ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR032718

Testing Hidden Assumptions of Representativeness in Reach-Scale Studies of Hyporheic Exchange

Paige S. Becker[1,2], Adam S. Ward[1,2], Skuyler P. Herzog[3], and Steven M. Wondzell[4]

[1]O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, IN, USA, [2]Department of Biological and Ecological Engineering, Oregon State University, Corvallis, OR, USA, [3]Natural Resources, Oregon State University Cascades, Bend, OR, USA, [4]Pacific Northwest Research Station, U.S. Forest Service, United States Department of Agriculture, Corvallis, OR, USA

Key Points:

- Reach length strategy and location both determine how representative a reach is compared to the larger stream segment in which it is located
- Researchers must clarify, test, and quantify assumptions of within- and between-reach variation
- Assumed behavior cannot be inferred from visual inspection and iteration between field and modeling studies minimize uncertainties in hyporheic exchange

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. S. Becker,
beckerpa@oregonstate.edu

Citation:

Becker, P. S., Ward, A. S., Herzog, S. P., & Wondzell, S. M. (2023). Testing hidden assumptions of representativeness in reach-scale studies of hyporheic exchange. *Water Resources Research*, *59*, e2022WR032718. https://doi.org/10.1029/2022WR032718

Received 3 MAY 2022
Accepted 18 DEC 2022

Abstract Field studies of hyporheic exchange in mountain systems are often conducted using short study reaches and a limited number of observations. It is common practice to assume these study reaches represent hyporheic exchange at larger scales or different sites and to infer general relationships among potential causal mechanisms from the limited number of observations. However, these assumptions of representativeness are rarely tested. In this study, we develop numerical models from four segments of mountain streams in different geomorphologic settings and extract shorter reaches to test how representative exchange metrics are in shorter reaches compared to their reference segments. We also map the locations of the representative reaches to determine if a pattern exists based on location. Finally, we compare variance of these shorter within-site reaches to 29 additional reaches across the same basin to understand the impacts of inferring causal mechanisms, for example, the expectation that wide and narrow valley bottoms will yield different hyporheic exchange patterns. Our results show that the location and length strategy of the study reach must be considered before assuming an exchange metric to be representative of anything other than the exact segment studied. Further, it is necessary to quantify within and between site variations before making causal inferences based on observable characteristics, such as valley width or stream morphology. Our findings have implications for future field practices and how those practices are translated into models.

Plain Language Summary Hydrology models are used to generalize processes and make predictions about how water moves through the environment. For example, models are used to study how water moves in and out of streambeds because processes occurring in the streambed can significantly impact the water quality. These models assume that data and observations from short sections of streams can represent longer portions of a stream or similar streams in different places. However, short sections of the stream may not accurately represent the conditions of the whole site that is being modeled, meaning that the assumptions and ultimately the models are misrepresenting the process. To test this, we compared how water was exchanged between a long segment of a stream and its subsurface to exchanges occurring in a shorter section of the stream and its subsurface. We found that only some of the shorter sections of the stream we modeled accurately represented the longer segment of the stream from which it was taken. These findings are important for future field studies as they demonstrate how scientists need to consider several aspects of the stream before they choose a section of a stream as the basis of their assumption for a model.

1. Introduction

The importance of hyporheic exchange (i.e., the movement of water and solutes between flowing surface waters and their adjacent subsurface domains) to a host of ecosystem services and functions (e.g., Findlay, 1995; Stanford & Ward, 1988; Wondzell, 2011) has motivated a desire to make predictions of these exchange fluxes at large scales (Boano et al., 2014; Harvey & Gooseff, 2015). However, our current understanding of hyporheic exchange is built on a foundation of observations from a small number of place- and time-specific studies that are subsequently generalized to inform predictions at larger scales (e.g., Magliozzi et al., 2018; Ward & Packman, 2019). Consequently, predictions at larger spatial scales or unstudied sites rely upon the untested (and often hidden) assumption that idiosyncratic field and model studies provide an understanding that can be used to inform predictions of exchange transferred at different spatial locations or hydrologic conditions (Ward et al., 2018a). In other words, there is a hidden assumption that reach-scale field and model studies are both *accurate* (i.e., there is not a statistical bias between the studied reach and the larger segment from which the study reach was selected) and *precise* (i.e., the findings are replicable with relatively small variability). For studies of hyporheic exchange, these factors manifest as a result of the particular location and length selected for a studied reach. Here, we adopt the working definition of "representative" to mean that interpreted metrics of exchange for a study reach do not significantly change as a function of the exact study location within a larger segment nor the strategy used to select a study within a river segment. To our knowledge, no prior study has quantitatively assessed whether reach-scale studies of hyporheic exchange are representative, despite this being a necessary condition for making meaningful interpretations of field data and extrapolation to network scales. Nonetheless, there have been several efforts to translate reach-scale findings into predictions across river networks (e.g., Cardenas, 2009; Covino et al., 2011; Gomez-Velez et al., 2015). Our goal in this study is to assess the extent to which reach-scale findings are representative of the larger segments within which they are located, with the long-term goal of validating or improving present upscaling techniques.

Common methods in studying hyporheic exchange assume study reaches (i.e., the exact places where field or model experiments are conducted, commonly tens to hundreds of meters in length (Montgomery & Buffington, 1998)); are representative of larger spatial scales (i.e., segments, each comprised of several reaches (Montgomery & Buffington, 1998)); though this is seldom explicitly stated. However, in the study of streams and rivers, practices for selecting study reach length and location are highly variable including systematic selection of equal lengths (e.g., Payn et al., 2009), random selection of locations (e.g., Anderson et al., 2005), intentionally biased selection to represent expected end-members (Wondzell, 2006), or to isolate other specific attributes (e.g., human impacts; Ward et al., 2018), and based on geomorphic characteristics (Leopold et al., 1964). Additionally, fixed study reach length (Payn et al., 2009; Wondzell et al., 2019), fixed transit timescales (Ward et al., 2018), and adaptive study reach lengths such as using a multiplier of Wetted Channel Widths (WCW) (Anderson et al., 2005; Day, 1977; Fitzpatrick et al., 1998; Frissell et al., 1986; Grant et al., 1990; Kasahara & Wondzell, 2003; Leopold et al., 1964; Montgomery & Buffington, 1997) have all been used in an attempt to control for expected variations that will occur as a function of reach selection (Schmadel et al., 2016). This breadth of approaches results in a seemingly arbitrary way to select a study reach, spanning from a few meters to hundreds of meters in studies with similar objectives (e.g., 7.6 m from Fabian et al. [2011] to 303 m from Zarnetske et al. [2011]). The lack of a common strategy among river scientists is particularly troubling because the hyporheic exchange is known to be controlled by processes occurring over a wide range of spatial scales (Wondzell et al., 2019), including regional groundwater gains and losses (Boano et al., 2008; Malzone et al., 2016), lateral inflows from hillslopes, geological discontinuities (Tonina & Buffington, 2009), and larger-scale features causing turnover of intermediate flow paths (Herzog et al., 2019) or down-valley flow (Ward et al., 2018a). Thus, if the reach length is too short, longer flow paths in the hyporheic zone are ignored (Findlay, 1995; Gooseff et al., 2003, 2006; Wondzell, 2006). The issue of too short of study reaches was evident in past studies where the same stream reach can show very different values of solute residence time and indicate different hydrological processes depending on if the water movement is studied along an entire segment or in its smaller reaches (Bencala, 1983).

This raises the question of if the reaches used in past studies were representative of hyporheic fluxes and transit times of the segments where they were studied. Despite it being known that the length and location of study reaches are associated with different hydrological processes (Kelleher et al., 2013) we still have no clear understanding of how metrics addressing hyporheic exchange change with scale and method by which reach lengths are selected. Put another way, we have a body of idiosyncratic studies that form the foundation of our understanding of exchange processes and inform our conceptual models, but how to interpret these consistently or synthesize our understanding remains an open and pressing issue (Ward & Packman, 2019). This presents a Catch-22: the potentially biased or incomplete results of field studies from short reaches are used to develop conceptual models, and these models are used as the basis for generating hypotheses about locations, timescales, and magnitudes of exchange which are—in turn—used to plan field studies. We expect that the assumption that reaches are representative of larger segments has given rise to a dominance of conceptual models focused on feature scale exchange while ignoring other scales and drivers of exchange.

Hyporheic exchange studies commonly infer causal mechanisms from relatively small bodies of empirical observations, with potentially conflicting results (Ward & Packman, 2019). In these cases, variation is attributed to visually apparent differences between study reaches without requiring a mechanistic understanding of whether between-reach differences are larger than the uncertainty associated with the arbitrary selection of study reaches. As an example, we critically consider the 40+ person-years of effort that we, the authors, have invested in studying within- and between-catchment differences in the paired study of WS01 and WS03 at the

10.1029/2022WR032718

H.J. Andrews Experimental Forest, two headwater basins with similar catchments but differences in valley width (Wondzell, 2006). Despite a body of publications critically comparing the wider and narrower study reaches within these basins (Nakamura & Swanson, 1993; Swanson & James, 1975; Wondzell, 2006), and even one citing the more and less constrained areas within WS03 (Ward et al., 2012), we have never critically asked if the study locations are representative of the wide- and narrow-valley conditions they were intended to represent when selected. Moreover, the selection of these sites was intentionally biased to represent reasonably wide and narrow segments of similar headwater basins that were reasonably accessible and logistically feasible to study hand-driven riparian wells (Wondzell, 2006). Once established, the study sites and their well networks were used in many subsequent studies, and results were often analyzed and interpreted as being representative of feature-scale exchange in mountainous headwater streams (e.g., Ward et al., 2016; Wondzell, 2011) or compared to randomly selected study reaches (Ward, Wondzell, et al., 2019). While we hope our own experience and history in these sites is a notable exception rather than a norm, we fear the latter may be true in a discipline where field-scale observations are costly, not readily repeatable in controlled conditions, and where inference of mechanisms from relatively small bodies of empirical observations is commonplace (Burt & McDonnell, 2015).

The overarching objective of this study is to assess the degree to which reach-scale studies are precise and accurate (i.e., "representative") observations of hyporheic exchange in the watersheds where they are conducted. Specifically, we ask (a) how do reach-scale simulations of hyporheic transit time, exchange flux, fraction of streambed upwelling, and reaction significance factor vary within headwater basins as a function of the study reach location and length strategy? and (b) is variation within an individual stream segment larger than variation between stream segments in different geologic settings? To answer these questions, we use groundwater flow models of the hyporheic zone in contrasting geologies at the H.J. Andrews Experimental Forest. We modeled hyporheic exchange along entire stream segments in four unique geologic settings, subdividing each stream segment into sets of fixed-length study reaches. We analyzed results using study lengths of both 20WCW and 100 m, both of which have been previously used in the basin (Anderson et al., 2005; Ward, Wondzell, et al., 2019). The 20WCW strategy is adaptive, accounting for the potential scaling of transport processes while the 100 m fixed length strategy removes bias selection from the study (Payn et al., 2009). We compared the hyporheic exchange metrics between these reach selection strategies to answer practical and theoretical questions including how representative are reach scale studies of larger segments, reach is defined as a part of the stream that exhibits similar bedforms and typically at the scale of 10–100s of m, and the segment is defined as a portion of the drainage network showing similar valley scale morphologies, typically hundreds to thousands of m in length (Montgomery & Buffington, 1998), does this change our current understanding, and how should we design field and model studies given the results of this study?

2. Methods

2.1. Field Characterization

This study was conducted at the H.J. Andrews Experimental Forest (HJA), Cascade Mountains, Oregon, USA, a 6,400-ha drainage basin with elevations ranging from 410 to 1,630 m above sea level. The forest receives an average of 2.1 m of rain a year (Segura et al., 2019). We studied streams spanning the three major landform types in the basin. In the lower elevations of the HJA, the geology is dominated by upper Oligocene-lower Miocene basaltic flows (named the Little Butte Formation), characterized by narrow V-shaped valleys with steep hillslopes. In the higher elevations, the Sardine Formation overtops the basaltic flows and landforms primarily consist of glacial cirques. Catchments in this landform type have characteristics of u-shaped valleys with uniform lateral tributary area and pool step morphology (Swanson & James, 1975; Ward, Wondzell, et al., 2019). Finally, deep-seated earth flows characterized by poorly developed channel networks and lack of lateral contributing area, forming parallel streams that are actively meandering, braiding, and downcutting are present at several locations in the basin (Caine & Swanson, 1989). Additional site details are available in a host of past studies (Anderson et al., 2005; Segura et al., 2019; Swanson & Jones, 2002).

For our primary study, we focused on four stream segments that spanned the three landform types (hereafter "reference segments"). Topographic surveys of the stream thalweg and channel surface profile of each segment were collected in 2015 and 2016, with surveyed lengths ranging from 247 to 542 m of stream centerline (Ward, Wondzell, et al., 2019; Ward, Zarnetske, et al., 2019). The full surveyed lengths of these four segments were used to quantify within-site variation as a function of study reach location and length strategy. Complementary valley

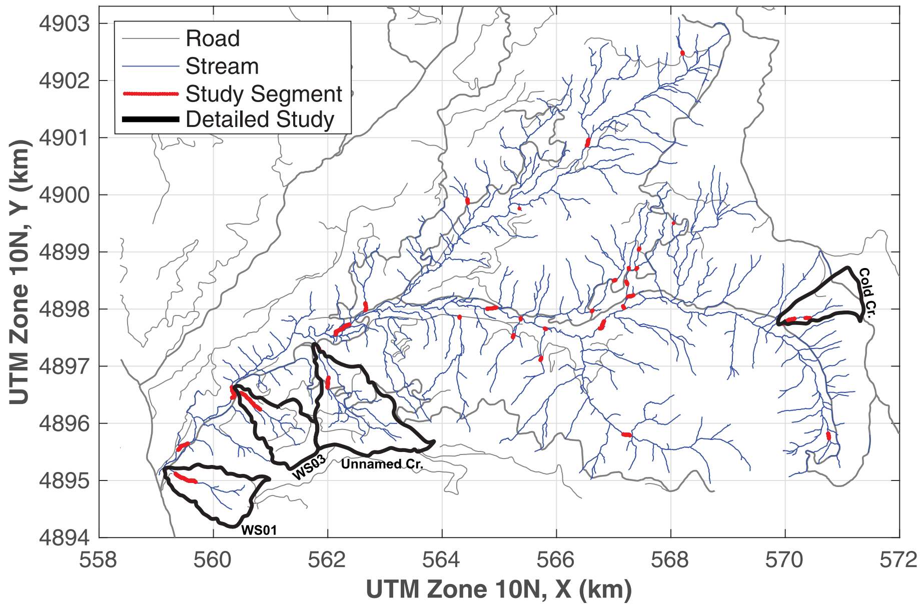

Figure 1. Maps of the HJ Andrews including the road network (gray), stream network (blue), and study sites (red). The four catchments outlined in black are those where more detailed surveys and simulations were included to assess within-site variation. Catchments, roads, and streams follow exactly those detailed in Ward, Wondzell, et al. (2019).

and channel morphology data were collected including wetted channel width, valley width, and several metrics derived from topographic analysis including drainage area, valley slope, and stream slope. Additionally, we have comparable data from 12 sites surveyed in 2004 by Anderson et al., 2005 and 17 sites surveyed by our team in 2019 that included a longitudinal profile equivalent in length to 20WCW across landform types for third order and smaller streams (Figure 1). These 29 reaches surveyed at 20WCW provide a basis to compare variation across low-order streams within the Lookout Creek basin to within-reach variation for the four reference segments.

2.2. Numerical Simulations

We constructed two-dimensional profile models of the stream along its centerline using COMSOL Multiphysics based on the surveyed streambed and water surface for each segment, following the same protocols as several past studies at the site (Gooseff et al., 2006; Herzog et al., 2019; Schmadel et al., 2017; Ward et al., 2018b). Briefly, surveyed streambed topography was used to define the shape of the sediment-water interface. No-flow boundaries were used to define the upstream, downstream, and bottom boundaries of the model domain. Average depth of the sediment for each reach was based on stream order, with two m for first-order, three m for second-order, and four m for third-order streams (Gooseff et al., 2006; Schmadel et al., 2017) and was offset from a linear regression line fit to the streambed topography. We note the planar bedrock assumption is consistent with prior studies and modeling efforts, but does inevitably impact the simulated flow field. Some sections of streambed were very shallow in the model (i.e., <10 cm), consistent with field observations of bedrock outcrops in many stream segments that cause turnover of down-valley flows (Herzog et al., 2019). Sediment was parameterized as homogeneous and isotropic with a hydraulic conductivity of 7×10^{-5} m/s (Kasahara & Wondzell, 2003), and a porosity of 0.2 (Schmadel et al., 2017). Sediment heterogeneity was not considered in this study, as past studies have shown that despite hydraulic conductivity spanning orders of magnitude, resultant spatial and temporal metrics did not show the same degree of variability (Ward et al., 2017). As with planar bedrock, variation in the flow field could result from representation of spatial heterogeneity. Hydraulic head at the streambed boundary was specified based on surveyed water surface elevation. A triangular mesh was constructed for each model with elements ranging from 0.0021 to 0.1 m in height (a summary of the computational mesh is provided in Becker et al. [2022]). Darcy's Law was solved at steady-state across the domain, yielding steady-state values for exchange fluxes, pore water velocities, and flow path geometries.

We characterized physical exchange processes at each site using three different model outputs (Gooseff et al., 2006; Herzog et al., 2019; Schmadel et al., 2017; Ward et al., 2018b). First, massless particles were released at the sediment-water interface every 0.1 m along the entire length of the domain and the position, velocities, and time elapsed since release were tracked for each particle until it exited the model domain. These data were used to construct hyporheic transit time distributions (TTD) for each segment. Next, we extracted flux perpendicular to the streambed at the location of each particle release and calculated the total downwelling flux per meter of streambed length (Q_{hef}, after Schmadel et al., 2017) as a measure of total hyporheic exchange flux at the segment scale. Finally, we tabulated the percent of particles that immediately upwelled along the reference segment to calculate the percentage of streambed length where upwelling occurs (P_{up}).

To quantify the potential transformation associated with fluxes and transit times at the reach scale, we calculated the reaction significance factors (RSF) (Harvey et al., 2019). RSF is the product of river connectivity and the Damköhler number, calculated as

$$\mathrm{RSF} = \frac{\tau_s}{\tau_r} \times \frac{L_c}{L_s} \tag{1}$$

where τ_s is the residence time in the storage zone (s) for each upwelling particle, L_c is the reference segment length (m), τ_r is the intrinsic reaction timescale in the storage zone (s), and L_s is the river turnover length, defined as the average downstream distance that a parcel of water travels in the river before entering the hyporheic zone (Harvey et al., 2019). We fixed τ_r at 10 hr following Harvey et al. (2019) and because this is a timescale representative of several important functions at our site (Ward et al., 2011). River turnover length (L_s) was calculated as

$$L_s = \frac{Q}{q_s \times w} \tag{2}$$

where Q is stream discharge (m^3 s^{-1}), q_s is hydrologic exchange flux normalized by streambed width (m s^{-1}), and w is surveyed channel width (m). Variables used to calculate RSF for each watershed can be found in Becker et al. (2022).

2.3. Data Analysis

The full surveyed and simulated lengths of the four reference segments (Ward, Wondzell, et al., 2019; Ward, Zarnetske, et al., 2019), 247 m for Cold Ck, 256 m for Unnamed Ck, 537 m for WS01, and 542 m for WS03, were used to test accuracy and precision of exchange metrics calculated from smaller reaches within each reference segment. We simulated the full length of each segment in a single COMSOL model. Then, we sub-sampled the output from the model simulation to characterize potential study reaches that were either 100-m long or 20WCW length (88.4 m in Cold Ck, 34.8 m in Unnamed Ck, 19 m in WS01, and 18.4 m in WS03). We treated these shorter reaches as moving windows which we "slid" along the total length of the simulated segment in 0.1 m increments (Figure S1 in Supporting Information S1). For each window location, we tabulated transit time distributions (TTD), total downward flux (Q_{hef}), percent upwelling per meter (P_{up}), and RSF for particles that downwell and return to the stream within the window to represent the approach of a researcher establishing a reach-scale study site within a longer stream segment. For example, there were 1,470 different 100-m reaches within Cold Creek, and each was compared to the full 247-m segment.

We used pairwise Kruskal-Wallis tests to compare each reach-scale distribution to the reference segment distributions. We report all p-values in this study and the supplemental material to allow readers to infer the significance of the relationships or differences rather than a binary interpretation that is implied by choosing a p-value threshold. Because the null hypothesis is that the reach- and segment-scale distributions will be identical, we interpret a $p_{kw} \geq 0.10$ as an indicator that the reach-scale is representative of the segment-scale. Summary statistics of the transit time distributions for each of the smaller reaches (i.e., mean, median, coefficient of variation, skewness) were also compared with their corresponding reference segment. For Q_{hef} and P_{up}, we calculated the percent difference between each reach-scale value and the reference segment, reporting the percentage of reach-scale values with less than 10% error relative to the reference segment. For the four reference segments (Cold Creek, Unnamed Creek, WS01, and WS03), we have four metrics to test the representation of 20WCW and 100-m reaches (p_{kw} for TTD and RSF, and percent difference for Q_{hef} and P_{up}) giving us 16 total metric-by-segment comparisons.

Next, we compared between-site variation for the reference reaches to variation across headwaters in the H.J. Andrews. Levene's test for equality of variance was used to compare the variances of the 29 surveyed sites across the basin, taken together as one population to represent variation at the scale of the fifth order river basin, to the population of 20WCW windows of the four reference segments which each represent within-site or within-segment variation. The null hypothesis of Levene is that the variances are equal, with a small p-value ($p_{\text{levene}} < 0.10$) indicating the rejection of the null hypothesis, meaning variances are significantly different between the two populations. If we find $p_{\text{levene}} \geq 0.10$, we interpret that variance across the 29 additional sites is equal to that within a given reference segment. Again, p-values for all tests are reported. RSF was not calculated for the additional 29 survey sites due to lack of discharge data at the time and date when surveys were conducted.

Finally, we tabulated the frequency that a given streambed location was included in a statistically representative sub-reach (defined as $p_{kw} > 0.1$, i.e., not significantly different, or as percent error less than 10%) for both 20WCW and 100-m reaches to determine if a pattern existed based on their location. For every reach of 20WCW or 100-m, a "1" was assigned to each particle location within that reach if it was representative, and a "0" was assigned if not. Then, the sum for each particle location was calculated and plotted against the location. This was then normalized by how many times each location was included in a moving window to yield the relative frequency of inclusion. Spearman's rank correlation was applied comparing representative locations between 20WCW, and 100-m reaches to determine if there was a correlation between location and frequency of representativity. Correlation coefficient (r), and p_{spearman} values are reported to determine if correlation exists. Presence of a correlation would indicate the same features would be systematically included in or excluded from representative sub-reaches regardless of study reach length strategy, indicating features or locations that are critical to include in a representative observation.

3. Results

3.1. How Representative Are Study Reaches of Longer Reference Segments?

The widely used study strategy of 20WCW lengths did not ensure that a representative TTD was measured. Among the four reference segments, we found study reach lengths of 20WCW were statistically indistinguishable from the full segment TTDs ($p_{kw} > 0.1$) in 12%–85% of the reaches considered (Figures 2a, 2e, 2i, and 2m). Similarly, 100-m reaches produced TTDs that were statistically indistinguishable from the reference segment in 21%–87% of cases ($p_{kw} > 0.1$; Figures 2a, 2e, 2i, and 2m). Performance was not consistent between catchments. For Unnamed Creek, WS01, and WS03, moving-window TTDs were not representative of the reference reach, as evidenced by large fractions of comparisons with $p_{kw} < 0.1$ (Figures 2e, 2i, and 2m). However, for Cold Creek the distribution was skewed toward high p_{kw} values (Figure 2a), indicating the TTD from any reach had a high probability of being representative of the longer reference segment. Additionally, longer study reaches were not necessarily more accurate nor precise than shorter reaches. For example, the 20WCW reaches had a greater probability of being representative than did the 100-m reaches for Cold Creek, WS01, and WS03 (Figures 2a, 2i, and 2m), while the inverse was true for Unnamed Creek (Figure 2e). Thus, within-site variability in TTDs was not solely dependent on the length strategy of the study reach and can vary between basins. In other words, the location selected for a study can be as important, or more important than, the selected length strategy to select the study reach for sampling accurate and precise TTDs.

Reaches were overall similarly representative for Q_{hef} compared to TTDs. Reach lengths of 100-m were indistinguishable from the reference segment for Q_{hef} 15%–72% of the time, while 20WCW reaches were representative 18%–72% of the time (Figures 2b, 2f, 2j, and 2n). The total range of error was smaller for 100-m lengths than 20WCW lengths, with nearly 100% of all 100-m reaches having less than 50% error compared to 61% of all 20WCW reaches. Thus, increasing study reach length was associated with increased precision and accuracy for estimating Q_{hef}.

Reach lengths of 100-m accurately predicted the fraction of streambed upwelling, P_{up}, (i.e., less than 10% difference from reference value) 60%–100% of the time, compared to 22%–100% of 20WCW reach windows across all four segments (Figures 2c, 2g, 2k, and 2o). As with estimates for flux, 100-m reaches had greater accuracy and precision than 20WCW reaches for percent upwelling (Figures 2c, 2g, 2k, and 2o). The total range of error decreased for the 100-m reach lengths compared to the 20WCW conditions, with 100% of all 100-m reaches

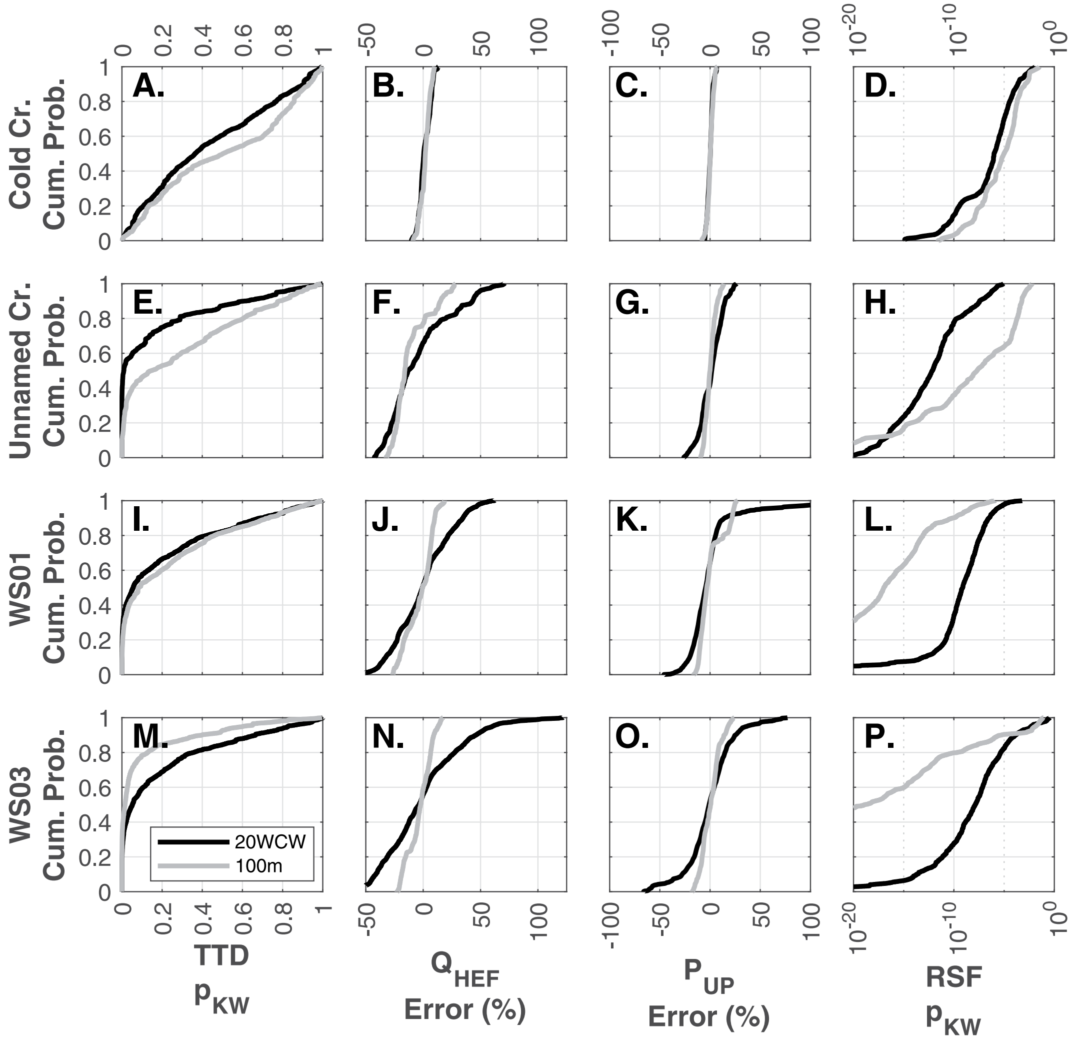

Figure 2. Cumulative distribution functions of both p_{kw} and percent error values for the four metrics: TTD (left column), Q_{hef} (middle-left column), P_{up} (middle-right column), and RSF (right column). Rows from top to bottom are Cold Creek, Unnamed Creek, WS01, and WS03. In all cases, results for 20WCW reaches are shown in solid black, and 100-m reaches are shown in solid gray. For panels (a, d, e, h, i, l, m, and p), a greater portion p_{kw} above 0.10 indicates better representativity compared to the reference reach. For panels (b, c, f, g, j, k, n, and o), a greater portion within ±10% (the narrower the line is), indicates more representativity to the reference reach.

having less than 50% error, compared to 89% of all 20WCW reaches. As with Q_{hef}, increasing reach length increased precision and accuracy for estimating P_{up}.

Finally, neither 20WCW nor 100-m reaches produced accurate estimates for RSF. Reach RSFs were statistically indistinguishable ($p_{kw} > 0.1$) to the reference segment 0% of cases for 20WCW for all reaches but WS03 (11% were considered representative) and less than 4% of cases for 100-m reach lengths (Figures 2d, 2h, 2l, and 2p). Thus, reach-scale RSF was neither length strategy nor location dependent within our studied sites and was ultimately not well predicted from reach-scale studies.

3.2. Are Some Locations or Features More Often Included in Representative Study Reaches?

Within a segment, some locations do contribute more frequently to representative reach distributions of HZ metrics compared to others (i.e., they are more often included in segments with $p_{kw} \geq 0.1$ or percent error <10%; locations with a higher *y*-axis value in Figure 3). Overall, we found little correlation between locations included for 100-m and 20WCW approaches (where small values of rho indicate a lower Spearman's Rank Correlation,

10.1029/2022WR032718

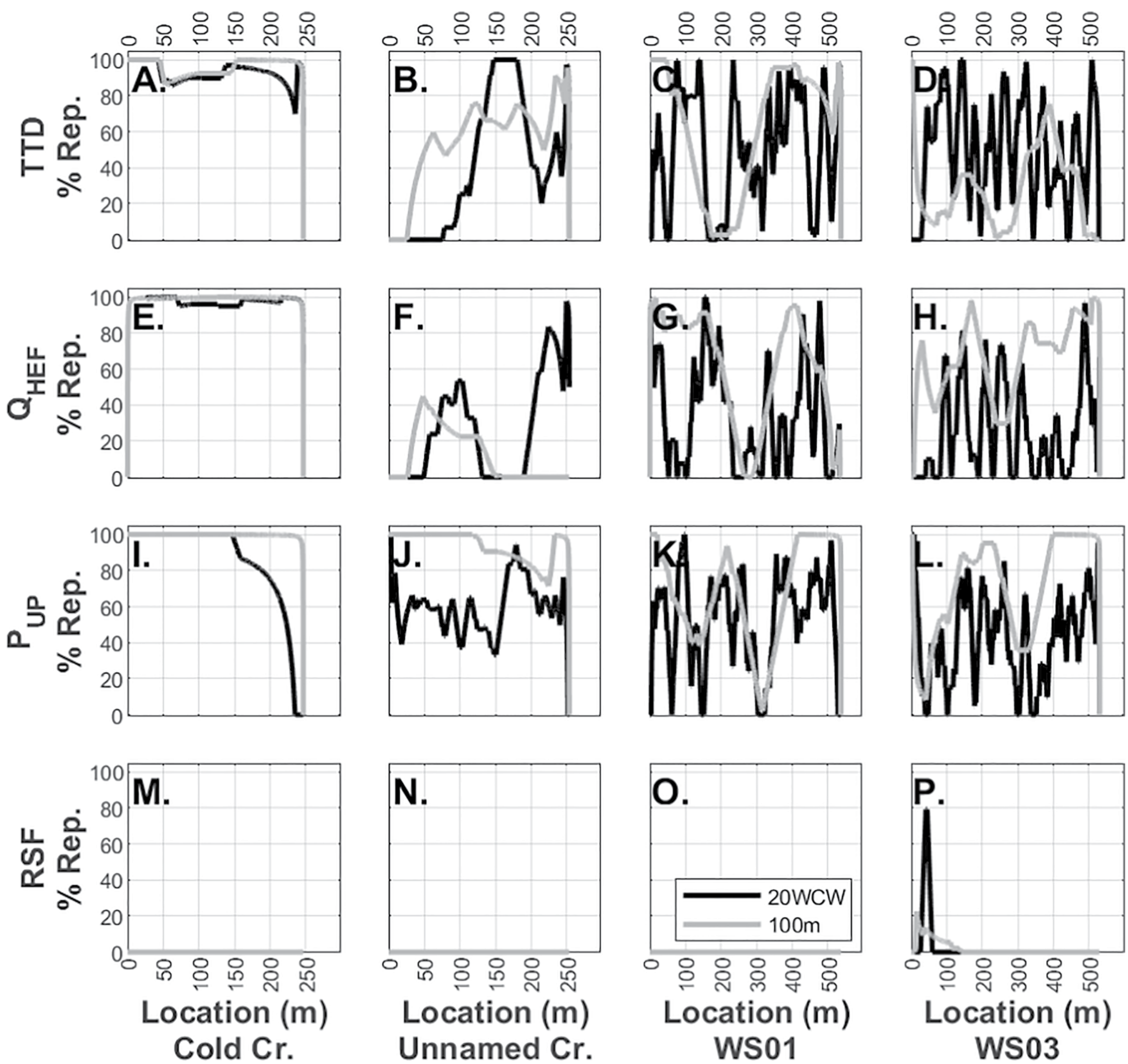

Figure 3. Locations and percent of total possible occurrences of representative reaches for the four metrics, based on $p_{kw} > 0.10$ for Transit Time Distributions and RSF, and error <10% for Q_{hef} and P_{up}. Rows from top to bottom are: TTD, Q_{hef}, P_{up}, and RSF. Columns from left to right are: Cold Creek, Unnamed Creek, WS01, and WS03. In all panels the black line represents 20WCW reaches, gray line represents 100 m reaches.

and p_{SR} indicates the p-value for the test that the two are correlated, Table 1). For RSF, we found no evidence of rank correlation between 100-m and 20WCW approaches ($p_{SR} << 0.001$). For TTD, Q_{hef}, and P_{up}, evidence of rank correlation was found only for Cold Creek ($r = 0.63$, 0.83, and 0.96, and $p_{SR} < 0.05$). Lack of spatial correlation (low r value) suggests there are not locations that are driving which reaches are representative within

Table 1
Spearman's Rank Coefficient, r, and p-Value for Occurrences of Representative Reaches for 20WCW and 100 m Reaches

Watershed	TTD (r, p-value)	Q_{hef} (r, p-value)	P_{up} (r, p-value)	RSF (r, p-value)
Cold Creek	0.63, $p << 0.001$	0.83, $p = 0$	0.96, $p = 0$	NA[a]
Unnamed Creek	0.10, $p << 0.001$	0.08, $p << 0.001$	0.47, $p << 0.001$	0.33, $p << 0.001$
WS01	0.01, $p = 0.3$	0.40, $p << 0.001$	0.36, $p << 0.001$	NA[a]
WS03	0.23, $p << 0.001$	−0.17, $p << 0.001$	−0.28, $p << 0.001$	0.12, $p << 0.001$

Note. r closer to 1 suggests strong correlation. Small p-value means rejecting the null hypothesis that there is no correlation between the two.
[a]Values of NA reflect cases where r cannot be calculated because values are uniform.

the segment. Taken together, we find little indication that a subset of features or locations are systematically included in or excluded from representative reaches.

3.3. Is Within- or Between-Segment Variation Greater for Hyporheic Exchange?

Between-site variation was greater than within-site variation for P_{up} in WS01 and WS03, for Median TTD for Unnamed, WS01, and WS03, and for Mean TTD at WS03 (Figure 4, Table 2). The coefficient of variation was greater across the 29 other sites compared to the coefficient of variation within Cold Creek, Unnamed Creek, WS01, and WS03 for exchange flux and median transit time, but not for percent upwelling (Table 2). For exchange flux, the range across the 29 other sites was greater than the range within the reference sites. For percent upwelling, the range in values and the interquartile range was well aligned with the four reference sites. Median and average transit times for the 29 sites tended to be larger than the reference sites, even though the ranges were not very different (Figures 4c and 4d).

The distributions for each exchange metric were significantly different (p_{levene} << 0.001) when comparing each reference segments to the 29 other surveyed reaches in the basin (Table 3), with the exception of WS03 Q_{hef} and Unnamed P_{up} (p_{levene} = 0.1 and 0.40). For Q_{hef}, variance within the four reference sites is less than the variance between sites. For P_{up}, Cold and Unnamed creeks have less variance than the 29 sites, while WS01 and WS03 have greater variance compared to the 29 other sites. For median travel times, within site variance is greater only in Cold Creek, while for mean travel times, within site variability is greater except in WS03 compared to the 29 other sites.

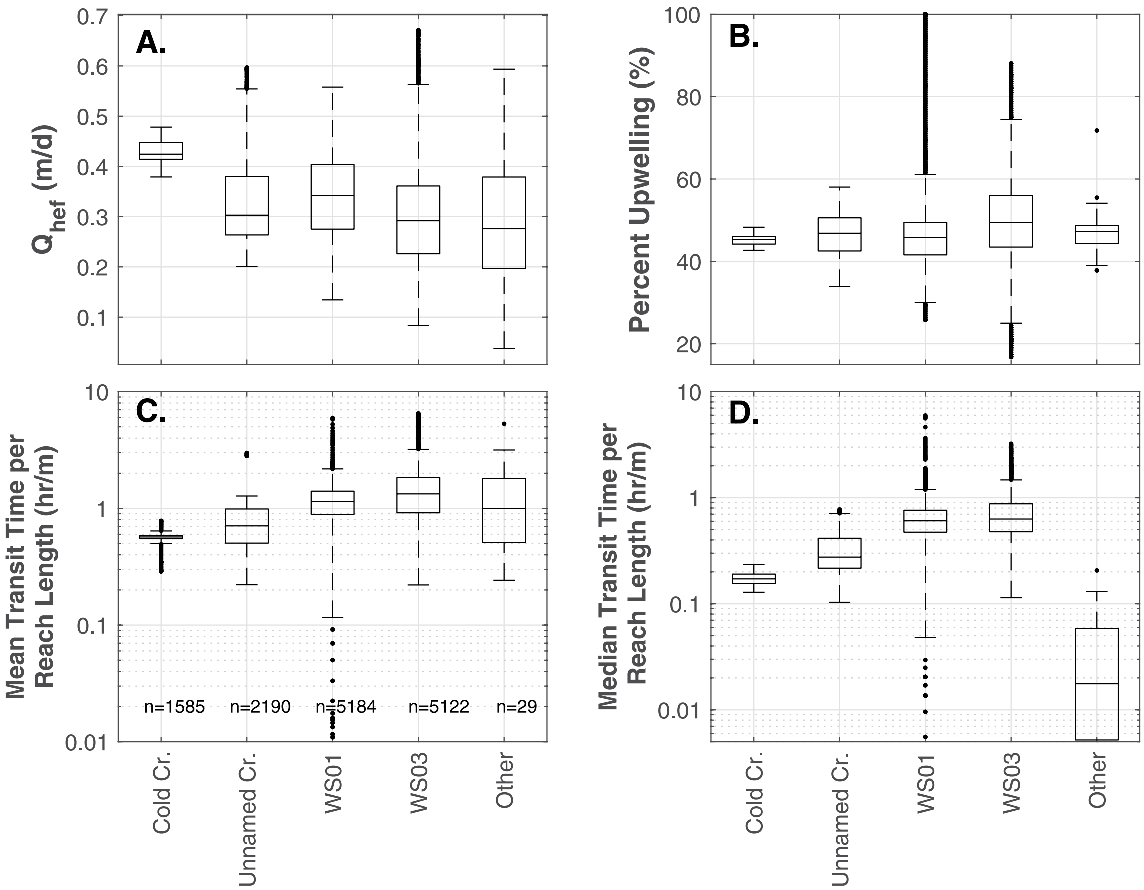

Figure 4. Comparison of exchange metrics between and within sites of a fifth order basin for 20WCW reaches. Panel (a) is showing Q_{hef} (m/s) for all 20WCW windows of the four reference segments and the 29 other sites. Panel (b) is percent of particles upwelling (P_{up}). Panel (c) is average transit time distribution normalized by reach length (hr/m), and *D* is median transit time distribution normalized by reach length (hr/m).

Table 2
Summary Results Comparing Within and Between Site Means, Skewness (γ), and Coefficient of Variation (CV) for 100 m and 20WCW Reaches

		100-m reaches				20WCW reaches				
		WS01	WS03	Cold	Unnamed	WS01	WS03	Cold	Unnamed	Others ($n = 29$)
Mean	Q_{hef} (m/s)	3.9e−6	3.4e−6	2.0e−6	3.6e−6	4.0e−6	3.5e−6	2.0e−6	3.9e−6	3.3e−6
	P_{up} (%)	48.0	50.1	45.2	46.1	47.7	49.5	45.3	46.5	47.6
	Median Transit time (s)	5.5e4	4.9e4	5.5e4	3.6e4	4.3e4	5.0e4	5.6e4	2.5e4	9.9e3
	RSF	8.82	1.4e5	0.34	5.74	0.92	2.6e4	0.29	0.91	NA
CV	Q_{hef}	0.12	0.11	0.09	0.17	0.13	0.34	0.10	0.28	0.54
	P_{up}	0.12	0.09	0.03	0.05	0.25	0.24	0.03	0.11	0.13
	TTD	0.07	0.12	0.13	0.10	0.68	0.55	0.11	0.26	1.82
	RSF	0.32	1.84	0.12	0.30	0.38	6.20	0.15	0.66	NA
γ	Q_{hef}	−0.411	−0.22	0.55	1.04	0.12	0.67	0.55	0.87	0.05
	P_{up}	0.93	0.27	0.10	0.05	2.69	0.01	−0.24	−0.11	1.72
	TTD	1.19	0.46	−0.09	0.74	6.64	1.86	0.15	0.05	2.99
	RSF	1.38	0.62	−0.29	1.66	0.81	7.68	−1.10	6.74	NA

Note. Results are presented for flux (Q_{hef}), fraction upwelling (P_{up}), median transit times, and RSF.

4. Discussion

4.1. Length Selection Strategy and Location Both Determine How Representative a Reach Is Compared to the Larger Segment

For 11 of 16 comparisons, we found 100-m study reaches were more likely to yield representative estimates for metrics of hyporheic exchange than 20WCW reaches (the exceptions being TTD for Cold Creek, WS01, and WS03; Q_{hef} for Unnamed Creek; and RSF for Unnamed Creek). Thus, we conclude that longer study reaches will be more likely to capture representative observations than short reaches will. However, this is far from a guarantee that a 100-m study reach would be representative. Location of the reach was also important because the locations of the representative reaches were not evenly distributed across the segments (Figure 3). This might be explained by the fact that if there is a short portion of the segment that is quite different than the rest of the segment, the effect of that location would be "averaged out" over longer reaches but will have a larger influence on parameters calculated for shorter reaches. Also, because the reaches overlapped, there is likely some spatial correlation occurring accounting for the times 100-m reaches did not perform better than 20WCW. The exact impact of the reach length strategy and location varied by metric. For example, in Unnamed Creek the locations of representative 20WCW for P_{up} were distributed across the entire segment (Figure 3j), but many of the locations for Q_{hef} were grouped together (Figure 3f). Additionally, RSF was rarely representative for the segments Unnamed Creek and WS03, and never representative for Cold Creek or WS01 and no pattern appears to exist regarding where

Table 3
p-Values Using Levene Test for Equality of Variances Comparing Cold, Unnamed, WS01, and WS03 to the Remaining 29 Sites Modeled at 20WCW

	Cold versus 29	Unnamed versus 29	WS01 versus 29	WS03 versus 29
Q_{hef}	$p << 0.001$ ↓	$p << 0.001$↓	$p << 0.001$↓	0.10 ↓
P_{up}	$p << 0.001$ ↓	0.397 ↓	0.10 ↑	$p << 0.001$ ↑
Median Travel Time	$p << 0.001$↓	0.01 ↑	0.35 ↑	0.02 ↑
Mean Travel Time	$p << 0.001$ ↓	$p << 0.001$ ↓	$p << 0.001$ ↓	0.06 ↑

Note. The Levene test was done for Q_{hef}, P_{up}, median transit times, and mean transit times. Smaller *p*-values indicate an increasingly strong rejection of the null hypothesis (i.e., increasingly likely that the variances are different). Arrows indicate how each watershed compares in variance to the other 29 sites. An up arrow (↑) indicates within-site variation was greater than between site (29 other sites) variation. A down arrow (↓) indicates within-site variation was less than between site variation.

10.1029/2022WR032718

RSF is to be representative (Figure 3p). We found that location is more likely to be important for TTD while length strategy is more likely to be important for Q_{hef} and P_{up}. Our results do not show any conclusive boundary regarding when reach length strategy versus exact location is more important, nor why the metrics differ. Thus, we acknowledge the variation and identify explanation as a fruitful future direction, which may require a larger and more diverse ensemble of studies to support robust conclusions.

Attempting to generalize findings based on 20WCW reaches is likely to bias our results if these studies are used as a basis to scale to entire segments or to even larger spatial extents of similar stream segments in other locations. Increasing the length of the study reach may help reduce the bias (e.g., range of error for Q_{hef} between 20WCW and 100-m reaches in all four reference segments), but the selected study reach also should account for location and the processes in question (Lee-Cullin et al., 2018). For example, in WS03, shorter reaches of 20WCW performed better than 100-m reaches in TTD, but the locations of the representative reaches were not evenly distributed along the segment. Thus, sampling based on location, rather than length strategy, could be enacted and the representativeness of those findings should be tested. Of course, this strategy presumes results are consistent between models and empirical data from the field, which remains - as yet - unknown. Caution must be exercised when taking results from studies of a single reach or a small number of reaches as the basis from which broad conclusions are drawn. As shown here, there is a high probability that behaviors measured at a single reach will not be representative of the behavior of a larger segment. The mismatch of reach length strategies and locations for representativeness suggests that using a single 20WCW reach and implicitly assuming it to be meaningful is likely to introduce substantial bias into our understanding of hyporheic exchange. This finding aligns with that of Poole et al. (2006) and Lee-Cullin et al. (2018), both of whom reported that simplification and low sampling resolution can influence our understanding at larger scales.

4.2. Assumed Behavior Cannot Be Inferred From Visual Inspection

Given the prevalence of 100-m and 20WCW reaches in our study that were not representative, and the frequency with which similar reach lengths are used in hyporheic studies reported in the literature, descriptions of expected patterns of hyporheic exchange in river networks are almost assuredly in error. We acknowledge this assertion is based on numerical simulations and field study will be needed to confirm our methods did not introduce bias. However, if our findings are representative of empirical observations, we must conclude we do not know the full effect of these errors as this phenomenon has not been previously investigated. Still, our results may explain at least some of the inconsistencies that have been reported in the literature (Ward & Packman, 2019). More importantly, these errors may substantially bias our understanding of the role of hyporheic exchange in stream networks.

To further illustrate this problem, we consider our own decades of work comparing the effects of valley bottom width on hyporheic exchange, specifically, our work comparing WS01 with a relatively wide valley with WS03 with a relatively narrow valley (e.g., Ward et al., 2016; Ward, Wondzell, et al., 2019; Wondzell, 2006). The study reaches were initially established in 1997 as part of a larger effort to identify the primary drivers of hyporheic exchange flows in distinctly different valley morphologies across a range of stream orders, including an exploration of how valley width interacts with hillslope inputs to influence hyporheic exchange. The initial focus of the work was on the mechanistic drivers of hyporheic exchange, and the empirical approach required a major initial investment to establish relatively dense well networks (~1 well per 10 m^2 in these headwater reaches). However, the need for dense well networks and the difficulty in installing them severely limited the length of reach it was possible to study. Nevertheless, once established, the dense well networks made the sites attractive for a wide variety of subsequent studies. Several of those studies explicitly compared and contrasted the two study reaches on the untested assumption that differences between them could be attributed to valley morphology (Voltz et al., 2013; Ward et al., 2016; Ward, Wondzell, et al., 2019).

Here, we examined the same study reaches, using a two-dimensional longitudinal model of the full stream segment in which those reaches are located. Of course, this 2-D model cannot capture the effects of differing valley-floor widths on hyporheic exchange, but they do allow us to systematically evaluate the potential representativeness of these specific study reaches. Our model analyses show that the hidden assumption about valley morphology driving hyporheic exchange is not representative of the longer segments of either WS01 or WS03 as simulated here. Instead, the results of our analysis show that the reaches in WS01 and WS03 in past studies (hereafter the "Wondzell reaches") do behave differently from each other but behave similarly within

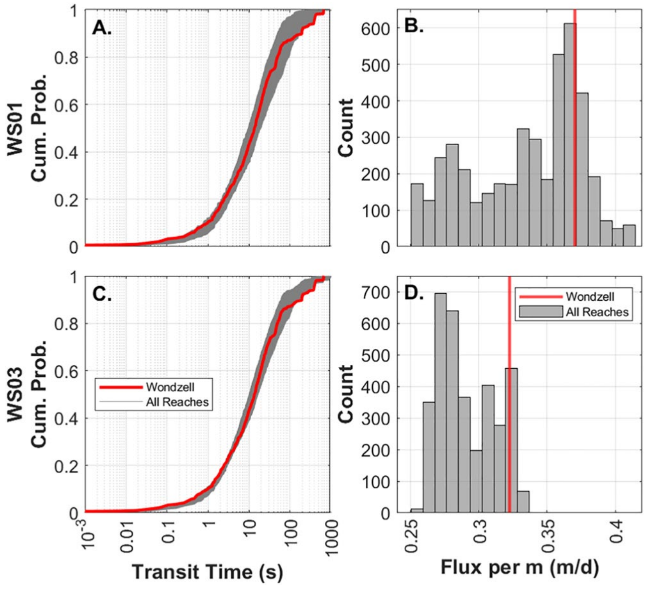

Figure 5. Comparison of the reaches studied by Wondzell (2006) versus other possible reaches of the same lengths in WS01 and WS03. Panels (a and c) are cumulative distribution plots of the transit time distributions. Red line is the TTD of the reach from Wondzell (2006). Gray lines are TTDs of other possible reaches along WS01 and WS03 of the same length. Panels (b and d) are distributions of Q_{hef} using the reach lengths studied by Wondzell (2006) across the reference segments for WS01 (top) and WS03 (bottom). The red line is the value for the reach studied by Wondzell (2006).

their respective watersheds. Both Wondzell reaches have a higher cumulative probability density for longer residence times compared to other reaches along the segment (Figures 5a and 5c). Similarly, the WS01 Wondzell reach is at the 76th percentile for flux while the WS03 Wondzell reach was at the 86th percentile for flux (Figures 5b and 5d). This indicates that both Wondzell reaches existed near the extremes within the larger segments, but those segments had similar flux ranges. Further, the TTD for both WS01 and WS03 show similar behavior. Had different reaches within these two segments been chosen, it is plausible that they would have behaved similarly and the conclusions previously made would not be representative of geologic impacts on hyporheic exchange. Critically, the assumptions that the reaches were (a) representative and (b) comparable to one another has propagated forward into subsequent empirical studies, mathematical models, and conceptual models. Taken together, this body of work (Voltz et al., 2013; Ward et al., 2016; Wondzell, 2006; Wondzell et al., 2019) has detailed how valley morphology is an important control on exchange. However, these findings are based on studies of single, relatively short reaches in WS01 and WS03 with no spatial replication of the tracer experiments and models. Past studies in WS01 and WS03 have provided the basis for a better understanding of controls on hyporheic flows. Additionally, the well fields and previous data continue to benefit studies taken at these places. However, our studies show that WS01 and WS03 do not behave so differently as previously thought and studying larger segments and at more sites could increase our understanding of geomorphic controls on hyporheic exchange. Again, we emphasize here that these findings are supported by Temnerud and Bishop (2005), who found that more variation exists in headwaters than along the river network. Indeed, they concluded generalized values used for landscapes should not be used to characterize headwater streams. While the setup of experimental study-sites is mostly controlled by technical difficulties, available workforce, and instruments, our results cast new light on the role that the chosen reach length strategy might have on investigating hyporheic transport. While short reaches (e.g., 20WCW) would be more feasible and manageable, they are not necessarily able to capture longer timescale flow paths nor are they assured to be representative. On the other hand, 100-m reaches (in locations where streams are less than 5-m wide) require more management effort to establish and maintain, but they will integrate longer processes and may allow for a more accurate representation of hyporheic exchanges. Future studies need to consider the tradeoff between reach length and the aim of the study before deciding where and how long reaches should be.

4.3. Variation Between and Within Sites Must Be Considered in Experimental Design

Our results show that, in headwater streams, study reaches of 20WCW and 100-m do not fully capture variation within a site and thus, should not be de facto considered representative of the segments within which they are located (Figure 4). One potential correction for this problem would be to both increase the length of study reaches and use several replicate reaches for each segment of stream studied. Between site variation was not consistently greater than within site variation across the four reference segments, indicating that there could be misinterpreted results about differences in basin characteristics being a causation rather than correlation. In this case, a better solution might be to simply pick more segments and locate one reasonably long study reach within each segment.

These findings align with other studies that explore local and large-scale variations in stream networks. For example, McGuire et al. (2014) explored stream chemistry characteristics and their spatial distribution and found that high variability or "patchiness" is seen at the fine scale, but that variability stabilizes at sufficiently large scales. McGuire et al.'s study also revealed that there is a nested type of heterogeneity across scales. This likely explains why shorter reaches show greater variability within a site than between sites. The 20WCW highlights small scale variations that longer reaches integrate. However, Lee-Cullin et al. (2018), found that there was little added value in increasing local scale sampling and that a single sampling array at a site can approximate variance of a site as well as three separate sampling arrays at a single site, and that point measurements are reasonably representative of plot measurements. Lee-Cullin's study did find that more variance existed within headwaters compared to third order streams, which appears to be the consensus among researchers (Likens & Buso, 2006; McGuire et al., 2014; Temnerud & Bishop, 2005; Zimmer et al., 2013), and emphasized the importance of quantifying the variance prior to using data for empirical and mechanistic modeling. If data input into a model has a different pattern of variance due to how the inputs were sampled, the conclusions will be different (Lee-Cullin et al., 2018).

4.4. Best Practices for Future Reach-Scale Studies of Hyporheic Exchange

Our study highlights the potential pitfalls in interpreting empirical, reach-scale studies as representative, which can be propagated forward into conceptual or numerical models that are used to predict hyporheic exchange at larger scales. We also recognize that researchers will always face resource limitations that place practical limits on study designs - limiting the number of study sites, and both their spatial extent and grain. The critical question is: How do we work within these constraints to ensure that study results cannot easily be misunderstood or interpreted as representative when they are not? Moving forward, we suggest studies of hyporheic exchange should embrace three key tenets: (a) clearly stating the aim of the study and to what degree assumptions or testing of representativeness are made; (b) quantify the within- and/or between-site variation to provide context for interpretations and how variation may impact interpreted results; and (c) integrate or iterate between field and modeling studies to understand the limits of both approaches.

First, clarifying the aim of the study and interpreting observations within the intended context will help prevent bad assumptions from becoming practice in hyporheic science. Not giving adequate consideration to scale issues can lead to inappropriate conclusions (Lee-Cullin et al., 2018). Without explicitly stating the goals or assumptions of a study, the interpretation can be too easily taken out of context. For example, this study focused explicitly on headwaters in heavily studied streams, examining them with 2-D groundwater flow models. The calculated fluxes and transit times should not be applied to other sites, nor should they be considered valid for questions based on 3-D interactions. However, the general conclusions made about how hydrologic research is practiced, should be transferable. Explicit statement of the assumptions also allows others to use the data, models, and conclusions in appropriate applications consistent with their intent.

Second, length selection strategy and location of reaches are critically important determinants for considering the variation within and between sites. By incorporating the variation into the location and length strategy used to select study reaches, we can reduce the bias that is potentially introduced in field studies with $n = 1$ observations. With limited observations, we are unable to determine whether a set of observations is independent of the chosen scale (spatially or temporally). To improve our understanding of hyporheic exchange, we need to increase observational resolution to enable us to test within- and between-site variation, potentially leveraging multiple scales where different drivers may be more relevant (Boano et al., 2014; Wondzell et al., 2019). This advance would pay dividends for modeling approaches, as empirical observations will shape our perceptual models of hyporheic zones and how these are translated into numerical approaches.

Finally, we need to come full circle with our studies and iterate between field and model studies. This study is an example of the iteration with field and modeling studies where we took field observations and implemented models to highlight how to enhance future field studies. An optimal strategy will use hypothesis and past studies to generate observational and experimental data, use these data to improve model representations, and learn from models to new hypotheses and update conceptual models. Through this study, we were able to map the features where representative reaches occur and how often. On this basis we could estimate if some locations or reach length strategies are more likely to yield robust empirical observations. For example, knowing in advance that reaches of 100-m are likely more representative of a larger segment than 20WCW would allow us to collect more representative observations in the field. However, models need field experimentation for validation, and field and modeling should be a continuously iterative process. Further, standardizing our methods and the results that are shared can help give more complete information for future research. It is, however, time and resource consuming to do context and background studies to select representative reaches. Thus, increasing length of the reaches and selecting the location based on processes of interest is generally recommended. Additionally, replicating the study both within the site and in other sites is important to better generalize the findings.

10.1029/2022WR032718

5. Conclusions

The focus of this study was to test the assumptions that reaches are representative of segments and to determine if within site variability is greater than between site variability for exchange metrics. Because the study was biased toward feature-scale exchanges and understanding how these features impact exchange metrics based on the length strategy and location of a potential study reach, we used two-dimensional profiles of the streams. While using 2D stream centerline profiles cannot capture all external drivers on hyporheic exchange flow - factors like valley width, channel sinuosity, and the presence of multiple channels - the analyses of model results presented here does provide a template and demonstrate the importance of evaluating the representativeness of a study reach when drawing broad conclusions. We also note that our models were simplified and idealized representations of field sites (e.g., fixed sediment depth, homogeneous conductivity, and fixed porosity for all sites). We highlight the limitations in our current practices and assumptions and show that the location and length strategy of our study needs to be considered before transferring or scaling findings. Similar studies could be done in 3D to consider impacts of valley morphology.

Despite their broad use in hydrologic studies, reaches of 20WCW are representative on average, only about 50% of the time for TTD, 76% of the time for Q_{hef}, 85% for percent upwelling, and 0% of the time for RSF. We found 100-m reaches almost always outperformed 20WCW reaches in being representative across all metrics and sites we tested (with the exception of WS01, Cold Ck, and Unnamed Ck for TTD, WS01 for P_{up}, and Cold Ck for Q_{hef}). These findings suggest that 20WCW is not a long enough reach to be representative of the hyporheic exchange at the segment scale. Study reach lengths of 100-m usually performed better than 20WCW reaches in the small headwater mountainous streams we examined. Thus, we recommend that appropriate lengths for representativeness be chosen with explicit consideration for the scale of the question being asked (Lee-Cullin et al., 2018).

Additionally, location of reaches must be considered within the context of and goals for a given study. Location can influence the conclusions drawn from studies based on a priori assumptions rather than clear documentation (e.g., Wondzell's reaches in Figure 5). Exchange throughout a stream segment is unlikely to be accurately represented by characteristic extremes such as a log jam or dramatic change in slope (Herzog et al., 2019). Additionally, when choosing multiple locations across a network, locations should be selected with similar morphologic characteristics. While access can be limiting, comparing end-member reaches in some streams to average reaches in other streams could result in mixed interpretations and conclusions.

Increased observations, with location and length strategy in mind, can reduce the bias introduced by experimental design. Based on previous literature, there is a need to increase observations and sampling among headwater streams given their high variability (Lee-Cullin et al., 2018), high prevalence in stream length (Temnerud & Bishop, 2005), and their outsized impact on stream quality (Alexander et al., 2007). As shown in Section 3.2, the 29 head waters sampled across the network behaved differently than the four reference streams ($p_{levene} < 0.1$) indicating that a small sample of head water streams likely will not provide accurate nor precise conclusions across a basin.

Finally, where possible, field hydrology and modeling studies need to be part of a cohesive iterative process so findings in the field properly translate to conceptual and numerical model processes. Improving our data, both quantitatively and qualitatively, can improve our models of these systems. We can then use these models to inform us of important locations along a stream that may be representative or turn over points such as that in Herzog et al. (2019), and observe those in the field. Bridging the gap between field experiments and models can help advance our understanding of hyporheic exchange. Improving our practices in the hydrological sciences and testing the common assumptions made is the first step to improving our predictive power in head water streams.

Data Availability Statement

Data associated with this manuscript are available in Becker et al. (2022).

Acknowledgments

The authors thank Tadd Bindas, Hunter Stanke, Billy Stansfield, Tatiana Fabiana, Nina Jung, Caroline Kryder-Reid, and Morgan Schultz for help with the surveys that took place in 2018 and 2019. Financial Support is from the National Science Foundation (Grant EAR-1652293). Data and facilities were provided by the H.J. Andrews Experimental Forest and Long Term Ecological Research (LTER) program, administered cooperatively by the USDA Forest Service Pacific Northwest Research Station, Oregon State University, and the Willamette National Forest. This material is based upon work supported by the National Science Foundation under the LTER Grants: LTER8 DEB-2025755 (2020–2026) and LTER7 DEB-1440409 (2012–2020). Wondzell and Ward were support in part by Department of Energy awards DE-SC000022 and DE-SC0019377. Ward's time was supported in part by the Burnell and Barbara Fischer fellowship from Indiana University and the Fulbright—University of Birmingham Scholar program. Any views or opinions expressed in this study are those of the authors and not positions of their employers.

References

Alexander, R. B., Boyer, E. W., Smith, R. A., Schwarz, G. E., & Moore, R. B. (2007). The role of headwater streams in downstream water quality. *Journal of the American Water Resources Association*, *43*(1), 41–59. https://doi.org/10.1111/j.1752-1688.2007.00005.x

Anderson, J. K., Wondzell, S. M., Gooseff, M. N., & Haggerty, R. (2005). Patterns in stream longitudinal profiles and implications for hyporheic exchange flow at the H.J. Andrews Experimental Forest, Oregon, USA. *Hydrological Processes*, *19*(15), 2931–2949. https://doi.org/10.1002/hyp.5791

Becker, P. S., Ward, A. S., Herzog, S. P., & Wondzell, S. M. (2022). Data files for "Testing hidden assumptions of representativeness in reach-scale studies of hyporheic exchange" [Dataset]. HydroShare. https://doi.org/10.4211/hs.826fbc5bb04e4674b2df002d979b5390

Bencala, K. E. (1983). Simulation of solute transport in a mountain pool-and-riffle stream with a kinetic mass transfer model for sorption. *Water Resources Research*, *19*(3), 732–738. https://doi.org/10.1029/WR019i003p00732

Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyophreic flow and transport processes: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, *52*(4), 603–679. https://doi.org/10.1002/2012RG000417.Received

Boano, F., Revelli, R., & Ridolfi, L. (2008). Reduction of the hyporheic zone volume due to the stream-aquifer interaction. *Geophysical Research Letters*, *35*(9), L09401. https://doi.org/10.1029/2008GL033554

Burt, T. P., & McDonnell, J. J. (2015). Whither field hydrology? The need for discovery science and outrageous hydrological hypotheses. *Water Resources Research*, *51*(8), 5919–5928. https://doi.org/10.1029/eo064i046p00929-04

Caine, N., & Swanson, F. J. (1989). Geomorphic coupling of hillslope and channel systems in two small mountain basins. *Zeitschrift Fur Geomorphologie*, *33*(2), 189–203. https://doi.org/10.1127/zfg/33/1989/189

Cardenas, M. B. (2009). A model for lateral hyporheic flow based on valley slope and channel sinuosity. *Water Resources Research*, *45*(1), W01501. https://doi.org/10.1029/2008WR007442

Covino, T., McGlynn, B., & Mallard, J. (2011). Stream-groundwater exchange and hydrologic turnover at the network scale. *Water Resources Research*, *47*(12), W12521. https://doi.org/10.1029/2011WR010942

Day, T. J. (1977). Field procedures and evaluation of a slug dilution gauging method in mountain streams. *Journal of Hydrology New Zealand*, *16*(2), 113–133.

Fabian, M. W., Endreny, T. A., Bottacin-Busolin, A., & Lautz, L. K. (2011). Seasonal variation in cascade-driven hyporheic exchange, northern Honduras. *Hydrological Processes*, *25*(10), 1630–1646. https://doi.org/10.1002/hyp.7924

Findlay, S. (1995). Importance of surface-subsurface exchange in stream ecosystems: The hyporheic zone. *Limnology and Oceanography*, *40*(1), 159–164. https://doi.org/10.4319/lo.1995.40.1.0159

Fitzpatrick, F. A., Waite, I. R., D'Arconte, P. J., Meador, M. R., Maupin, M. A., & Gurtz, M. E. (1998). *Revised methods for characterizing stream habitat in National Water-Quality Assessment Program: Water-resources investigations report*. US Department of the Interior, US Geological Survey.

Frissell, C. A., Liss, W. J., Warren, C. E., & Hurley, M. D. (1986). A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management*, *10*(2), 199–214. https://doi.org/10.1007/BF01867358

Gomez-Velez, J. D., Harvey, J. W., Cardenas, M. B., & Kiel, B. (2015). Denitrification in the Mississippi River network controlled by flow through river bedforms. *Nature Geoscience*, *8*(12), 941–945. https://doi.org/10.1038/ngeo2567

Gooseff, M. N., Anderson, J. K., Wondzell, S. M., LaNier, J., & Haggerty, R. (2006). A modelling study of hyporheic exchange pattern and the sequence, size, and spacing of stream bedforms in mountain stream networks, Oregon, USA. *Hydrological Processes*, *20*(11), 2443–2457. https://doi.org/10.1002/hyp.6349

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

Grant, G. E., Swanson, F. J., & Wolman, M. G. (1990). Pattern and origin of stepped-bed morphology in high-gradient streams, Western Cascades, Oregon. *Bulletin of the Geological Society of America*, *102*(3), 340–352. https://doi.org/10.1130/0016-7606(1990)102<0340:PAOOSB>2.3.CO;2

Harvey, J., Gomez-Velez, J., Schmadel, N., Scott, D., Boyer, E., Alexander, R., et al. (2019). How hydrologic connectivity regulates water quality in river corridors. *Journal of the American Water Resources Association*, *55*(2), 369–381. https://doi.org/10.1111/1752-1688.12691

Harvey, J., & Gooseff, M. (2015). River corridor science: Hydrologic exchange and ecological consequences from bedforms to basins. *Water Resources Research*, *51*(9), 6893–6922. https://doi.org/10.1002/2015WR017617

Herzog, S. P., Ward, A. S., & Wondzell, S. M. (2019). Multiscale feature-feature interactions control patterns of hyporheic exchange in a simulated headwater mountain stream. *Water Resources Research*, *55*(12), 10976–10992. https://doi.org/10.1029/2019WR025763

Kasahara, T., & Wondzell, S. M. (2003). Geomorphic controls on hyporheic exchange flow in mountain streams. *Water Resources Research*, *39*(1), SBH 3-1–SBH 3-14. https://doi.org/10.1029/2002wr001386

Kelleher, C., Wagener, T., McGlynn, B., Ward, A. S., Gooseff, M. N., & Payn, R. A. (2013). Identifiability of transient storage model parameters along a mountain stream. *Water Resources Research*, *49*(9), 5290–5306. https://doi.org/10.1002/wrcr.20413

Lee-Cullin, J. A., Zarnetske, J. P., Ruhala, S. S., & Plont, S. (2018). Toward measuring biogeochemistry within the stream-groundwater interface at the network scale: An initial assessment of two spatial sampling strategies. *Limnology and Oceanography: Methods*, *16*(11), 722–733. https://doi.org/10.1002/lom3.10277

Leopold, L. B., Wolman, M. G., & Miller, J. P. (1964). *Fluvial processes in geomorphology*. W. H. Freeman and Company.

Likens, G. E., & Buso, D. C. (2006). Variation in streamwater chemistry throughout the Hubbard Brook Valley. *Biogeochemistry*, *78*(1), 1–30. https://doi.org/10.1007/s10533-005-2024-2

Magliozzi, C., Grabowski, R., Packman, A. I., & Krause, S. (2018). Toward a conceptual framework of hyporheic exchange across spatial scales. *Hydrology and Earth System Sciences Discussions*, *22*(12), 6163–6185. https://doi.org/10.5194/hess-22-6163-2018

Malzone, J. M., Lowry, C. S., & Ward, A. S. (2016). Response of the hyporheic zone to transient groundwater fluctuations on the annual and storm event time scales. *Water Resources Research*, *52*(7), 5301–5321. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

McGuire, K. J., Torgersen, C. E., Likens, G. E., Buso, D. C., Lowe, W. H., & Bailey, S. W. (2014). Network analysis reveals multiscale controls on streamwater chemistry. *Proceedings of the National Academy of Sciences*, *111*(19), 7030–7035. https://doi.org/10.1073/PNAS.1404820111

Montgomery, D. R., & Buffington, J. M. (1997). Channel-reach morphology in mountain drainage basins. *Bulletin of the Geological Society of America*, *109*(5), 596–611. https://doi.org/10.1130/0016-7606(1997)109<0596:CRMIMD>2.3.CO;2

Montgomery, D. R., & Buffington, J. M. (1998). Channel processes, classification, and response. *River Ecology and Management*, 13–42. https://doi.org/10.1007/978-1-4612-1652-0_2

Nakamura, F., & Swanson, F. J. (1993). Effects of coarse woody debris on morphology and sediment storage of a mountain stream system in western Oregon. *Earth Surface Processes and Landforms*, *18*(1), 43–61. https://doi.org/10.1002/esp.3290180104

Payn, R. A., Gooseff, M. N., McGlynn, B. L., Bencala, K. E., & Wondzell, S. M. (2009). Channel water balance and exchange with subsurface flow along a mountain headwater stream in Montana, United States. *Water Resources Research*, *45*(11), W11427. https://doi.org/10.1029/2008WR007644

Poole, G. C., Stanford, J. A., Running, S. W., & Frissell, C. A. (2006). Multiscale geomorphic drivers of groundwater flow paths: Subsurface hydrologic dynamics and hyporheic habitat diversity. *Journal of the North American Benthological Society*, *25*(2), 288–303. https://doi.org/10.1899/0887-3593(2006)25[288:MGDOGF]2.0.CO;2

Schmadel, N. M., Ward, A. S., Lowry, C. S., & Malzone, J. M. (2016). Hyporheic exchange controlled by dynamic hydrologic boundary conditions. *Geophysical Research Letters*, *43*(9), 4408–4417. https://doi.org/10.1002/2016GL068286

Schmadel, N. M., Ward, A. S., & Wondzell, S. M. (2017). Hydrologic controls on hyporheic exchange in a headwater mountain stream. *Journal of the American Water Resources Association*, *53*(7), 6260–6278. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Segura, C., Noone, D., Warren, D., Jones, J. A., Tenny, J., & Ganio, L. M. (2019). Climate, landforms, and geology affect baseflow sources in a mountain catchment. *Water Resources Research*, *55*(7), 5238–5254. https://doi.org/10.1029/2018WR023551

Stanford, J. A., & Ward, J. V. (1988). The hyporheic habitat of river ecosystems. *Nature*, *335*(6185), 64–66. https://doi.org/10.1038/335064a0

Swanson, F. J., & James, M. E. (1975). Geology and geomorphology of the H.J. Andrews experimental forest, Western Cascades, Oregon. *Research Paper PNW*, *188*, 14.

Swanson, F. J., & Jones, J. A. (2002). Geomorphology and hydrology of the H.J. Andrews Experimental Forest, Blue River, Oregon. *Field guide to geologic processes in Cascadia* (pp. 289–314).

Temnerud, J., & Bishop, K. (2005). Spatial variation of streamwater chemistry in two Swedish boreal catchments: Implications for environmental assessment. *Environmental Science and Technology*, *39*(6), 1463–1469. https://doi.org/10.1021/es040045q

Tonina, D., & Buffington, J. M. (2009). Hyporheic exchange in mountain rivers I: Mechanics and environmental effects. *Geography Compass*, *3*(3), 1063–1086. https://doi.org/10.1111/j.1749-8198.2009.00226.x

Voltz, T., Gooseff, M., Ward, A. S., Singha, K., Fitzgerald, M., & Wagener, T. (2013). Riparian hydraulic gradient and stream-groundwater exchange dynamics in steep headwater valleys. *Journal of Geophysical Research: Earth Surface*, *118*(2), 953–969. https://doi.org/10.1002/jgrf.20074

Ward, A. S., Fitzgerald, M., Gooseff, M. N., Voltz, T. J., Binley, A. M., & Singha, K. (2012). Hydrologic and geomorphic controls on hyporheic exchange during base flow recession in a headwater mountain stream. *Water Resources Research*, *48*(4), W04513. https://doi.org/10.1029/2011WR011461

Ward, A. S., Gooseff, M. N., & Johnson, P. A. (2011). How can subsurface modifications to hydraulic conductivity be designed as stream restoration structures? Analysis of Vaux's conceptual models to enhance hyporheic exchange. *Water Resources Research*, *47*(8), 1–13. https://doi.org/10.1029/2010WR010028

Ward, A. S., Morgan, J. A., White, J. R., & Royer, T. V. (2018). Streambed restoration to remove fine sediment alters reach-scale transient storage in a low-gradient fifth-order river, Indiana, USA. *Hydrological Processes*, *32*(12), 1786–1800. https://doi.org/10.1002/hyp.11518

Ward, A. S., & Packman, A. I. (2019). Advancing our predictive understanding of river corridor exchange. *Wiley Interdisciplinary Reviews: Water*, *6*(1), e1327. https://doi.org/10.1002/wat2.1327

Ward, A. S., Schmadel, N. M., & Wondzell, S. M. (2018a). Simulation of dynamic expansion, contraction, and connectivity in a mountain stream network. *Advances in Water Resources*, *114*, 64–82. https://doi.org/10.1016/j.advwatres.2018.01.018

Ward, A. S., Schmadel, N. M., & Wondzell, S. M. (2018b). Time-variable transit time distributions in the hyporheic zone of a headwater mountain stream. *Water Resources Research*, *54*(3), 2017–2036. https://doi.org/10.1002/2017WR021502

Ward, A. S., Schmadel, N. M., Wondzell, S. M., Gooseff, M. N., & Singha, K. (2017). Dynamic hyporheic and riparian flow path geometry through base flow recession in two headwater mountain stream corridors. *Water Resources Research*, *53*(5), 3988–4003. https://doi.org/10.1002/2016WR019875

Ward, A. S., Schmadel, N. M., Wondzell, S. M., Harman, C. J., Gooseff, M. N., & Singha, K. (2016). Hydrogeomorphic controls on hyporheic and riparian transport in two headwater mountain streams during base flow recession. *Water Resources Research*, *52*(2), 1479–1497. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Ward, A. S., Wondzell, S. M., Schmadel, N. M., Herzog, S., Zarnetske, J. P., Baranov, V., et al. (2019). Spatial and temporal variation in river corridor exchange across a 5th-order mountain stream network. *Hydrology and Earth System Sciences*, *23*(12), 5199–5225. https://doi.org/10.5194/hess-23-5199-2019

Ward, A. S., Zarnetske, J. P., Baranov, V., Blaen, P. J., Brekenfeld, N., Chu, R., et al. (2019). Co-located contemporaneous mapping of morphological, hydrological, chemical, and biological conditions in a 5th-order mountain stream network, Oregon, USA. *Earth System Science Data*, *11*(4), 1567–1581. https://doi.org/10.5194/essd-11-1567-2019

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. https://doi.org/10.1002/hyp.5902

Wondzell, S. M. (2011). The role of the hyporheic zone across stream networks. *Hydrological Processes*, *25*(22), 3525–3532. https://doi.org/10.1002/hyp.8119

Wondzell, S. M., Herzog, S. P., Gooseff, M. N., Ward, A. S., & Schmadel, N. M. (2019). Geomorphic controls on hyporheic exchange across scales—Watersheds to particles. *Treatise on Geomorphology*, 409–429. https://doi.org/10.1016/B978-0-12-409548-9.12135-9

Zarnetske, J. P., Haggerty, R., Wondzell, S. M., & Baker, M. A. (2011). Dynamics of nitrate production and removal as a function of residence time in the hyporheic zone. *Journal of Geophysical Research*, *116*(1), G01025. https://doi.org/10.1029/2010JG001356

Zimmer, M. A., Bailey, S. W., Mcguire, K. J., & Bullen, T. D. (2013). Fine scale variations of surface water chemistry in an ephemeral to perennial drainage network. *Hydrological Processes*, *27*(24), 3438–3451. https://doi.org/10.1002/hyp.9449