Stringer Creek

Field experiment and modeling In August 2005, a series of conservative solute-tracer injections were completed along Stringer Creek in the Tenderfoot Creek Experimental Forest in Montana, USA (Payn et al. 2009). Stringer Creek is a 2nd-order stream draining a 5.5-km^2 catchment. The stream itself includes ~2600 m of valley length. Here, we focus on results of 2 instantaneous injections of NaCl into the stream channel for valley segments with downstream coordinates 100 and 2500 m upstream from the permanent gauge. These segments are representative of 2 primary typologies of breakthrough curves identified by Kelleher et al. (2013) in their study of the same data set. In-stream tracer breakthrough curves were recorded as specific conductance at the up- and downstream ends of each 100-m valley segment. Additional details about the field site and experiments were published by Payn et al. (2009), Patil et al. (2013), Ward et al. (2013b), and Kelleher et al. (2013).

For each segment, we applied OTIS-P as detailed in the user manual (Runkel 1998) and iterated to a best-fit parameter set. The STARPAC tool yielded local estimates of parameter certainty as 95% CIs. We also applied the OTIS-MCAT. Parameter ranges were set as $10^{-3} \leq D \leq 5$ m^2/s, $0.01 \leq A_s \leq 10$ m^2, and $10^{-7} \leq \alpha \leq 0.1$/s for both segments. We used modal advective velocity and discharge to estimate a channel area, with parameter bounds set as ±50% around this estimate for each segment (~$0.11 \leq A \leq 0.33$ m^2 and $0.031 \leq A \leq 0.095$ m^2 for the 100- and 2500-m segments, respectively). For each experiment we sampled 100,000 uniformly sampled realizations and assessed model fitness based on *RMSE*.

Interpretation of parameter identifiability For the Stringer Creek data, both OTIS-P and OTIS-MCAT produce best-fit parameter values with narrow 95% CIs (Table 2). If using only 1 of these methods, the interpretation of these results would be that these parameters can be interpreted reasonably or that the user has high confidence in these parameter values. Taken together, these CIs add an element of uncertainty to the interpretation. Values for A, the most certain parameter, are similar between the 2 methods. However, the 95% CIs are nonoverlapping for D, A_s, and α. Thus, 2 unique methods have produced best-fit parameters with comparable CIs, but uniquely different values. In this case, the additional insight provided by OTIS-MCAT aids interpretation of parameters. Based on dotty plots, the parameters D and A are identifiable (i.e., both sensitive and certain) for both study segments (Fig. 8A, B, I, J). However, both A_s and α appear somewhat insensitive (Fig. 8C, D, K, L). Based on the probability distributions generated from the dotty plots (Fig. 8E–H, M–P), we recommend that A_s be interpreted only with acknowledgement of questionable certainty and that α not be interpreted because it appears both insensitive and uncertain.

This example highlights the utility of OTIS-MCAT for providing context for the interpretation of parameters and confirming local uncertainty analyses reported by OTIS-P. The difference between OTIS-P optimized parameters and OTIS-MCAT parameter estimates may be surprising, given that these approaches use the same data, same model, and equivalent objective functions. The key difference is the way these estimates are obtained because each is geared toward a different outcome. OTIS-MCAT conducts a search of the global parameter space. It may not find a better set of parameter estimates than OTIS-P because the primary goal of OTIS-MCAT is to characterize broadly the relationship between parameter estimates and model performance across a range of plausible values for each parameter. In contrast, OTIS-P uses an algorithm based on the gradient between parameter values and model performance to iterate toward a best fit, searching a portion of the parameter space but geared toward producing a best set of parameter estimates corresponding to high model performance. Un-

Table 2. Comparison of best-fit transient-storage model parameters and 95% confidence intervals using multiple fitting techniques for Stringer Creek. Abbreviations are as in Table 1.

	100 m		2500 m	
Parameter	OTIS-P[A]	OTIS-MCAT[B]	OTIS-P[A]	OTIS-MCAT[B]
D (m^2/s)	0.0932 (0.0893–0.0970)	0.184 (0.178–0.191)	0.0502 (0.0460–0.0543)	0.109 (0.104–0.113)
A (m^2)	0.203 (0.201–0.205)	0.217 (0.214–0.219)	0.0584 (0.0576–0.0592)	0.0619 (0.0610–0.0623)
A_S (m^2)	0.0336 (0.0322–0.0351)	1.04 (0.856–1.221)	9.56×10^{-3} (8.79×10^{-3}–10.3×10^{-3})	1.098 (0.920–1.276)
α (10^{-3}/s)	1.95 (1.79–2.11)	5.16 (3.88–6.44)	0.0596 (0.0499–0.0693)	3.75 (2.62–4.89)

[a] Optimization with OTIS-P (Runkel 1998) using STARPAC (Donaldson and Tryon 1990). Objective function is minimization of residual sum of squares.

[b] Monte-Carlo simulation with OTIS-MCAT. Mean and 95% CI for the mean reported for the top 0.5% of model runs based on minimizing *RMSE* between observed and predicted breakthrough curves.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

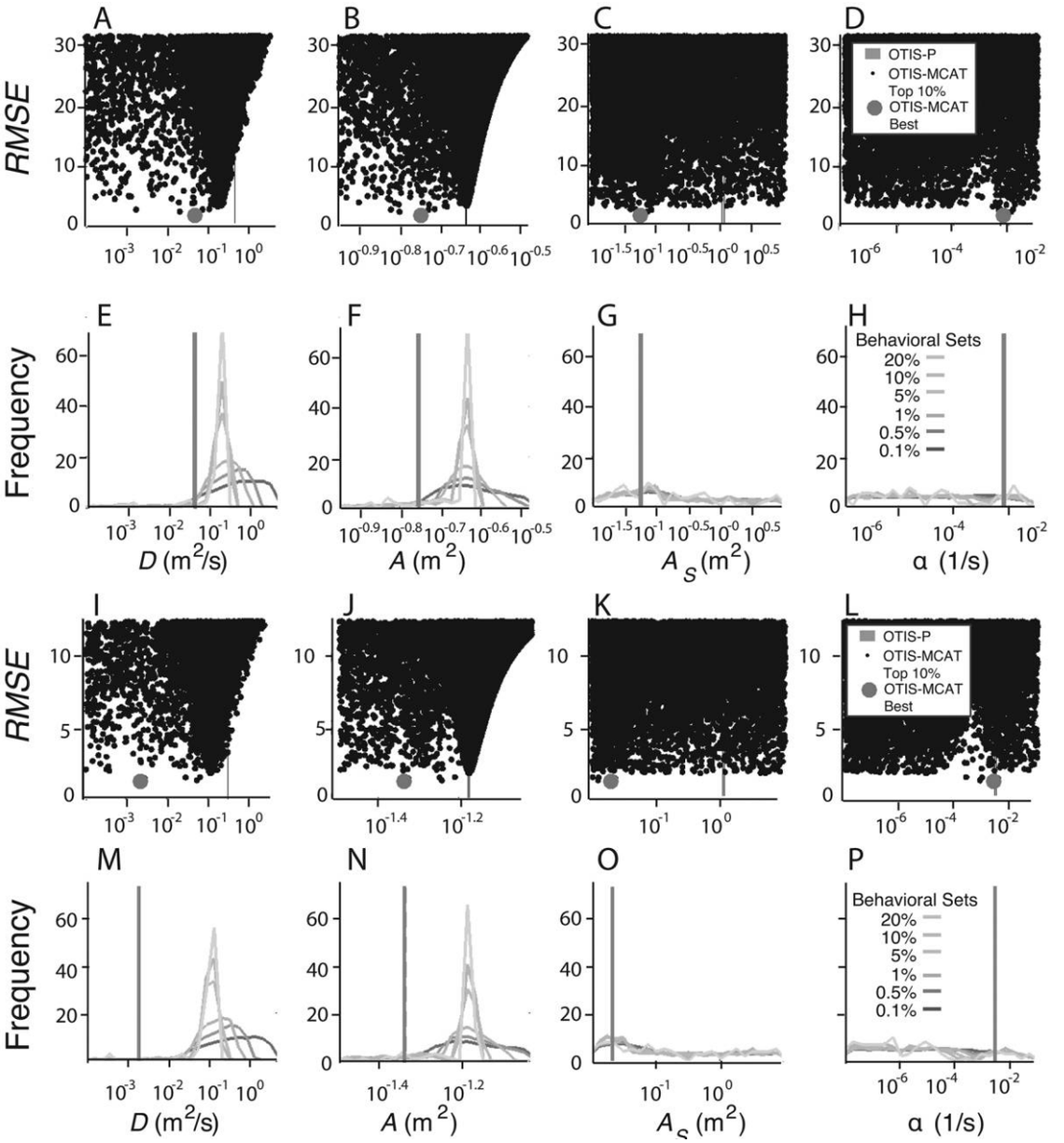

Figure 8. Dotty plots displaying parameter values for D (A, I), A (B, J), A_s (C, K), and α (D, L) corresponding to model simulations for experiments in Tenderfoot Creek Experimental Forest at reaches 100 (A–D) and 2500 m (I–L) upstream from the permanent gauge. For each parameter per reach, dotty plots display parameter values vs *RMSE* for the top 10% of runs. The mean (vertical dashed line) and 95% confidence intervals (shaded regions) are shown for OTIS-P optimized parameters (note that the confidence intervals are extremely narrow). We also show probability density functions for each parameter (E–H, M–P) corresponding to the top 20, 10, 5, 1, 0.5, and 0.1% of values, as selected by each objective function. The parameter value corresponding to the best objective function is shown as a grey solid line. Abbreviations are as in Fig. 3.

certainty estimation in this formulation is secondary and is done after the algorithm has ceased iterating. Thus, these methods provide different but complementary information. We may trust the parameter estimates obtained from OTIS-P, but OTIS-MCAT is perhaps a better tool for assessing the certainty of those estimates. Likewise, OTIS-P probably will deliver a parameter estimate with better performance than any parameter set investigated via OTIS-MCAT. In short, finding a set of parameters that performs well does not inherently prove they are identifiable or certain.

Two plausible explanations exist for the observed mismatch in parameter certainty delivered by these 2 tools. First, OTIS-P may be converging on a local minimum with

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

steep local gradients. The optimization algorithm implemented in OTIS-P would lend itself to locating any steep gradient within the parameter space. In contrast, OTIS-MCAT searches broadly and without regard to optimization (i.e., a brute force approach to identifying best-fit parameters). Second, OTIS-MCAT may be undersampling at the location where OTIS-P converges. A very narrow "spike" of high likelihood could be missed by the OTIS-MCAT's random generation of points. A strategy to test this case would be to run an additional suite of OTIS-MCAT simulations with more narrowly defined parameter ranges around the OTIS-P best-fit solution. The apparent disagreement between the 2 tools indicates that OTIS-P parameters are generally situated in a position where minor changes do not strongly effect the model performance. Still, this characteristic is complemented by OTIS-MCAT, thereby enabling an understanding of where that parameter range exists in a global context.

DISCUSSION

Recommendations and synthesis

Best practices in reporting TSM parameters OTIS-MCAT users have several key considerations when interpreting model results. First, users should consider whether the full TSM is required to represent the observed data, or whether the advection–dispersion equation alone is sufficient. Scott et al. (2003) demonstrated a technique to assess a model with and without transient-storage terms. OTIS-MCAT also can be applied to approximate this test by fixing values for $\alpha = 0$. In this case, improved model performance when α and A_S are varied alongside A and D relative to when α and A_S are fixed would indicate that the increase in model complexity and parameterization requirements is justified to better represent key processes influencing transient storage within a stream reach.

Second, users must select the objective function(s) for which they will interpret models. Selection of *RMSE* as an objective function will enable more-direct comparison with estimates obtained from OTIS-P, given similarity in the form of *RMSE* and the objective function used by OTIS-P. In addition, *RMSE* probably will give good model fits across the entire breakthrough curve. For a different objective or outcome, such as emphasizing accurate prediction of late-time behavior (e.g., t_{99}), OTIS-MCAT and OTIS-P results may diverge. This divergence is not an indication that OTIS-P is incorrect, rather that the optimal parameter values vary for different objective functions because they emphasize or weight different parts of the breakthrough curve. As such, optimized model parameters must be reported on the basis of the objective function used to select them. Not all objective functions will provide the same parameter estimates, and the best values for one objective function probably will not correspond to the best values for another objective function. Thus, reducing uncertainty in one objective function may not guarantee a reduction in uncertainty for another objective function. For example, parameters selected for minimizing *RMSE* may not minimize error in t_{99}, whereas optimized parameters for t_{99} may increase uncertainty in peak concentration and average arrival time, and other performance metrics related to advection and dispersion. This situation is evidenced by the nonoverlapping or minimally overlapping parameter ranges for behavioral runs in Fig. 7A–D when considering different objective functions.

Last, we recommend that researchers and practitioners consider results of global uncertainty analysis within the context of local uncertainty obtained from optimized parameter estimates, based on lines of evidence from both OTIS-P and OTIS-MCAT (Fig. 5A–P). As demonstrated in the case studies, OTIS-P output contains useful information for interpreting parameter best estimates and certainty with respect to the RSS objective function implicit to the OTIS-P tool. By interpreting this information alongside global uncertainty output from OTIS-MCAT, users can decide whether their parameter values are certain and, therefore, interpretable with respect to *RMSE*/*RSS* and a broad range of alternatives. Uncertain parameters should not be compared across experiments nor further interpreted. A suite of nonparametric metrics is calculated within OTIS-MCAT to describe the observed breakthrough curve as an alternative to reporting TSM parameters. Each of these summary metrics (e.g., temporal moments, t_{peak}) can be used to describe the observed downstream breakthrough curve without invoking the TSM and its assumptions. Examples of work based on this strategy can be found in the literature (Koestel et al. 2011, Mason et al. 2012, Ward et al. 2013b).

CONCLUSIONS

Across the ecology, biogeochemistry, and hydrology literature, a need exists to extract meaningful and trustworthy values to describe key transport processes for in- and near-stream transport. Values for these parameters are regularly obtained by solute transport modeling with the OTIS model. However, obtaining reliable estimates for parameters when using this method can be complicated. We addressed this challenge by describing and testing the established OTIS-P model, which provides estimates of uncertainty in OTIS parameters and results, and comparing it with the new OTIS-MCAT tool that provides a global Monte-Carlo-based uncertainty estimation process. Thus, researchers and practitioners have improved ability to arrive at a 'best-fit' parameter set that not only represents the observed data, but corresponds to parameter values (with associated CIs). These more robust parameter sets will support defensible conclusions about stream transport processes on the basis of the OTIS model parameters.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

ACKNOWLEDGEMENTS

Author contributions: ASW and CAK primarily wrote this manuscript and conducted the Monte-Carlo simulations. SJKM developed an early version of the Monte-Carlo framework in Matlab. TW and NM developed the MCAT. BM and RAP collected field data from Stringer Creek. RLR developed the OTIS code. All authors contributed to editing and revising the manuscript and to combining the multiple tools and methods into a usable workflow.

A portion of ASW's time was supported by National Science Foundation (NSF) Grant No. EAR 1331906, by NSF Grant No. EAR 1505309, and by US Department of Agriculture (USDA) Grant No. 2013-67019-21365. This research was also supported in part by Lilly Endowment, Inc., through its support for the Indiana University (IU) Pervasive Technology Institute, and in part by the Indiana METACyt Initiative. The Indiana METACyt Initiative at IU is also supported in part by Lilly Endowment, Inc. RLR was supported by the USGS Toxic Substances Hydrology Program. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US Government. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF, or the authors' institutions. The authors report no conflicts of interest.

LITERATURE CITED

Avanzino, R. J., G. W. Zellweger, V. C. Kennedy, S. M. Zand, and K. E. Bencala. 1984. Results of a solute transport experiment at Uvas Creek, September 1972, California. Open-File Report 84-236. US Geological Survey, Menlo Park, California.

Beer, B. T., P. C. Young, and T. Beer. 1983. Longitudinal dispersion in natural streams. Journal of Environmental Engineering 109:1049–1067.

Bencala, K. E. 1983. Simulation of solute transport in a mountain pool-and-riffle stream with a kinetic mass transfer model for sorption. Water Resources Research 19:732–738.

Bencala, K. E., and R. A. Walters. 1983. Simulation of solute transport in a mountain pool-and-riffle stream: a transient storage model. Water Resources Research 19:718–724.

Beven, K. J., and A. M. Binley. 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6:279–298.

Briggs, M. A., M. N. Gooseff, B. J. Peterson, K. Morkeski, W. M. Wollheim, and C. S. Hopkinson. 2010. Surface and hyporheic transient storage dynamics throughout a coastal stream network. Water Resources Research 46:W06516.

Camacho, L. A., and R. A. González. 2008. Calibration and predictive ability analysis of longitudinal solute transport models in mountain streams. Environmental Fluid Mechanics 8:597–604.

Choi, J., J. W. Harvey, and M. H. Conklin. 2000. Characterizing multiple timescales of stream and storage zone interaction that affect solute fate and transport in streams. Water Resources Research 36:1511–1518.

Courant, R., K. Friedrichs, and H. Lewy. 1928. Über die partiellen Differenzengleichungen der mathematischen Physik. Mathematische Annalen 100:32–74.

Danckwerts, P. 1953. Continuous flow systems. Distribution of residence times. Chemical Engineering Science 2:1–13.

Dennis, J. E., D. M. Gay, and R. E. Walsh. 1981. An adaptive nonlinear least-squares algorithm. ACM Transactions on Mathematical Software (TOMS) 7:348–368.

Donaldson, J. R., and P. V Tryon. 1990. User's Guide to STARPAC: the standards time series and regression package. Internal Report NBSIR 86–3448. National Institute of Standards and Technology, Washington, DC.

Drummond, J. D., T. P. Covino, A. F. Aubeneau, D. Leong, S. Patil, R. Schumer, and A. I. Packman. 2012. Effects of solute breakthrough curve tail truncation on residence time estimates: a synthesis of solute tracer injection studies. Journal of Geophysical Research: Biogeosciences 117:1–11.

Duan, Q., S. Sorooshian, and V. K. Gupta. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology 158:265–284.

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

González-Pinzón, R., A. S. Ward, C. E. Hatch, A. N. Wlostowski, K. Singha, M. N. Gooseff, J. W. Harvey, O. A. Cirpka, and J. T. Brock. 2015. A field comparison of multiple techniques to quantify groundwater–surface-water interactions. Freshwater Science 34:139–160.

Gooseff, M. N., K. E. Bencala, D. T. Scott, R. L. Runkel, and D. M. McKnight. 2005. Sensitivity analysis of conservative and reactive stream transient storage models applied to field data from multiple-reach experiments. Advances in Water Resources 28:479–492.

Gooseff, M. N., M. A. Briggs, K. E. Bencala, B. L. McGlynn, and D. T. Scott. 2013. Do transient storage parameters directly scale in longer, combined stream reaches? Reach length dependence of transient storage interpretations. Journal of Hydrology 483:16–25.

Gupta, A., and V. Cvetkovic. 2000. Temporal moment analysis of tracer discharge in streams: combined effect of physicochemical mass transfer and morphology. Water Resources Research 36:2985–2997.

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

Haggerty, R., S. M. Wondzell, and M. A. Johnson. 2002. Power-law residence time distribution in the hyporheic zone of a 2nd-order mountain stream. Geophysical Research Letters 29:13.

Harvey, J. W., and M. N. Gooseff. 2015. River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. Water Resources Research Special Issue:1–30.

Harvey, J. W., B. J. Wagner, and K. E. Bencala. 1996. Evaluating the reliability of the stream tracer approach to characterize stream-subsurface water exchange. Water Resources Research 32:2441–2451.

Hill, M. C., and C. R. Tiedeman. 2007. Effective groundwater model calibration: with analysis of data, sensitivities, predictions, and uncertainty. 1st edition. Wiley Interscience, Hoboken, New Jersey.

Jackman, A. P., R. A. Walters, and V. C. Kennedy. 1984. Transport and concentration controls for chloride, strontium, potassium

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

and lead in Uvas Creek, a small cobble-bed stream in Santa Clara County, California, U.S.A.: 2. Mathematical modeling. Journal of Hydrology 75:111–141.

Kalbus, E., F. Reinstorf, and M. Schirmer. 2006. Measuring methods for groundwater–surface water interactions: a review. Hydrology and Earth System Sciences 10:873–887.

Kazezyilmaz-Alhan, C. M., and M. A. Medina. 2006. Stream solute transport incorporating hyporheic zone processes. Journal of Hydrology 329:26–38.

Kelleher, C. A., T. Wagener, B. McGlynn, A. S. Ward, M. N. Gooseff, and R. A. Payn. 2013. Identifiability of transient storage model parameters along a mountain stream. Water Resources Research 49:5290–5306.

Kennedy, V. C., A. P. Jackman, S. M. Zand, G. W. Zellweger, R. J. Avanzino, and R. A. Walters. 1984. Transport and concentration controls for chloride, strontium, potassium and lead in Uvas Creek, a small cobble-bed stream in Santa Clara County, California, USA: 2. Mathematical modeling. Journal of Hydrology 75:67–110.

Kerr, P. C., M. N. Gooseff, and D. Bolster. 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.

Koestel, J. K., J. Moeys, and N. J. Jarvis. 2011. Evaluation of nonparametric shape measures for solute breakthrough curves. Vadose Zone Journal 10:1261.

Kumar, A., and D. Dalal. 2010. Analysis of solute transport in rivers with transient storage and lateral inflow: an analytical study. Acta Geophysica 58:1094–1114.

Mason, S. J. K., B. L. McGlynn, and G. C. Poole. 2012. Hydrologic response to channel reconfiguration on Silver Bow Creek, Montana. Journal of Hydrology 438–439:125–136.

Mueller Price, J., B. P. Bledsoe, and D. W. Baker. 2015. Influences of sudden changes in discharge and physical stream characteristics on transient storage and nitrate uptake in an urban stream. Hydrological Processes 29:1466–1479.

Neilson, B. T., D. K. Stevens, S. C. Chapra, and C. Bandaragoda. 2010. Two-zone transient storage modeling using temperature and solute data with multiobjective calibration: 2. temperature and solute. Water Resources Research 46:1–17.

Patil, S., T. P. Covino, A. I. Packman, B. L. McGlynn, J. D. Drummond, R. A. Payn, and R. Schumer. 2013. Intrastream variability in solute transport: hydrologic and geomorphic controls on solute retention. Journal of Geophysical Research: Earth Surface 118:413–422.

Payn, R. A., M. N. Gooseff, B. L. McGlynn, K. E. Bencala, and S. M. Wondzell. 2009. Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States. Water Resources Research 45:W11427.

Pianosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener. 2016. Sensitivity analysis of environmental models: a systematic review with practical workflow. Environmental Modelling and Software 79:214–232.

Poeter, E. P., and M. C. Hill. 1998. Documentation of UCODE, a computer code for universal inverse modeling. Water Resources Investigation Report 98-4080. US Geological Survey, Denver, Colorado.

Rosenberry, D. O., and J. W. LaBaugh. 2008. Field techniques for estimating water fluxes between surface water and ground water. Techniques and Methods 4-D2:128. US Geological Survey, Reston, Virginia.

Runkel, R. L. 1998. One-dimensional Transport with Inflow and Storage (OTIS): a solute transport model for streams and rivers. Water-Resources Investigation Report 98-4018. US Geological Survey, Reston, Virginia.

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.

Runkel, R. L., K. E. Bencala, and B. A. Kimball. 1999. Modeling solute transport and geochemistry in streams and rivers using OTIS and OTEQ. Pages 120–127 *in* D. W. Morganwalp and H. T. Buxton (editors). U.S. Geological Survey Toxic Substances Hydrology Program—Proceedings of the Technical Meeting, Charleston, South Carolina, March 8–12, 1999. Volume 1. Contamination from hardrock mining. Water-Resources Investigations Report 99-4018A. US Geological Survey, Reston, Virginia.

Runkel, R. L., and S. C. Chapra. 1994. Reply to "Comment on 'An efficient numerical solution of the transient storage equations for solute transport in small streams' by Dawes and Short". Water Resources Research 30:2863–2865.

Runkel, R. L., D. M. McKnight, and E. D. Andrews. 1998. Analysis of transient storage subject to unsteady flow: diel flow variation in an Antarctic stream. Journal of the North American Benthological Society 17:143–154.

Schmid, B. H. 2003. Temporal moments routing in streams and rivers with transient storage. Advances in Water Resources 26:1021–1027.

Scott, D. T., M. N. Gooseff, K. E. Bencala, and R. L. Runkel. 2003. Automated calibration of a stream solute transport model: implications for interpretation of biogeochemical parameters. Journal of the North American Benthological Society 22:492–510.

Stream Solute Workshop. 1990. Concepts and methods for assessing solute dynamics in stream ecosystems. Journal of the North American Benthological Society 9:95–119.

Thackston, E. L., and K. B. J. Schnelle. 1970. Predicting effects of dead zones on stream mixing. Journal of the Sanitary Engineering Division 96:319–331.

Wagener, T., L. A. Camacho, and H. S. Wheater. 2002. Dynamic identifiability analysis of the transient storage model for solute transport in rivers. Journal of Hydroinformatics 4: 199–212.

Wagener, T., and J. Kollat. 2007. Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox. Environmental Modelling and Software 22:1021–1033.

Wagener, T., H. S. Wheater, and H. V. Gupta. 2004. Rainfall-runoff modelling in gauged and ungauged catchments. Imperial College Press, London, UK.

Wagner, B. J., and S. M. Gorelick. 1986. A statistical methodology for estimating transport parameters: theory and applications to one-dimensional advective-dispersive systems. Water Resources Research 22:1303–1315.

Wagner, B. J., and J. W. Harvey. 1997. Experimental design for estimating parameters of rate-limited mass transfer: analysis of stream tracer studies. Water Resources Research 33:1731–1741.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Ward, A. S., M. N. Gooseff, T. J. Voltz, M. Fitzgerald, K. Singha, and J. P. Zarnetske. 2013a. How does rapidly changing discharge during storm events affect transient storage and channel water balance in a headwater mountain stream? Water Resources Research 49:5473–5486.

Ward, A. S., R. A. Payn, M. N. Gooseff, B. L. McGlynn, K. E. Bencala, C. A. Kelleher, S. M. Wondzell, and T. Wagener. 2013b. Variations in surface water–ground water interactions along a headwater mountain stream: comparisons between transient storage and water balance analyses. Water Resources Research 49:3359–3374.

Wlostowski, A. N., M. N. Gooseff, and T. Wagener. 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 49:1184–1188.

Wörman, A. 1998. Analytical solution and timescale for transport of reacting solutes in rivers and streams. Water Resources Research 34:2703–2716.

Wörman, A. 2000. Comparison of models for transient storage of solutes in small streams. Water Resources Research 36:455–468.

Wörman, A., A. I. Packman, H. Johansson, and K. Jonsson. 2002. Effect of flow-induced exchange in hyporheic zones on longitudinal transport of solutes in streams and rivers. Water Resources Research 38:1–15.

Young, P. C., and S. G. Wallis. 1993. Solute transport and dispersion in channels. Pages 129–174 *in* K. J. Beven and M. J. Kirkby (editors). Channel network hydrology. John Wiley, New York.

Zand, S. M., V. C. Kennedy, G. W. Zellweger, and R. J. Avanzino. 1976. Solute transport and modeling of water quality in a small stream. Journal of Research of the US Geological Survey 4:233–240.

This content downloaded from 152.003.043.166 on July 18, 2017 08:33:19 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).