Article:

Schmadel, NM, Ward, AS, Kurz, MJ et al. (15 more authors) (2016) Stream solute tracer timescales changing with discharge and reach length confound process interpretation. Water Resources Research, 52 (4). pp. 3227-3245. ISSN 0043-1397

https://doi.org/10.1002/2015WR018062

Stream solute tracer timescales changing with discharge and reach length confound process interpretation

Noah M. Schmadel[1]
Adam S. Ward[1]
Marie J. Kurz[2]
Jan H. Fleckenstein[2]
Jay P. Zarnetske[3]
David M. Hannah[4]
Theresa Blume[5]
Michael Vieweg[2]
Phillip J. Blaen[4]
Christian Schmidt[2]
Julia L.A. Knapp[6]
Megan J. Klaar[4]
Paul Romeijn[4]
Thibault Datry[7]
Toralf Keller[2]
Silvia Folegot[4]
Amaia I. Marruedo Arricibita[8]
Stefan Krause[4]

[1]School of Public and Environmental Affairs, Indiana University, Bloomington, Indiana, USA
[2]Department of Hydrogeology, Helmholtz Centre for Environmental Research-UFZ, Leipzig, Germany
[3]Department of Geological Sciences, Michigan State University, East Lansing, Michigan, USA
[4]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, Edgbaston, United Kingdom
[5]Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam, Germany
[6]Center for Applied Geoscience, University of Tübingen, Tübingen, Germany
[7]IRSTEA, UR-MALY, Centre de Lyon-Villeurbanne, Villeurbanne, France
[8]Department of Ecohydrology, Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany

Correspondence To:
Noah M. Schmadel
School of Public and Environmental Affairs, Indiana University
400 MSB-II, Bloomington, IN 47405
nschmadel@gmail.com

Key points:

- Advection is the primary control on observed stream solute tracer responses.
- The influence of spatial heterogeneity in morphology is muted by advection.
- Interpretation of solute transport requires consideration of tracer timescales.

Abstract

Improved understanding of stream solute transport requires meaningful comparison of processes across a wide range of discharge conditions and spatial scales. At reach scales where solute tracer tests are commonly used to assess transport behavior, such comparison is still confounded due to the challenge of separating dispersive and transient storage processes from the influence of the advective timescale that varies with discharge and reach length. To better resolve interpretation of these processes from field-based tracer observations, we conducted recurrent conservative solute tracer tests along a 1-km study reach during a storm discharge period and further discretized the study reach into six segments of similar length but different channel morphologies. The resulting suite of data, spanning an order of magnitude in advective timescales, enabled us to (1) characterize relationships between tracer response and discharge in individual segments and (2) determine how combining the segments into longer reaches influences interpretation of dispersion and transient storage from tracer tests. We found that the advective timescale was the primary control on the shape of the observed tracer response. Most segments responded similarly to discharge, implying that the influence of morphologic heterogeneity was muted relative to advection. Comparison of tracer data across combined segments demonstrated that increased advective timescales could be misinterpreted as a change in dispersion or transient storage. Taken together, our results stress the importance of characterizing the influence of changing advective timescales on solute tracer responses before such reach scale observations can be used to infer solute transport at larger network scales.

Key words: stream solute transport, transient storage, conservative tracer, storm event, statistical moments, advective timescale

1. Introduction

Meaningful comparison of stream solute transport processes across different discharge conditions and spatial scales (e.g., morphologic unit to reach to network) is necessary to accurately represent and predict solute transport through stream networks [e.g., Covino et al., 2011; Kelleher et al., 2013]. Such comparison remains a persistent methodological and conceptual problem due to the uncertain distinction between spatial variability of solute transport processes, such as dispersion and transient storage, and changes caused by varying discharge conditions [e.g., González-Pinzón et al., 2013]. At the reach scale, on the order of tens to thousands of meters, solute transport processes are commonly interpreted through stream solute tracer tests [e.g., Stream Solute Workshop, 1990]. Direct comparison of observed tracer responses, however, is impossible without separating the spatially variable solute transport process from the unique timescale (commonly transit time distribution) of each test [Harvey et al., 1996]. Unfortunately, this separation is not straightforward because the tracer response changes primarily as a function of the down-stream advective transport time (commonly modal transport time, hereafter advective time), which shifts with discharge and reach length selection [e.g., Ward et al., 2013a].

Traditionally, comparison of solute tracer responses is performed by standardizing reach lengths through dimensionless numbers that relate physical processes to advective times [Runkel, 2002; Wagner and Harvey, 1997]. This standardization is thought to yield the appropriate “window of detection” inherent in tracer studies [i.e., the time from tracer first arrival to last detection; Harvey and Wagner, 2000] and allow for assessment of rapid exchanges between the stream and connected subsurface—the solute transport process often deemed most important to stream ecosystem functions [e.g., Boulton et al., 2010; Hester and Gooseff, 2010]. Subsurface and surface tracer exchange flowpaths that return to the stream within the window of detection, but have residence times slower than the advective time, are defined as short-term storage (commonly “transient storage”) [e.g., Harvey et al., 1996]. Conversely, tracer flowpaths that return outside of the window of detection (i.e., not recovered) are considered long-term storage and reflected in a tracer-based channel water balance [Payn et al., 2009; Schmadel et al., 2014a]. However, the window of detection changes with discharge and, consequently, defines an arbitrary boundary between short- and long-term storage regardless of the actual physical flowpaths [e.g., Ward et al., 2013a; Wondzell, 2006]. The window of detection may also be influenced by the tracer type and associated resolution of observations. For example, some tracers can provide more late-time tailing information than others due to differences in detection limit sensitivity (e.g., fluorescent tracers in comparison to salt tracers) [Drummond et al., 2012]. While the tracer selection and interaction between the advective time and window of detection can influence the interpretation of short-term storage, the storage flowpaths themselves can change with discharge [Wondzell, 2011; Zarnetske et al., 2007], further complicating meaningful comparison of tracer responses and thus impeding accurate conceptualization of solute transport processes.

Recent studies have shown, with the availability of high-frequency discharge observations, that analyzing solute transport processes across different discharge conditions is essential to improve conceptualization of how transient storage and dispersion change or compete with advection [Dudley-Southern and Binley, 2015; Ward et al., 2013a; Zimmer and Lautz, 2014]. A more refined conceptual understanding of stream solute transport across a range of discharges can better facilitate the upscaling of reach scale observations to infer solute transport at larger network scales. Because solute transport processes are dynamic and spatially variable along a stream network, current upscaling strategies use relationships between replicate reach scale processes (e.g., long-term storage assessed from tracer observations) and discharge [e.g., Covino et al., 2011; Stewart et al., 2011]. While there are reported relationships between solute transport processes and discharge—both for long-term [Covino et al., 2011; Ward et al., 2013a; Ward et al., 2013b] and short-term storage [Schmid et al., 2010; Ward et al., 2013a; Ward et al., 2013b]—varying advective times reflected in reach scale observations can lead to misinterpreting a change in discharge as an apparent change in transient storage or dispersion [e.g., Gooseff et al., 2007]. For example, when fixing study reach lengths throughout a watershed (e.g., standard 100-m or 200-m reaches [Payn et al., 2009; Ward et al., 2013b]), headwater reaches with lower discharge typically have larger windows of detection and advective times compared to higher discharge, downstream reaches. An apparent process relationship with discharge may be incorrectly inferred because a common assumption is that tracer studies across fixed reach lengths are directly comparable [Covino et al., 2011; Payn et al., 2009; Ward et al., 2013b], despite preexisting knowledge that different discharges in those fixed reaches will yield different windows of detection. The applicability of this assumption is uncertain as few studies have compared stream solute tracer studies across different discharges to characterize the influence of changing advective times and windows of detection on interpreting the processes of interest.

In this study, we directly examine the extent to which solute transport process information assessed from observed conservative solute tracer responses is comparable between different advective times and windows of detection. The objectives of this study are (1) to characterize relationships between physical solute transport processes and discharge in reach segments with distinct channel morphologies, and (2) to determine how study reach length influences the interpretation of these processes. To achieve these objectives, we conducted seven conservative solute tracer injections during a storm discharge period with in-stream responses recorded at seven downstream locations—enabling the comparison of up to 21 reach segments of different spatial scales, windows of detection, and advective times for each injection. We co-injected salt and fluorescein dye tracers for each injection, which also allowed us to test how tracer type and the associated data resolution influences interpreting solute transport.

2. Background on the Challenge of Interpreting Physical Solute Transport Processes from Stream Solute Tracers

The stream solute transport timescale, defined here as the advective time and window of detection, can be directly estimated from an observed in-stream conservative solute tracer time series (hereafter breakthrough curve, or BTC, Figure 1a). Following an instantaneous tracer injection, the elapsed time from injection to peak concentration is commonly interpreted as the down-stream advective time (t_{ad}) [e.g., Haggerty et al., 2002]. The window of detection (t_w) of an observed BTC is often quantified as the elapsed time from tracer first arrival (t_1) to 99% of recovered signal above background (t_{99}) [e.g., Mason et al., 2012; Ward et al., 2013a]. The elapsed time from t_{ad} to t_{99} describes the persistence of tailing in an observed BTC, which is an indicator of transient storage (hereafter the transient storage index, or TSI) [Mason et al., 2012]. The unique shape of a BTC is controlled by complex interactions of solute transport processes. For example, dispersive processes that contribute to spreading of the BTC include molecular diffusion and turbulent mixing [e.g., Fischer, 1975]. However, a BTC typically provides a reach-average representation of solute transport processes. Therefore, the processes most commonly interpreted from the shape of a BTC are down-stream advection, longitudinal dispersion, and transient storage [e.g., Ward et al., 2013a].

The extent to which down-stream advection or processes of interest (i.e., dispersion and transient storage) control the shape of the BTC remains unclear as different combinations can result in similar shapes. At one extreme, if dispersion and transient storage are constant between high and low discharges in the same reach, differences between the BTCs are controlled solely by differences in advective times (Figure 1b). For example, during lower discharges that result in smaller advective velocities, there will be more time for dispersion and transient storage to act on the tracer. At the other extreme, if advective times are constant (e.g., discharges are equivalent) between two different reaches, variation in the shape of the BTC and, therefore, the transport timescale is controlled by differences in dispersion and transient storage (Figure 1d). In practice, in-stream tracer observations reflect interacting advective times with dispersive and transient storage processes, which can yield BTCs with similar shapes for different reasons (Figure 1c). Therefore, we anticipate that an understanding of the proportional controls of advective timescales relative to dispersion and transient storage is necessary to compare stream solute tracer studies across discharges, reaches, and spatial scales.

3. Methods

3.1. Site Description and Discharge Measurements

The Selke River originates in the Harz Mountains in central Germany, flows through steep gradient, deeply incised valleys, and transitions to lower gradients underlain by alluvial deposits up to 10 m thick [Trauth et al., 2015] (Figure 2a). The study reach is a ~1-km section situated immediately downstream of this transition in largely agricultural lowlands

(51°43’21.4”N, 11°18’17.6”E). Continuous stream discharge was generated from stage—measured at 10-minute intervals (LTC Levellogger Junior M10, Solinst, Georgetown, Canada, accuracy +/- 0.1% of reading) about 15 m upstream of site E (Figure 2b)—applied to a site-specific power function rating curve developed by researchers at Helmholtz Centre for Environmental Research. The confidence bounds of this rating curve were produced following Schmadel et al. [2010] to provide the 95% confidence intervals of the discharge estimates (see Figure S1 in supporting information S1 for details). Discharge within the study reach increased from 0.27 to 2.35 $m^3 s^{-1}$ with stage remaining below bankfull during the July 2014 storm event used in this study (Figure 2c). The reach-specific rating curve was corroborated with continuous stream discharge recorded 4.6 km upstream at the Meisdorf federal gaging station. This station reports that discharge can range from 0.2 to 16 $m^3 s^{-1}$ during storm events and seasonal snowmelt, and the long-term (1921 to 2014) annual mean discharge is 1.5 $m^3 s^{-1}$.

The study reach has general channel morphologic characteristics of meandering glides and shallow riffles, pool-riffle sequences, point bars, and in-stream gravel bars (Figure 2b). The streambed primarily consists of medium sand to coarse gravel. Visually different morphologic characteristics are expected to manifest as distinct solute transport controls across changing discharge conditions [e.g., Gostner et al., 2013]. Therefore, seven in-stream monitoring sites were selected within the study reach to bracket sections of visually distinct morphologies while preserving similar lengths. Labeled moving downstream from A to G (Figure 2b), these sites delineate the six individual reach segments (100 to 152 m). The three upstream most segments, AB, BC, and CD, consist primarily of meandering glides connected by shallow riffles visible under low-flow conditions. Moving downstream, the morphology becomes more complex: segment DE contains one in-stream bar, two point bars, and a pronounced pool-riffle sequence;

segment EF contains a pool-riffle sequence at the downstream end; and segment FG contains two pool-riffle sequences and a large in-stream and point bar. The segment-wise average streambed slopes are about 0.7, 0.4, 0.1, 0.4, 0.4 and 0.8% for segments AB, BC, CD, DE, EF, and FG, respectively, estimated from a 1-m digital elevation model acquired from the Land Survey Administration Saxony-Anhalt dated April 2009. The reach average streambed slope is 0.5% and active channel width is 9 m.

The individual segments can be combined into a total of 21 unique segments of different lengths (Figure 2d). These 21 combinations were used to test if combining shorter segments into longer reaches integrates differences in processes or only generates apparent differences due to changing advective timescales and windows of detection. Similar to the varying reach length analysis of Gooseff et al. [2013], we recombined individual segments into every possible unique combination based on available tracer data.

3.2. In-Stream Solute Tracer Observations

Seven instantaneous conservative solute tracer injections were performed over the storm discharge period (July 7-11, 2014). For all injections, dissolved salt (NaCl) and fluorescein dye tracers were co-injected at the same location and measured at sites A through G, resulting in 49 observed BTCs each. In-stream responses of the salt tracer were observed at 30-second intervals (Schlumberger CTD diver 10m, Schlumberger Water Services, Delft, Netherlands, accuracy 1% of reading) using fluid specific conductance as a surrogate for salt concentration. Background specific conductance was corrected to zero. In some cases, the background specific conductance drifted during the measurement period, which was observed by an additional sensor placed about 10 m upstream of the injection site. This background signal was shifted downstream based on the advective time observed from the corresponding BTC and subtracted to correct for background (hereafter salt BTC). Because instantaneous and constant rate tracer injection techniques are expected to provide similar transit time distributions [Payn et al., 2008], instantaneous tracer injections were chosen over constant rate injections to minimize the influence of dynamic discharge conditions on measurements, similar to Ward et al. [2013a]. In-stream responses of the fluorescein dye tracer (hereafter fluorescein BTC) were measured at 10-second intervals using flow-through fluorometers (Albilia Sarl GGUN-FL30, Switzerland) at the same monitoring sites. No background correction was performed for the fluorescein BTCs because the instruments were calibrated in the field with stream water to reduce background interference and no dye was present prior to injection.

The observed tracer BTC from these types of experiments is an incomplete temporal signal due to detection limits and signal to noise interferences at late times [e.g., Drummond et al., 2012], but can provide meaningful dispersion and short-term storage information [e.g., Bellin et al., 2015; González-Pinzón et al., 2013]. In this study, we truncated each observed BTC to t_{99} to reduce the subjective selection of a response due to the injected tracer above background from noise at late times, limiting interpretation to observed short-term storage [after Mason et al., 2012; Ward et al., 2013a; Ward et al., 2013b]. An approximation to the window of detection ($t_w = t_{99} - t_1$) was quantified by solving equation (1) for t_{99},

$$0.99 = \frac{\int_{t=t_1}^{t=t_{99}} C(t)dt}{\int_{t=t_1}^{t=t_{CLIP}} C(t)dt}, \quad (1)$$

where C is the observed, background corrected BTC (g m^{-3}), t is the time after injection (s), t_1 is the time from injection to tracer first arrival (s), and t_{CLIP} is the time at which the observed BTC was initially clipped based on visual inspection. All BTCs (i.e., salt and fluorescein) used hereafter were truncated to t_{99}.

We compared the salt and fluorescein BTCs to examine how much temporal information, such as t_w, varies given differing tracer sensitivities and resolutions. In this comparison, we computed the associated temporal metrics (including those in Figure 1a and statistical moments as described in section 3.4.1) and examined the slope of the trend-line constructed through linear regression between those metrics corresponding to the fluorescein and salt BTCs. If a slope of unity and intercept of zero were within the associated 95% t-based confidence intervals (i.e., the ratio of fluorescein metrics to salt metrics is 1:1), the two BTCs were considered to provide similar temporal information. For this trend-line and those presented hereafter, we also tabulated the coefficient of determination (R^2).

3.3. Net Change in Discharge and Unrecovered Tracer Mass

Sections of the study reach have been reported as losing during baseflow conditions, but the hydraulic gradients between the stream and adjacent alluvial aquifer may alternate seasonally [Schmidt et al., 2012]. We used the salt BTCs and rating curve estimates to examine the general pattern of net changes to discharge and mass losses through the study reach during the storm discharge period. Specifically, we estimated discharge at site A via dilution gaging [after Kilpatrick and Cobb, 1985],

$$Q = \frac{M}{\int_{t=t_1}^{t=t_{99}} C(t)\,dt}, \tag{2}$$

where Q is the stream discharge ($m^3\ s^{-1}$) and M is the tracer mass injected (g) which ranged from 9 to 16 kg of NaCl in this study. The section of the study reach between the injection and site A was treated as a mixing length where we assume no tracer mass was lost. If some mass was lost over this section, Q will be slightly overestimated. We quantified the net change in discharge (downstream minus upstream) using estimates at sites A (dilution gaging) and E (rating curve), normalized by the upstream discharge to express as percent change. Assuming that the error in dilution gaging is roughly 8% [after Schmadel et al., 2010] and using the 95% confidence intervals of the site E discharge estimates (Figure 2c), the 95% confidence intervals of these net changes were approximated. A net change in discharge is considered significant if zero is outside of these intervals. Likewise, tracer mass recoveries were estimated from the discharge estimates near site E,

$$M_R = Q_D \int_{t=t_1}^{t=t_{99}} C_D(t)\,dt\,, \qquad (3)$$

where M_R is the tracer mass recovered (g), Q_D is the downstream stream discharge (i.e., site E) ($m^3\ s^{-1}$), and C_D is the observed tracer concentration at site E from the tracer released at the injection site ($g\ m^{-3}$). The percent unrecovered mass was estimated as the difference $M_R - M$ normalized by M. Again, because 95% confidence intervals were approximated for discharge estimates, we estimated the confidence intervals of unrecovered tracer mass. These confidence interval estimates do not include uncertainty due to $\int C_D(t)dt$ in equation (3). Although this uncertainty is expected to be small relative to the uncertainty in Q_D, a potential artifact of objectively truncating the BTC to t_{99} is artificially low M_R estimates.

3.4. Temporal Metrics

We estimated temporal metrics (including those in Figure 1a and statistical moments as described below) of each fluorescein and salt BTC to compare field-based tracer observations and interpret relationships between solute transport processes and discharge. Following this reach scale analysis, we used the temporal metrics for each individual reach segment to examine how segment-wise processes changed with discharge. Lastly, we used the same metrics of each segment combination (Figure 2d) to test whether combining shorter segments into longer reaches influenced the interpretation of how solute transport processes change with discharge.

3.4.1. Relationships between Reach Scale Temporal Metrics and Discharge

Using the t_{ad} and t_{99} estimates, TSI was estimated (Figure 1a). To compare metrics across discharges, t_w and TSI were normalized by t_{ad} to remove the variation in the BTC caused by advection [after Gooseff et al., 2007; Ward et al., 2013a],

$$t_{w,norm} = t_w/t_{ad}, \tag{4}$$

$$TSI_{norm} = TSI/t_{ad}, \tag{5}$$

where $t_{w,norm}$ defines the window of detection relative to advective time and TSI_{norm} reflects the persistence of tailing relative to advective time. These metrics provide the overall influences of dispersion and short-term storage on the shape of the BTC independent of advective time assuming that t_w and TSI are linearly related to t_{ad}. While this assumption is appropriate in advective-dominated stream systems [Gooseff et al., 2007; Ward et al., 2013a], other types of normalization may be necessary to better account for potential nonlinearities [e.g., Gelhar et al., 1992]. Each observed BTC was normalized to express only the available temporal signature,

$$c(t) = \frac{C(t)}{\int_{t=t_1}^{t=t_{99}} C(t)dt}, \tag{6}$$

where c reflects the recovered transit time distribution and the integral of C(t) with respect to time represents the zeroth temporal moment. We estimated the statistical moments of the transit time distributions by empirically calculating temporal moments of the recovered BTCs. Specifically, we estimated the first temporal moment (M_1),

$$M_1 = \int_{t=t_1}^{t=t_{99}} tc(t)dt\,, \tag{7}$$

which provides an estimate of mean arrival time. The nth-order temporal moment centered about M_1 (hereafter central moment) was estimated as

$$\mu_n = \int_{t=t_1}^{t=t_{99}} (t - M_1)^n c(t)dt \;\text{ for } n > 1. \tag{8}$$

The second central moment (μ_2) provides the temporal variance, a metric of symmetrical spreading; the third central moment (μ_3) reflects the temporal extent of late-time tailing. We computed the normalized metrics of the coefficient of variation (CV) and skewness (γ) to further compare BTCs across discharges,

$$CV = \frac{\mu_2^{1/2}}{M_1}\,, \tag{9}$$

$$\gamma = \frac{\mu_3}{\mu_2^{3/2}}\,, \tag{10}$$

where CV expresses the rate of symmetrical spreading relative to mean arrival time and γ reflects the extent of late-time tailing relative to symmetrical spreading (see Table S1 in supporting information S1 for a summary of all temporal metrics used). Therefore, these metrics should provide an indication of how dispersion and short-term storage processes change with discharge, respectively.

Relationships between discharge and the normalized reach scale metrics ($t_{w,norm}$, TSI_{norm}, CV, and γ) for each monitoring site were constructed through linear regression. The discharges used for this relationship were those estimated from the rating curve near site E. Significant slopes were those whose 95% t-based confidence interval did not include zero.

3.4.2. Relationships between Segment-Wise Temporal Metrics, Discharge, and Length

The transfer function, g(t), reflects the unique temporal signature of a solute traveling through a given reach segment independent of the input signal. When a single tracer injection is observed at different locations within a reach, the transfer function can be related to the upstream and downstream transit time distributions by convolution, assuming linear, time-invariant transport from one segment to the next,

$$c_{ds}(t) = \int_{t=t_1}^{t=t_{99}} g(\tau) c_{us}(t-\tau) d\tau , \tag{11}$$

where c_{ds} is the observed transit time distribution at a downstream monitoring site and c_{us} is the observed transit time distribution at the upstream monitoring site (input signal). To isolate segment-wise responses to discharge, we quantified metrics of the temporal characteristics of segment-specific transfer functions [after Ward et al., 2016] rather than apply a more complicated deconvolution scheme [e.g., Cirpka et al., 2007].

We estimated the change in window of detection between consecutive downstream and upstream BTCs (Δt_w) as an estimate of the transfer function window of detection,

$$\Delta t_w = t_{w,ds} - t_{w,us}. \tag{12}$$

To allow for comparison between segments, segment lengths, and discharges, we normalized this metric by the change in the advective time,

$$\Delta t_{w,norm} = \Delta t_w / \Delta t_{ad}, \tag{13}$$

where $\Delta t_{ad} = t_{peak,ds} - t_{peak,us}$, and $t_{peak,ds}$ and $t_{peak,us}$ are the times to peak concentrations of the downstream and upstream BTCs, respectively. Next, we estimated the change in TSI and normalized by the corresponding change in advective time,

$$\Delta TSI_{norm} = (TSI_{ds} - TSI_{us})/\Delta t_{ad}. \tag{14}$$

We used the change in t_{ad}, t_w, and TSI between downstream and upstream BTCs as surrogates to the corresponding metrics of the actual transfer function. Because the segment-wise temporal moments are theoretically linearly additive to produce the overall reach estimate [e.g., Riml and Wörman, 2011], and have been interpreted as such in existing field studies [Ward et al., 2014], we isolated statistical moments of the transfer function while eliminating the need to solve equation (11),

$$\Delta M_1 = M_{1,ds} - M_{1,us}, \tag{15}$$

$$\Delta \mu_n = \mu_{n,ds} - \mu_{n,us} \text{ for } n > 1. \tag{16}$$

Again, to compare between segments, segment lengths, and discharges, we computed the coefficient of variation and skewness of the transfer function,

$$\Delta CV = \frac{\Delta \mu_2^{1/2}}{\Delta M_1}, \tag{17}$$

$$\Delta \gamma = \frac{\Delta \mu_3}{\Delta \mu_2^{3/2}}. \tag{18}$$

We used the normalized segment metrics ($\Delta t_{w,norm}$, ΔTSI_{norm}, ΔCV, and $\Delta \gamma$) to investigate (1) relationships with discharge, (2) the spatial variability of segment response to discharge, and (3) whether estimates of these metrics were dependent on segment length.

Relationships with discharge were investigated by constructing linear trend-lines between the normalized metrics calculated for each segment (and combination of segments) and discharge near site E. Discharge was anticipated to vary slightly between segments (i.e., potentially a net losing stream), but if a relationship is present, discharge near site E will suffice because the large magnitude of change in discharge due to the storm event overwhelms the potential spatial change over the study reach. The metrics estimated for each of the seven discharges that reflect a single segment provided the sample population for the associated trend-line. Additionally, we pooled all normalized segment-wise metrics and examined the corresponding slope with discharge to provide an indication of the overall stream system response to changing discharge.

Spatial variability between segment responses to discharge was examined by comparing each metric ($\Delta t_{w,norm}$, ΔTSI_{norm}, ΔCV, and $\Delta\gamma$) sample population that reflected the individual segments (i.e., not the combined segments). This comparison was completed through a parametric test of comparing means in a one-way analysis of variance (ANOVA). We additionally completed a nonparametric test of comparing medians in a Kruskal-Wallis one-way ANOVA. The nonparametric test does not require the assumption of normally distributed residuals like the parametric version, so it provides additional information to prevent over-interpreting results from small sample sizes where the parent distribution is unknown. A 95% confidence level where $p < 0.05$ indicates means (denoted by p_{ANOVA}) or medians (denoted by p_{KW}) are different and segments are a significant source of variability. While these p-values indicate whether means and medians of at least one segment are different from all of the others, these tests were also repeated to compare each segment distribution to another on a paired basis. We selected these statistical tests because they are among the simplest to apply, acknowledging that the small sample size limits the ability to identify underlying distributions or satisfy the assumptions of additional statistical tests.

To determine how the normalized transfer function metrics ($\Delta t_{w,norm}$, ΔTSI_{norm}, ΔCV, and $\Delta\gamma$) depended on segment length, we grouped corresponding estimates according to the number of combined segments. Each group provided the sample populations that were compared using the parametric and nonparametric tests outlined above. We only considered the sample populations for one, two, and three combined reach segments (Figure 2d) that contained more than 20 values representing different spatial scales and discharges to reduce over-interpretation of results. In this case, a p_{ANOVA} and p_{KW} of less than 0.05 indicates that the normalized transfer function metrics are dependent on the transport timescale (i.e., Δt_{ad} and Δt_w) due to reach length selection. Hence, if this condition is recognized, interpreting how solute transport processes change with discharge is influenced by reach length selection.

4. Results

4.1. In-Stream Solute Tracer Observations

The influence of time-varying discharge conditions on the observed BTCs was assumed negligible due to the relatively short measurement periods. For the seven tracer injections measured at sites A through G, the duration of each corresponding measurement period from tracer injection to last detection at G ranged from 40-150 minutes (see Figure S2 in supporting information S1 for all observed BTCs). The possible change in discharge was between 1% and 6% over these measurement periods (red lines in Figure 2c).

Through the storm discharge period, some of the salt and fluorescein BTCs (49 each possible) were deemed erroneous based on initial quality control (e.g., debris potentially blocking the sensor, power failure, or data with visually high noise compared to other observations). This quality control resulted in a total of 29 co-located salt and fluorescein BTCs. In addition to these co-located BTCs, there were 12 instances where only salt BTCs were obtained and 4 instances where only fluorescein BTCs were obtained. A total of 5 fluorescein BTCs were omitted based on visually high noise. Data availability is summarized in Figure S2 and Table S2 in supporting information S1.

The co-located salt and fluorescein BTCs provided similar reach scale temporal metrics (t_w, TSI, M_1, μ_2, and μ_3) (Figure 3, left column). All the slopes between these metrics estimated from salt BTCs and those estimated from fluorescein BTCs were statistically the same as 1 (i.e., 1 falls within the associated 95% confidence interval). The intercepts were also statistically the same as zero for all these metrics. The fluorescein tracer produced longer detectable tailing, resulting, mostly, in higher t_w and TSI estimates (above the 1:1 line in Figure 3, left column). The higher order central moment estimates began to deviate further from the 1:1 line at higher values (i.e., those associated lower discharges). All slopes and intercepts of the normalized metrics ($t_{w,norm}$, TSI_{norm}, CV, and γ) were statistically less than 1 and greater than zero, respectively (Figure 3, right column). However, the largest average difference between these metrics estimated from salt and fluorescein BTCs was ~10%, indicating that both tracers provide similar temporal metric estimates following objective truncation to t_{99}.

As both tracers provided similar temporal metric estimates yet incomplete datasets, we used fluorescein BTCs to fill in gaps in the salt BTC dataset, resulting in a total of 45 BTCs (41 salt and 4 fluorescein) to allow for greater coverage of advective times and windows of detection, and thus a more complete analysis. Fluorescein BTCs were selected to fill in gaps in the salt BTC dataset, and not vice versa, because there were more salt BTCs available and salt is more readily available and commonly used in stream tracer studies.

4.2. Net Change in Discharge and Unrecovered Tracer Mass

The study reach was generally net losing through the storm discharge period (Figure 4a). The net change in discharge from site A to E ranged from -15 to -20% and was significant for every injection with the exception of injection 3 (where the discharge estimate at site E was highest) and injection 1 (where there was no salt BTC available at site A). A similar pattern occurred for the unrecovered tracer mass estimates (Figure 4b). For injections 4 to 7, unrecovered mass was significant and ranged from -15 to -25%, providing further evidence that this study reach was a net losing stream.

4.3. Relationships between Reach Scale Temporal Metrics and Discharge

Regardless of the tracer type, the normalized metrics were mostly insensitive to changing discharge where statistically significant relationships were the exception rather than the norm. The reach scale metrics of the transit time distributions (t_w, TSI, M_1, μ_2, and μ_3) all decreased with increasing discharge (Figure 5, left column, estimated from 41 salt and 4 fluorescein BTCs). There was an increasing trend between these metrics and downstream location (bottom-to-top order of lines in the left column of Figure 5, representing increased length). The normalized metrics ($t_{w,norm}$, TSI_{norm}, CV, and γ) along the study reach were not significantly related to discharge in most cases (Figure 5, right column). The exceptions are that CV was significantly negatively related to discharge only at sites C and D and γ was significantly positively related to discharge at site A (Table 1). Repeated analyses using only the salt BTCs and using only the fluorescein BTCs produced similar results that most relationships with discharge were not significant (see Figures S3 and S4 and Tables S3 and S4 in supporting information S1 for the results of these analyses).

4.4. Relationships between Segment-Wise Temporal Metrics, Discharge, and Length

4.4.1. Segment-Wise Responses to Changing Discharge

The normalized transfer function metrics ($\Delta t_{w,norm}$, ΔTSI_{norm}, ΔCV, and $\Delta\gamma$) indicated that some reach segments were more sensitive to changing discharge than others where more variability in the distribution is interpreted as a higher sensitivity (Figure 6). Segment EF was consistently the most sensitive to discharge based on the interquartile range. For $\Delta t_{w,norm}$, there was no significant difference between the segment-wise means and medians ($p_{ANOVA} = 0.30$ and $p_{KW} = 0.12$). For ΔTSI_{norm}, the means were statistically the same ($p_{ANOVA} = 0.30$), but there was evidence that the medians were different ($p_{KW} = 0.04$). The segment-wise means and medians of $\Delta\gamma$ were statistically the same ($p_{ANOVA} = 0.11$ and $p_{KW} = 0.06$). Note that some negative $\Delta\gamma$ estimates resulted from isolating the central moments of the transfer functions (equation (16)). The variation of ΔCV means and medians between segments was significant ($p_{ANOVA} < 0.001$ and $p_{KW} = 0.01$). When each segment distribution was compared to another on a paired basis, $\Delta t_{w,norm}$ was different between segments AB and BC, BC and CD, and BC and DE (also visualized by non-overlapping notches in Figure 6); ΔTSI_{norm} of segment AB was different from segments BC, CD, and DE. The mean and median ΔCV of segment AB was different from all other segments. The mean and median ΔCV between all segments other than AB were the same. The means and medians of $\Delta\gamma$ were statistically the same between each segment with the exception of segments AB and FG. See Table S5 in supporting information S1 for p-values of all these pairwise comparisons.

An apparent relationship between the normalized transfer function metrics and discharge was not clearly recognized. The slopes of the relationship between $\Delta t_{w,norm}$ and discharge for each segment and combination were generally negative (15 out of 21 segments; see Table S6 in supporting information S1 for slopes and R^2 values). However, the slope between this metric and discharge was only significant for segments AC and AD. The slopes of the relationship between ΔTSI_{norm} and discharge were generally negative (15 out of 21 segments), but none were significant. The slopes of the relationship between ΔCV and discharge were also generally negative (19 of 21 segments) with only two significant within segments AD and BD (negative slopes). A relationship between $\Delta\gamma$ and discharge was unclear, with variable slopes (9 out of 21 segments with positive slopes, 12 out of 21 with negative slopes) that were not significant. Pooling all 21 segments indicated that $\Delta t_{w,norm}$, ΔTSI_{norm}, and ΔCV overall had a slight downward trend with discharge while the relationship between $\Delta\gamma$ and discharge was not significant (see Table S6 in supporting information S1 for slopes).

4.4.2. Segment-Wise Metrics Dependence on Length

We found a general pattern of decreasing variability in the transfer function metric estimates for longer segments (Figure 7). This decrease was due to both increased averaging of the heterogeneity in longer segments and an overall decrease in the number of samples available (i.e., where fewer extremes in the underlying distribution are reflected in longer segments). We statistically compared only the distributions associated with one to three combined segments where the numbers of individual observations in the sample populations were greater than 20. The means and medians of the $\Delta t_{w,norm}$ and ΔTSI_{norm} distributions did not significantly change when reach segments were combined (i.e., $p_{ANOVA} = 0.53$ and 0.44 and $p_{KW} = 0.76$ and 0.96, respectively). The means and medians of the ΔCV distributions were significantly different (i.e., p_{ANOVA} and $p_{KW} < 0.001$). By combining segments in this study reach under the seven discharge conditions, the mean and median of ΔCV significantly increased over longer lengths. While the medians of $\Delta\gamma$ distributions decreased slightly over longer lengths ($p_{KW} = 0.04$), the means were statistically the same ($p_{ANOVA} = 0.81$).

5. Discussion

5.1. A Changing Transport Timescale Complicates Comparison of Stream Solute Tracer Studies

The extent to which dispersive and transient storage processes change with discharge can be easily misinterpreted from in-stream solute tracer tests based on different transport timescales (defined as the advective time and window of detection). For example, similar to findings from other tracer studies [e.g., Schmid et al., 2010], magnitudes of several BTC metrics decreased with increasing discharge (Figure 5, left column). A possible interpretation of this result is that physical processes have changed (e.g., activation or deactivation of storage flowpaths), which is to be expected as varying discharge alters the turbulent energy, wetted geometry, hydraulic gradients, and connections to storage flowpaths [Leopold and Maddock, 1953]. However, through the suite of transport timescales of this study, the normalized temporal metrics ($t_{w,norm}$, TSI_{norm}, CV and, γ) were mostly insensitive to changes in discharge (Figure 5, right column and Table 1). This lack of a strong relationship indicates that determining how differences in advective times and windows of detection control the shape of the BTC is needed when interpreting changes in physical solute transport processes. However, we recommend that similar tracer tests should be repeated across more reaches of different stream systems and discharge conditions to test the robustness of this conclusion.

We anticipate general changes to some transport processes due to varying discharge based on previous studies. The relative influence of dispersive processes can rise with discharge due to increased turbulent mixing [D'Angelo et al., 1993]. Likewise, we expect CV to increase with discharge. Subsurface short-term storage volume can increase with discharge in stream systems with relatively coarse (e.g., sand to gravel) streambeds [Dudley-Southern and Binley, 2015; Schmid et al., 2010; Zimmer and Lautz, 2014], which would likely cause an increase in prolonged tracer tailing or γ. Conversely, streams with low subsurface short-term storage potential (i.e., low streambed hydraulic conductivity and valley slope) are less likely to undergo substantial storage flowpath changes with discharge [Wondzell, 2011]. In this latter case, an increase in stream discharge should result in reduced tailing and corresponding decrease in γ because transport of the tracer is expected to be more sensitive to advective timescales than changing subsurface short-term storage processes. Similarly, based on a fluid-mechanics classification scheme after Jackson et al. [2013], we expect surface short-term storage to generally increase with discharge. The Selke study reach is not constrained by its valley and is set in highly permeable gravels [Schmidt et al., 2012]; therefore, we expected the reach-scale extent of short-term storage to increase with discharge, which would likely cause an increase in γ. Contrary to expectation, our estimates of CV were mostly insensitive to changes in discharge with some evidence of a negative relationship with discharge (sites C and D, Table 1). Although γ was mostly insensitive to changes in discharge, there was some indication of a positive relationship at site A, suggesting that the extent of storage increased with discharge. The lack of a clear relationship with discharge indicates that these metrics either scaled directly with the transport timescale or short-term storage and dispersion did not change appreciably despite the order of magnitude change in observed discharge. Direct scaling with the transport timescale is the more likely conclusion because a negative response in CV suggests that this metric is still influenced by changing transport timescales rather than changes in dispersive processes.

As expected, the advective time of the transfer function was inversely proportional to discharge and directly proportional to reach length—the longest advective time occurred at the lowest discharge and longest segment length (Figure 8a). Similarly, the window of detection expanded with an increase in advective time (Figure 8b). Theoretical temporal moments (i.e., those derived from commonly used transient storage models) suggest that both the coefficient of variation and skewness decrease nonlinearly with increasing reach length (and advective time) as the predicted BTC becomes more symmetrical [Schmid, 2002]. Contrary to theory, we observed an increasing trend in ΔCV with advective time (Figure 8c), which corresponds to an increase in reach length (Figure 7). Consistent with theory, $\Delta\gamma$ did have a slight decreasing trend with advective time (Figure 8d), but did not show a significant change with increased reach length (Figure 7). While Ward et al. [2013a] found that comparison of tracer observations across different discharges and reaches requires normalizing by associated advective times, the increasing trend in the coefficient of variation indicates that an expanding window of detection associated with longer advective times still limits direct comparison. A larger window of detection allows more time for processes like dispersion and short-term storage to act on the tracer. Consequently, tracer observations made at short advective times are more sensitive to the influence of truncation, or lack of an ability to measure the entire tail, than those made at longer advective times. An estimate of the coefficient of variation is clearly more sensitive to truncation than an estimate of skewness. This issue of truncation may also partially explain why dispersion in aquifers assessed from tracers typically increases with the spatial scale of observations [e.g., Gelhar et al., 1992]. Our results provide evidence that when tracer observations are compared, the influence of differences in the advective time and window of detection must be explicitly considered to prevent misinterpreting changes in solute transport processes. However, any change in the coefficient of variation or skewness at a given advective time is likely due to the influence of heterogeneity in channel morphology and process change associated with discharge.

In addition, other sources of uncertainty may restrict comparisons of tracer observations between different discharge conditions. A potential source of uncertainty is that short-term storage flowpaths may change differently on the rising and falling limbs of the hydrograph [e.g., Ward et al., 2013a; Zimmer and Lautz, 2014]. For example, storage flowpaths within the Selke study reach might be activated during the rising limb and subsequently deactivated during the falling limb. To better understand this influence, more tracer tests along the rising limb would be needed (Figure 2c). Furthermore, segment-wise tracer mass loss, as the study reach is anticipated to be net losing (Figure 4b), might change with discharge. Fortunately, if the stream only loses water without having significant gains, mass losses will not have substantially influenced the temporal information contained in the BTC because in-stream concentrations remain unchanged. Testing this mass loss influence would require tracer injections at the limits of each segment (such as those by Payn et al. [2009]). Considerations of dynamic mixing length and data resolution would be necessary to perform finer scale dilution gaging and mass recovery techniques during high discharges.

5.2. Sensitivity of Tracer Detection is Less Important at High Discharges

Detection of fluorescein was slightly more sensitive than detection of salt in late-time tailing based on generally larger windows of detection (t_w). Accordingly, the higher order moments (μ_2 and μ_3) were generally larger for the fluorescein tracer (Figure 3, left column). These differences also caused the normalized metrics ($t_{w,norm}$, TSI_{norm}, CV and, γ) to be generally larger than those estimated from the salt tracer. Despite these differences, however, similar conclusions were drawn from both tracers: changes to the transport timescales confound the ability to observe changes in other processes because deviations between the metrics of the two different tracer types were small relative to changes in advective time. The metrics provided by the fluorescein and salt tracers were in close agreement at high discharges, indicating that the ability to detect short-term storage processes through tracer tests is limited at high discharges regardless of the tracer type selected.

An artifact of incomplete tracer response due to detection limits or signal to noise interferences in late-time tailing is artificially low temporal moment estimates [e.g., Drummond et al., 2012]. Consequently, objective truncation limits this study to observable short-term storage and may partially explain the lack of a robust trend between ΔCV and $\Delta\gamma$ and discharge. Although we expect dispersion and short-term storage to change with discharge, this truncation may have reduced the influence of these processes relative to advection in the observations. One potential way to address this late-time detection issue is to apply statistical techniques to better isolate the tracer signal from noise [e.g., Aquino et al., 2015; Drummond et al., 2012]. Other sensing techniques and tracers may also help address this truncation issue. For example, highly sensitive tracers (e.g., synthetic DNA (deoxyribonucleic acid) molecules) have been shown to reflect much longer residence time storage flowpaths [Foppen et al., 2013; Foppen et al., 2011]. The use of such tracers thereby allows for a better determination of which portion of mass loss may be long-term storage and, in turn, which portion is groundwater recharge. Otherwise, modeling techniques to approximate the late-time tail may be necessary to provide a more complete representation of the actual transit time distribution and support a better understanding of relationships between short- and long-term storage processes and discharge [e.g., Drummond et al., 2012; Stonedahl et al., 2012].

5.3. Temporal Metrics Do Not Reflect Observed Morphologic Patterns

Heterogeneity in stream morphologic characteristics is expected to cause differences in short-term storage [e.g., Gostner et al., 2013; Wondzell and Gooseff, 2013]. Likewise, this heterogeneity can cause dispersive processes to respond differently to changes in discharge. For example, increased turbulence due to higher discharge over features like riffles can increase dispersion, but higher discharge through slower moving sections like pools can reduce dispersion [Dyer and Thoms, 2006]. Based on this understanding, dispersion in sections with riffles (e.g., segment BC, Figure 2b) should increase with discharge compared to sections with pool-riffle sequences (e.g., segment DE, Figure 2b). The response of the window of detection ($\Delta t_{w,norm}$) to changing discharge was different between segments BC and DE where DE was consistently higher (Figure 6 and Table S5 in supporting information S1), which corresponds to the more expected spreading of the tracer over riffle features. However, the symmetrical spreading relative to mean arrival time (ΔCV) between segments BC and DE did not respond significantly differently to changing discharge, suggesting that the influence of spatially variable morphology was muted relative to changing advective timescales. The estimates of ΔCV within segment AB were consistently the largest, the least sensitive to changing discharge, and different from all other reach segments (Figure 6 and Table S5 in supporting information S1). Segment AB is a glide section with the steepest streambed slope (0.7%). We believe that the relatively uniform straight character and steep gradient of this segment (Figure 2b) caused the highest symmetrical spreading relative to mean arrival time while remaining the least sensitive to changing turbulent energy. Still, the means and medians of the normalized segment metrics were statistically equivalent between most segments, suggesting that the influence of segment spatial heterogeneity was muted or obscured by changes to transport timescales inherent in tracer observations. Only the most drastic differences in morphology appear to be reflected by tracer observations in our study.

5.4. Reach Scale Tracer Study Limitations May Complicate Upscaling

Appropriately representing the heterogeneity and dynamics of solute transport processes across a network is critical to predict solute transport to receiving water bodies [e.g., Kiel and Cardenas, 2014; Wollheim et al., 2008]. The approach of many previous studies to examine solute transport processes throughout a network has been to conduct solute tracer tests over fixed reaches of equal length regardless of differences in transport timescales or morphologic characteristics [e.g., Covino et al., 2011; Gooseff et al., 2013; Stewart et al., 2011], with some calling for a length-normalized metric to compare processes between reaches or discharges [Runkel, 2002]. This study indicates that a change in the transport timescale (i.e., set by the interacting advective time and window of detection) resulting from a change in discharge or reach length selection can artificially manifest as spatial heterogeneity or process change. Therefore, we argue that fixing reaches of equal length or using length-normalized metrics to compare stream solute tracer observations may not be the best approach. Rather, since we observed variability between segment morphologies and combining segments influences the interpretation of some transport processes, setting reach lengths and monitoring sites for tracer experiments should focus more on an understanding of transport timescales. One possible approach would be to fix the advective timescale between reaches, adjusting reach length accordingly. Ultimately, design of comparable tracer experiments may require preliminary studies or modeling efforts to initially quantify the range of transport timescales. Only comparable tracer observations will clarify relationships between solute transport processes and discharge. Improving the ability to compare tracer observations will likely lead to more accurate process-discharge relationships.

Spatial patterns of stream morphology (e.g., streambed slope and channel width) throughout a network may also be an important consideration when representing solute transport processes within a network model [e.g., Zarnetske et al., 2007]. Fortunately, it may be possible to represent the heterogeneity of advective velocity at long reach scales by identifying the spatial pattern of the stream channel from imagery [Schmadel et al., 2014b]. Advances in upscaling from reach-based studies to networks will require consideration of within-reach spatial heterogeneity, spatial patterns in discharge and advective velocity, more sophisticated techniques to map the inter-connected surface and subsurface waters, and a comprehensive analysis of tracer tests conducted over a range of discharges, spatial scales, and geologic settings. For example, tracer studies could be paired with independent and complementary measures of transport, such as through geophysics [e.g., Toran et al., 2013; Ward et al., 2012] or well networks in the adjacent aquifer [e.g., Voltz et al., 2013; Zarnetske et al., 2011], to better develop relationships between transient storage processes and discharge. However, from this study, one clear way forward is to better acknowledge and quantify the influence of changing advective timescales and windows of detection on the interpretation of processes such as dispersion and transient storage in our future reach scale solute tracer studies.

6. Conclusions

Using a suite of transport timescales (advective times and windows of detection) reflected in conservative stream solute tracer responses observed during a storm discharge period, we demonstrate that changes in advective times dominate and, consequently, mask other transport processes like dispersion and transient storage. While a possible interpretation of the tracer data could be that solute transport processes changed with discharge, we found through further analysis that the differences in the tracer data were generated primarily by variation in advective times. Furthermore, most individual segments within the original study reach did not respond differently to changes in discharge, suggesting that the influence of distinct channel morphologies was muted by the differences in advective times between tracer tests. Only the largest differences in morphologies were reflected in the tracer observations. We also found that through combining segments into longer reaches, differences of transport timescales could manifest as an incorrect interpretation of how processes change with discharge.

Based on these findings, this study provides general recommendations for future tracer studies. First, while a high-sensitivity tracer (fluorescein) provided more late-time tailing information than that obtained from a lower-sensitivity tracer (salt), both independently provided the same conclusion that the impact of changing advective times obscured other solute transport process changes. Furthermore, differences between these tracer types were smaller at higher discharges, indicating that tracer selection is less important than considering the influence of changing timescales on the interpretation of changing processes. Second, the influence of changing advective times and windows of detection must be established before tracer tests can be compared. Otherwise, a change in the transport timescale could be misinterpreted as a change in dispersion or transient storage. We show that this influence can be approximated by normalizing temporal metrics directly assessed from the observations by characteristics of the transport timescale including advective time and mean arrival time. However, the normalized metrics did not provide a sufficient correction to completely isolate dispersion and transient storage. Therefore, development of more complete relationships between processes and discharge may require better approximations of late-time tailing through other methods. Lastly, we recommend that selection of study reaches for design of tracer studies should be less influenced by length scales, such as fixed reaches of equal length proposed in previous studies, and more so by the transport timescales, which can be approximated from preliminary tests or modeling efforts. This recommendation is based on evidence that differences between windows of detection still limit the comparison of tracer responses despite normalization by characteristics such as the advective time. Improving the ability to compare tracer observations will clarify relationships between solute transport processes and discharge necessary to use reach scale observations to infer solute transport at larger network scales.

Acknowledgments:

The data collection campaign was funded by The Leverhulme Trust through the project “Where rivers, groundwater and disciplines meet: A hyporheic research network” with additional support from EU FP7-ITN “Interfaces: Ecohydrological interfaces as critical hotspots for transformations of ecosystem exchange fluxes and biogeochemical cycles” Grant No. 607150, and from the authors’ institutions. Tools for solute tracer time series analyses were developed by Ward and others with support provided in part by the National Science Foundation (NSF) Grant No. EAR-1331906 for the Critical Zone Observatory for Intensively Managed Landscapes (IML-CZO), a multi-institutional collaborative effort. A portion of Ward’s time in analysis of transport dynamics in agricultural landscapes was supported by NSF Grant No. EAR-1505309. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily represent the official views of sponsoring agencies. Data presented in this study are available upon request to the corresponding author. The authors declare no conflicts of interest.

References

Aquino, T., A. Aubeneau, and D. Bolster (2015), Peak and tail scaling of breakthrough curves in hydrologic tracer tests, Advances in Water Resources, 78, 1-8.

Bellin, A., D. Tonina, and A. Marzadri (2015), Breakthrough curve moments scaling in hyporheic exchange, Water Resources Research, 51(2), 1353-1358.

Boulton, A. J., T. Datry, T. Kasahara, M. Mutz, and J. A. Stanford (2010), Ecology and management of the hyporheic zone: stream–groundwater interactions of running waters and their floodplains, Journal of the North American Benthological Society, 29(1), 26-40.

Cirpka, O. A., M. N. Fienen, M. Hofer, E. Hoehn, A. Tessarini, R. Kipfer, and P. K. Kitanidis (2007), Analyzing Bank Filtration by Deconvoluting Time Series of Electric Conductivity, Ground Water, 45(3), 318-328.

Covino, T., B. McGlynn, and J. Mallard (2011), Stream-groundwater exchange and hydrologic turnover at the network scale, Water Resources Research, 47(12).

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, Journal of the North American Benthological Society, 12(3), 223-235.

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(G3).

Dudley-Southern, M., and A. Binley (2015), Temporal responses of groundwater-surface water exchange to successive storm events, Water Resources Research, 51(2), 1112-1126.

Dyer, F. J., and M. C. Thoms (2006), Managing river flows for hydraulic diversity: an example of an upland regulated gravel-bed river, River Research and Applications, 22(2), 257-267.

Fischer, H. B. (1975), Discussion of "Simple method for predicting dispersion in streams" by R.S. McQuivey and T.N. Keefer, Journal of Environmental Engineering Division American Society of Civil Engineers, 101(3), 453-455.

Foppen, J. W., J. Seopa, N. Bakobie, and T. Bogaard (2013), Development of a methodology for the application of synthetic DNA in stream tracer injection experiments, Water Resources Research, 49(9), 5369-5380.

Foppen, J. W., C. Orup, R. Adell, V. Poulalion, and S. Uhlenbrook (2011), Using multiple artificial DNA tracers in hydrology, Hydrological Processes, 25(19), 3101-3106.

Gelhar, L. W., C. Welty, and K. R. Rehfeldt (1992), A critical review of data on field-scale dispersion in aquifers, Water Resources Research, 28(7), 1955-1974.

González-Pinzón, R., R. Haggerty, and M. Dentz (2013), Scaling and predicting solute transport processes in streams, Water Resources Research, 49(7), 4071-4088.

Gooseff, M. N., R. O. Hall, and J. L. Tank (2007), Relating transient storage to channel complexity in streams of varying land use in Jackson Hole, Wyoming, Water Resources Research, 43(1), W01417.

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(0), 16-25.

Gostner, W., P. Parasiewicz, and A. J. Schleiss (2013), A case study on spatial and temporal hydraulic variability in an alpine gravel-bed stream based on the hydromorphological index of diversity, Ecohydrology, 6(4), 652-667.

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), 18-11-18-14.
Harvey, J. W., and B. J. Wagner (2000), Quantifying hydrologic interactions between streams and their subsurface zones, in Streams and Ground Waters, edited by J. B. Jones and P. J. Mulholland, pp. 3-44, Academic Press, San Diego, CA.
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(8), 2441-2451.
Hester, E. T., and M. N. Gooseff (2010), Moving Beyond the Banks: Hyporheic Restoration Is Fundamental to Restoring Ecological Services and Functions of Streams, Environmental Science & Technology, 44(5), 1521-1525.
Jackson, T. R., R. Haggerty, and S. V. Apte (2013), A fluid-mechanics based classification scheme for surface transient storage in riverine environments: quantitatively separating surface from hyporheic transient storage, Hydrol. Earth Syst. Sci., 17(7), 2747-2779.
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 Resources Research, 49(9), 5290-5306.
Kiel, B. A., and M. B. Cardenas (2014), Lateral hyporheic exchange throughout the Mississippi River network, Nature Geosci, 7(6), 413-417.
Kilpatrick, F. A., and E. D. Cobb (1985), Measurement of discharge using tracers, U.S. Geological Survey. Techniques of Water-Resources Investigations, Book 3 (Chapter A16, 52pp.).
Leopold, L. B., and T. Maddock (1953), The Hydraulic Geometry of Stream Channels and Some Physiographic Implications, Geological Survey Professional Paper 252, U.S. Department of the Interior, Washington, D.C.
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(0), 125-136.
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(11).
Payn, R. A., M. N. Gooseff, D. A. Benson, O. A. Cirpka, J. P. Zarnetske, W. B. Bowden, J. P. McNamara, and J. H. Bradford (2008), Comparison of instantaneous and constant-rate stream tracer experiments through non-parametric analysis of residence time distributions, Water Resources Research, 44(6).
Riml, J., and A. Wörman (2011), Response functions for in-stream solute transport in river networks, Water Resources Research, 47(6), W06502.
Runkel, R. L. (2002), A new metric for determining the importance of transient storage, Journal of the North American Benthological Society, 21(4), 529-543.
Schmadel, N. M., B. T. Neilson, and D. K. Stevens (2010), Approaches to estimate uncertainty in longitudinal channel water balances, Journal of Hydrology, 394(3–4), 357-369.
Schmadel, N. M., B. T. Neilson, and T. Kasahara (2014a), Deducing the spatial variability of exchange within a longitudinal channel water balance, Hydrological Processes, 28(7), 3088-3103.

Schmadel, N. M., B. T. Neilson, J. E. Heavilin, D. K. Stevens, and A. Wörman (2014b), The influence of spatially variable stream hydraulics on reach scale solute transport modeling, Water Resources Research, 10.1002/2014WR015440.

Schmid, B. (2002), Persistence of Skewness in Longitudinal Dispersion Data: Can the Dead Zone Model Explain It After All?, Journal of Hydraulic Engineering, 128(9), 848-854.

Schmid, B. H., I. Innocenti, and U. Sanfilippo (2010), Characterizing solute transport with transient storage across a range of flow rates: The evidence of repeated tracer experiments in Austrian and Italian streams, Advances in Water Resources, 33(11), 1340-1346.

Schmidt, C., A. Musolff, N. Trauth, M. Vieweg, and J. H. Fleckenstein (2012), Transient analysis of fluctuations of electrical conductivity as tracer in the stream bed, Hydrol. Earth Syst. Sci., 16(10), 3689-3697.

Stewart, R. J., W. M. Wollheim, M. N. Gooseff, M. A. Briggs, J. M. Jacobs, B. J. Peterson, and C. S. Hopkinson (2011), Separation of river network–scale nitrogen removal among the main channel and two transient storage compartments, Water Resources Research, 47(10), W00J10.

Stonedahl, S. H., J. W. Harvey, J. Detty, A. Aubeneau, and A. I. Packman (2012), Physical controls and predictability of stream hyporheic flow evaluated with a multiscale model, Water Resources Research, 48(10), W10513.

Stream Solute Workshop (1990), Concepts and Methods for Assessing Solute Dynamics in Stream Ecosystems, Journal of the North American Benthological Society, 9(2), 95-119.

Toran, L., J. E. Nyquist, A. C. Fang, R. J. Ryan, and D. O. Rosenberry (2013), Observing lingering hyporheic storage using electrical resistivity: variations around stream restoration structures, Crabby Creek, PA, Hydrological Processes, 27(10), 1411-1425.

Trauth, N., C. Schmidt, M. Vieweg, S. E. Oswald, and J. H. Fleckenstein (2015), Hydraulic controls of in-stream gravel bar hyporheic exchange and reactions, Water Resources Research, 51(4), 2243-2263.

Voltz, T., M. Gooseff, A. S. Ward, K. Singha, M. Fitzgerald, and T. Wagener (2013), Riparian hydraulic gradient and stream-groundwater exchange dynamics in steep headwater valleys, Journal of Geophysical Research: Earth Surface, 118(2), 953-969.

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(7), 1731-1741.

Ward, A. S., M. N. Gooseff, M. Fitzgerald, T. J. Voltz, and K. Singha (2014), Spatially distributed characterization of hyporheic solute transport during baseflow recession in a headwater mountain stream using electrical geophysical imaging, Journal of Hydrology, 517(0), 362-377.

Ward, A. S., M. Fitzgerald, M. N. Gooseff, T. J. Voltz, A. M. Binley, and K. Singha (2012), Hydrologic and geomorphic controls on hyporheic exchange during base flow recession in a headwater mountain stream, Water Resources Research, 48(4), W04513.

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(9), 5473-5486.

Ward, A. S., N. M. Schmadel, S. M. Wondzell, C. Harman, M. N. Gooseff, and K. Singha (2016), Hydrogeomorphic controls on hyporheic and riparian transport in two headwater

mountain streams during base flow recession, Water Resources Research, 10.1002/2015WR018225.
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(6), 3359-3374.
Wollheim, W. M., B. J. Peterson, S. M. Thomas, C. H. Hopkinson, and C. J. Vörösmarty (2008), Dynamics of N removal over annual time periods in a suburban river network, Journal of Geophysical Research: Biogeosciences, 113(G3).
Wondzell, S. M. (2006), Effect of morphology and discharge on hyporheic exchange flows in two small streams in the Cascade Mountains of Oregon, USA, Hydrological Processes, 20(2), 267-287.
Wondzell, S. M. (2011), The role of the hyporheic zone across stream networks, Hydrological Processes, 25(22), 3525-3532.
Wondzell, S. M., and M. N. Gooseff (2013), 9.13 Geomorphic Controls on Hyporheic Exchange Across Scales: Watersheds to Particles, in Treatise on Geomorphology, edited by J. F. Shroder, pp. 203-218, Academic Press, San Diego.
Zarnetske, J. P., R. Haggerty, S. M. Wondzell, and M. A. Baker (2011), Dynamics of nitrate production and removal as a function of residence time in the hyporheic zone, Journal of Geophysical Research: Biogeosciences, 116(G1).
Zarnetske, J. P., M. N. Gooseff, T. R. Brosten, J. H. Bradford, J. P. McNamara, and W. B. Bowden (2007), Transient storage as a function of geomorphology, discharge, and permafrost active layer conditions in Arctic tundra streams, Water Resources Research, 43(7), W07410.
Zimmer, M. A., and L. K. Lautz (2014), Temporal and spatial response of hyporheic zone geochemistry to a storm event, Hydrological Processes, 28(4), 2324-2337.

Table 1. The reach scale temporal metrics, window of detection ($t_{w,norm}$) and transient storage index (TSI_{norm}) normalized by the advective time, coefficient of variation (CV), and skewness (γ), of the transit time distributions measured at in-stream monitoring sites A through G and their relationships (slope) with changing stream discharge (Q) (also see Figure 5, right column). Note that 41 salt and 4 fluorescein breakthrough curves were used to estimate these metrics and relationships. A positive or negative slope indicates the metrics increase or decrease with an increasing Q, respectively. A slope is considered significant (bold text) if zero is outside of the associated 95% confidence interval.

Site	x (m)	Number of Qs	$t_{w,norm}$		TSI_{norm}		CV		γ	
			Slope with Q	R^2	Slope with Q	R^2	Slope with Q	R^2	Slope with Q	R^2
A	194	6	2.0E-02	0.07	-1.4E-02	0.03	-1.8E-02	0.60	**4.6E-01**	0.97
B	294	7	-5.4E-02	0.40	-1.5E-02	2.2E-02	-6.8E-03	0.37	-5.5E-04	6.0E-06
C	428	7	-8.4E-02	0.53	-6.2E-02	0.48	**-1.1E-02**	0.76	-8.2E-03	4.7E-03
D	559	7	-8.6E-02	0.42	-7.2E-02	0.30	**-1.2E-02**	0.60	-2.0E-02	2.6E-02
E	667	5	-1.2E-01	0.62	-9.5E-02	0.54	-1.6E-02	0.72	-1.1E-01	0.43
F	819	6	3.9E-02	0.09	5.1E-02	0.11	-1.2E-03	3.5E-03	1.8E-01	0.31
G	928	7	-7.6E-02	0.24	-5.3E-02	0.10	-8.6E-03	0.18	4.4E-03	3.5E-04

Bold indicates significant value

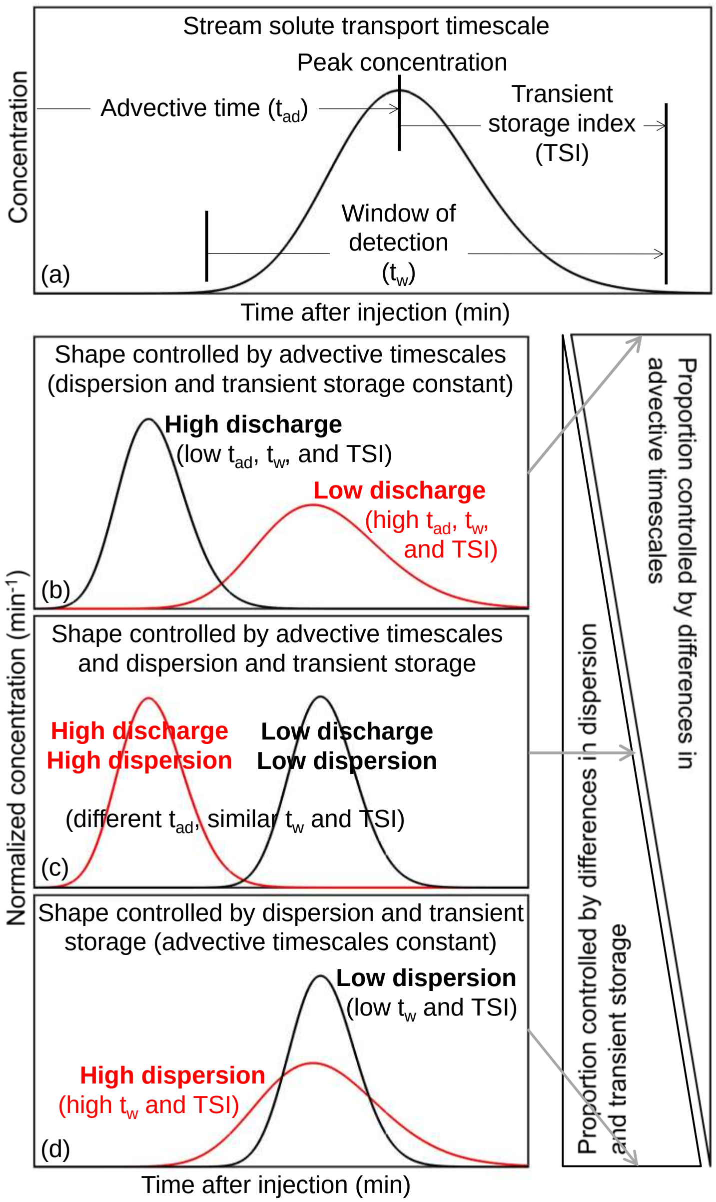

Figure 1. (a) The stream solute transport timescale, defined here as the advective time (t_{ad}) and window of detection (t_w), can be directly estimated from an observed in-stream conservative solute tracer time series (breakthrough curve, or BTC). Following an instantaneous injection, the elapsed time from tracer injection to peak concentration describes t_{ad}. The elapsed time from tracer first arrival to last detection describes t_w. The elapsed time from t_{ad} to tracer last detection provides an indicator of transient storage (transient storage index, or TSI). Below is an illustration of the challenge of interpreting solute transport processes from stream solute tracers. (b) At one extreme, if underlying processes (e.g., dispersion and transient storage) are constant but t_{ad} is different between high and low discharge conditions, the shape of the BTC is controlled by differences in advective timescales. (d) At the other extreme, if t_{ad} is constant between two reaches, the shape of the BTC is controlled by differences in the underlying processes. (c) In practice, shapes of observed BTCs are controlled by different proportions of advective timescales and these underlying processes.

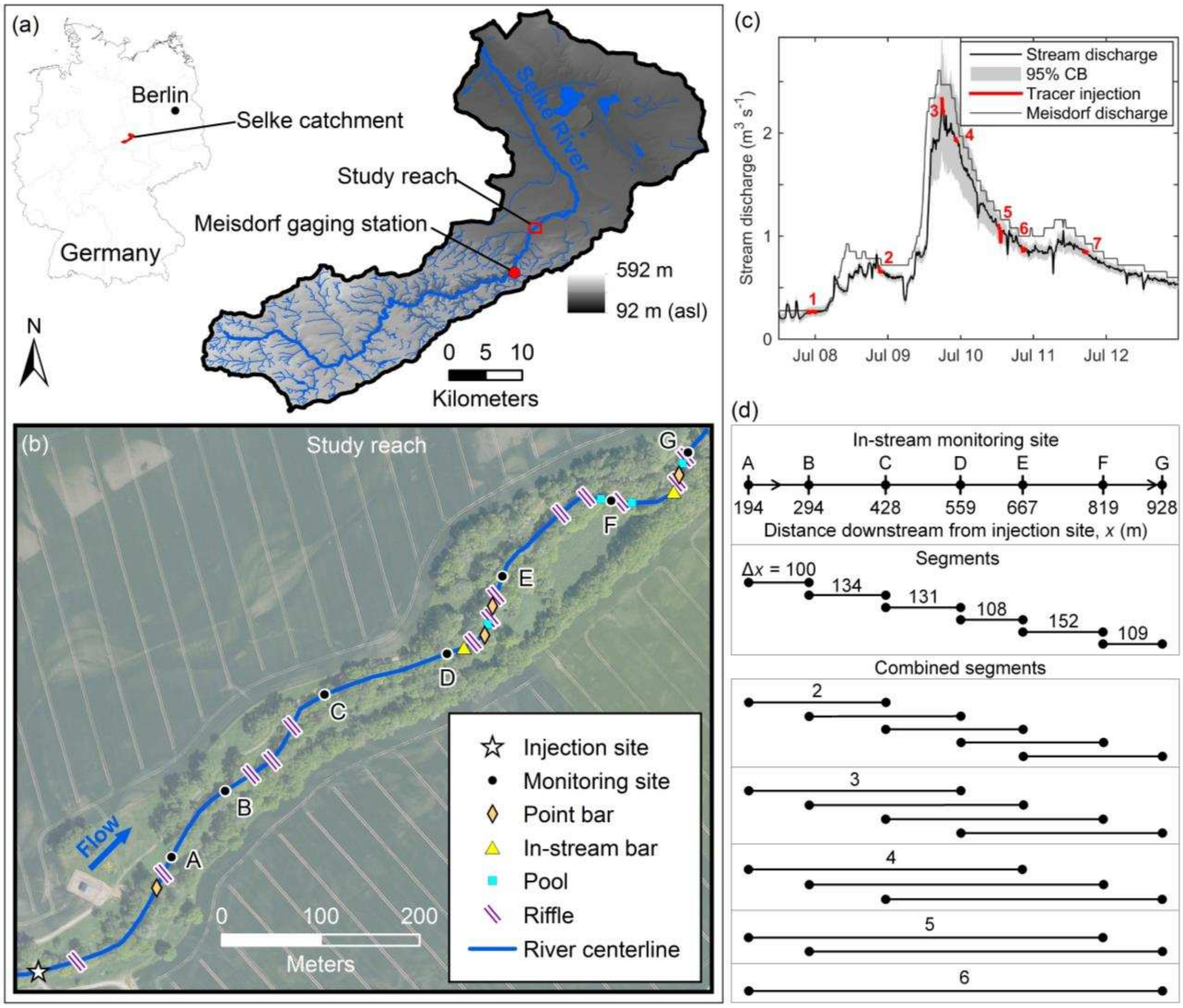

Figure 2. (a) The Selke River catchment located in central Germany and the location of the study reach and Meisdorf federal gaging station. (b) In-stream monitoring sites and general channel morphologic characteristics of the study reach. (c) Stream discharge estimated within the study reach from an established rating curve (15 m upstream of site E) and the associated 95% confidence bounds (CB), timing of tracer test durations (elapsed time from injection to last detection at site G), and discharge at the Meisdorf gaging station located ~4.6 km upstream of the injection site. (d) The monitoring sites delineate six reach segments, which were recombined into every possible unique combination (grouped by the number of individual segments) to produce segments of various lengths and transport timescales.

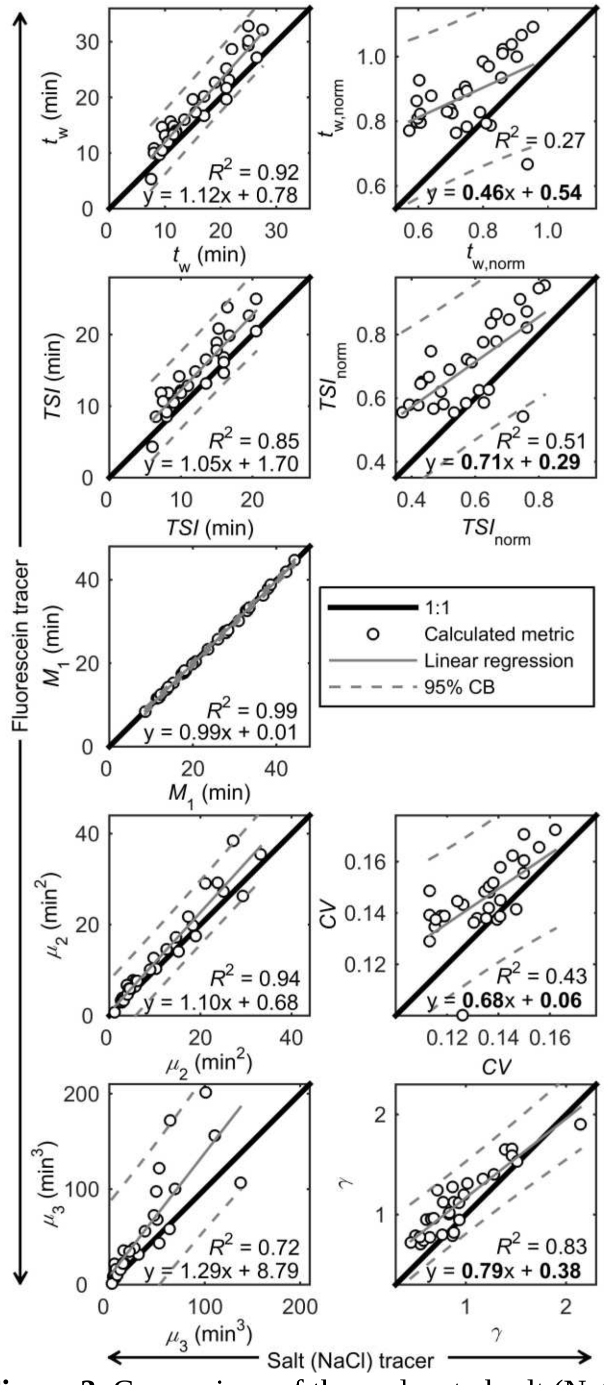

Figure 3. Comparison of the co-located salt (NaCl) and fluorescein breakthrough curves measured at in-stream monitoring sites A through G (i.e. 29 each salt and fluorescein). In the left column, the metrics compared are the window of detection (t_w), transient storage index (TSI), mean arrival time (M_1), variance (μ_2), and third central moment (μ_3). In the right column, the normalized metrics are compared, which include t_w and TSI normalized by the associated advective time ($t_{w,norm}$ and TSI_{norm}), the coefficient of variation (CV), and skewness (γ). Corresponding linear trend-lines are shown along with their slopes, intercepts, coefficients of determination (R^2), and 95% confidence bounds (CB). Bold text indicates that either the slope is statistically different than 1 or the intercept is statistically different than zero.

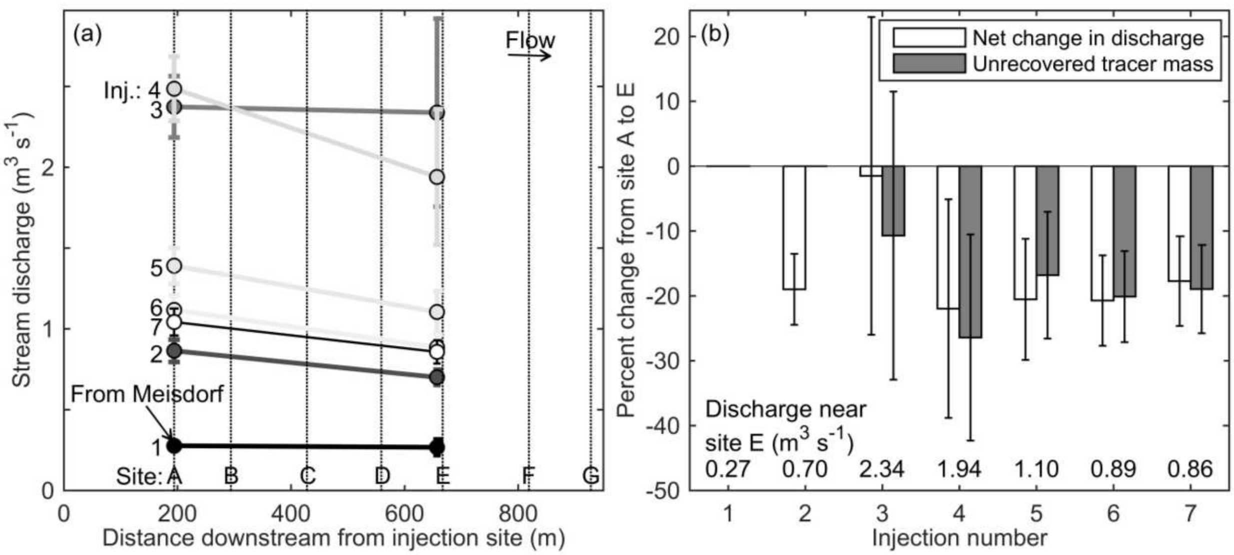

Figure 4. (a) Discharge estimates at site A (estimated from dilution gaging) and near site E (estimated from established rating curve). (b) Percent net change in discharge from site A to E and percent unrecovered salt tracer mass from the injection to site E. Error bars are the estimated 95% confidence intervals.

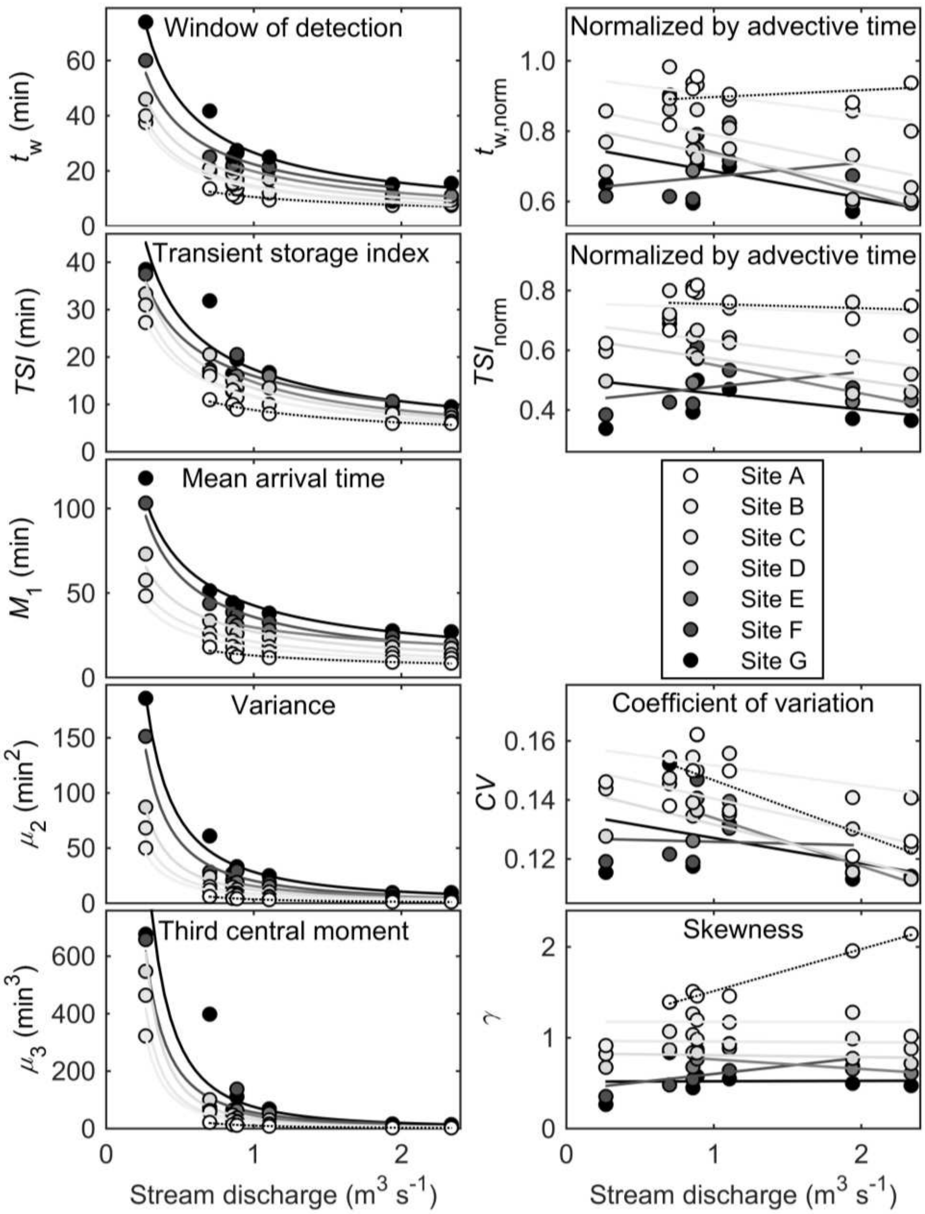

Figure 5. In the left column, the reach scale temporal metrics of the transit time distributions measured at in-stream monitoring sites A through G. Specifically, these metrics are the window of detection (t_w), transient storage index (TSI), mean arrival time (M_1), variance (μ_2), and third central moment (μ_3). Note that 41 salt and 4 fluorescein breakthrough curves were used to estimate these metrics. Each metric decreases with increasing discharge and increases with increasing travel distance (shown by a general site-specific exponential function and white to black color, respectively). To better compare across discharges, shown in the right column are t_w and TSI normalized by the advective time ($t_{w,norm}$ and TSI_{norm}), the coefficient of variation (CV), and skewness (γ). Linear trend-lines were estimated for each site and a slope is considered significant if zero is outside of the associated 95% confidence interval (see Table 1).

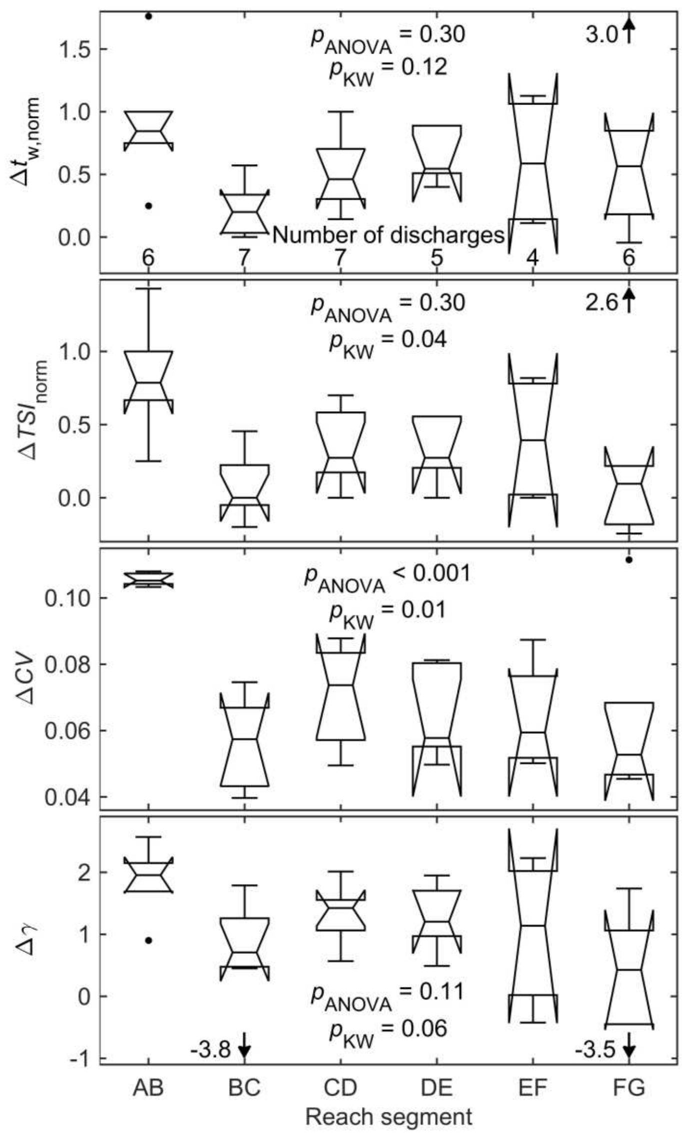

Figure 6. Segment-wise distributions (box-and-whisker plots of the quartiles where the notches are the 95% confidence intervals about the medians) of the transfer function metrics that express the window of detection ($\Delta t_{w,norm}$) and transient storage index (ΔTSI_{norm}) normalized by the change in advective time, coefficient of variation (ΔCV), and skewness ($\Delta\gamma$) over changing discharge. Note that 41 salt and 4 fluorescein breakthrough curves were used to estimate these metrics, resulting in 35 unique values (i.e., segment-wise estimates corresponding to number of discharges). Distribution means were compared through a one-way analysis of variance (ANOVA) and medians were compared through a Kruskal-Wallis one-way ANOVA. For mean and median comparisons, a $p_{ANOVA} < 0.05$ or $p_{KW} < 0.05$ indicates the mean or median of at least one segment is statistically different from all others, respectively. See Table S5 in supporting information S1 for segment-to-segment pairwise comparisons.

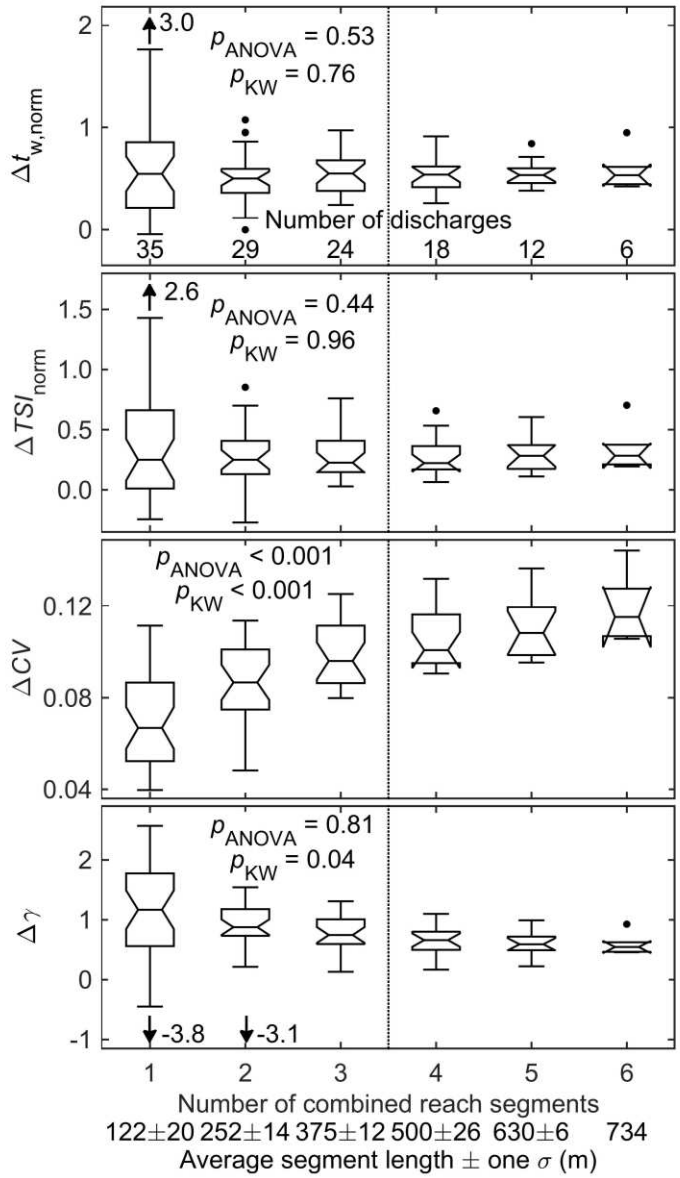

Figure 7. The temporal metrics of the transfer function, the window of detection ($\Delta t_{w,norm}$) and transient storage index (ΔTSI_{norm}) normalized by the advective time, coefficient of variation (ΔCV), and skewness ($\Delta\gamma$), lumped together for each number of individual segments combined together (see Figure 2d). The corresponding distributions are represented as box-and-whisker plots of the quartiles where the notches are the 95% confidence intervals about the medians. Each distribution contains unique estimates representing different segment lengths and discharges (denoted as the number of discharges). Distribution means were compared through a one-way analysis of variance (ANOVA) and medians through a Kruskal-Wallis one-way ANOVA. A $p_{ANOVA} < 0.05$ or $p_{KW} < 0.05$ indicates the mean or median of at least one distribution is statistically different from all others, respectively. Only the distributions representing the original segments, two combined segments, and three combined segments (see Figure 2d) were compared (vertical dashed line).

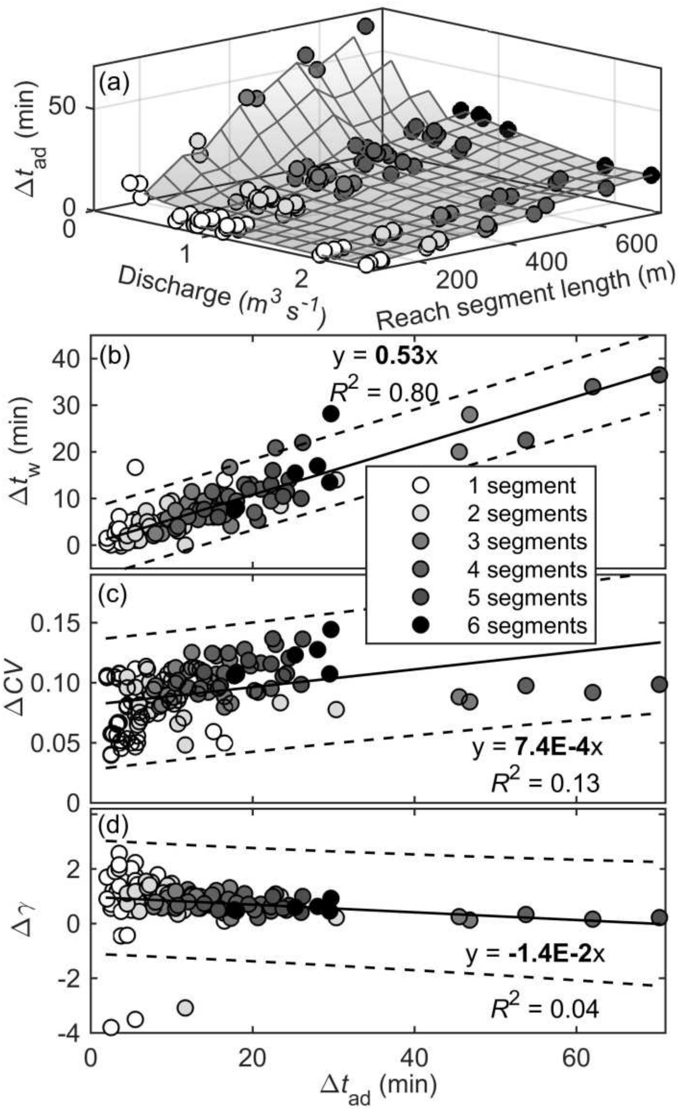

Figure 8. (a) The advective time of the transfer function (Δt_{ad}) relative to discharge and reach segment length. A surface was linearly interpolated between these estimates to illustrate the general trend. (b) The window of detection of the transfer function (Δt_w) relative to Δt_{ad} for each segment combination. A linear trend-line (solid line) and associated 95% confidence bounds (dashed lines) were approximated. The slope of this trend-line and coefficient of determination (R^2) are shown. (c) The coefficient of variation (ΔCV) and (d) skewness ($\Delta \gamma$) for each segment combination relative to the corresponding Δt_{ad} estimates and corresponding trend-line and 95% CBs. Bold text indicates a significant value (i.e., zero is outside of the associated 95% confidence interval).