WATER RESOURCES RESEARCH, VOL. 49, 5290–5306, doi:10.1002/wrcr.20413, 2013

Identifiability of transient storage model parameters along a mountain stream

C. Kelleher,[1,3] T. Wagener,[1,2] B. McGlynn,[3] A. S. Ward,[4] M. N. Gooseff,[5] and R. A. Payn[6]

Received 15 October 2012; revised 9 May 2013; accepted 10 July 2013; published 4 September 2013.

[1] Transient storage models are widely used in combination with tracer experiments to characterize stream reaches via calibrated parameter estimates. These parameters quantify the main transport and storage processes. However, it is implicitly assumed that calibrated parameters are uniquely identifiable and hence provide a unique characterization of the stream. We investigate parameter identifiability along with the stream conditions that control identifiability for 10 breakthrough curves (BTC) for 100 m pulse injections along Stringer Creek, Montana, USA. Identifiability is assessed through global, variance-based sensitivity analysis of the one-dimensional transport with inflow and storage model (OTIS). Results indicate that the main channel area parameter A and the dispersion coefficient D were the most sensitive parameters and, therefore, likely to be identifiable across all timescales and reaches. Identifiability of transient storage zone size A_s fell into two categories along Stringer Creek. A_s was identifiable for lower elevation regions, corresponding to a constrained valley, higher stream slopes, and in-channel roughness, but not for upper stream regions, corresponding to a wider valley floor, flatter stream slopes, and low roughness. The storage zone exchange parameter α was nonidentifiable across all study reaches. Our results suggest that only some of the processes represented in the model will be relevant and, therefore, identifiable for pulse injection data. As such, calibrated parameter estimates should be accompanied by an assessment of parameter sensitivity or uncertainty. We also show that parameter identifiability varies with stream setting along Stringer Creek, suggesting that physical characteristics directly influence the identification of dominant stream processes.

Citation: Kelleher, C., 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 Resour. Res.*, *49*, 5290–5306, doi:10.1002/wrcr.20413.

1. Introduction

[2] Transient storage models (TSM) are conceptually simple, widely used tools that have been applied to analyze solute transport in streams around the world [e.g., *Bencala and Walters*, 1983; *Runkel*, 1998; *Edwardson et al*., 2003; *Gooseff et al*., 2003a; *Martinez and Wise*, 2003; *Keefe et al*., 2004]. These models route mass in a one-dimensional framework via main-channel processes and exchange with a temporally static, spatially uniform transient storage zone. The transient storage zone is a conceptual representation of slower flow paths in the stream, which includes in-channel areas of slow flow, dead zones, and the hyporheic zone [*Bencala and Walters*, 1983]. The hyporheic zone is a spatially and temporally variable region where channel water interacts with groundwater [*Findlay*, 1995; *Boulton et al*., 1998]. Transient storage models can be used to predict mass transport [*Runkel*, 1998], though recent research has focused on the relationship between model parameters and physical stream characteristics to understand nutrient cycling [*Edwardson et al*., 2003], ecological functioning [*Tate et al*., 1995; *Mulholland et al*., 1997], geomorphological setting [*Wondzell*, 2006; *Gooseff et al*., 2007], and hydrological functioning [*Harvey et al*., 1996; *Gooseff et al*., 2003a]. Specifically, model parameters that characterize transient storage, including hyporheic exchange, are used to assess stream disturbance and biogeochemical functioning [*Edwardson et al*., 2003; *Lautz et al*., 2006; *Lautz and Siegel*, 2007], and to interpret flow path timescales [*Harvey et al*., 1996; *Gooseff et al*., 2003a]. Though multiple formulations of transient storage models exist, the one-dimensional transport with inflow and storage (OTIS) model [*Runkel*, 1998] is the most commonly applied transient storage model, and the focus of this study.

[1]Department of Civil and Environmental Engineering, Pennsylvania State University, University Park, Pennsylvania, USA.
[2]Department of Civil Engineering, University of Bristol, Bristol, UK.
[3]Division of Earth and Ocean Sciences, Nicholas School of the Environment, Duke University, Durham, North Carolina, USA.
[4]Department of Earth and Environmental Science, University of Iowa, Iowa City, Iowa, USA.
[5]Department of Civil and Environmental Engineering, Colorado State University, Fort Collins, Colorado, USA.
[6]Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, Montana, USA.

Corresponding author: C. Kelleher, Division of Earth and Ocean Sciences, Nicholas School of the Environment, Box 90328, Duke University, Durham, NC 27708, USA. (christakelleher@gmail.com)

©2013. American Geophysical Union. All Rights Reserved.
0043-1397/13/10.1002/wrcr.20413

[3] Transient storage models represent stream processes through conceptual parameters which must be inferred indirectly from other data sources because they cannot be feasibly measured at reach scales [*Wagener and Gupta*, 2005]. Within OTIS, model parameters are tied to three processes affecting mass transport: advection, dispersion, and transient storage (Figure 1a). To quantify these processes, reach-representative parameters are estimated by optimizing the model to fit solute concentration data from tracer experiments [*Stream Solute Workshop*, 1990]. Stream tracer injection experiment data and flow measurements are used as inputs and boundary conditions to the transient storage model. The model parameters are calibrated to the observed downstream tracer concentrations, referred to as the solute breakthrough curve (BTC) [*Runkel*, 1998]. The resulting parameter estimates quantify the magnitude of processes inferred for a given stream [e.g., *D'Angelo et al*., 1993; *Runkel*, 2002; *Edwardson et al*., 2003].

[4] The crux of transient storage model calibration is that the parameter estimates are assumed to be uniquely identified from the experimental data (i.e., a unique single set of parameters can be determined that characterizes reach response for a given tracer experiment). This assumption is rarely tested. Lack of parameter identifiability is a problem common to all types of environmental models [*Beven*, 1989; *Beven and Binley*, 1992]. Past studies have suggested that TSM parameter nonidentifiability is a result of equifinality, as different combinations of parameters can reproduce observed BTCs with the same degree of accuracy [*Harvey and Bencala*, 1993; *Harvey et al*., 1996; *Harvey and Wagner*, 2000; *Wagener et al*., 2002]. Despite this recognized problem, researchers and practitioners commonly interpret calibrated parameter values without testing their uniqueness. Only unique, identifiable parameters should be used to formulate reliable conclusions relating parameter estimates from transient storage models to the characteristics of a stream [*Harvey et al*., 1996; *Harvey and Wagner*, 2000].

[5] The aim of this study is twofold: (1) to demonstrate a method for quantifying TSM parameter identifiability and interactions and (2) to use these quantities to link physical stream attributes to the dominance of flow processes. We, therefore, analyzed TSM parameter uniqueness for multiple reaches with varying physical settings along a single stream. The uniqueness of parameter estimates was tested via global, variance-based sensitivity analysis for four OTIS model parameters: the dispersion coefficient D, the main channel area A, the storage zone area A_s, and the exchange rate between the main channel and storage zone α. Global sensitivity analysis, both applied to the entire BTC and at each time step across the BTC, quantifies the individual effects of the parameters and their interactions.

[6] Experiments were performed in Stringer Creek, located within the Tenderfoot Creek Experimental Forest in Montana, USA. We analyzed pulse injection tracer test data from ten 100 m reaches selected along the valley from the initialization of Stringer Creek to its confluence with Tenderfoot Creek, to test how stream characteristics relate to parameter sensitivity. Tracer tests were performed under relatively constant flow conditions and base flow contributions. The 10 reaches selected along Stringer Creek also exhibit variability in vegetation, geology, and hydrology. A change in valley structure at 1200 m separates the upper from the lower portions of Stringer Creek and corresponds to general differences in physical setting (stream slope, valley morphology, riparian vegetation, and the presence of woody debris in the stream channel) [*Payn et al*., 2009, 2012; *Patil et al*., 2013; *Ward et al*., 2013]. Upstream of 1200 m, the valley is less constrained and characterized by wide valley floors and meadows. Downstream of this point, stream slopes and hillslope slopes are greater, the stream channel has more woody debris, and the valley is much more constrained than in the upper region. As the magnitude of stream storage and transport processes are a function of the physical setting within a given stream reach, we

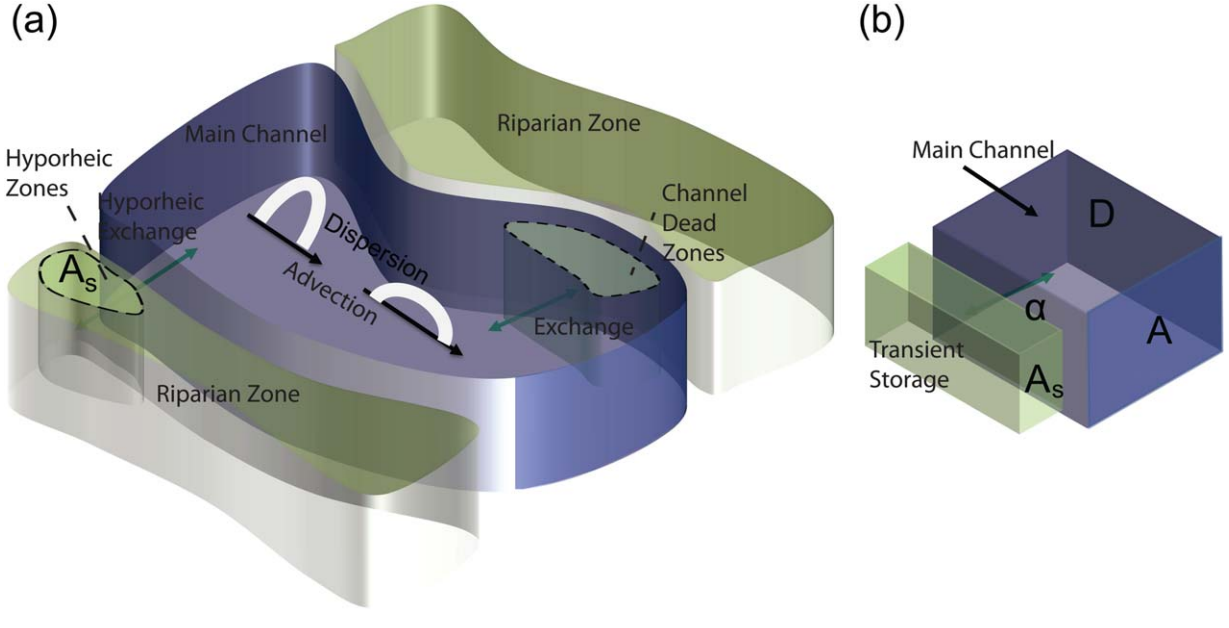

Figure 1. Conceptual diagrams of (a) stream reach processes affecting solute transport and (b) the conceptualization of these processes as implemented in the OTIS model via four calibration parameters.

hypothesize that physical setting likely influences our ability to identify the parameters describing these processes; therefore, highlighting which processes are dominant in a given stream. We seek to identify the conditions (e.g., stream characteristics) associated with identifiable parameters for this series of reaches. Understanding parameter identifiability across these reaches will help clarify whether stream characteristics (e.g., stream slope, amount of in-stream vegetation) are associated with identifiable transient storage parameters.

2. The OTIS Model

[7] The one-dimensional transport with inflow and storage (OTIS) model, the most commonly used transient storage model [*Runkel*, 1998], estimates solute concentrations via mass balance equations for an advective main channel with an adjacent, static storage zone [*Thackston and Schnelle*, 1970; *Bencala and Walters*, 1983]. Within the main channel, solute transport is governed by advection, dispersion, and exchange with the transient storage zone; within the storage zone, solute transport is a function of the rate and quantity of water exchanged with the main channel and the concentration gradient between the transient storage zone and main channel (Figure 1b) [*Runkel*, 1998]. The storage zone represents the lumped effects of water exchange with surface and subsurface exchange, including exchange with the hyporheic zone, channel dead zones, eddies, or pools, and groundwater [*Harvey and Wagner*, 2000; *Payn et al*., 2009]. Readers should refer to *Bencala and Walters* [1983] for the detailed derivation of model equations and *Runkel* [1998] for details on the solution scheme.

[8] The OTIS model operates from the following equations

$$\frac{\partial C}{\partial t} = \underbrace{-\frac{Q}{A}\frac{\partial C}{\partial x}}_{\text{Advection}} + \underbrace{\frac{1}{A}\frac{\partial}{\partial x}\left(AD\frac{\partial C}{\partial x}\right)}_{\text{Dispersion}} + \underbrace{\frac{q_{Lin}}{A}(C_L - C)}_{\text{LateralFlux}} + \underbrace{\alpha(C_s - C)}_{\text{TransientStorage}} \quad (1)$$

$$\frac{\partial C_s}{\partial t} = \underbrace{\alpha\frac{A}{A_s}(C_s - C)}_{\text{TransientStorage}} \quad (2)$$

where t is time (s), x is distance (m), C is the solute concentration in the stream (mg L^{-1}), Q is the volumetric flow rate (m^3 s^{-1}), A is the cross-sectional area of the stream channel (m^2), D is the dispersion coefficient (m^2 s^{-1}), q_{Lin} is the lateral volumetric groundwater inflow length (per length of stream) (m^2 s^{-1}), C_L is the solute concentration in the lateral inflow (mg L^{-1}), C_s is the solute concentration in the storage zone (mg L^{-1}), A_s is the cross-sectional area of the storage zone (m^2), and α is the stream-storage exchange coefficient (s^{-1}) [*Stream Solute Workshop*, 1990; *Runkel*, 1998]. A, A_s, D, and α are typically calibration parameters (Table 1).

[9] Though OTIS only has four calibration parameters, interactions will likely occur among all parameters based on the forms of equations (1) and (2). Parameters are said to interact when their values are interdependent, meaning that their effects on a given model output are not additive and, therefore, not independent [*Saltelli et al*., 2008]. For OTIS parameters, the level and cause of interactions will differ for each experiment. Parameter interactions for the OTIS model are acknowledged in early publications, where stepwise, manual calibration recommends iterative parameter adjustment because parameter values (e.g., A and D) are interrelated [*Stream Solute Workshop*, 1990]. High correlations between model parameters found by later studies also corroborate this point [*Gooseff et al*., 2005; *Wagener et al*., 2002]. In terms of equations (1) and (2), A, present in every term of both equations, is expected to interact with all parameters. A_s and α, which are inversely proportional in equation (2), are also expected to interact. Past work also shows that D and the storage zone parameters A_s and α will interact, either for well-mixed conditions when either process can simulate concentration in the tail or for cases when storage zone processes are insensitive [*Harvey and Wagner*, 2000].

2.1. OTIS Calibration

[10] OTIS parameters are usually estimated by calibration to an experimentally obtained BTC [*Edwardson et al*., 2003; *Harvey and Wagner*, 2000; *Ryan et al*., 2004; *Wondzell*, 2006]. The most widely used optimization method for calibration of the OTIS model is OTIS-P, an adaptive, nonlinear least squares algorithm based on the NL2SOL software package [*Dennis et al*., 1981] and implemented within the Standards Time Series and Regressions Package (STARPAC) [*Donaldson and Tryon*, 1987]. In this package, error is quantified by the residual sum of squares R_{SS}, which is calculated according to

$$R_{ss} = \sum_{k=1}^{n} w_k (O_k - S_k)^2 \quad (3)$$

where O_k and S_k are the observed and simulated kth values, n is the total number of observations, and w_k is the weight for the kth observation [*Wagner and Gorelick*, 1986].

Table 1. Model Parameter Names, Abbreviations, and Units[a]

				Bounds	
Analysis	Parameter	Units	Symbol	Lower	Upper
1, 2, and 3	Dispersion coefficient	L^2T^{-1}	D	0.001	1
1, 2, and 3	Storage zone area	L^2	A_s	0.001	0.01
1, 2, and 3	Surface-subsurface exchange	T^{-1}	α	10^{-5}	0.001
1	Channel area	L^2	A	0.001	1
2 and 3	Channel area	L^2	A	See Table 3	

[a]The values for both the wide and constrained runs are included. Refer to Table 3 for the constrained channel area parameter ranges for each reach.

Weights are not required, but can be included to remove the emphasis on fitting to larger values at the BTC peak, given that the squared error values are minimized [*Wagner and Gorelick*, 1986].

[11] R_{SS} is minimized iteratively within the OTIS-P framework. Parameter estimates are updated based on approximates of the Hessian (second derivative) matrix for the residual sum of squares and partial derivatives of model-predicted concentrations, which are constrained by a parameter trust region that grows and shrinks based on the level of R_{SS} fit [*Donaldson and Tryon*, 1987]. The algorithm converges to a final parameter set when there is a minimal change in (1) parameter value or (2) objective function value, based on user-defined thresholds for both criteria [*Donaldson and Tryon*, 1987; *Runkel*, 1998].

[12] Like all local search algorithms, convergence to a final parameter set is sensitive to the choice of an initial parameter set, and can lead to a local instead of global minimum solution [*Dennis et al*., 1981]. OTIS-P may not converge to a final solution for all applications. Singular convergence can occur when the data or complexity of an application cannot be used to estimate all transient storage parameters [*Donaldson and Tryon*, 1987; *Ryan et al*., 2004]. When OTIS-P does not converge, the estimated parameter values are not reliable, and should not be reported. Beyond assessing convergence, OTIS-P provides statistical information on parameter certainty calculated within the neighborhood of the parameter estimate. These values, though available as output from an OTIS-P run, are rarely used or reported in the literature. Statistics from OTIS-P are based on the variance-covariance matrix for the solution achieved by the nonlinear least squares optimization algorithm. The variance-covariance matrix represents an approximation of parameter uncertainty within the neighborhood of the solution [*Donaldson and Tryon*, 1987]. Though correlation coefficients from the variance-covariance may give an indication of uniqueness, they are a local, not global, assessment [*Hill and Tiedeman*, 2007]. Therefore, these values cannot be used to infer global behavior or uniqueness across the feasible parameter space. One approach to assess parameter uniqueness for similar models is usually tested by running the least-squares optimization multiple times from different initial parameter values; if the algorithm converges to the same parameter set, the set is said to be unique [*Hill and Tiedeman*, 2007]. While this method has a better chance of assessing parameter uniqueness than a single optimization run, it still does not guarantee complete exploration of the parameter space, and any statistics will be local values.

2.2. OTIS Parameter Uncertainty and Sensitivity

[13] Environmental models are simplified mathematical representations of physical systems used to describe observed phenomena [*Wagener and Gupta*, 2005]. Parameter values in these models are often conceptualized physical processes or system properties that cannot be directly measured, and therefore, must be estimated via calibration. Ideally, calibration produces a single set of parameter values that best reproduce the observed data [*Cobelli and DiStefano*, 1980; *Sorooshian and Gupta*, 1983]. However, it is also possible that more than one parameter set will reproduce the observations equally well with respect to a particular performance measure [*Cobelli and DiStefano*, 1980]. This condition is sometimes referred to as equifinality [*Beven*, 1989; *Beven and Binley*, 1992]. When equifinality occurs, it is usually the result of parameters that are insensitive [*Johnston and Pilgrim*, 1976] and/or interactive [*Ibbitt and O'Donnell*, 1971], and therefore nonidentifiable. Parameter equifinality and identifiability are especially problematic for complex models with many parameters [*Beven and Binley*, 1992], though equifinality can also occur for simple models with fewer parameters [*Ibbitt and O'Donnell*, 1971; *Pickup*, 1977; *Beven*, 1989; *Wagener et al*., 2002]. Parameter identifiability is affected by the model structure, data uncertainty, and the combination thereof [*Sorooshian and Gupta*, 1983; *Beck*, 1987; *Beven*, 2008]. The portion of the parameter space being sampled can also influence the interpretation of identifiability and sensitivity [*Saltelli et al*., 2006].

[14] Past studies have shown that OTIS model parameters can be nonidentifiable [*Harvey and Bencala*, 1993; *Harvey et al*., 1996; *Wagner and Harvey*, 1997; *Harvey and Wagner*, 2000; *Wagener et al*., 2002]. Sensitivity analyses addressing nonidentifiability have concluded that sensitivity for each of the parameters occurs for different parts of the BTC. In these studies, the main channel area parameter A is usually well identified, while identifiability for D varies [*Wagener et al*., 2002]. A_s and α are typically the least identifiable parameters [*Harvey et al*., 1996; *Wagner and Harvey*, 1997].

[15] Beyond testing parameter identifiability, these studies have attempted to answer the question, *What conditions control identifiability?* As A and D are typically well identified, research to assess controls on identifiability has focused on the storage zone parameters A_s and α. These controls can be explained in terms of the Damköhler number (DaI), defined as the ratio of average advective transit time to the trnaisnet storage zone interaction timescale and given by *Bahr and Rubin* [1987] as

$$DaI = \underbrace{\left(\alpha\left(1+\frac{A}{A_s}\right)\right)}_{\text{StorageZone}} \cdot \underbrace{\frac{L}{v}}_{\text{Advection}} \tag{4}$$

[16] Average advective transit time is a function of reach length (L), which depends on the experimental design, and the average reach water velocity (v), both of which can be determined prior to performing a tracer experiment [*Harvey and Wagner*, 2000; *Wondzell*, 2006]. Storage zone interaction time is a function of calibrated parameter values for A, A_s, and α, which must be estimated after performing a tracer experiment. *Wagner and Harvey* [1997] previously investigated parameter sensitivities via constant rate injections. They found that A_s and α were best identified for DaI values near 1 and 0.1 respectively, representative of a ratio where the timescale of advection is reasonably approximated by the timescale of transient storage zone reactivity. Nonidentifiability occurs when one process overwhelms the other, governed by the following conditions:

[17] 1. $DaI >> 1$, which occurs when storage zone exchange rates are much greater than advective velocities. This can occur for slow moving streams with rapid transient storage exchange. For this condition, the tracer mixes

quickly in the stream and storage zone, generating a tail signal that can be matched using D or using A_s and α [*Harvey and Wagner*, 2000]. The influences of the parameters are indistinguishable from one another therefore interact, resulting in parameter interaction and nonidentifiability.

[18] 2. $DaI << 1$, which occurs when advective velocities are much larger than storage zone exchange rates. This occurs for streams with high velocities and lower exchange rates, resulting in long time scales of transient storage exchange [*Wagner and Harvey*, 1997]. For this condition, the model is more sensitive to the main-channel processes and less sensitive to the transient storage zone processes and its parameters A_s and α [*Harvey et al*., 1996].

[19] Both conditions can also be influenced by the experimental design, via constraining the reach length to *DaI* values, corresponding to recommendations from *Wagner and Harvey* [1997] and following *Harvey and Wagner* [2000], [*Wagner and Harvey*, 1997; *Harvey and Wagner*, 2000], or through the sampling frequency and the duration of the experiment, which control the length and timescales of flowpaths that can be detected [*Harvey et al*., 1996; *Harvey and Wagner*, 2000; *Gooseff et al*., 2003a; *Scott et al*., 2003; *Wondzell*, 2006]. There is no recommendation for a specific *DaI* range that constitutes certain versus uncertain parameter sets; these guidelines instead offer an interpretation of how transport processes and experimental design will influence parameter uncertainty. The *DaI* recommendations from *Wagner and Harvey* [1997] are applicable to high gradient streams for either constant rate or pulse injections, though the authors show that parameter sensitivity (quantified in terms of the coefficient of variation) is greater for a constant rate injection with sampling on the rise, plateau, and fall of the breakthrough curve as compared to a breakthrough curve generated from a pulse injection.

[20] Other researchers have applied the *DaI* recommendations from *Wagner and Harvey* [1997] to assess parameter reliability and sensitivity in their own studies [*Fellows et al*., 2001; *Martinez and Wise*, 2003]. However, this approach again assumes that calibrated parameter values are unique. If the *DaI* value indicates high uncertainty in parameter estimates, the estimated parameter values cannot be used, despite the fact that a tracer experiment has already been performed. Confidence in this value is especially limited when it is used with uncertain parameters, as it implies that the *DaI* value is also uncertain because it is based on uncertain parameter estimates. For these reasons, *DaI* is not used directly in this study. However, we do believe that the conditions described above, where *DaI* is much greater or much less than 1, should represent cases where certain transient storage parameters will not be identifiable.

3. Study Area and Experimental Data

[21] This study uses BTC data from stream tracer tests along Stringer Creek, which drains a subcatchment of the Tenderfoot Creek Experimental Forest (TCEF), Montana, USA (Figure 2). TCEF has a predominantly continental climate, with an average annual precipitation of 840 mm. Precipitation occurs mainly as snow, and the majority (70%) falls between November and May [*Farnes et al*., 1995; *Jencso et al*., 2009]. Streamflow is snowmelt dominated, reaching a peak between late April and early June during the melt period. Flow generally declines through the summer and early fall [*Jencso et al*., 2009; *Nippgen et al*., 2011].

[22] Stringer Creek is a generally gaining, montane headwater stream. We refer to locations along Stringer Creek by valley distance from the gauge near the confluence with Tenderfoot Creek, where 0 m is at the gauge and 2700 m is near the origin of flow during base flow conditions. Physical characteristics of the stream vary along the valley, predominantly due to the gain in stream flow and to a change in underlying bedrock at approximately 1600 m [*Payn et al*., 2009, 2012]. With reference to this location, granite-gneiss bedrock underlies the valley downstream and sandstone bedrock underlies the valley upstream. The transition in bedrock corresponds to a substantial change in valley structure near 1200 m. Upstream of this location, down-valley slope averages 6%. Relative to downstream regions, the upstream valley is less constrained (wider valley floor) and hillslopes have lower relief. Meadows dominate the riparian landscape in the upstream valley, and there is less large woody debris in the stream channel compared to downstream regions. In contrast, the down-valley slope downstream of 1200 m averages 9%. Downstream valleys are more constrained (narrower valley floor) with higher relief hillslopes, and there are more riparian trees and more large woody debris in the stream channel.

[23] Our analysis was based on tracer test data collected in August 2005 using instantaneous tracer releases. *Payn et al*. [2009] conducted individual tracer tests every 100 m along the valley. Salt (NaCl) was used as a conservative tracer, and tracer concentrations were measured by in situ electrical conductivity measurements. Mixing lengths across the reaches varied between a valley distance of 5–30 m and experiments were planned and conducted to limit the possibility of incomplete mixing by considering stream channel structure as well as repeating experiments in which incomplete mixing was observed in the breakthrough curve [*Payn et al*., 2009]. Full details of the experimental design are available in *Payn et al*. [2009] and *Payn et al*. [2012], and other analysis with the 100 m data sets are reported in *Ward et al*. [2013].

[24] We selected 10 of the 26 BTCs for 100 m reaches collected along the 2700 m length of valley (Figure 2). Reaches were selected to span the entire length of the valley, to include a characteristic set of BTC shapes and durations, and to include a representative set of varying physical features along the valley (Figure 3). Each study reach is referenced by the valley distance at the downstream end of the reach, and reaches include: 2500, 2400, 2100, 1300, 700, 600, 400, 300, 200, and 100 m. These locations reflect the longitudinal spatial variability in stream flow, velocity, channel area, elevation, slope, gross hydrologic loss/gain, and net loss/gain (Figure 2b). Physical stream setting encompasses the variability in physical characteristics displayed in Figure 2b. These characteristics were quantified for each reach from the tracer tests or from a light detection and ranging (LIDAR) derived 1 m digital elevation model (DEM) of TCEF (Table 2). We compare model analyses from BTCs among these sites to explore how parameter identifiability changes with the variability

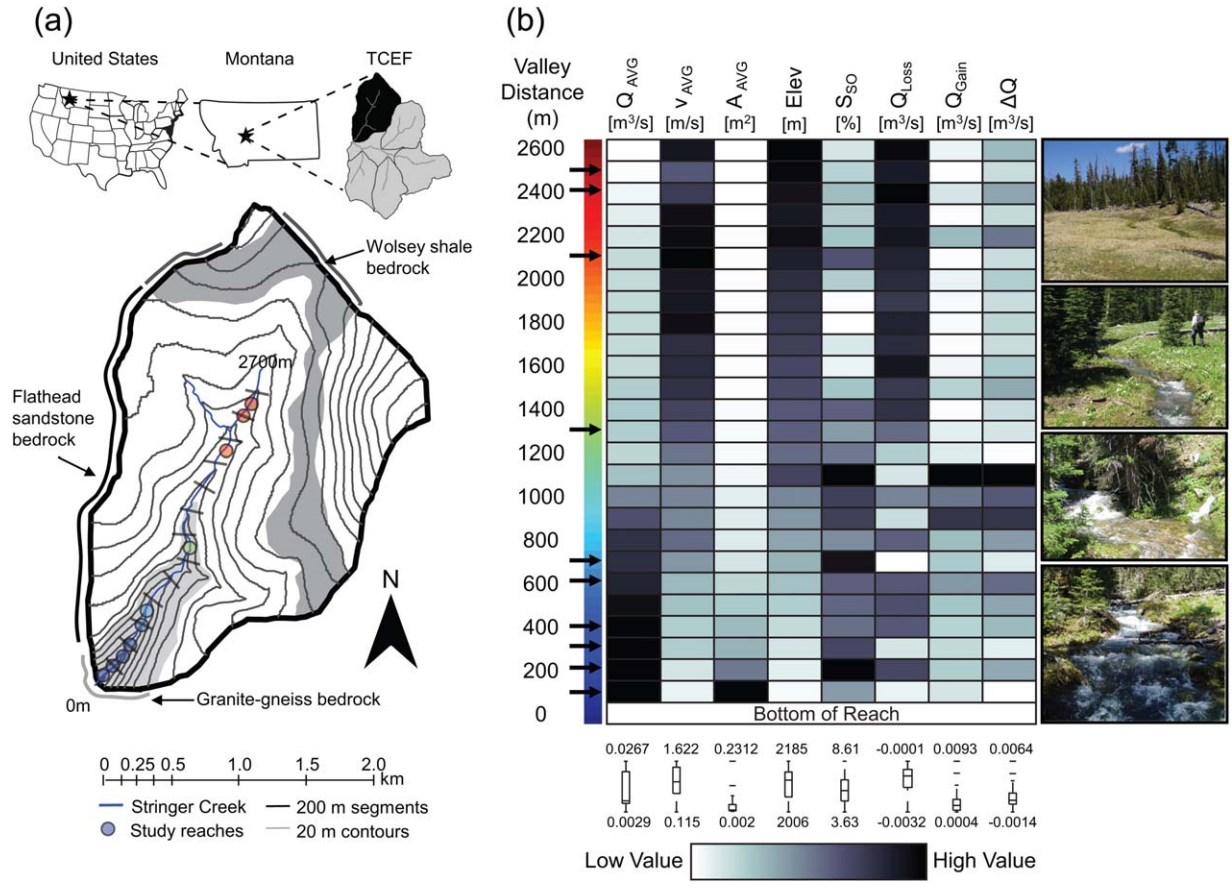

Figure 2. The study area and 10 study reaches from a plan view. The map shows the changes in geologic type, elevation contours, and Stringer Creek (modified from *Payn et al*. [2009]). Color indicates the reach location, and corresponds to the color bar in Figure 2b. (b) Values of watershed metrics from the top of the reach (2700 m) to its confluence with Tenderfoot Creek (0 m). Reach locations are all indicated using valley distances, and do not include sinuosity in the active channel [*Payn et al*., 2012]. Color indicates magnitude, normalized to the minimum and maximum value across the reaches for each metric. Arrows indicate the 10 reaches included in the analysis. Photographs correspond to approximate locations along Stringer Creek and were taken by R. Payn and C. Kelleher.

in these characteristics. It is generally regarded that physical stream setting, including the influence of differences in geology [*Morrice et al*., 1997], geomorphology [*Harvey and Bencala*, 1993; *Gooseff et al*., 2007], streamflow [*Harvey et al*., 1996], and obstructions to sediment transport [*Gooseff et al*., 2003b; *Ensign and Doyle*, 2005], has some influence on transient storage parameter estimates. Instead, we seek to move beyond the scope of these past studies, to understand the relationship between physical setting and parameter identifiability. The Stringer Creek system provides a range of settings to test this relationship.

4. Methods

4.1. Sobol' Sensitivity Analysis

[25] Sensitivity analysis assesses how variability in model output is ascribed to variability in model factors [*Saltelli et al*., 2004]. Model factors include, but are not limited to, model inputs, initial states, parameters, and other user-defined properties that will influence the model output. We quantify sensitivity using Sobol' sensitivity analysis, a global, variance-based technique [*Sobol'*, 1993, 2001]. Past studies indicate that Sobol's method effectively characterizes sensitivities for models of low to intermediate levels of complexity [*Fieberg and Jenkins*, 2005; *Tang et al*., 2007; *Saltelli et al*., 2008]. Sobol's method [*Sobol'*, 1993, 2001] ascribes the total variance in model output to variance in model factors. Model output is often measured by an objective function which computes the difference between observed and simulated output, though other model output can be used. The total variance in the model output $D(y)$ is divided into contributions from individual factors and factor interactions, represented as

$$D(y) = \sum_i D_i + \sum_{i<j} D_{ij} + \sum_{i<j<k} D_{ijk} + D_{12\ldots m} \qquad (5)$$

where y is the distribution of model output, D_i is the measure of the sensitivity to model output y due to the ith component of a given input factor p_i, D_{ij} is the portion of output variance due to the interaction of parameters p_i and p_j, and m represents the total number of factors being investigated.

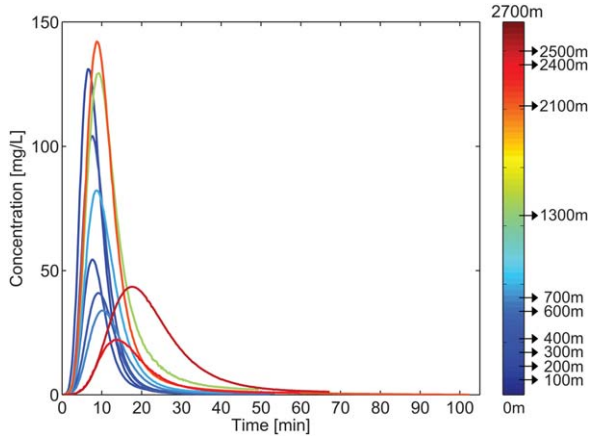

Figure 3. The 10 study breakthrough curves, obtained experimentally from tracer tests. Color indicates the reach location along Stringer Creek, with 0 m being the bottom of the stream, and 2700 m being the top. Reach colors correspond to locations in two. Indicated locations are all measured by valley distances.

From equation (5), first- and total-order Sobol' sensitivity indices are calculated for each model factor in terms of the amount of total output variance reduced by that factor plus its interactions with other factors. This can be written as

$$\text{First-order index} : S_i = \frac{D_i}{D} \tag{6}$$

$$\text{Total-order index} : S_{T_i} = 1 - \frac{D_{\sim i}}{D} \tag{7}$$

where $D_{\sim i}$ represents the average variance from all factors but p_i. S_{Ti} represents the variance in the output due to the factor p_i and its interactions. It represents the sum of an interactions index and the first-order index S_i, which measures the effect of factor p_i alone. The ratio of first-order to total-order index is important to factor identifiability, while the magnitude of the total-order index determines sensitivity. Factors with high first- and total-order indices are likely to be identifiable. Factors with a low first-order index but high total-order index are dominated by interactions and are likely nonidentifiable. The magnitude of the total-order index is high for a sensitive parameter and low for an insensitive parameter. Insensitive parameters, defined as having a low total-order index, are nonidentifiable.

[26] Factor sets are sampled from user-defined ranges according to the Sobol' quasi-random sampling sequence [*Sobol'*, 1967; *Bratley and Fox*, 1988; *Sobol'*, 1994]. Sobol' indices are calculated for a total of N model runs and corresponding model outputs for each run. *Saltelli* [2002] recommends

$$N = s(2m + 2) \tag{8}$$

where m is the number of factors being investigated and s is the quasi-random sample size. *Sobol'* [1967] suggests values of s that satisfy $s = 2^b$ where b is an integer to achieve stability in Sobol' indices.

4.2. Objective Functions

[27] The Sobol' method quantifies indices based on the variance in model output. The root-mean squared error (RMSE) is calculated to assess sensitivity to the entire simulated and observed time series'. RMSE is calculated as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(O_k - S_k)^2} \tag{9}$$

where O_k and S_k are the observed and model simulated values at each time step k and n is the total number of time steps. RMSE is a slightly different but equivalent form of R_{SS} (equation (3)), used commonly in OTIS model calibration [*Runkel*, 1998]. As RMSE is not a normalized error metric, we also report model performance in terms of the coefficient of determination r^2. The r^2 value, which assesses the fraction of variability in the observed BTC that is described by variability in the simulated BTC, is directly comparable across reaches. This is calculated as

$$r^2 = 1 - \frac{\sum_{l=1}^{n}(O_l - S_l)^2}{\sum_{l=1}^{n}(O_l - \overline{O})^2} \tag{10}$$

where $\overline{(O)}$ is the mean of the observed time series [*Devore*, 2000].

Table 2. Names, Abbreviations, Units, and Calculation Methods for Site Characteristics[a]

Characteristic	Abbreviation	Units	Calculation
Average flow	Q_{avg}	m^3s^{-1}	Averaged streamflow at beginning and end of the reach
Flow change	ΔQ	m^3s^{-1}	The difference between the flow at the beginning and end of each reach
Average velocity	v_{avg}	ms^{-1}	Reach length divided by the time between the pulse peak at the injection point and the pulse peak at the downstream measurement location
Average area	A_{avg}	m^2	Average flow divided by average velocity
Elevation	*Elev*	m	Average value from the DEM across the reach
Stream slope	S_o	mm^{-1}	The difference in DEM elevation from the point of injection to the point of measurement, divided by the reach length
Average loss	Q_{loss}	m^2s^{-1}	The average of the minimum and maximum loss, calculated from [*Payn et al.*, 2009]
Average gain	Q_{gain}	m^2s^{-1}	The average of the minimum and maximum gain, calculated from [*Payn et al.*, 2009]

[a]Site characteristics are visualized in Figure 2.

4.3. Fuzzy Performance

[28] Sensitivity analysis sampling strategies can include combinations of unreasonable model factors which generate poorly performing simulations that would normally be removed from an uncertainty analysis or calibration. When these solutions are included, the model factors which produce the largest errors have the highest sensitivity indices. This obscures the influence of other model factors, leading to a false assessment of model controls. Fuzzy metrics are alternative objective functions which allow users to determine which simulations should be included or excluded for an analysis [*Pappenberger and Beven*, 2004]. Applying a fuzzy performance metric evaluates the sensitivity of parameters to simulations which approximate the observations within an envelope of error, minimizing the effect of simulations outside of this envelope.

[29] The concept of fuzzy sets was first introduced by *Zadeh* [1965] as a method for expanding binary to gradational membership. Values for this objective function M vary between 0 (nonmember) and 1 (member). Previous applications of this method [*Aronica et al*., 1998; *Pappenberger and Beven*, 2004; *Cloke et al*., 2008] determine values for M by comparing simulation values to two error widths $E1$ and $E2$ from the observed values. As seen in Figure 4, if the simulation value

[30] 1. is within a width of E_1 from the observations, $M = 1$,

[31] 2. is outside a width of E_2 from the observations, $M = 0$,

[32] 3. is between a width of E_1 and E_2, M is linearly interpolated between 0 and 1 based on the distance of the simulation value from E_1, normalized to the difference between E_1 and E_2.

[33] Typically, values and widths are selected based on measurement error/uncertainty or user-specific knowledge [*Pappenberger and Beven*, 2004; *Cloke et al*., 2008]. In this study, fuzzy metric performance was calculated at each time step for each model simulation (Figure 4). Because the observations span multiple orders of magnitude, we constructed the membership envelope in $\log_{10}$ space. We assigned values of $E_1 = 0.5$ and $E_2 = 1$, so that any simulated values greater or less than an order of magnitude from the observed value are rejected. These thresholds are wide so that only very poor solutions are excluded. Before calculating the fuzzy metric, we additionally removed all unstable model simulations, identified as oscillating simulations with more than one peak. Any peaks beyond the maximum concentration were classified as oscillating solutions if the difference between the peak value and the value at the previous and next time step was greater than 0.0048 mg/L, the smallest value of detection for concentration across the 10 reaches. Fuzzy metric values for an unstable simulation were set to $M = 0$ across the entire BTC.

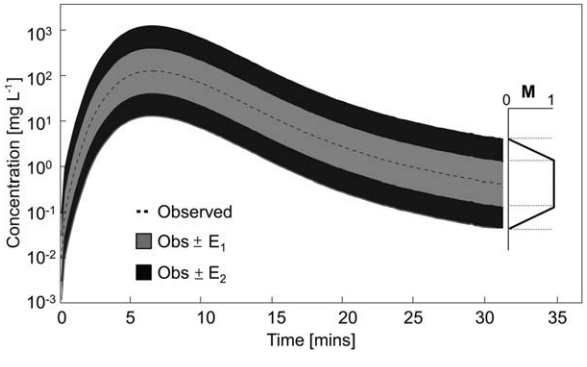

Figure 4. The fuzzy metric objective function for the BTC. The dotted black line indicates the observed BTC. The gray-shaded region indicates behavioral samples, where $M = 1$. The black region indicates acceptable values approximated to a value M between 0 and 1. The white region represents nonbehavioral space, where simulated values were assigned $M = 0$.

4.4. Model Setup for Current Study

[34] The OTIS model was implemented for a conservative solute, steady flow, and a constant dispersion coefficient for each river reach. Solute mass input was a measured time series, which was recorded during the tracer injection. The model was run at a 2 second time step and Sobol' analysis was performed at 6 s intervals. Q_{in} was set to the flow at the location of the solute injection, which was calculated via dilution gaging following *Payn et al*. [2009]. Lateral inflow (q_{Lin}) and lateral outflow (q_{Lout}) were calculated from water balance data from *Payn et al*. [2009] based on the net change in flow over the reach and tracer mass recovery. Lateral inflow can be treated as a calibration parameter [*Wagner and Harvey*, 1997; *Gooseff et al*., 2003a]. In this application, these values were constrained with independent estimates based on the evidence of both gross gains and losses in the study reaches. C_L was set to zero, as NaCl concentrations were determined using electrical conductivity measurements which were corrected for background values [*Payn et al*., 2009]. A description of the tracer experiments can be found in *Payn et al*. [2009].

[35] In this study, parameter identifiability is determined based on parameter sensitivities. A parameter is likely to be nonidentifiable for two cases:

[36] 1. It has a low total-order index, in which case the parameter is insensitive, or,

[37] 2. It has a high total-order index but a low first-order index, indicating that the parameter influences model output through interactions with other parameters.

[38] A parameter is likely to be identifiable if it has a high first-order and total-order index, indicating that it influences model output through changes in its value alone. *Tang et al*. [2007] used a threshold index of 0.1 and 0.01 for a complex, distributed watershed model to distinguish highly sensitive and sensitive parameters. We distinguish between high and low sensitivities as those indices greater or less than 0.25, given that our model is parsimonious with only four parameters. It is also important to note that a model parameter can be best identified for a value of zero, indicating that it is the absence of a process that is important [*Wagener et al*., 2003]. For an application where this does occur, the parameter would be identifiable, corresponding to a high first-order sensitivity index.

[39] Sensitivity indices were computed for A, A_s, D, and α (Table 1). Results are reported for three analyses, each with varying parameter ranges and objective functions. For the first analysis, wide ranges were used for all parameters. Wide ranges were initially used because sensitivity

analyses are dependent on the associated parameter ranges. Sensitivity indices were calculated based on RMSE for the entire BTC. Ranges, specified in Table 1, were defined from an understanding of the system and from values used in previously performed sensitivity analysis studies [*Wagner and Harvey*, 1997].

[40] For the second analysis, parameter bounds were kept wide for A_s, D, and α (Table 1), but the ranges for A were narrowed. The results from the first analysis were used to constrain A, based on the best 1000 runs (quantified by RMSE and r^2) (Table 3). In the first analysis, sensitivity indices for A were much higher than for any other parameter, so A was constrained to test how the sensitivities for the other three parameters would change. The top 1000 runs for RMSE and r^2 produced similar ranges for A for all reaches except at location 2500 m. Again, indices were calculated from RMSE of the entire BTC. For the final analysis, fuzzy metric values were calculated from the simulations produced by the second analysis. Sobol' indices were computed for the fuzzy metric values at each time step in the BTC.

5. Results

5.1. Sensitivity Indices for the Complete BTC

[41] Sensitivity indices for the entire BTC were computed for all 10 reaches (Figure 5). Main channel area A is the most sensitive parameter for the first 2 analyses and across all 10 reaches. For the first analysis, where wide bounds were used for all parameters, the total-order indices for A were between 0.83 and 1.12 (Figure 5a). D was slightly sensitive, and A_s and α had total-order sensitivities below 0.1 for all reaches. The differences between first and total-order indices also indicate that interactions accounted for 34 and 100% of the total-order indices for D. A had the highest first-order indices across all parameters. First-order values for A represented 63–89% of the total-order index. For the second analysis, A and D were most sensitive, with total-order indices between 0.38 and 1.17 for A and 0.25 and 0.70 for D (Figure 5b). Total-order indices for A_s and α were much smaller than values for A and D, but larger than the total-order values for A_s and α in the first analysis. A_s was slightly sensitive and had total-order index values between 0.08 and 0.19 across all reaches. Interactions accounted for nearly all of the total-order indices for A_s (72–89%) and α (87–100%), and for variable portions of the total-order indices for D (29–82%) and A (19–64%).

[42] To illustrate that the parameter ranges did generate parameter sets that approximated the observations, we reported the level of parameter fit for the top 1000 runs for both wide and constrained analyses in Table 3. The top 1000 runs achieved high levels of fit for all reaches for both analyses. The best values for both analyses and across all reaches were 0.99. The lowest r^2 values for the top 1000 runs ranged from 0.578 to 0.853 for the first analysis to 0.693–0.981 for the second analysis. These levels of fit demonstrate that model runs provide reasonable predictions of the observed data.

5.2. Fuzzy Performance Sensitivity Indices

[43] Though indices for RMSE were computed for the first and second analysis in a moving window framework, the resulting indices were similar through time and across all parameters and locations. Thus, the fuzzy metric was implemented to filter poor simulations and to avoid misrepresentation of parameter sensitivities. In general, first-order fuzzy metric sensitivity indices were small, indicating that parameter interactions control output variance (Figure 6). Total-order indices across all reaches were highest for A and D. Though the patterns of periods when parameters were most sensitive differ between reaches, D and A were more sensitive than A_s and α in the rising limb, the peak of the BTC, and occasionally through the falling limb. In contrast, A_s and α were more sensitive on the falling limb and in the tail of the BTC. Applying the fuzzy metric also removed all unstable solutions, which occurred for only three reaches (100, 200, and 2500 m) and represented less than 0.3% of simulations.

[44] We classify high sensitivities as those greater than 0.25, and low sensitivities as those less than 0.25, corresponding to sensitive and insensitive parameters. Across all reaches, first and total-order sensitivities indicated high sensitivity to A and D. A_s and α had high total-order sensitivities for all reaches but 2500 m, and 2400 m, and 2500 m, respectively. In contrast, A_s had high first-order sensitivities for reaches near the bottom of the stream (100, 200,

Table 3. Model Fits for the Top 1000 Runs for the First, Second, and Third Analyses[a]

		Analysis 1				Analysis 2 and 3					
		r^2, Top 1000		RMSE, Top 1000		Channel Area, A (m^2)		r^2, Top 1000		RMSE, Top 1000	
Location	Reach Length	Best	Worst	Best	Worst	Lower	Upper	Best	Worst	Best	Worst
100	103.89	0.999	0.807	1.93	22.56	0.192	0.277	0.999	0.970	1.00	9.98
200	106.58	0.999	0.821	0.58	7.68	0.193	0.272	0.999	0.974	0.50	3.20
300	106.39	0.999	0.853	1.08	15.68	0.180	0.256	0.999	0.981	1.23	6.34
400	106.75	0.998	0.835	0.60	5.80	0.223	0.322	0.999	0.972	0.43	2.58
600	109.4	0.998	0.836	0.49	5.05	0.211	0.314	0.999	0.971	0.43	2.29
700	108.18	0.998	0.818	1.30	13.34	0.180	0.269	0.999	0.970	1.26	5.89
1300	109.86	0.995	0.647	4.31	28.09	0.077	0.164	0.998	0.909	3.06	15.80
2100	124.67	0.997	0.578	2.91	23.20	0.057	0.110	0.996	0.916	2.21	11.95
2400	128.64	0.984	0.602	1.21	6.23	0.005	0.167	0.998	0.764	0.32	4.83
2500	128.09	0.994	0.702	1.57	12.19	0.042	0.620	0.995	0.693	2.14	12.41

[a]The constrained ranges for the main-channel area parameter A were taken from the top 1000 runs by RMSE and r^2 for the first analysis. These values were computed for each reach.

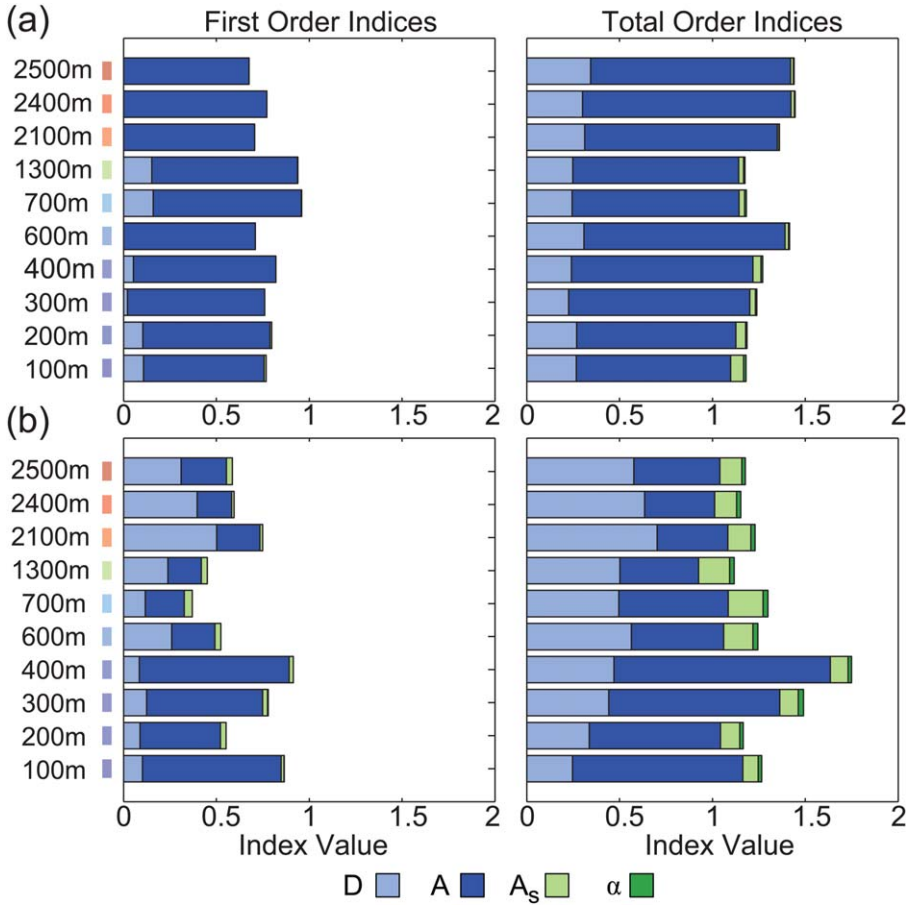

Figure 5. First- and total-order indices for RMSE for the entire breakthrough curve, for (a) the first analysis, with wide parameter ranges and (b) the second analysis, with parameter ranges constrained for A based on the first analysis. Reach colors correspond to locations in Figure 2. Index colors correspond to parameters.

300, 400, 600, and 700 m) and the reach at 2100 m. First-order sensitivities for α were low across all reaches, indicating that α influences model output through interactions. Given these patterns, first-order indices were grouped into two different types of responses. The reaches near the top of the stream (1300, 2100, 2400, and 2500 m) had higher first-order indices for D and A, whereas A_s, when sensitive, influenced model output through interactions. The exception is at location 2100 m, where A_s was sensitive through the tail. For the reaches toward the bottom of the stream (100, 200, 300, 400, 600, and 700 m), first-order sensitivities were highest for A_s and A in the BTC tail, and D on the rising limb. For these reaches, interactions were largely responsible for all parameter sensitivities.

6. Discussion

6.1. The Importance of Parameter Ranges and Performance Objectives in Assessing Parameter Sensitivities

[45] Though the magnitudes of indices for the entire BTC varied slightly across the four parameters, general controls across all reaches remained the same independent of whether or not A is constrained. The main channel area parameter overwhelmingly controlled the largest portion of model output variance for both analyses, indicating that it is the main control on output variance in advection dominated systems (Figure 5). A also had a high first-order index, indicating that the parameter has a strong independent influence on model output, though it still interacted to some extent with other parameters. The sensitivity indices of other parameters increased when A was constrained (Figure 5b). D and A still controlled the majority of model output variance across all reaches. Sensitivity of the output to A_s was limited to interaction with other parameters.

[46] These results suggest that the model output for the entire BTC provided the most information about the cross-sectional area of the advective channel. A is likely identifiable and will be accurately estimated using optimization tools such as OTIS-P. These results are a function of the performance metric RMSE, which amplifies large errors and is, therefore, primarily influenced by errors in BTC peak timing and magnitude. Calibration with RMSE, or by extension R_{SS}, will therefore be sensitive to A, and will fit the BTC peak. High sensitivity to A across all reaches is also a function of stream setting, given the moderate to steep stream slope gradient (average stream slopes from a 1 m DEM ranging from 2.3 to 6.3%). The first- and total-order sensitivity indices for A_s and α were low. As such, A_s and α are likely to be nonidentifiable. Calibration to the

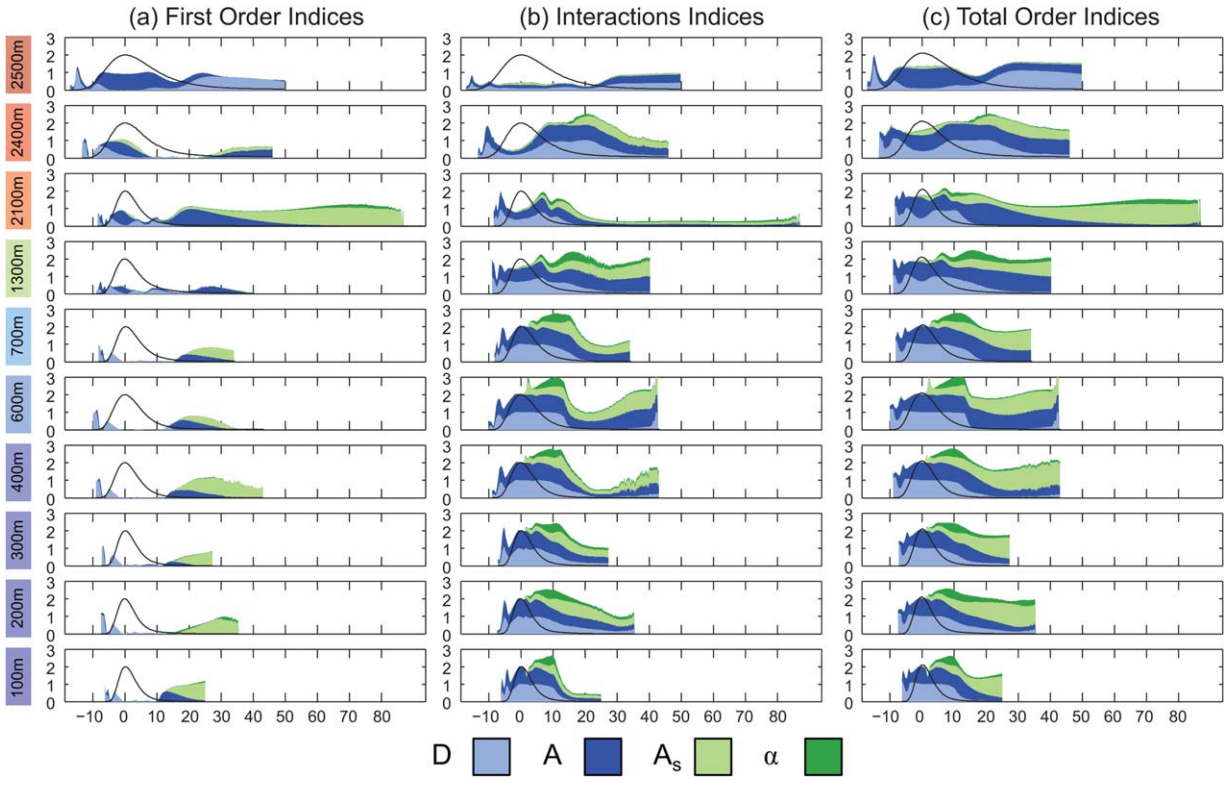

Figure 6. Sensitivity indices for the fuzzy metric objective function applied to the area-constrained simulations, with indices plotted as a function of time. Indices are (a) first order, (b) interactions, and (c) total order. Color corresponds to the parameter.

entire BTC will provide little unique information about transient storage for pulse injections in advective systems.

[47] Sensitivity indices calculated for the fuzzy metric indicated that A and D controlled the majority of output variance through the entire BTC for all reaches. For reaches at the bottom of the stream, A exhibited lower first-order sensitivities than A_s or D, though is still likely to be identified with first-order index values greater than 0.25 across all reaches for variable portions of the BTC. Constraining parameter bounds for A resulted in parameter sets that well approximate the BTC, such that changes in A may not lead to much variability in the fuzzy metric. Instead, A influenced model output across a constrained range through interactions with other parameters. A_s and α had larger total-order sensitivity indices than in the previous application to the entire BTC and variable first-order indices across the reaches.

[48] All parameters across all reaches largely influence model output through interactions. This supports conclusions of past research that OTIS parameters are highly interactive [*Wagener et al*., 2002; *Gooseff et al*., 2005]. To our knowledge, this level of interaction has not been previously quantified. This high level of interaction indicates that OTIS parameters for the reaches in this study are most likely nonidentifiable, except when first-order indices are high (greater than 0.25). Nonidentifiability does not imply a value of zero for a parameter estimate; it indicates that the value is interacting with other parameters, such that its value depends on the value of another parameter, or that the value is unimportant with respect to a given objective function.

6.2. Parameter Controls Across Reaches

[49] Reach parameter sensitivities were generalized into two types of responses, which were grouped based on similarities in sensitivities to the fuzzy metric through time and changes in physical setting in Stringer Creek. The discontinuous change in valley structure near 1200 m discussed previously corresponds to similarities in fuzzy metric sensitivities, which divided the reaches into two stream types, Type I (below 1200 m) and Type II (above 1200 m). Among many differences in valley, riparian, and channel structure, this change corresponds to differences in the down-valley slope, the amount of in-stream woody debris in the active channel, and the topography of hillslope terrain [*Payn et al*., 2012]. To explore how this discontinuity related to parameter sensitivities, fuzzy metric sensitivity indices were averaged across reaches and different periods of the BTC for each stream type and visualized in Circos plots (Figures 7 and 8; [*Krzywinski et al*., 2009]).

[50] Type I streams included locations 100, 200, 300, 400, 600, and 700 m. In these reaches, A and D had high (<0.25) total-order sensitivity indices on the rising limb, peak, and falling limb (Figures 8 and 9a). Additionally, interactions between A an d D dominated all parts of the BTC but the tail. Dispersion played a greater role than advection, with larger total-order sensitivity indices for D versus A in all but the falling limb and tail. A_s had high total-order sensitivity indices through the falling limb and tail and was most sensitive to model output on the falling limb. In the falling limb, A_s and α interacted primarily with

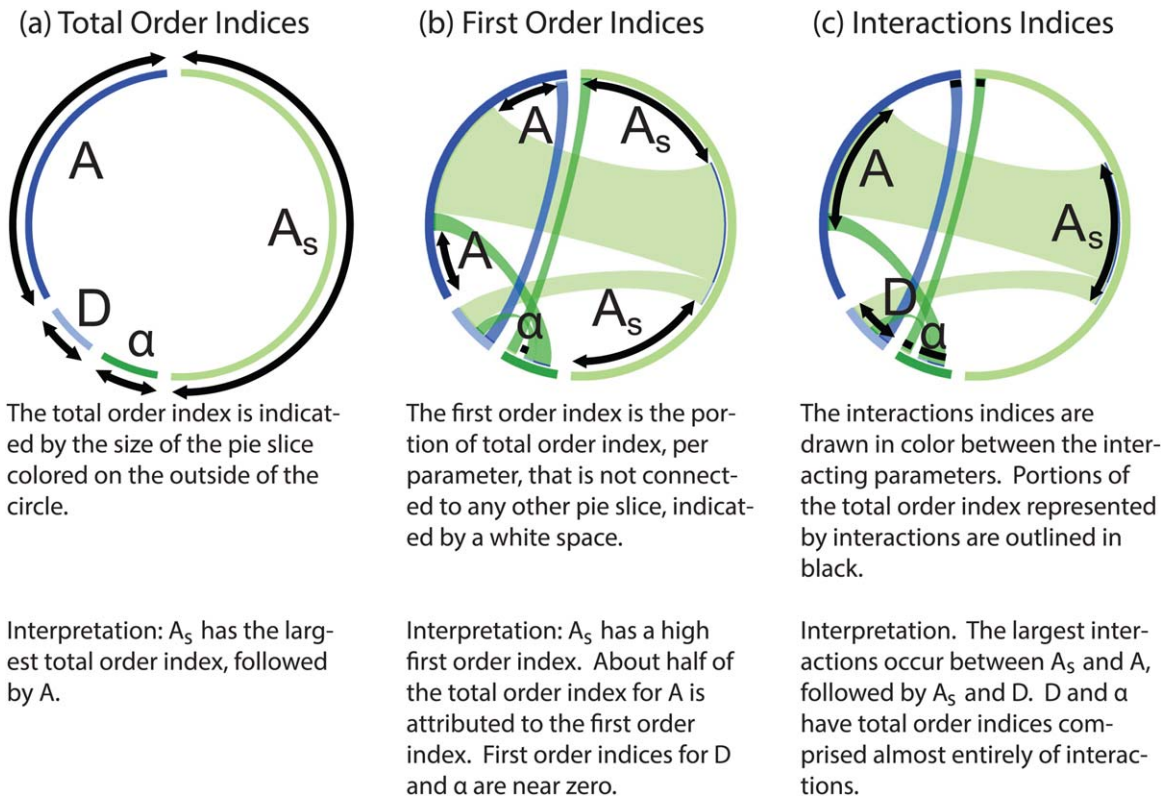

Figure 7. An introduction to reading a Circos plot (http://circos.ca/) [*Krzywinski et al.*, 2009]. Each plot contains information about the (a) total order, (b) first order, and (c) interactions indices for all parameters. An explanation for how to read and interpret a Circos plot is provided.

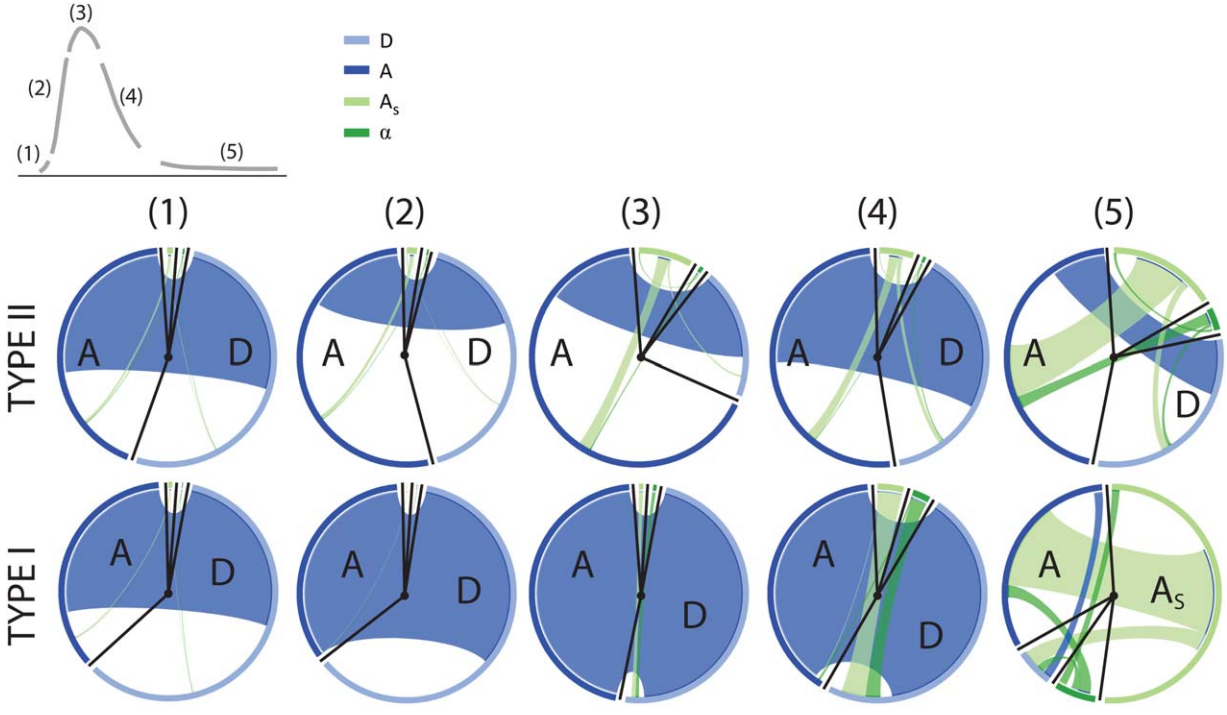

Figure 8. Circos plots of the sensitivity indices for the fuzzy metric objective function averaged across the two stream types and across the periods of the BTC shown in the figure.

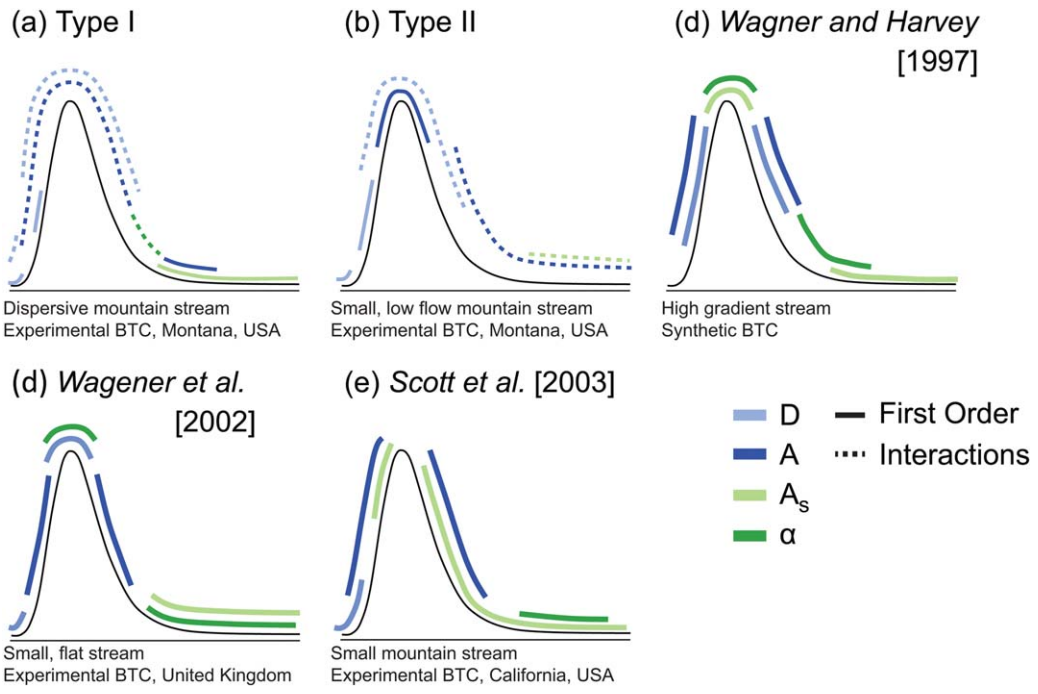

Figure 9. The two generalized response types from Figure 6, (a) Type I and (b) Type II from this present study, and results from (c) *Wagner and Harvey* [1997], (d) *Wagener et al*. [2002], and (e) *Scott et al*. [2003]. Color corresponds to the parameter, and line type distinguishes whether the parameter influences model output by itself or through interactions.

D but also with *A*. Across these reaches, *D*, *A*, and A_s are likely identifiable parameters given their high first-order indices.

[51] Type II streams include BTCs from 1300, 2100, 2400, and 2500 m (Figures 8 and 9b). In contrast to Type I streams, *A* and *D* were sensitive to model output through all parts of the BTC. Advection dominated dispersion, with larger sensitivity indices for *A* compared to *D*. *A* and *D* may be identifiable, given their high first-order indices. Transient storage parameters A_s and α had low first-order indices for Type II streams. As such, they are nonidentifiable for this stream type. The only exception is the reach at 2100 m, which had a high first-order index for A_s through the tail of the BTC.

6.3. Controls on Parameter Sensitivities

[52] Differences in model sensitivities for the two generalized stream types along Stringer Creek are likely related to the physical characteristics of the stream systems (Figure 10). The relationships between sensitivity and physical setting also point to controls on parameter identifiability. We expect sensitive parameters to be tied to dominant reach processes and use the physical characteristics of Stringer Creek to interpret this.

[53] The highly advective nature of the Stringer system and results from Figure 5 indicate that *A* should be identifiable for all reach types. The differences in sensitivity between *A* and *D* and the influence of A_s and α appear to be variable and organized by the physical context of the stream channel. For Type I streams, model output is more sensitive to *D* than *A* in the range of the observed tracer data. These reaches are subject to greater dispersion in more tortuous flow through step-pool sequences created by fallen trees and boulders. In contrast, model output for Type II streams is more sensitive to *A* than *D*. Flow in Type II streams is much less tortuous than in Type I streams, likely indicating that dispersion is less important to solute transport, resulting in lower sensitivity of output to *D*.

[54] Sensitivity of the storage zone parameters appears to be related to both stream size and stream type. A_s and α are not identifiable in Type II reaches. Transient storage for these reaches is most likely occurring to a small extent, though not absent, and parameters related to this process are, therefore, insensitive. This is supported by the stream setting in these locations. The riparian zone in upper portions of the watershed is almost entirely meadow across a wide valley floor, resulting in far fewer impediments to water flow or sediment transport in the active channel of Type II reaches. A_s may be identifiable for Type I streams. In contrast to Type I streams, boulders and large woody debris are common in the active channel of the lower portions of the watershed (Figure 10). The presence of these features has the potential to result in transient storage. Identifiability of A_s appears to be a precursor for the identifiability of α for instantaneous tracer release data, due to the strong interactions between these two parameters through the falling limb and into the tail.

[55] Overall, these results indicate that transient storage models can be used to characterize a subset of processes in a given reach for pulse tracer data. In this context, it is unlikely that all four parameters will be identifiable from instantaneous releases. Instead, dominant processes related to physical setting will correspond to the most identifiable parameters, and only these values can be used to formulate reach-level conclusions. However, we do demonstrate that

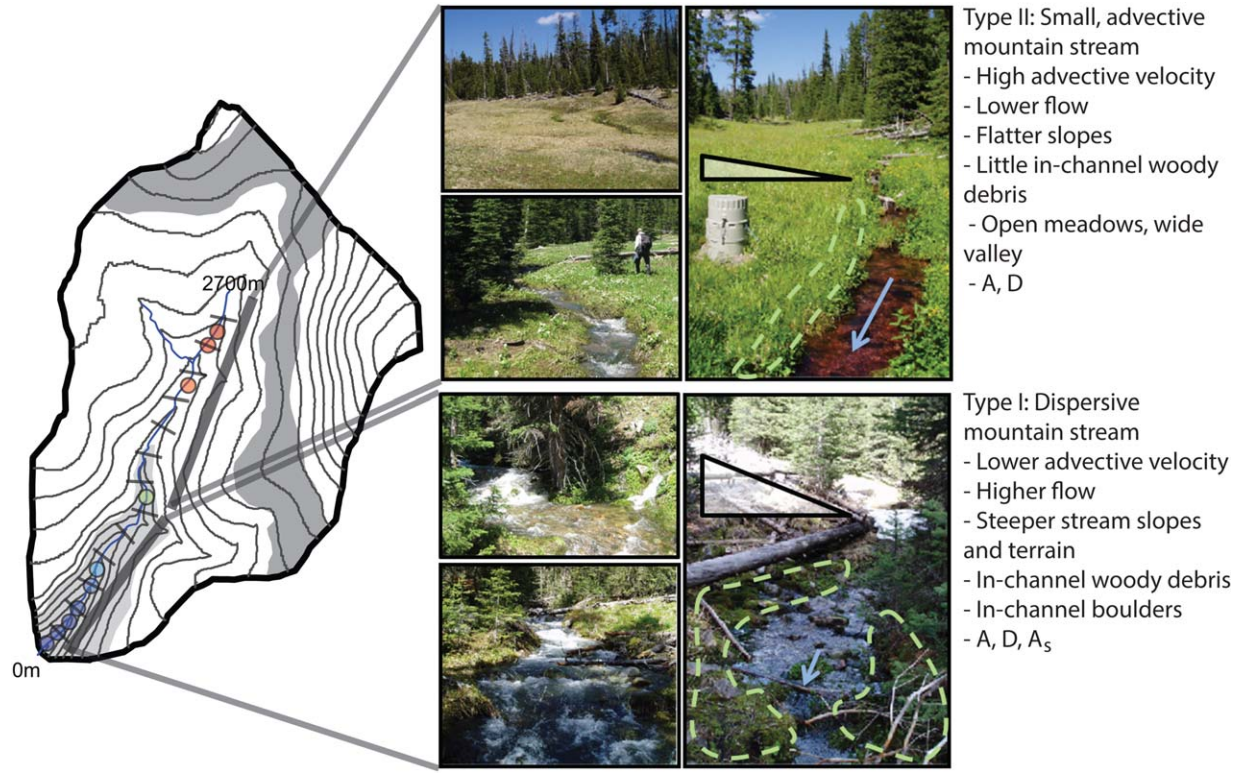

Figure 10. Photographs and summaries of the general traits of the two stream types along Stringer Creek. Photos taken by R. Payn and C. Kelleher.

pulse data can inform estimates of A_s, but that this information is contained within the tail of the BTC, and cannot be estimated from a calibration to the entire curve. This suggests that there is still valuable information on transient storage processes contained within a pulse release BTC for an advection dominated system.

6.4. Comparison to Previous Studies

[56] Several studies have suggested both similar and different conclusions regarding the identifiability of transient storage parameters from instantaneous release tracer data (Figure 9). *Wagner and Harvey* [1997] (Figure 9c), *Wagener et al*. [2002] (Figure 9d), and *Scott et al*. [2003] (Figure 9e) have all examined time-varying parameter sensitivities or surrogate measures of parameter information content for tracer experiments in various stream settings. All three stream studies found sensitivity to A_s and α along either the falling limb and/or tail of the BTC. This corroborates results for Type I streams from this analysis. Additionally, sensitivity to D and A along the rising limb and falling limb for all three studies was similar to the responses found for Type I and Type II streams. The most apparent difference between these past studies and current results occurs at the peak of the BTC. All three studies found sensitivity at this peak to either A_s or α, which was not consistent with our analysis. Interestingly, none of the studies found sensitivity to A at the BTC peak, which occurs for both stream types from this study.

[57] Differences in peak controls may be caused by a number of different factors. While the stream system studied by *Wagner and Harvey* [1997] was similar to the Stringer system, their study used different methods for characterizing sensitivity, in terms of the number of parameter sets, the algorithm used to generate parameter sets, the type of data investigated, and the method used to calculate sensitivity indices. The United Kingdom stream studied by *Wagener et al*. [2002] was much different than the Stringer Creek system, with flat terrain and stream slopes. Also important to note is that the generalized response from *Wagener et al*. [2002] does not show that A is the most sensitive parameter through all parts of the BTC except the tail. This is again similar to the results of the Stringer Creek sensitivity analysis. Of these three studies, reaches along the Stringer Creek valley would be most similar to the reaches in Uvas Creek investigated by *Scott et al*. [2003]. Results of *Scott et al*. [2003] have similarities to both Type I and Type II streams. Overall, results of previous studies corroborate findings for the Stringer System that different parameters will be sensitive across different parts of the BTC, and that this will vary across stream reaches. These differences can be attributed, at least in part, to difference in physical settings among these sites.

7. Conclusions

[58] While parameter estimates from the transient storage model have been used to characterize streams in terms of the magnitude of these processes, this approach can be flawed when estimated parameters are not uniquely identifiable. The sensitivity analysis performed in this study corroborates this point, with parameters A and D generally dominating across 10 reaches in an advective mountain

stream. Storage zone parameter α, which other studies have shown to be much less certain than A and D, was not sensitive across the analyzed reaches. This suggests that α cannot be used to characterize high gradient mountain streams similar to Stringer Creek. Model output was sensitive to A_s through the BTC tail for models of reaches with stream features generating in-stream transient storage, suggesting that pulse injections have the potential to identify this parameter. Estimates for A_s should be reported with caution, as stream characteristics will influence whether or not this parameter is identifiable.

[59] Given that the concentration data from this study were observed for a particular experimental setup and for particular conditions, there are certain limitations to the study conclusions. Past synthetic work from *Wagner and Harvey* [1997] suggests that parameter uncertainty is also a function of the type of tracer release and study reach length. *Wlostowski et al*. [2013] also show for experimental BTCs that the type of tracer release will influence which parameters can be identified as well as where along the BTC this information is contained. For the experiments analyzed in this study, all data were taken from instantaneous tracer releases and for 100 m reach lengths. Future work with this approach should test how parameter identifiability varies for constant rate versus instantaneous tracer injections, as well as how it varies across different reach lengths [e.g., *Wlostowski et al*., 2013]. This study also uses reach level data from a single stream; all conclusions are given within the context of the physical characteristics and setting which govern this stream. These results represent one application to a particular type of stream in one environment. We expect that streams with more variable stream characteristics than those seen in Stringer Creek will exhibit different patterns and levels of parameter identifiability. For instance, transient storage zone parameters may be identifiable for braided streams or streams known to experience shorter time or length scales for hyporheic exchange, as transient storage may be expected to be a dominant process in these types of streams. While these limitations merit future investigation, this study, to our knowledge, represents the most comprehensive sensitivity analysis and investigation of parameter identifiability for the application of transient storage models to date.

[60] The one-dimensional form of the transient storage model makes a number of approximations to the stream system. Assuming a constant value for D and A_s along the reach length simplifies a two-dimensional flow field into a one-dimensional application. For reaches where significant variation in physical characteristics along a given length of stream exists, crude approximations to the velocity flow field may compromise characterization of patterns of hyporheic exchange [*Cardenas et al*., 2004]. The alternative to the one-dimensional TSM is a two- or three-dimensional model, e.g., MODFLOW, MODPATH [e.g., *Gooseff et al*., 2006], or COMSOL [e.g., *Cardenas*, 2009]. While efficient numerical solvers exist for two-dimensional applications [*Hill and Tiedeman*, 2007], the applicability of a 2-D or 3-D model to a given site depends on whether there is enough information about a field site to constrain the model [*Harvey and Wagner*, 2000]. Aside from experimental characterization of the field site, equifinality becomes an even greater problem as the number of parameters in a model increases [*Beven*, 2006].

[61] The wider implications of this study are that modeling of instantaneous tracer data may not be capable of characterizing all stream processes related to the model parameters for all stream types. Given that this is an application to a series of reaches along a single stream, these results do not demonstrate that storage zone parameters are always nonidentifiable. Instead, we recommend that model parameter estimates should be reported along with sensitivity estimates as provided here or other types of uncertainty estimates. Only identifiable parameter values should be used to infer reach functioning.

[62] While these results suggest that care should be used in the reporting of these parameters, they highlight that the transient storage model is an excellent tool for identifying the dominant processes across stream reaches. They also show that there is still much information that can be gained through a pulse injection, and that the amount of information that can be abstracted will depend on the characteristics of a given reach. The relationship between parameter identifiability and physical setting suggests that detection of important stream processes may be better framed in terms of parameter sensitivity and identifiability. Parameter identifiability indicates that a given process is important for a given reach, while a lack of identifiability can be attributed to nonsensitivity or interactions with other parameters (which can be distinguished). In this application, the difference in dominant controls across reaches suggests that there is also an opportunity for an a priori classification of streams if we can relate sensitive processes to physical reach characteristics. Such a classification system would allow researchers to determine which transient storage model parameters are likely identifiable prior to performing a tracer test. In a broader sense, such a classification system will undoubtedly improve our understanding of transient storage processes and their relationship to physical setting. Broad characterization of stream reaches and the processes governing solute transport will be especially important as the anthropogenic footprint on ecosystems continues to grow, especially within small stream systems [*Freeman et al*., 2007] where transient storage models are often applied [*Stanley and Jones*, 2000].

[63] **Acknowledgments.** The authors would like to acknowledge Seth Kurt-Mason for discussions that helped contribute to the manuscript. Partial support for this project was provided by an EPA STAR Early Career Award (R834196) to Thorsten Wagener, and by an EPA STAR Graduate Fellowship Award to Christa Kelleher. The authors would also like to thank Rob Payn, Kelsey Jencso, and other field researchers from Montana State University for preparing and sharing tracer tests and terrain metrics for Stringer Creek. Sobol' analysis code was generated by Y. Tang, and generously shared by the Patrick Reed group at Penn State. Data for this study were provided by NSF grants EAR 03–37650, EAR-0943640, and EAR-0837937. The authors would like to thank the U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station for providing runoff data and logistical support. Any opinions, findings, and conclusions or recommendations expressed in this study are those of the authors and do not necessarily reflect the views of the funding institution.

References

Aronica, G., B. Hankin, and K. Beven (1998), Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data, *Adv. Water Resour.*, *22*(4), 349–365, doi:10.1016/S0309–1708(98)00017-7.

Bahr, J. M., and J. Rubin (1987), Direct comparison of kinetic and local equilibrium formulations for solute transport affected by surface reactions, *Water Resour. Res. 23*(3), 438–452, doi:198710.1029/WR023i003p00438.

Beck, M. B. (1987), Water quality modeling: A review of the analysis of uncertainty, *Water Resour. Res.*, *23*(8), 1393–1442, doi:198710.1029/WR023i008p01393.

Bencala, K. E., and R. A. Walters (1983), Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model, *Water Resour. Res.*, *19*(3), 718–724, doi:198310.1029/WR019i003p00718.

Beven, K. (1989), Changing ideas in hydrology the case of physically-based models, *J. Hydrol.*, *105*(1–2), 157–172, doi:10.1016/0022-1694(89)90101-7.

Beven, K. (2006), A manifesto for the equifinality thesis, *J. Hydrol.*, *320*(12), 18–36, doi:10.1016/j.jhydrol.2005.07.007.

Beven, K. J. (2008), *Environmental Modelling: An Uncertain Future?*, Routledge, London.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, *6*(3), 279–298, doi:10.1002/hyp.3360060305.

Boulton, A. J., S. Findlay, P. Marmonier, E. H. Stanley, and H. M. Valett (1998), The functional significance of the hyporheic zone in streams and rivers, *Annu. Rev. Ecol. Syst.*, *29*, 59–81.

Bratley, P., and B. L. Fox (1988), Algorithm 659: Implementing Sobol's quasirandom sequence generator, *ACM Trans. Math. Software*, *14*(1), 88–100, doi:10.1145/42288.214372.

Cardenas, M. B. (2009), A model for lateral hyporheic flow based on valley slope and channel sinuosity, *Water Resour. Res.*, *45*, W01501, doi:10.1029/2008WR007442.

Cardenas, M. B., J. L. Wilson, and V. A. Zlotnik (2004), A modelling study of hyporheic exchange pattern and the sequence, size, and spacing of stream bedforms in mountain stream networks, Oregon, USA, *Water Resour. Res.*, *40*, W08307, doi:10.1029/2004WR003008.

Cloke, H. L., F. Pappenberger, and J. P. Renaud (2008), Multi–method global sensitivity analysis (MMGSA) for modelling floodplain hydrological processes, *Hydrol. Processes*, *22*(11), 1660–1674, doi:10.1002/hyp.6734.

Cobelli, C., and J. J. DiStefano (1980), Parameter and structural identifiability concepts and ambiguities: A critical review and analysis, *Am. J. Physiology Regul. Integrative Comparative Physiol.*, *239*(1), R7–R24.

D'Angelo, D. J., J. R. Webster, S. V. Gregory, and J. L. Meyer (1993), Transient storage in Appalachian and cascade mountain streams as related to hydraulic characteristics, *J. North Am. Benthol. Soc.*, *12*(3), 223–235, doi:10.2307/1467457.

Dennis, J. E., Jr., D. M. Gay, and R. E. Welsch (1981), Algorithm 573: NL2SOL—An adaptive nonlinear least-squares algorithm, *ACM Trans. Math. Software*, *7*(3), 369–383, doi:10.1145/355958.355966.

Devore, J. L. (2000), *Probability and Statistics for Engineering and the Sciences*, Duxbury, Pacific Grove, Calif.

Donaldson, J., and P. Tryon (1987), *User's Guide to STARPAC: The Standard Time Series and Regression Package: Version 2.07*, Sci. Comput. Div., Natl. Cent. for Atmos. Res, Boulder, Colo.

Edwardson, K. J., W. B. Bowden, C. Dahm, and J. Morrice (2003), The hydraulic characteristics and geochemistry of hyporheic and parafluvial zones in Arctic tundra streams, north slope, Alaska, *Adv. Water Resour.*, *26*(9), 907–923, doi:10.1016/S0309-1708(03)00078-2.

Ensign, S. H., and M. W. Doyle (2005), In-channel transient storage and associated nutrient retention: Evidence from experimental manipulations, *Limnol. Oceanogr.*, *50*(6), 1740–1751.

Farnes, P. E., R. Shearer, W. McCaughey, and K. Hansen (1995), Comparisons of hydrology, geology, and physical characteristics between tenderfoot Creek experimental forest (East side) Montana, and Coram experimental forest (West side) Montana, Final Rep. RJVA-INT-92734, For. Sci. Lab., Intermountain Res. Stn., For. Serv., U.S. Dep. of Agric., Bozeman, Mont.

Fellows, C. S., H. M. Valett, and C. N. Dahm (2001), Whole-stream metabolism in two montane streams: Contribution of the hyporheic zone, *Limnol. Oceanogr.*, *46*(3), 523–531.

Fieberg, J., and K. J. Jenkins (2005), Assessing uncertainty in ecological systems using global sensitivity analyses: A case example of simulated wolf reintroduction effects on elk, *Ecol. Modell.*, *187*(23), 259–280, doi:10.1016/j.ecolmodel.2005.01.042.

Findlay, S. (1995), Importance of surface-subsurface exchange in stream ecosystems: The hyporheic zone, *Limnol. Oceanogr.*, *40*(1), 159–164.

Freeman, M. C., C. M. Pringle, and C. R. Jackson (2007), Hydrologic connectivity and the contribution of stream headwaters to ecological integrity at regional scales, *J. Am. Water Resour. Assoc.*, *43*(1), 5–14, doi:10.1111/j.1752-1688.2007.00002.x.

Gooseff, M., K. Bencala, D. Scott, R. Runkel, and D. McKnight (2005), Sensitivity analysis of conservative and reactive stream transient storage models applied to field data from multiple-reach experiments, *Adv. Water Resour.*, *28*(5), 479–492, doi:10.1016/j.advwatres.2004.11.012.

Gooseff, M. N., D. M. McKnight, R. L. Runkel, and B. H. Vaughn (2003a), Determining long time–scale hyporheic zone flow paths in Antarctic streams, *Hydrol. Processes*, *17*(9), 1691–1710, doi:10.1002/hyp.1210.

Gooseff, M. N., S. M. Wondzell, R. Haggerty, and J. Anderson (2003b), Comparing transient storage modeling and residence time distribution (RTD) analysis in geomorphically varied reaches in the Lookout Creek Basin, Oregon, *USA, Adv. Water Resour.*, *26*(9), 925–937, doi:10.1016/S0309-1708(03)00105-2.

Gooseff, M. N., J. K. Anderson, S. M. Wondzell, J. LaNier, and R. Haggerty (2006), A modelling study of hyporheic exchange pattern and the sequence, size, and spacing of stream bedforms in mountain stream networks, Oregon, USA, *Hydrol. Processes*, *20*, 2443–2457, doi:10.1002/hyp.6349.

Gooseff, M. N., R. O. Hall, Jr., and J. L. Tank (2007), Relating transient storage to channel complexity in streams of varying land use in Jackson Hole, Wyoming, *Water Resour. Res.*, *43*, W01417, doi:200710.1029/2005WR004626.

Harvey, J., and B. Wagner (2000), Quantifying hydrologic interactions between streams and their subsurface hyporheic zones, in *Streams and Ground Waters*, edited by J. Jones and P. Mulholland, pp. 3–44,Academic, San Diego, Calif.

Harvey, J. W., and K. E. Bencala (1993), The effect of streambed topography on surface-subsurface water exchange in mountain catchments, *Water Resour. Res.*, *29*(1), 89–98, doi:199310.1029/92WR01960.

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 Resour. Res.*, *32*(8), 2441–2451, doi:199610.1029/96WR01268.

Hill, M. C., and C. R. Tiedeman (2007), *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Predictions, and Uncertainty*, 1st ed., Wiley-Interscience, Hoboken, N. J.

Ibbitt, R., and T. O'Donnell (1971), Designing conceptual catchment models for automatic fitting methods, *Int. Assoc. Hydrol. Sci. Publ.*, *101*(2), 461–475.

Jencso, K. G., B. L. McGlynn, M. N. Gooseff, S. M. Wondzell, K. E. Bencala, and L. A. Marshall (2009), Hydrologic connectivity between landscapes and streams: Transferring reach- and plot- scale understanding to the catchment scale, *Water Resour. Res.*, *45*, W04428, doi:200910.1029/2008WR007225.

Johnston, P. R., and D. H. Pilgrim (1976), Parameter optimization for watershed models, *Water Resour. Res.*, *12*(3), 477–486, doi:197610.1029/WR012i003p00477.

Keefe, S. H., L. B. Barber, R. L. Runkel, J. N. Ryan, D. M. McKnight, and R. D. Wass (2004), Conservative and reactive solute transport in constructed wetlands, *Water Resour. Res.*, *40*, W01201, doi:200410.1029/2003WR002130.

Krzywinski, M. I., J. E. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra (2009), Circos: An information aesthetic for comparative genomics, *Genome Res.*, *19*(9), 1639–1645, doi:10.1101/gr.092759.109.

Lautz, L. K., and D. I. Siegel (2007), The effect of transient storage on nitrate uptake lengths in streams: An inter-site comparison, *Hydrol. Processes*, *21*(26), 3533–3548, doi:10.1002/hyp.6569.

Lautz, L. K., D. I. Siegel, and R. L. Bauer (2006), Impact of debris dams on hyporheic interaction along a semi-arid stream, *Hydrol. Processes*, *20*(1), 183–196, doi:10.1002/hyp.5910.

Martinez, C. J., and W. R. Wise (2003), Analysis of constructed treatment wetland hydraulics with the transient storage model OTIS, *Ecol. Eng.*, *20*(3), 211–222, doi:10.1016/S0925-8574(03)00029-6.

Morrice, J. A., H. M. Valett, C. N. Dahm, and M. E. Campana (1997), Alluvial characteristics, groundwater-surface water exchange and hydrological retention in headwater streams, *Hydrol. Processes*, *11*(3), 253–267, doi:10.1002/(SICI)1099-1085(19970315)11:3¡253::AID-HYP439¿3.0.CO;2-J.

Mulholland, P. J., E. R. Marzolf, J. R. Webster, and D. R. Hart (1997), Evidence that hyporheic zones increase heterotrophic metabolism and phosphorus uptake in forest streams, *Limnol. Oceanogr.*, *42*(3), 443–451, doi:10.4319/lo.1997.42.3.0443.

Nippgen, F., B. L. McGlynn, L. A. Marshall, and R. E. Emanuel (2011), Landscape structure and climate influences on hydrologic response, *Water Resour. Res.*, *47*, W12528, doi:201110.1029/2011WR011161.

Pappenberger, F., and K. J. Beven (2004), Functional classification and evaluation of hydrographs based on multicomponent mapping (Mx), *Int. J. River Basin Manage.*, *2*(2), 89–100, doi:10.1080/15715124.2004.9635224.

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, *J. Geophys. Res. Earth Surf.*, doi:10.1029/2012JF002455, in press.

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 Resour. Res.*, *45*, W11427, doi:200910.1029/2008WR007644.

Payn, R. A., M. N. Gooseff, B. L. McGlynn, K. E. Bencala, and S. M. Wondzell (2012), Exploring changes in the spatial distribution of stream baseflow generation during a seasonal recession, *Water Resour. Res.*, *48*, W04519, doi:201210.1029/2011WR011552.

Pickup, G. (1977), Testing the efficiency of algorithms and strategies for automatic calibration of rainfall-runoff models, *Hydrol. Sci. Bull.*, *22*, 257–274.

Runkel, R. L. (1998), One-dimensional transport with Inflow and storage (OTIS): A solute transport model for streams and rivers, Water Resour. Invest. Rep. 98–4018, U.S. Dep. of the Inter., U.S. Geol. Surv., Denver, Colo. [Available at http://water.usgs.gov/software/OTIS.].

Runkel, R. L. (2002), A new metric for determining the importance of transient storage, *J. North Am. Benthol. Soc.*, *21*(4), 529–543, doi:10.2307/1468428.

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, *J. North Am. Benthol. Soc.*, *17*(2), 143–154, doi:10.2307/1467958.

Ryan, R. J., A. I. Packman, and C. Welty (2004), Estimation of solute transport and storage parameters in a stream with anthropogenically produced unsteady flow and industrial bromide input, *Water Resour. Res.*, *40*, W01602, doi:200410.1029/2003WR002458.

Saltelli, A. (2002), Making best use of model evaluations to compute sensitivity indices, *Comput. Phys. Commun.*, *145*(2), 280–297, doi:10.1016/S0010–4655(02)00280-1.

Saltelli, A., F. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley, Hoboken, N. J.

Saltelli, A., M. Ratto, S. Tarantola, and F. Campolongo (2006), Sensitivity analysis practices: Strategies for model-based inference, *Reliab. Eng. Syst. Safety*, *91*(10–11), 1109–1125, doi:10.1016/j.ress.2005.11.014.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), *Global Sensitivity Analysis: The Primer*, John Wiley, Hoboken, N. J.

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, *J. North Am. Benthol. Soc.*, *22*(4), 492–510, doi:10.2307/1468348.

Sobol', I. (1967), On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys.*, *7*, 407–417.

Sobol', I. (2001), Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, *Math. Comput. Simul.*, *55*(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6.

Sobol', I. M. (1993), Sensitivity analysis for non-linear mathematical models, *Math Modell. Comput. Exp.*, *1*(4), 407–414.

Sobol', I. M. (1994), *A Primer for the Monte Carlo Method*, 1st ed., CRC Press, Boca Raton, Fla.

Sorooshian, S., and V. K. Gupta (1983), Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness, *Water Resour. Res.*, *19*(1), 260–268, doi:198310.1029/WR019i001p00260.

Stanley, E., and J. Jones (2000), Surface-subsurface interactions: Past, present, and future, in *Streams and Groundwaters*, edited by J. Jones and P. Mulholland, pp. 405–415, Academic, San Diego, Calif.

Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2007), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, *11*(2), 793–817, doi:10.5194/hess-11–793-2007.

Tate, C. M., R. E. Broshears, and D. M. McKnight (1995), Phosphate dynamics in an acidic mountain stream: Interactions involving algal uptake, sorption by iron oxide, and photoreduction, *Limnol. Oceanogr.*, *40*(5), 938–946, doi:10.2307/2838633.

Thackston, E. L., and K. B. Schnelle (1970), Predicting effects of dead zones on stream mixing, *J. Sanit. Eng. Div.*, *96*(2), 319–331.

Wagener, T., and H. Gupta (2005), Model identification for hydrological forecasting under uncertainty, *Stochastic Environ. Res. Risk Assess.*, *19*(6), 378–387, doi:10.1007/s00477-005-0006-5.

Wagener, T., L. A. Camacho, and H. S. Wheater (2002), Dynamic identifiability analysis of the transient storage model for solute transport in rivers, *J. Hydroinformatics*, *94*(3), 199–211, doi:10.1073/pnas.94.17.9171.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Wagner, B. J., and S. M. Gorelick (1986), A statistical methodology for estimating transport parameters: Theory and applications to one-dimensional advectivec-dispersive systems, *Water Resour. Res.*, *22*(8), 1303–1315, doi:198610.1029/WR022i008p01303.

Wagner, B. J., and J. W. Harvey (1997), Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies, *Water Resour. Res.*, *33*(7), 1731–1741, doi:199710.1029/97WR01067.

Ward, A. S., R. A. Payn, M. N. Gooseff, B. L. McGlynn, K. E. Bencala, C. A. Kelleher, S. M. Wondzell, and T. Wagener (2013), Variations in surface water ground water interactions along a headwater mountain stream: Comparisons between transient storage and water balance analyses, *Water Resour. Res.*, doi:10.1002/wrcr.20148, in press.

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 Resour. Res.*, *49*(2), 1184–1188, doi:10.1002/wrcr.20103.

Wondzell, S. M. (2006), Effect of morphology and discharge on hyporheic exchange flows in two small streams in the Cascade Mountains of Oregon, USA, *Hydrol. Processes*, *20*(2), 267–287, doi:10.1002/hyp.5902.

Stream Solute Workshop (1990), Concepts and methods for assessing solute dynamics in stream ecosystems, *J. North Am. Benthol. Soc.*, *9*(2), 95–119, doi:10.2307/1467445.

Zadeh, L. (1965), Fuzzy sets, *Inf. Control*, *8*(3), 338–353, doi:10.1016/S0019-9958(65)90241-X.