AN ABSTRACT OF THE DISSERTATION OF

Paige Sheridan Becker for the degree of Doctor of Philosophy in Water Resources Engineering presented on August 17th, 2023.

Title: Scaling of Hyporheic Exchange Processes and Finding the Intermediate Sweet Spot

Abstract approved: _____

Adam Scott Ward

Hyporheic zones are important regions that reside below and along the sides of streams. Within this region, several ecosystem services are provided including stream temperature regulation, habitats for a large variety of species, pollutant removal, and nutrient cycling. Exchange between the hyporheic zone and stream occurs across multiple scales, but historically studies of hyporheic exchange have been focused on the reach scale, which is typically tens to hundreds of meters in length, and features multiple individual geomorphic features such as steps, pools, or riffles. At the reach scale, the shortest flow paths dominate, being driven by individual geomorphic features. Reach-scale studies also provide the basis for scaling up hyporheic exchange processes in models, combining reaches in series. However, this results in the exclusion of the segment scale. The segment scale is at the length of multiple reaches, having similar morphologies and often defined by major changes in geomorphology such as outcropping bedrock, waterfalls, or stream confluences. At the segment scale, longer flow paths emerge that span multiple reaches. These longer flow paths (coined intermediate-length flow paths for this body of work) are thought to be important for ecological purposes, being an optimal length for hyporheic turnover and increased residence time in the hyporheic zone. Despite their importance, these longer flow paths are often ignored in hyporheic studies.

This dissertation explores the importance of the segment scale and inclusion of intermediate-length flow paths in hyporheic exchange and how incorporating the segment scale and intermediate-length flow paths in studies changes our understanding of hyporheic processes. The first study (Chapter 2) assesses how well reach-scale studies represent hyporheic exchange at larger scales. This study quantitatively assesses how assumptions about representativeness

hold up when comparing the reach scale to the segment scale. Results demonstrate that selection strategy and location of a study reach both determine if a study is representative, and that the assumption of representativeness gets propagated into conceptual models and future studies resulting in an understanding of hyporheic exchange being based on reach-scale interactions. The second study (Chapter 3) analyzes the model performance of incorporating the segment scale and intermediate-length flow paths to predict hyporheic exchange. The study compares the truncation of transit time distributions as well as spatial discretization to a model that includes intermediate-length flow paths. The results show that using the segment scale and including intermediate-length flow paths improves prediction for the fate and transport of solutes in the hyporheic zone. The third study (Chapter 4) is a proof-of-concept experiment in measuring intermediate-length flow paths in the field. Using a combination of tracers and channel water balance, this study empirically measures intermediate flow paths and quantifies their relative makeup in fluxes and through the hyporheic zone. Results show that intermediate-length flow paths in the subsurface, which are normally ignored or missed in reach-scale studies, make up a significant portion of the total water balance. Together, this body of work quantifies intermediate flow paths within the hyporheic zone and assesses their importance when modeling and upscaling hyporheic exchange studies.

SCALING OF HYPORHEIC EXCHANGE PROCESSES AND FINDING THE INTERMEDIATE SWEET SPOT

by

Paige Sheridan Becker

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of

the requirements for the

degree of

Doctor of Philosophy

Presented August 17, 2023

Commencement June 2024

Doctor of Philosophy dissertation of Paige Sheridan Becker presented on August 17, 2023

APPROVED:

Major Professor, representing Water Resources Engineering

Director of the Department of Water Resources Graduate Program

---

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Paige Sheridan Becker, Author

ACKNOWLEDGEMENTS

My thanks first and foremost are to Adam Ward. Without your patience, guidance, high-fives (both virtual and in person), coffee dates, random kind notes, and endless support through this Series of Unusual of Ph.D. Events, this work would not have been possible. You made me a better scientist, writer, and person. You believed in me when I doubted myself and brought excitement to my research when I needed it the most. I also doubt anyone else would be willing to handle my sass to the extent you have. Thank you for everything.

Thanks to Steve Wondzell for your help in the field and on paper, for offering pickled possums on rainy days, and for sitting through two qualifying exams. Thanks to Skuyler Herzog who has been a mentor and friend, and first introduced me to the wonderful hyporheic zone.

Thanks to Tadd Bindas, Hunter Stanke, Billy Stansfield, Tatiana Fabiana, Nina Jung, Caroline Kryder-Reid, and Morgan Schultz for their long hours and hard work in the field. Thanks to Molly Cain, Lienne Sethna, Michelle Schwartz, Mikaela Cherry, and Mackenzie Fiss as we struggled through our PhDs together.

Thanks to everyone at the H.J. Andrews Experimental Forest LTER for their support as I trudged through the forest dumping salt and dye. Thanks to the U.S. Department of Energy, Oak Ridge National Laboratory, and Scott Painter, Saubhagya Rathore, Eric Pierce, and Marie Kurz. Thanks to past and present committee members. This has all been very weird.

Thanks to team Unconfined Unicorns for your love and mentorship and for being incredible women in the hydrology world to look up to. Thanks also to my many Twitter friends and H3S friends. Your virtual support means so much.

Thanks to the administrative faculty at both Indiana University and Oregon State University: Lorann, Katherine, Catherine, and Jen. Your patience, help, and ability to navigate university red tape have been essential in getting me to this point and in making my life much easier. Thanks to Mary Santelmann, the original steely-eyed missile woman. Your help in my transfer, nominating me for awards, and enthusiasm for my dissertation are so appreciated.

This work was supported by: the National Science Foundation (grant no. 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). DOE Office of Science Graduate Student Research Program (SCGSR) Solicitation 2, 2020.

Indigenous peoples have been in a relationship for thousands of years with the forests, streams, and meadows we now call the Blue River watershed. In the Kalapuya Treaty of 1855 (aka Treaty of Dayton, Willamette Valley Treaty), the Kalapuya were forced to cede this land to the US Government. We continue to learn about, recognize, and value the attributes of the Blue River watershed that reflect the enduring relationship between Indigenous people and the land. We strive to be mindful of this relationship and to integrate it in our research, our decision-making, and our actions.

Oregon State University in Corvallis, Oregon, is located within the traditional homelands of the Marys River or Ampinefu Band of Kalapuya. Following the Willamette Valley Treaty of 1855, Kalapuya people were forcibly removed to reservations in Western Oregon. Today, living descendants of these people are a part of the Confederated Tribes of Grand Ronde Community of Oregon (grandronde.org) and the Confederated Tribes of the Siletz Indians (ctsi.nsn.us).

Thanks to my doctor for caring and ensuring I had the support I needed, including my happy pills, to get through this. Thanks to my therapist for listening, providing guidance, and helping me create boundaries.

Thanks to Regis, my work-life balance manager, even though he'll never read this. While I am your emotional support animal, your need for love, cuddles, and walks kept me going through a really hard time.

Thanks to my Mom for always loving me, helping me with far too many moves, and being a phone call away.

Thanks are owed to my family and friends. You were the best cheerleaders rooting for me despite not knowing what I was doing or where I was living most of the time. You never doubted me for a second. You all were there through the worst and best times and supported me every

step of the way. I love you all so much.

Finally, thanks are to the future, Dr. Becker. You did this. You earned this. You worked so hard and pushed yourself. This dissertation is a celebration of you! When you doubt yourself, remember how you conducted a field season mostly by yourself, with a smile on your face. Remember you came out on the other side, with great science to share. Be proud of yourself. Thank you for showing up and doing the work and being the badass you are.

DEDICATION

This is dedicated in memory of my grandparents. My Grandma broke glass ceilings and inspired me to be my best self and she is the source of my stubbornness. My Grandpa inspired me to be an engineer, building rockets together and encouraging my curiosity about the world. He also was the epitome of “treat yo’ self” by having us indulge in cookies or ice cream frequently. I would not be the person I am without you two. Miss you and love you both dearly.

CONTRIBUTION OF AUTHORS

Drs. Adam Ward, Steve Wondzell, and Skuyler Herzog contributed to “Testing Hidden Assumptions of Representativeness in Reach-Scale Studies of Hyporheic Exchange” and “Disentangling Underflow and Window of Detection Limitations in the River Corridor of a Headwater Stream” chapters.

Drs. Adam Ward, Saubhagya Rathore, and Scott Painter contributed to the “Impacts of Process Truncation and Timescale Reduction on Hyporheic Exchange in a Headwater Mountain Stream”

EPIGRAPH

This is the water that moves under the stream, in cobble beds and old sandbars. It edges up the toe slope to the forest, a wide unseen river that flows beneath the eddies and the splash. A deep invisible river, known to roots and rocks. The water and the land intimate beyond our knowing. It is hyporheic flow that I'm listening for.

Dr. Robin Wall Kimmerer

Braiding Sweetgrass

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (CONTINUED)

PAGE

LIST OF FIGURES

FIGURE PAGE

LIST OF TABLES

Table Page

LIST OF APPENDIX TABLES

Table Page

2. INTRODUCTION

2.1. SCOPE AND CONTEXT

The hyporheic zone is an "invisible" river that flows below and along streams in the subsurface. Traian Orghidan first coined the hyporheic zone in 1959, deriving its name from two Greek words, *hypo* (below) and *rheos* (flow), (Orghidan, 1959). The hyporheic zone serves many purposes including habitat for aquatic organisms, a redd for salmon eggs, stream temperature regulation, nutrient cycling, and pollutant removal resulting in it being coined a "river's liver" (Fischer et al., 2005; Orghidan, 1959; Peter et al., 2018; Soulsby et al., 2009). The hyporheic zone is a part of the river corridor with water, energy, and solutes being exchanged between the hyporheic zone and the stream across scales from individual geologic features such as a step or dune to entire catchments (Boano et al., 2014; Vannote et al., 1980). Drivers of exchange between the stream and hyporheic zone include but are not limited to, geologic features like steps, pools, and riffles (Anderson et al., 2005), Lateral inflows from hillslopes and down-valley flow (Water et al., 2018), and regional groundwater gains and losses (Boano et al., 2008; Malzone et al., 2016). These spatial scales also influence transit times which range from seconds to years (Boano et al., 2014). Hyporheic exchange is one of the most studied processes under the general category of river corridor exchange (Harvey & Gooseff, 2015).

Over the decades, a growing number of studies on the hyporheic zone have been conducted across disciplines to understand the chemical, physical, and biological functions of the hyporheic zone (Ward, 2016). There have also been efforts to advance the predictive power or hyporheic zone exchanges by describing physical processes in numerical models (Ward & Packman, 2019), with these efforts focused on scaling processes to entire river networks (multiple catchments), such as the Upper Mississippi Basin (Gomez-Velez et al., 2015). In addition to modeling efforts, there have been efforts to engineer hyporheic zones within urban streams to utilize their ecosystem services (e.g., Herzog et al., 2016), emphasizing the value of hyporheic zones.

Flow in the river corridor occurs at nested scales (Boano et al., 2014; Stonedahl et al., 2012), with common classifications to differentiate them based on time scales, spatial scales, or analysis of geologic features (Frissell et al., 1986; Montgomery and Buffington, 1998). Typically, the river corridor is organized as reaches, segments, and watersheds. Reaches

are defined as similar bedforms over multiple channel widths in length; segments are defined as portions of the stream exhibiting similar valley-scale morphologies and governing geomorphic processes; and watersheds are defined by topographic highs and lows and clear drainage divisions (Montgomery and Buffington, 1998). Groundwater studies have used local, intermediate, and regional to define regimes with local regimes being defined by maxima and minima of individual features, regional intermediate regimes being multi-feature flow paths, and regional regimes being defined by watershed divisions (Tóth, 1963). For the purposes of this dissertation, we have adapted Tóth's terminology to apply to hyporheic flows and use local flow paths occurring at the reach scale, intermediate flow paths occurring at the segment scale, and regional flow paths occurring at the catchment or watershed scale.

Despite the nested scales in which exchange occurs, many of the studies focused on hyporheic exchange occur at the reach scale (10s to 100s of meters in length, or 10-20 wetted channel widths (WCW) in length, Montgomery and Buffington, 1998). The reach-scale has been the scale of focus for a couple of reasons. For one, it is a practical scale to conduct field experiments, when resources and access can be limited. Additionally, the reach scale has been used for relating stream morphology to channel processes and habitat characteristics (e.g., Day, 1977; Leopold et al., 1964, etc.). The reach-scale used in field studies is commonly propagated into models to predict exchange in individual catchments and whole networks, resulting in an implicit assumption that only reach-scale drivers of exchange matter at large scales and that no new processes emerge. Stonedahl et al., (2013) however, show that new processes may emerge at larger scales, and processes from one scale are not necessarily additive to represent larger scales. Thus, the importance of intermediate flow paths, and the consequences of using only reach-scale studies to expand to catchments and beyond are unknown and largely understudied.

This body of work investigates the importance of intermediate flow paths at the segment scale in hyporheic exchange in three ways: (1) testing how well reach-scale studies represent hyporheic exchange at the segment scale, (2) testing the consequences of ignoring intermediate flow paths in models of conservative and reactive transport at the scale of study reaches and segments, and (3) conducting a proof-of-concept field experiment to measure and quantify intermediate flow paths at the segment scale.

2.2. Testing Representativeness of Intermediate Flow Paths

Practices for selecting study reaches are highly variable, and include using a multiple such as a wetted channel width (Anderson et al., 2005), biased selection towards a certain attribute like an end member (Wondzell, 2006), fixed length reaches (Payn et al., 2009), fixed transit time scales (Ward, Wondzell, et al., 2019), and random selection (Gooseff et al., 2006). Because the selection strategy is highly variable, the location and length are also variable, with the length ranging from tens to thousands of meters. Despite the variation, there is a hidden assumption that these studies represent hyporheic exchange across sites or conditions and are sometimes used as the basis for building catchment scale models (Ward et al., 2018a; Ward et al., 2020).

Additionally, some studies have made comparisons between observable characteristics of a catchment (e.g., visually steep vs. low-relief or wide vs. narrow valleys) and then assumed that differences in exchange are attributable to those observed characteristics. That assumption, in turn, is translated into conceptual models without verification that the observed characteristics are drivers of differences in exchange (i.e. correlation is confused with causation). Models used to predict exchange rely on untested assumptions, but it is unkown if those assumptions are precise or accurate, and thus are representative of conditions we care about (Ward et al., 2018a; Ward et al., 2020).

Thus, in Chapter 2, I explore how representative reach-scale studies are of segment-scale processes within which the reach scale is located. Without understanding how representative our studies are of the processes we care about, we may be missing key drivers of exchange. This study quantifies how representative two strategies for selecting study reaches (fixed length and adaptive length) are for representing segment-scale processes in the hyporheic exchange, and what the consequences are for assuming representativeness. We can better inform our practices by understanding how representative reach-scale studies are of the larger and longer scales.

2.3. Consequences of Ignoring Intermediate-Length Flow Paths in Hyporheic Exchange Models

Having established the limitations of applying reach-scale understanding to segments, I next turn to a more detailed assessment of the consequences of this limitation. Accurate prediction of the fate and transport of constituents in the stream and hyporheic zone will help us

understand their influence on water quality. Numerical models are useful tools for making predictions at large scales (e.g., NEXSS and ATS, Gomez-Velez et al., 2015; and Painter, 2018), but these models are often based on reach-scale processes (Becker et al., 2023). Knowing that exchange and hyporheic turnover (Herzog et al., 2019), occur at larger scales, such as at the segment scale, is likely that models are underpredicting the transport and reactions occurring in the river corridor. In Chapter 3, I use a numerical, hyporheic-exchange model to understand the consequences of using the reach scale to understand segment-scale processes, thus ignoring intermediate flow paths. I first assess the impacts of truncating the transit time distributions, incrementally cutting off longer flow paths and quantify differences in transport and fate of physical and chemical processes.

Additionally, this study explores if the segment can be 'intelligently' divided based on known turnover points that define distinct groupings of features (after Herzog et al., 2019), and how leveraging the knowledge of turnover cells may improve prediction when the full transit-time distribution is not available. Understanding the consequences of including or ignoring reach- versus segment-scale flow paths to predict exchange helps inform our knowledge of whether or predictions are likely to be precise and to what extent.

2.4. Quantifying Underflow in the Field

Measuring exchange beyond the reach scale can be challenging. For one, tracers used for field measurements have a window of detection, meaning tracers have a limited scale at which they can be directly measured (Bencala et al., 2011; Payn et al., 2009; Wagner & Harvey, 1997). Further, subsurface flows often get lumped with other fluxes that occur along the stream, and cannot be isolated (Payn et al., 2009). However, Payn et al., (2009) also showed that by using a channel water balance and accounting for gross gains and losses along a stream segment, we can better parse out subsurface flow paths that flow across multiple reaches, beyond the window of detection. Combining studies across multiple reaches, and intermediate flow paths we can hope to measure intermediate flow paths.

Chapter 4 combines the concepts of reach-scale channel water balances to measure intermediate subsurface flow paths and parse them out from other fluxes normally lumped together. Two segments (lengths of 173.2 m and 303.6 m) lying between sections of outcropping bedrock in WS01 of the HJ Andrews Experimental Forest were chosen for this study. In these

segments, the bedrock pushes subsurface flows into the stream and thus, with a water balance accounting for evapotranspiration and lateral inputs, subsurface flows can be isolated and quantified. We compare reach- and segment-scale water balances to understand what proportion of total fluxes are subsurface flows, and how the water balances vary across the segments. This study is a proof-of-concept experiment to measure and quantify intermediate subsurface flow paths in the hyporheic zone.

2.5. SYNTHESIS

Taken together, Chapters 2-4 systematically explore these key questions in the river corridor: (1) How do decisions to study at one scale bias our understanding of other scales? And (2) How important are oft-overlooked intermediate flow paths that occur at the segment scale as we strive to represent river corridor exchange outcomes in river networks? These three independent studies each advance our understanding of river corridor exchange, and together provide a cohesive understanding of intermediate flow paths. Ultimately, these studies combined, highlight the importance of incorporating intermediate flow paths and segment-scale processes. Ultimately, these studies combined highlight the importance of including the segment scale in hyporheic studies and improve our understanding and predictive power when scaling to whole networks.

3. TESTING HIDDEN ASSUMPTIONS OF REPRESENTATIVENESS IN REACH-SCALE STUDIES OF HYPORHEIC EXCHANGE

PAIGE S. BECKER
ADAM S. WARD
SKUYLER P. HERZOG
STEVEN M. WONDZELL

PUBLISHED IN WATER RESOURCES RESEARCH
VOLUME 59, ISSUE 1
HTTPS://DOI.ORG/10.1029/2022WR032718

3.1. 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.

3.2. 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, Schmadel, 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 10s to 100s of m 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, Wondzell, Gooseff, & Haggerty, 2005), intentionally biased selection to represent expected end-members (Wondzell, 2006), or to isolate other specific attributes (e.g., human impacts; Ward, Morgan, White, & Royer, 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, Morgan, 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, Schmadel, et al., 2018a). Thus, if the reach length is too short, longer flow paths in the hyporheic zone are ignored (Gooseff et al., 2003, 2006; J. W. Harvey et al., 1996; 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 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 (1) 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 (2) 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 20 WCW and 100 m, both of which have been previously used in the basin. (Anderson et al., 2005; Ward, Zarnetske, et al., 2019). The 20 WCW 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 10s to 100s of m, and the segment is defined as a portion of the drainage network showing similar valley scale morphologies, typically 100s to 1000s 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?

3.3. Methods

3.3.1. Field Characterization

This study was conducted at the H.J. Andrews Experimental Forest (HJA), Cascade Mountains, Oregon, USA, a 6400-ha drainage basin with elevations ranging from 410 m to 1630 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 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 3rd order and smaller streams (Figure 2.1). 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.

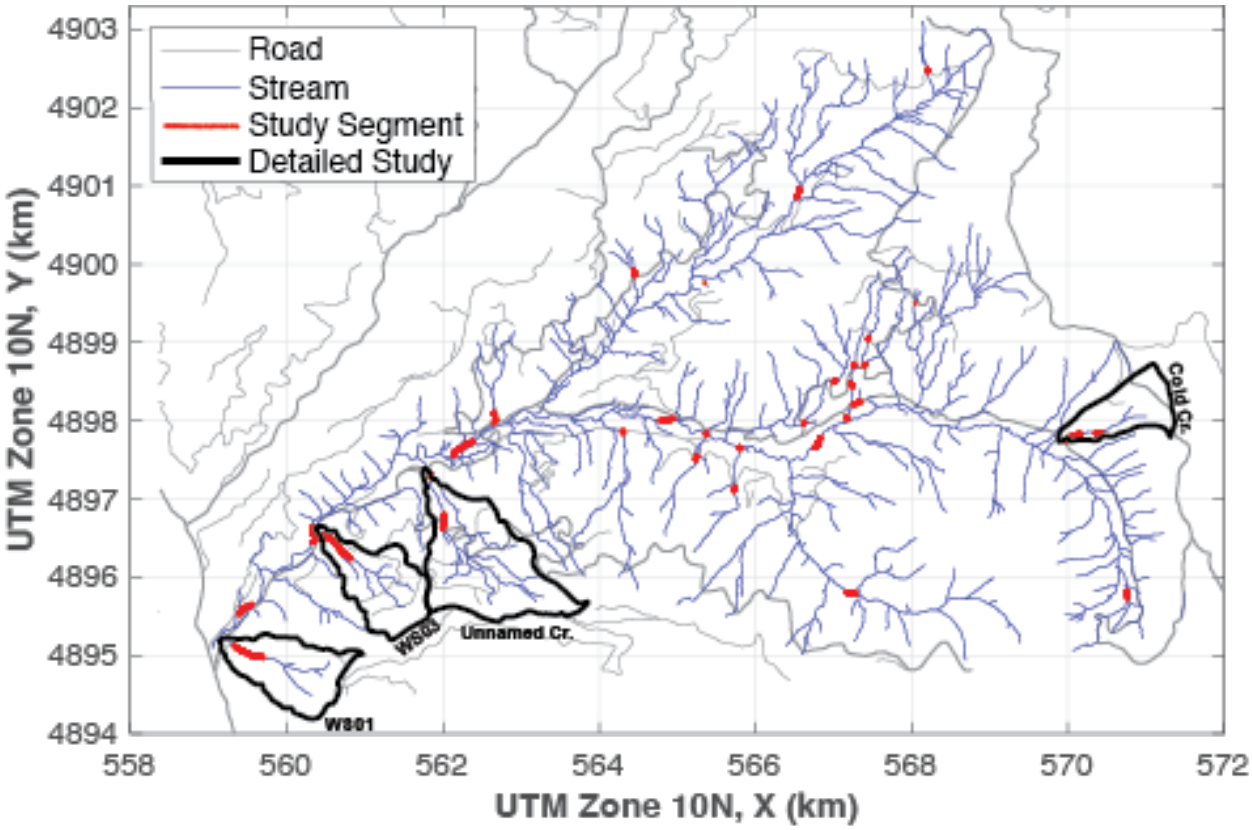

Figure 2.1: Maps of the HJ Andrews including the road network (grey), stream network (blue), and study sites (red). The four catchments outlined in black are those where more detailed surveys and simulations were included to assess within-site variation. Catchments, roads, and streams follow exactly those detailed in Ward et al. (2019).

3.3.2. 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, Schmadel, 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 meters for first-order, three meters for second-order, and four meters 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) (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 m to 0.1 m in height (a summary of the computational mesh is provided in section 7.1 and 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, Schmadel, 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, Ward, & Wondzell, 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*) (J. Harvey et al., 2019). *RSF* is the product of river connectivity and the Damköhler number, calculated as

$$RSF = \frac{\tau_s}{\tau_r} * \frac{L_c}{L_s} \quad (1)$$

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 (J. 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

$$L_s = \frac{Q}{q_s * w} \quad (2)$$

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 Table 7.3 and Becker et al (2022).

3.3.3. 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 (SI Figure 1). 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 5th 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 ($p_{levene} < 0.10$) indicating the rejection of the null hypothesis, meaning variances are significantly different between the two populations. If we find $p_{levene} \geq 0.10$, 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.

3.4. Results

3.4.1. 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% to 85% of the reaches considered (Figure 2.2A, 2.2E, 2.2I, 2.2M). Similarly, 100-m reaches produced *TTDs* that were statistically indistinguishable from the reference segment in 21% to 87% of cases ($p_{kw} > 0.1$; Fig. 2.2A, 2.2E, 2.2I, 2.2M). 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 $p_{kw} < 0.1$ (Fig. 2.2E, 2.2I, 2.2M). However, for Cold Creek the distribution was skewed towards high p_{kw} values (Fig. 2.2A), 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 (Fig. 2.2A, 2.2I, 2.2M), while the inverse was true for Unnamed Creek (Fig 2.2E). 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% to 72% of the time, while 20WCW reaches were representative 18 to 72% of the time (Figure 2.2B, 2.2F, 2.2J, 2.2N). 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 the 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% to 100% of the time, compared to 22% to 100% of 20WCW reach windows across all four segments (Figure 2.2C, 2.2G, 2.2K, 2.2O). As with estimates for flux, 100-m reaches had greater accuracy and precision than 20WCW reaches for percent upwelling (Figure 2.2C, 2.2G, 2.2K, 2.2O). The total range of error decreased for the 100-m reach lengths compared to the 20WCW conditions, with 100% of all 100-m reaches 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 in 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 (Figure 2.2D, 2.2H, 2.2L, 2.2P). 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.

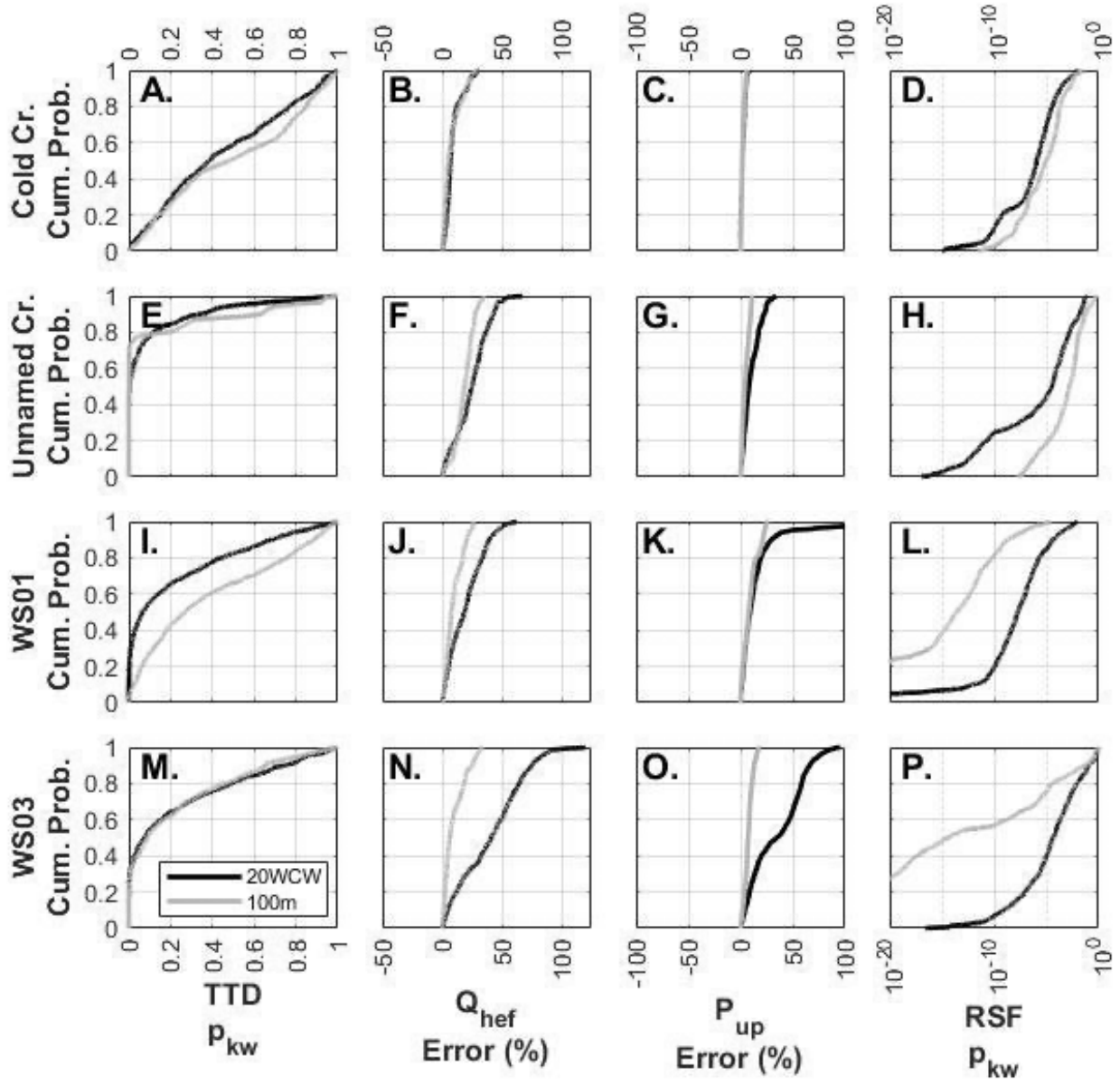

Figure 2.2: Cumulative distribution functions of both p_{kw} and percent error values for the four metrics: *TTD* (left column), Q_{hef} (middle-left column), P_{up} (middle-right column), and *RSF* (right column). Rows from top to bottom are Cold Creek, Unnamed Creek, WS01, and WS03. In all cases, results 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.

2.4.2 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 Fig. 2.3). 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, and p_{SR} indicates the p-value for the test that the two are correlated). 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 the segment. Taken together, we find little indication that a subset of features or locations are systematically included in or excluded from representative reaches.

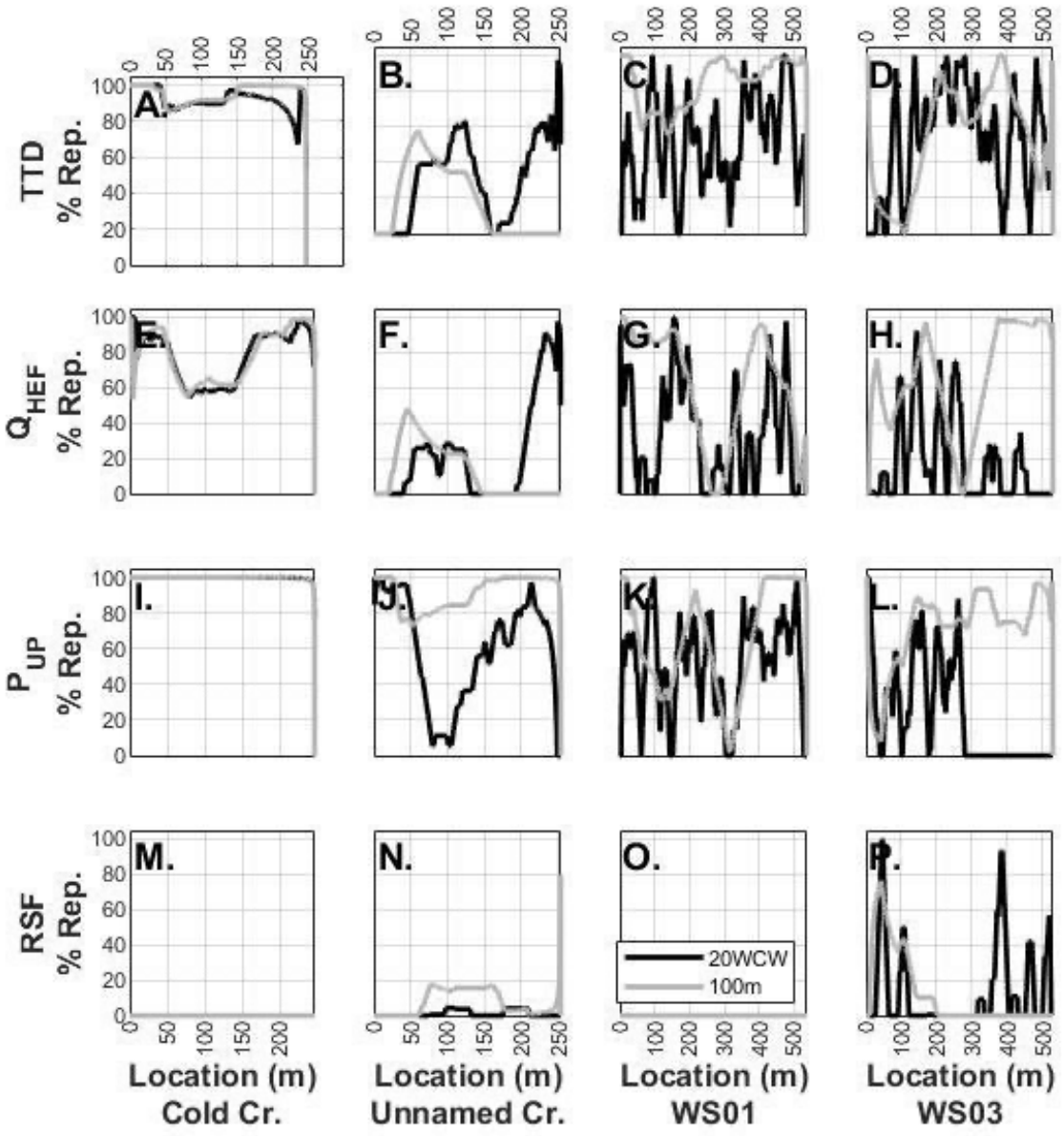

Figure 2.3: Locations and percent of total possible occurrences of representative reaches for the four metrics, based on $p_{kw} > 0.10$ for Transit Time Distributions and *RSF*, and error < 10% for Q_{hef} and P_{up}. Rows from top to bottom are: *TTD*, Q_{hef}, P_{up}, and *RSF*. Columns from left 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.

Table 2.1: Spearman's rank coefficient, *r*, and p-value for occurrences of representative reaches for 20WCW and 100-m reaches. *r* closer to 1 suggests strong correlation. Small p-value means rejecting the null hypothesis that there is no correlation between the two.

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*
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*
WS03	0.23, p << 0.001	-0.17, p << 0.001	-0.28, p << 0.001	0.12, p << 0.001

* Values of NA reflect cases where *r* cannot be calculated because values are uniform.

2.4.3 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 2.4, Table 2.2). 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 2.2). 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 (Figure 2.4C and 2.4D).

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 2.3), 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.

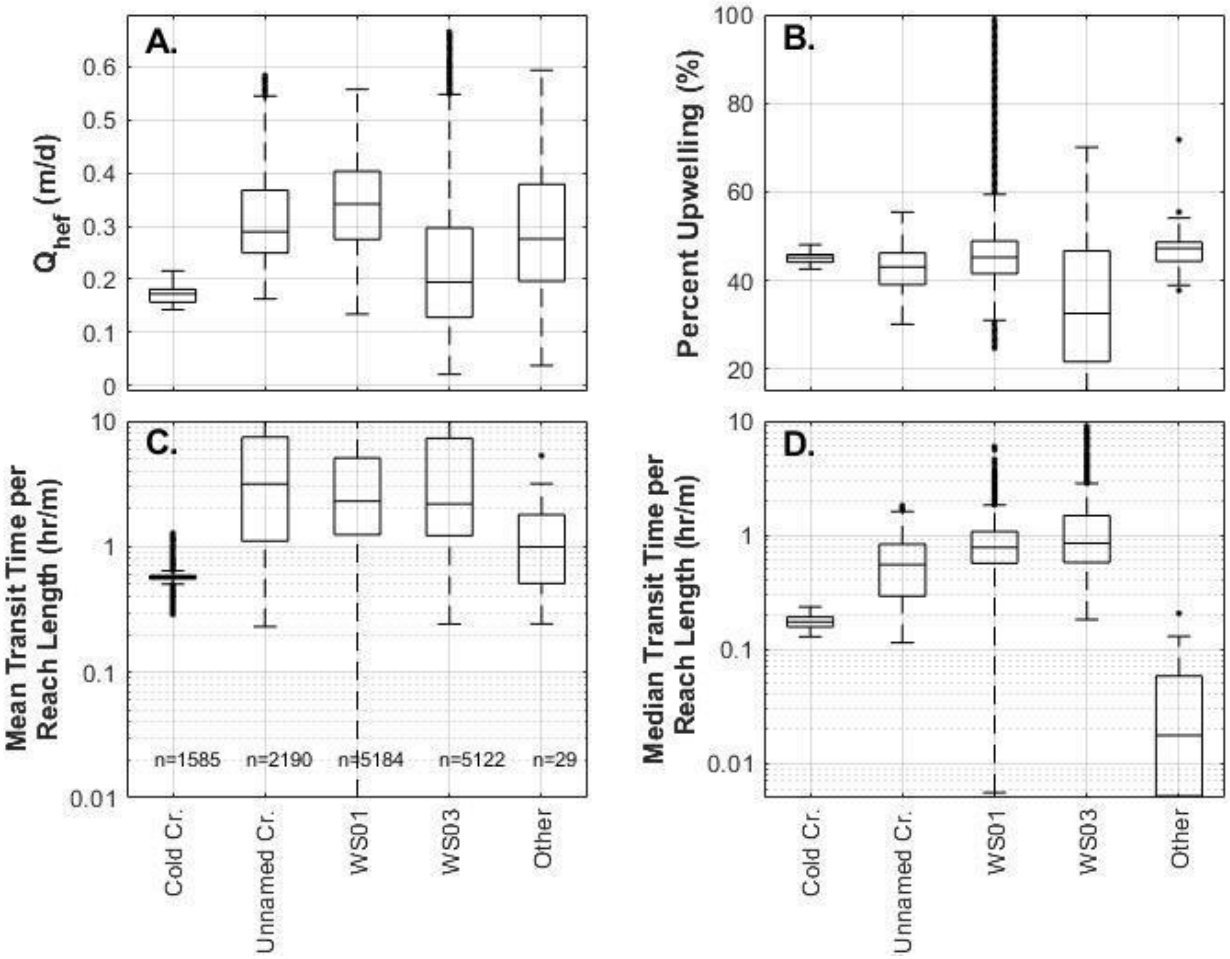

Figure 2.4: Comparison of exchange metrics between and within sites of a 5th order basin for 20WCW reaches. A. is showing Q_{hef} (m/s) for all 20WCW windows of the 4 reference segments and the 29 other sites. B. is percent of particles upwelling (P_{up}), C. is average transit time distribution normalized by reach length (hr/m), and D is median transit time distribution normalized by reach length (hr/m).

Table 2.2: Summary results comparing within and between site means, skewness (γ), and coefficient of variation (*CV*) for 100-m and 20WCW reaches. Results are presented for flux (Q_{hef}), fraction upwelling (P_{up}), median transit times, and *RSF*.

		100-m reaches				20WCW reaches				
		WS01	**WS03**	**Cold**	**Unmd.**	**WS01**	**WS03**	**Cold**	**Unmd.**	**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	.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}	-0.41	-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

Table 2.3: p-values using Levene Test for Equality of variances comparing Cold, Unnamed, WS01, and WS03 to the remaining 29 sites modeled at 20WCW. 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.

	Cold vs. 29	Unnamed vs. 29	WS01 vs. 29	WS03 vs. 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 ↑

3.5. Discussion

3.5.1. 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 (Fig. 2.3). 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 (Fig. 2.3J), but many of the locations for Q_{hef} were grouped together (Figure 2.3F). 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 RSF is to be representative (Fig. 2.3P). 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 vs. 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.

3.5.2. 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 their respective watersheds. Both Wondzell reaches have a higher cumulative probability density for longer residence times compared to other reaches along the segment (Figure 2.5A and 2.5C). Similarly, the WS01 Wondzell reach is at the 76th percentile for flux while the WS03 Wondzell reach was at the 86th percentile for flux, (Fig. 2.5B and 2.5D). 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 & 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., 20 WCW) would be more feasible and manageable, they are not necessarily able to to capture longer timescale flowpaths 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.

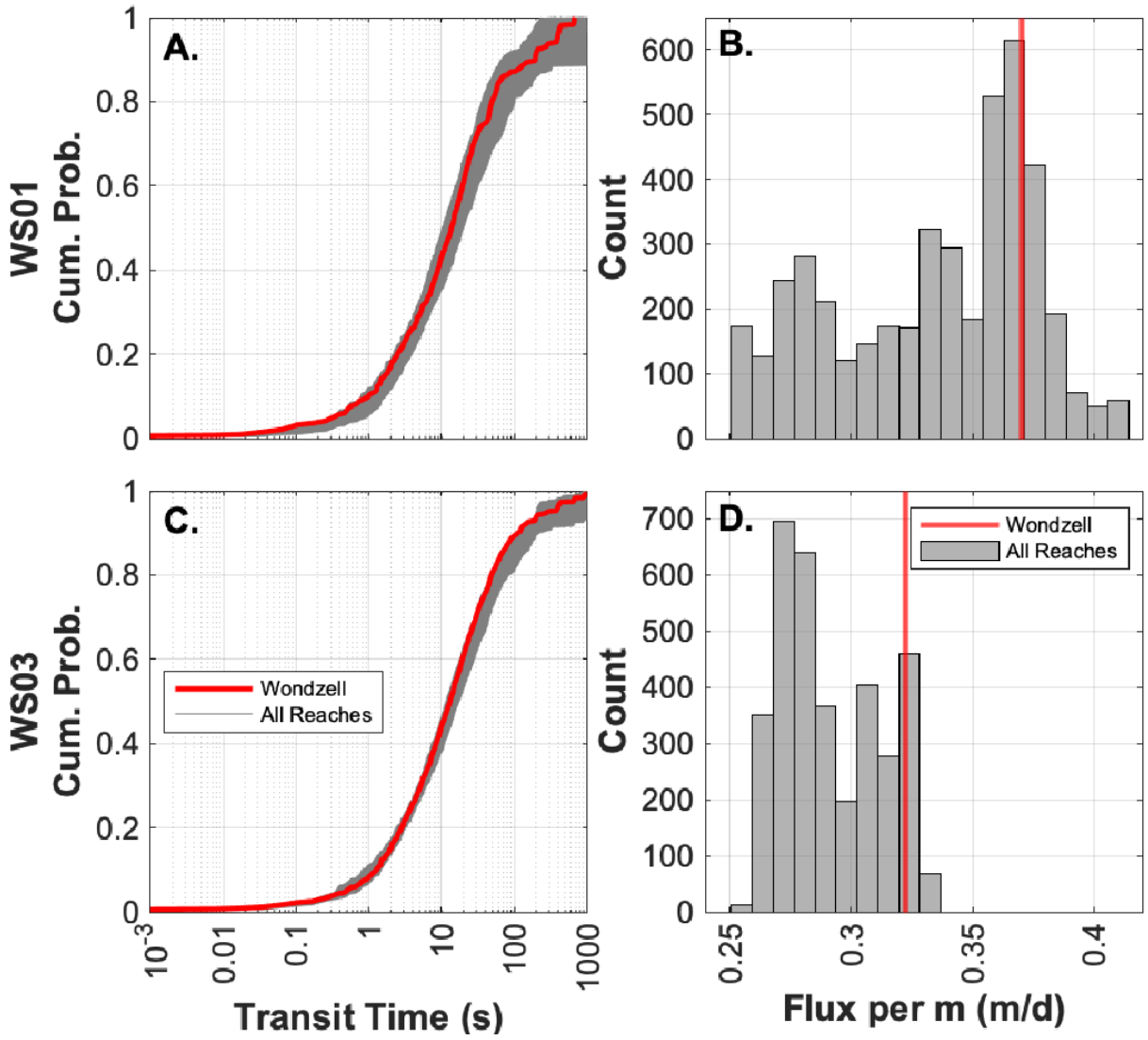

Figure 2.5: Comparison of the reaches studied by Wondzell (2006) versus other possible reaches of the same lengths in WS01 and WS03. 2.5A and 2.5C are cumulative distribution plots of the transit time distributions. Red line is the TTD of the reach from Wondzell, 2006. Gray lines are TTDs of other possible reaches along WS01 and WS03 of the same length. 2.5B and 2.5D are distributions of *Qhef* 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.

3.5.3. 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 2.4). 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 head waters 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).

3.5.4. 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: (1) clearly stating the aim of the study and to what degree assumptions or testing of representativeness are made; (2) quantify the within- and/or between-site variation to provide context for interpretations and how variation may impact interpreted results; and (3) 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.

3.6. 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 Fig. 2.5). 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 2.4.2, the 29 headwaters sampled across the network behaved differently than the four reference streams ($p_{levene} < 0.1$)

indicating that a small sample of headwater 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 turnover 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 headwater streams.

3.7. Acknowledgments

The authors thank Tadd Bindas, Hunter Stanke, Billy Stansfield, Tatiana Fabiana, Nina Jung, Caroline Kryder-Reid, and Morgan Schultz, and for help with the surveys that took place in 2018 and 2019. Financial Support is from the National Science Foundation (grant no. 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 supported in part by Department of Energy awards DE0SC000022 and DE-SC0019377. Ward's time was supported in party 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.

3.8. Open Data

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

4. IMPACTS OF TIMESCALE REDUCTION AND PROCESS TRUNCATION AND ON HYPORHEIC EXCHANGE IN A HEADWATER MOUNTAIN STREAM

4.1. ABSTRACT

Intermediate-length flow paths are subsurface flow paths that span multiple features and reaches of a stream. However, current field methods and models do not account for this intermediate-length flow paths resulting in a lack of understanding about where and when accounting for these subsurface flow paths is important for predicting hyporheic exchange. Therefore, there is a need to understand how incorporating underflow changes our understanding of processes in the hyporheic zone. This study explores the impacts of truncating time scales and discretizing segments into reaches to understand the impacts of not including intermediate-length flow paths. We used the Advanced Terrestrial Simulator with the Advection Dispersion Equation with Lagrangian Subgrids (ATS-ADELS) to model the river corridor with different forms of truncations, one to mimic window of detections issue from solute tracers, and one to mimic truncation through study-reach location in the form of discretization strategies.

4.2. INTRODUCTION

The hyporheic zone is an important region of streams for numerous ecosystem services and functions including pollutant degradation, regulating stream temperature, and providing habitat for aquatic species (Boano et al., 2014; Harvey & Gooseff, 2015). Recently, several modeling studies have extended hyporheic processes representation to the scale of river networks (e.g., Gomez-Velez & Harvey, 2014; Ward & Packman, 2019). These models are typically based on processes that were studied at the reach scale (10s to 100s of m, Montgomery & Buffington, 1997; Ward, 2016), aggregating the processes to represent networks as a series of smaller reaches. Additionally, the field studies that underlie these processes are inherently limited by the scale of study such as processes emerging at larger scales are not present at the reach scale, and limitations of the techniques themselves (e.g., stream solute tracers are known to be biased toward the shortest and fastest flow paths; (Payn et al., 2009; Ward et al., 2023). Consequently, longer subsurface flow paths are comparatively understudied (Ward et al., 2023), despite a recognition that they exist and might be biogeochemically important (Alexander et al., 2000). Nonetheless, these intermediate-length flow paths (i.e., flows that extend multiple reaches in the subsurface) have been functionally omitted from network-scale representations. Here, we explore

the consequences of ignoring or omitting intermediate-scale flow paths in our representation of river corridors, including both conservative and reactive transport.

Present efforts to represent river corridor exchange at the scale of river networks rely on a critical assumption that networks can be successfully represented as a series of reaches. For example, Kiel & Bayani Cardenas, (2014) represented meander-driven exchange at the scale of river basins, and Gomez-Velez et al., (2015) did the same for bedform-driven exchange. The impact of the reach-scale process has been described based on the reactions of interest and potential for reaction (Harvey et al., 2019) and included in network-scale simulations of nutrient transport and fate (Boyer et al., 2006; Runkel & Broshears, 1991).

However, this success comes at a cost. Reach-to-network scaling, or representing networks as a series of reaches has inherent limitations. Using reaches implicitly excluded the representation of process dynamics that emerge at larger scales (Becker et al., 2023). Underflow, for example, is the down-valley flow of water in riparian aquifers and along the river corridor (Toth, 1963). Flow paths of underflow are longer than the traditional reach scale (Becker et al., 2023; Herzog et al., 2019) and known to be biogeochemically important (Harvey et al., 2019) but are omitted when the reach scale is used to form the building block for scaling studies.

In addition to the assumptions inherent in reach-to-network scaling, our reach-scale strategies also suffer from at least two distinct forms of truncation that limit which processes and timescales are represented. First, field-based truncation due to the window of detection issues with common field methods biases empirical observations toward the shortest and fastest flow paths (Drummond et al., 2012; Harvey et al., 1996; Payn et al., 2009; Ward et al., 2023). Inherent limitations of detection in the stream, commonly the window of detection (WoD), mean that the longest elapsed timescale from the release of a solute tracer to the study stream reaches is the latest time that tracer mass is discernible from background concentrations at the downstream end of the study reach (Ward et al., 2023). Functionally, some of the unrecovered tracer is just beyond this WoD timescale, but the unknown fate of the mass means those timescales cannot be represented at the reach scale, so they will not emerge when aggregating at the network scale. The second form of truncation at the reach scale is due to the selection of study reaches. Due to flow path geometry, there are some flows that never get traced by the solute injection, as well as flows that get traced at the upstream end of the reach but bypassing in-stream observations at the end of the study reach (Payn et al., 2009). Thus, while these flow paths are understood to exist,

they are not captured in traditional reach-scale empirical nor modeling studies, functionally omitting their impact on both reach- and network-scale studies.

To evaluate the impact of reach-scale truncation on behavior at the larger segment scale, we consider the role of down-valley flow in the subsurface along intermediate flow paths (commonly 'underflow'; Castro & Hornberger, 1991). Common 2-D approaches represent underflow at the length of the domain, with focused downwelling at the upgradient end of the model domain and upwelling at the down-gradient end of the model domain (e.g.,Gooseff et al., 2006; Herzog et al., 2019; Schmadel et al., 2017), setting a maximum timescale for representation. Other approaches may be parameterized solely from solute tracer data (e.g., Harman et al., 2016; Harman & Kim, 2014; Jan et al., 2021; Painter, 2018), which have the capability of representing longer flow paths but lack empirical data to parameterize transit time distributions.

This study explores how truncation changes our understanding of hyporheic exchange at the scale of segments. Specifically, we ask:

1. How does truncation of the transit-time distribution change our understanding of conservative transport processes at the segment scale?
2. How does reach selection truncation change our understanding of conservative transport processes at the segment scale?
3. How do these truncations change our understanding of reactive transport processes at the segment scale?

To answer these questions, we conduct a series of simulations using the ATS-ADELS hyporheic model (after Painter, 2018), parameterizing model geometry and transit times from pre-existing finite-element model (Herzog et al., 2019). By simulating conservative and reactive solute transport in ATS with pre-established model parameters and sampling from a well-studied distributed model, we can isolate the role of model discretization in a realistic set of model experiments. We chose the segment (i.e., a combination of several reaches, (Montgomery & Buffington, 1998) as our scale of interest for two primary reasons. First, this is the scale at which reaches are first aggregated to describe transport behavior. Second, underflow is known to occur between reaches, but segments might be commonly selected using natural breaks to divide them (e.g., bedrock outcrops or geologic discontinuities selected as logical divisions).

4.3. Methods

4.3.1. Modeling Inputs for ATS-ADELS

Models presented in this study are based on field and modeling work completed in Watershed 1 (WS01) at the H.J Andrews Experimental Forest (HJA) in the central cascades of Oregon, United States. First, we replicated exactly the simulation of the entire study segment detailed in Herzog et al., (2019) using COMSOL Multiphysics. Briefly, streambed topography and water surface elevations were surveyed to parameterize the top boundary of the model. The model encompasses the stream and subsurface between two bedrock outcrops, ensuring no down-valley flow exists at the up or downstream end of the simulation domain. Key model outputs include flux and transit time for downwelling flow paths every 10 cm along the surface of the model domain. (See additional details in Becker et al., 2023; Herzog et al., 2019; Schmadel et al., 2017).

COMSOL model results were used to generate the inputs for the reduced complexity Advection Dispersion Equation with Lagrangian Subgrids (ADELS) model. Briefly, this model represents hyporheic zone transport using an ensemble of travel times, coupling the subsurface with the channel (Painter 2018). Here, we implement ADELS following existing strategies to simulate stream-hyporheic systems (Painter 2018, Jan 2021, and Rathore 2021) including a one-dimensional advection-dispersion-reaction equation for the surface channel coupled to a one-dimensional advection-reaction model in each subgrid (i.e., discretization) model. The subgrid model represents an ensemble of streamlines that are diverted into the hyporheic zone before returning to the channel (Painter 2018, A. Jan et al., 2021, Rathore et al., 2021). For each ADELS simulation, COMSOL results were used to calculate flux-weighted transit time distributions to describe hyporheic transport. Following Painter, (2018) and Rathore et al., (2021), the cumulative distribution of flux-weighted transit times was separated by evenly spaced quantiles and the difference between those quantiles (DTs) was calculated and used as the input for the model. The quantile differences calculated above make up the ensemble of streamlines.

Table 3.1: Model input parameters

Variable	Value	Unit
Segment length	300	m
Discharge	0.01	m3/s
Cross-sectional area	0.464	m2
Hyporheic exchange (alpha)	1.7e-5	1/s
Water flux (q)	2.155e-3	m/s
Cell length	3	m
number of quantile differences (# of DTs)	20	(unitless)
Dispersion	0.01	m2/s

4.3.2. Model Experiment 1: Conservative Transport

With ADELS parameterized for the site, we conducted a series of model experiments of conservative solute transport. For each experiment, we stimulated a slug release of a conservative solute tracer into the upstream end of the model domain and tabulated concentration through time at the downstream end to produce a breakthrough curve (BTC). Key model parameters are detailed in Table 1. First, we simulated the entire 300-m segment between bedrock outcrops including the complete TTD for the study. This baseline model represents all timescales included in the COMSOL model and serves as the basis for comparison of the effect of truncation both temporally and spatially.

To understand the impacts of truncating a TTD, we truncated the entire distribution to include only 99%, 95%, 90%, 80%, and 70% of the fastest flow paths. This approximates the impact of a segment-scale solute tracer study conducted with increasing mass loss (i.e., a window of detection that misses increasingly longer flow paths). The baseline model was parameterized with each of these truncated transit times. DTs from those transit time distributions and breakthrough curves of the conservative tracer were extracted for each simulation.

Table 3.2: Modeling cases

Discretization type	**Subgrid lengths**	**Model Name**
Single subgrid, full transit time distribution	300 m	Baseline Model
WoD truncation	300 m	WoD_99
WoD truncation	300 m	WoD_95
WoD truncation	300 m	WoD_90
WoD truncation	300 m	WoD_80
WoD truncation	300 m	WoD_70
Spatial Discretization Homogeneous discretization (even division of the study segment)	60.2 m for each.	Equal_disc
Spatial Discretization Heterogeneous discretization (based on turnover cells)	64.7 m, 54.9 m, 75.5 m, 51.9 m, and 53.9 m	TO_disc

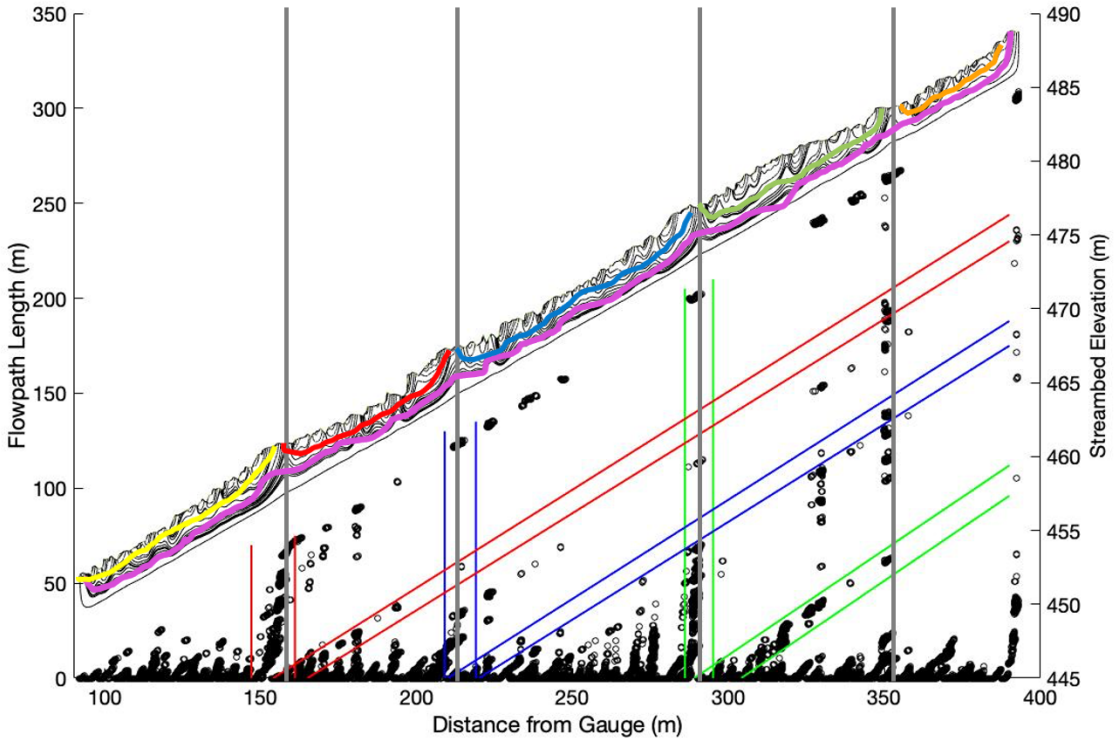

Figure 3.1: Segment of WS01 used for ATS-ADELS model showing the turnover locations from Herzog et al., 2019. The intermediate-length flow paths are highlighted in purple, flowing along the entire segment. Reach-scale flow paths are highlighted in yellow, red, blue, green, and orange to highlight the flow paths driven by turnover points.

4.3.3. Break-Through Curve Analysis

Breakthrough curves extracted from the stream at the downstream end of the model were analyzed using standard techniques for stream solute tracers, including percent mass arrival times (t05,t10, etc. where t05 represents the time at which 5% of the cumulative mass passed the monitoring location) and peak concentration (C_{peak}). Additionally, we used central temporal moments as proxies for advection (median arrival time), dispersion (coefficient of variance), and transient storage (skewness), (Wlostowski et al., 2017).

4.3.4. Model Experiment 2: Reactive Transport

To assess the impacts and implications of water quality, the truncations and modeling cases were also coupled with PFLOTRAN to model denitrification in the stream. Initial conditions,

reactions, and rates were based on values used by Jan et al., (2021). Concentrations and total mass nitrate (NO3-) were calculated at the downstream end of the stream. Percent changes in concentration were calculated between the baseline and model cases to compare impacts on water quality. Inputs for the PFLOTRAN coupling are based on previously published work by Jan et al., (2021) and can be found in the Appendix, section 7.2.

4.4. Results

4.4.1. Effects of Break-Through Curve Truncation

Truncating the transit time distributions input into the model had little effect on advective processes, such as the first 75% of mass moving through the reach (Figure 3.2), and the first central moment (Table 3.3). The difference in peak concentration (c_peak) between the full transit time distribution and using only the 70% fastest flow paths was 21%. Differences in percent mass arrival times varied little between modeling cases for a majority of the solute mass, with the difference in arrival times being less than half of a day for the first 75% of the mass (Figure 3.2). However, BTC metrics that are influenced more by late time tailing were more sensitive to truncation. For example, skewness decreases dramatically with truncation, with a 93% difference between the complete distribution and using only 70% of the fastest transit times (Figure 3.2).

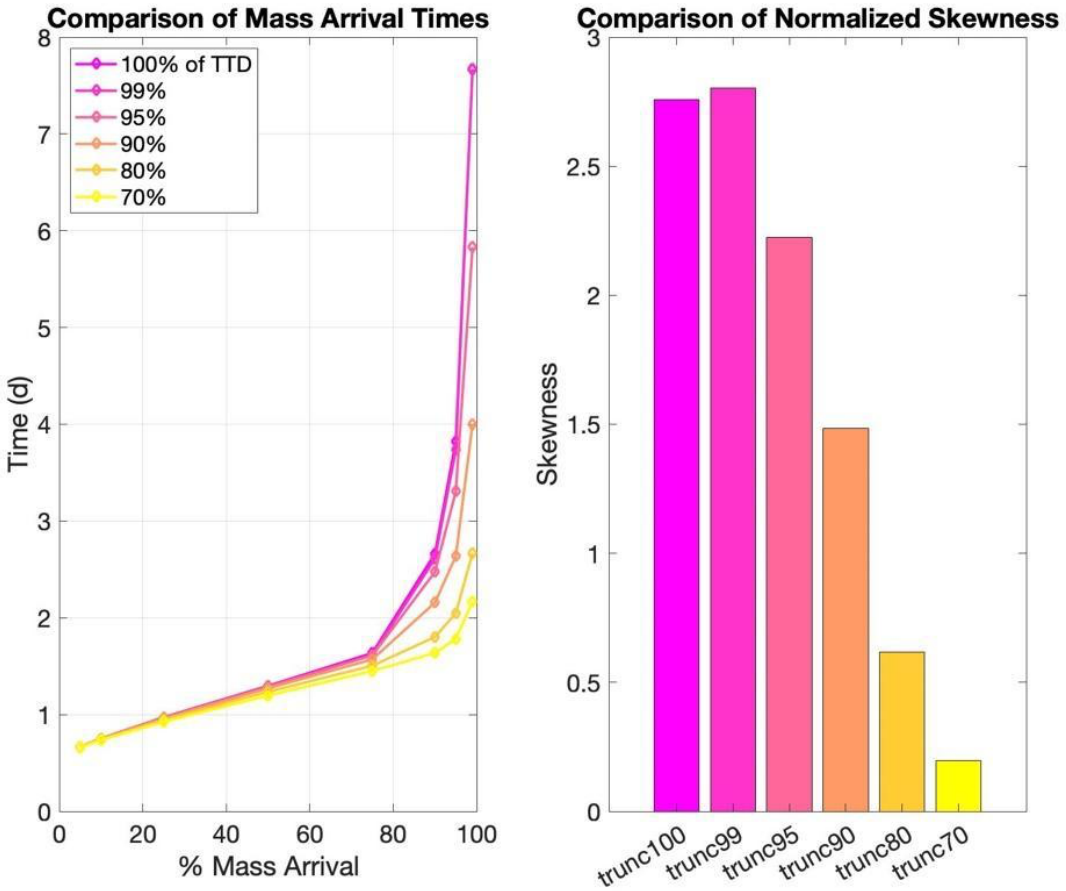

Figure 3.2: Comparison of mass arrival times and skewness between cases

Table 3.3: C_{peak}, M1, CV, and Skewness from BTC truncations in order of increasing temporal moments

Truncation	C_{peak} (uM)	1st moment (d)	CV	Skewness
100	29.15	1.56	0.677	2.76
99	29.11	1.55	0.672	2.80
95	29.91	1.48	0.564	2.22
90	31.04	1.37	0.435	1.48
80	33.63	1.26	0.329	0.62
70	35.39	1.20	0.288	0.20

4.4.2. Effects of Discretization Truncation

Discretization truncation has a much smaller impact as WoD truncation, with the most notable deviation from baseline occurring for the late-time tailing of the BTC (Figure 3.3). Using turnover points as a way to discretize the model produced a better prediction of transport in relation to the complete model compared to the arbitrary discretization. Comparing mass arrival times, a similar pattern as BTC truncation occurs where the first 75% of the mass is predicted to move through the stream at the same time (Figure 3.3), but as the last 25% of the mass moves through the system, the times start to diverge. Additionally, for all temporal moments except skewness, turnover discretization better matches the complete discretization compared to the arbitrary (equal) discretization choice. However, using equally spaced subgrids results in a skewness closer to the single subgrid modeling case. The arbitrary discretization model case had a difference of 7% from the single subgrid skewness whereas the turnover discretization had a difference of 11%.

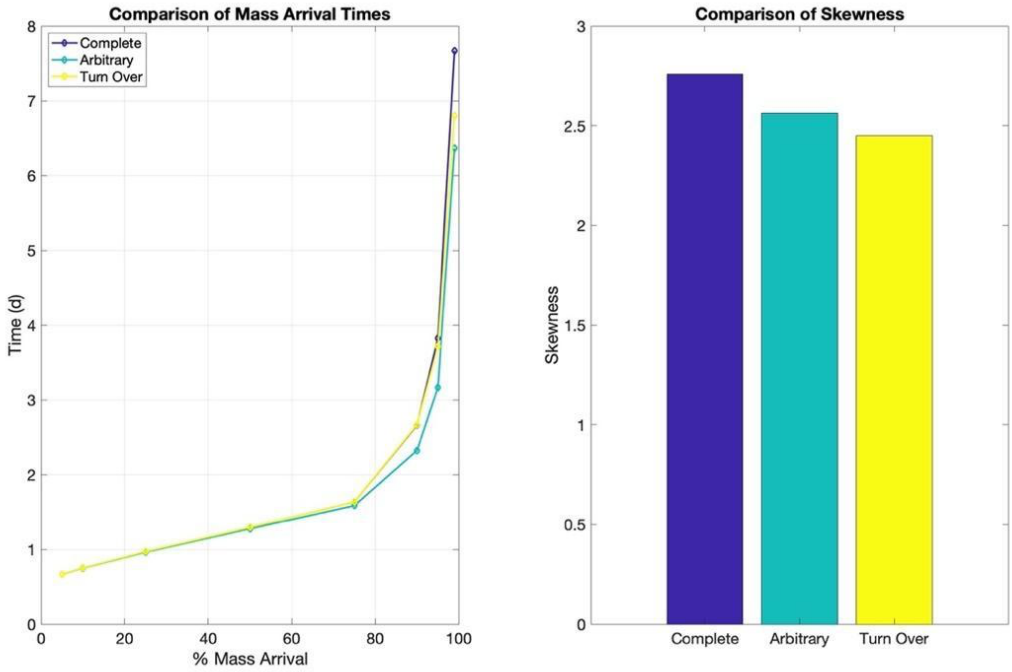

Figure 3.3: Comparison of mass arrival times and skewness between cases

Table 3.4: C_{peak}, M1, CV, and Skewness from BTC truncations in order of increasing temporal moments

Discretization	C_{peak} (uM)	1st moment (d)	CV	Skewness
Single Subgrid (complete)	29.15	1.56	0.677	2.76
Equal Subgrids	30.50	1.46	0.571	2.56
Turnover Subgrids	29.41	1.54	0.632	2.45

4.4.3. Coupling with a Biogeochemical Reaction

Truncations of the transit time distributions – both BTC and discretization truncation - have little impact on the segment-scale biogeochemistry of denitrification. In comparison to the entire distribution, using 70% of the fastest flow paths had less than a 3% difference in total Nitrate mass at the outlet, with the percent difference decreasing as more of the flow paths were

included (Figure 3.4). Additionally, changing discretization had little effect on denitrification between cases when compared to the baseline modeling case. The equal subgrid modeling case had less than 1% difference from the baseline model. Discretization based on turnover points did marginally better than equally spaced subgrids but were not significantly different.

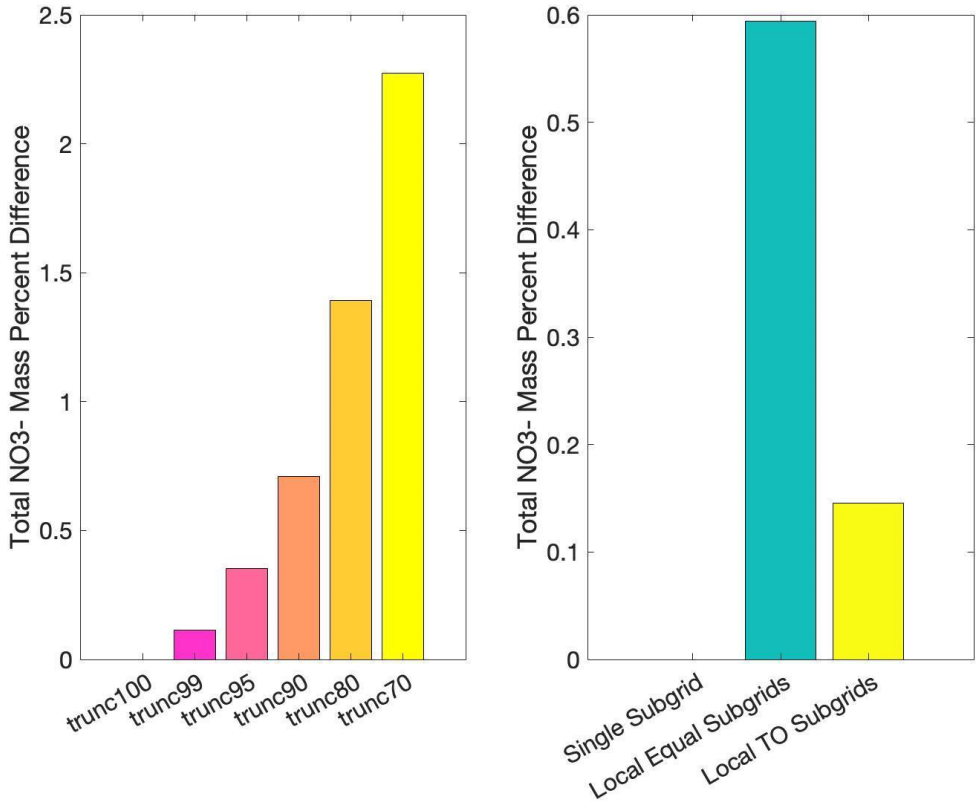

Figure 3.4: Total NO3- mass percent difference between modeling cases. Trunc100 and Single Subgrid are zero due to being the baseline model for comparison. Note the scales are different.

4.5. Discussion

4.5.1. The Inclusion of Intermediate-Length Flow Paths Improves the Prediction of Late-Time Tailing

Both WoD and discretization truncation of TTDs resulted in under-prediction of late-time mass arrival. Our model results show that the incorporation of intermediate flow paths changes the timescales of transport through the segment, manifesting as changes when mass arrives. By characterizing only 70% of the fastest flow paths, for example, our interpretation would be functionally unchanged for primarily advective transport (e.g., M1 differs by less than nine hours), but our expectation for the 99% of the mass is 132 hours longer for the full transit time than the truncated transit time distribution. As the transit time is truncated more, the tailing behaviors of mass exhibit less of an exponential relationship and appear to be more of a linear relationship. Thus, expanding this across a whole network (combining multiple 300 m segments) would greatly change the prediction of transport for mass through the system.

Discretization truncation of underflow based on location showed less deviation from the truth compared to temporal truncation. Both arbitrary discretization and turnover points resulted in differences in the prediction of late arrival times for mass through the hyporheic zone. By using arbitrary and informed discretization, our interpretation of advective transport would be largely unchanged (e.g., M1 differs by less than two and half hours), but our expectation for the 99% of the mass is 31 hours longer for the undiscretized model versus the arbitrary discretization. Skewness is the only temporal moment where arbitrary discretization had a better prediction than using turnover points for discretization. These results suggest that the geomorphic features controlling where turnover happens are important and should be considered when discretizing the model. Linking back to the turnover points used by Herzog et al, these results should not come as a surprise. It is well known that geomorphic features drive both local and intermediate scales (Gomez-Velez & Harvey, 2014; Herzog et al., 2019; Toth, 1963). Using geomorphic features rather than arbitrary discretization can help improve our predictive power of hyporheic exchange.

While our study considered the impact of truncation at relatively small scales, these deviations would only grow as we aggregate truncations along a study reach. Because reaches are convolved in series to represent segments and networks, systematic truncation errors will grow as scale increases. These findings are consistent with previous studies examining the

impacts of using truncated transit time distributions due to solute tracers. Stewart et al., (2010) found that our view of how catchments store and transmit water is skewed due to the removal of the long tails in transit time distributions. Additionally, Drummond et al., (2012) found that hyporheic exchange is underestimated when truncation of the tail is not accounted for.

4.5.2. The Inclusion of Intermediate-Length Flow Paths is Not Mission-Critical for Denitrification in our Study System

Given that the inclusion of intermediate flow paths dramatically changed the prediction of the conservative mass transport through the hyporheic zone, it was expected that a biogeochemical reaction would behave similarly, if not more exaggerated, due to the reliance on long flow paths to reach an anoxic state. Stewart et al., 2010 found that predictions for nitrate concentrations were dramatically different when using a truncated breakthrough curve versus the inclusion of the late-time tailing. Additionally, a study on hyporheic exchange for the upper Mississippi River basin found that "channel geomorphology is a principal control on hyporheic exchange and biogeochemical potential of rivers" (Gomez-Velez et al., 2015), suggesting that incorporating turnover points rather than arbitrary discretization would matter for biogeochemical processes, such as denitrification. However, results in this study show that the truncation and discretization choices have little impact on this particular set of parameters to model denitrification. Values for the denitrification coupling were from Jan et al. (2021) and applied to physical conditions that align with HJ Andrews Experimental Forest. Expanding this model to larger scales or with different reactions that require long-time scales would likely give the expected results.

4.5.3. Future Work and the Goldilocks Zone of Timescales and Reaction Rates

Although the denitrification model had little effect from omitting underflow, it is expected that omitting the underflow would be most consequential when reaction rates align with the timescales that are often truncated. For proof of concept using the Damkohler number, the ratio of transit time to reaction times, a small Damkohler number or $Da << 1$ means the hyporheic zone is reaction limited, and the reactant does not have enough time in the hyporheic zone for a reaction to occur. A large Damkohler number ($Da >> 1$) is transport-limited, so the reaction

occurs regardless of the time in the hyporheic zone (J. Harvey et al., 2019). If the hyporheic zone is reaction-limited, omission or inclusion of underflow does not matter because the reaction rates would be slower than any of the possible flow paths and transit times available. Additionally, if the hyporheic zone is transport-limited, the inclusion of underflow will not matter because long timescales are not needed for the reaction to happen. This concept will be explored in future work using a first-order reaction and changing reaction rates across multiple orders of magnitudes with the different truncation models.

4.6. Conclusions

As hyporheic models continue to be developed and advanced across space and scale, our results suggest it is important to 'tap the brakes' as we represent networks and segments as convolutions of reaches, particularly to ensure larger-scale simulations are reflecting the processes and scales of interest (Becker et al., 2023). First, it would be beneficial to do forward modeling of the conditions and processes of interest to understand if the inclusion of intermediate flow paths impacts the results in a meaningful way. As shown, under certain conditions, the inclusion of the intermediate flow paths does matter, particularly for the prediction of mass transport in the hyporheic zone. However, the inclusion of intermediate flow paths is less important for the denitrification processes tested here, but knowing a priori to field studies or expanded modeling studies, can help improve our predictive power of hyporheic exchange across scales.

Further, the discretization of models should be based on geomorphic features expected to control exchange rather than equal or arbitrary discretization for processes. Put plainly, the selection of turnover points has a clear advantage over arbitrary discretization. Studies have shown that geomorphology drives exchange (Gomez-Velez & Harvey, 2014; Herzog et al., 2019; Leopold et al., 1964), so models should match these findings and be discretized based on geomorphology rather than the mathematical convenience of equal-length segments. The turnover points from Herzog et al., that were tested in this study were based on field survey studies from a heavily studied watershed. This is not always possible, however, remote sensing allows for the determination of changes in geomorphology that would induce exchange, such as changes in valley width, slope, tributaries, step-pool riffles (Anderson et al., 2005), bends (Stonedahl et al., 2013), and more. Many models, at large scales especially, use arbitrary or

equally discretized segments of a stream, such as 1km discretization (Gomez-Velez et al., 2015). The 300 m used in this model had a difference of a day between the equal discretization and local flow paths vs the whole transit times. This difference can get compounded as the model is sued for larger scales. For a single 1000 m reach, the difference in the time scale for mass transport might only be about 3 days, but when this is applied to the Mississippi River's length, the mass transport prediction is greatly exacerbated.

This study looked to explore the importance of incorporating intermediate flow paths in models for predicting hyporheic exchange. We looked at both the impacts of truncating transit time distributions, as is commonly done when using solute tracers and breakthrough curves. We found that mass transport and tailing are greatly underpredicted when transit times are truncated. We also explored how the discretization of the model impacts predictions. We tested how both equally spaced discretization and discretization based on geomorphology, or turnover points changed our prediction of hyporheic exchange. Results started to diverge from the single "underflow" model using spatial truncation. Results showed that using the turnover points rather than arbitrary discretization performed better than equal spacing, suggesting that models should be discretized based on geomorphic features. Further, aggregate transit times, rather than local transit times did better, likely due to the incorporation of longer flow paths.

Despite the physical transport being different between modeling cases, biogeochemical reactions did not show significant differences for either truncation or discretization for these particular reactions and parameters. The inclusion of intermediate flow paths, and discretization from geomorphic features that drive turnover though is important for improved production of hyporheic exchange and can have important implications when producing models at a large scale.

4.7. Acknowledgments

This work was possible thanks to the U.S. Department of Energy's Office of Science Graduate Student Research Program (SCGSR), 2020 Solicitation 2.

Thanks to Scott Painter, Saubhagya Rathore, and the Watershed Systems Modeling team at Oak Ridge National Laboratory for their support in conceptualizing this research and for helping execute and troubleshoot ATS-ADELS.

5. ESTIMATING UNDERFLOW IN THE RIVER CORRIDOR OF A HEADWATER MOUNTAIN STREAM

5.1. ABSTRACT

Underflow, the downstream flow of water in the subsurface in a valley bottom, bounded by areas of low permeability, is important for the ecological function and services of headwater streams. However, we lack quantitative evidence of the magnitude of this flux compared to other sources and sinks in the river corridor. The presence of underflow is commonly invoked to explain the mass lost in empirical solute tracer studies or as a result of boundary conditions in models, but seldom studied directly. In this study, we strategically use the geology of p in WS01 of the HJ Andrews Experimental Forest to quantify underflow in the river corridor. By conducting water balances between bedrock outcrops - where 100% of down-valley flow must be in the stream - we can quantify underflow within the study segment. Moreover, we can compare our estimates to other known sources (e.g., lateral inflows, precipitation) and sinks (e.g., evapotranspiration), comparing underflow to other hydrological fluxes known to be important in the river corridor. Additionally, a segment-scale solute tracer studies provides an independent estimate to confirm our perceptual model of the river corridor. We estimated that underflow could account for 60% of total flows through this river corridor. Further, from channel water balances, we determined that both segments were losing streams. Accounting for other losses, such as evapotranspiration, there are still losses in water along the segments, suggesting there might be deeper groundwater losses, not previously thought to occur in WS01.

5.2. INTRODUCTION

Underflow, or the downstream flow of water in the subsurface of a stream's valley, is known to occur in the river corridor (Harvey and Gooseff, 2015; Larkin and Sharp, 1992), and is thought to be important for the transport of water through the catchment (Castro & Hornberger, 1991; Voltz et al., 2013; Ward et al., 2018a). Underflow is differentiated from hyporheic flow based on spatial scale. Whereas hyporheic flow paths are commonly associated with an individual morphologic feature such as a bedform, pool-step-riffle, or meander bend (Gooseff et al., 2006; Stonedahl et al., 2010), underflow represents larger flow paths that span multiple features (Herzog et al., 2019). Underflow in the river corridor is akin to Toth's (1963) intermediate flow paths, where the geometry and timescale are not directly coupled to individual

surficial features. Instead of primarily occurring in response to surficial features (the most commonly studied scale and driver of hyporheic flow; Ward, 2015), underflow instead integrates multiple features and is equally responsive to subsurface heterogeneity (e.g., geologic structure) as surficial features (Herzog et al., 2019). In this study, we seek to empirically document the magnitude and spatial patterns in the recharge to and discharge from underflow, comparing its magnitude to other fluxes known to be important to reach- and segment-scale water balances.

Underflow is commonly invoked to explain stream-centric observations, particularly with respect to solute transport and fate, though actual empirical evidence of the flux is relatively limited. In catchment-scale water balances, fluxes to and from deep groundwater are commonly difficult to measure, and thus either assumed to be zero or invoked to close the water balance (Safeeq et al., 2021). We contend that underflow is treated similarly to groundwater fluxes in channel water balances at the scales of reaches (multiple geomorphic features such as a series of steps, pools, and riffles, Anderson et al., 2005) and segments (divided by major changes in the stream such as outcropping bedrocks, waterfalls, or stream confluences, Montgomery & Buffington, 1998). Like groundwater, the key controls of underflow are relatively well understood (Darcy's Law), but subsurface heterogeneity and limitations on direct observations make fluxes difficult to measure. We invoke underflow in various ways rather than providing a magnitude for the flux through the subsurface. For example, in conservative solute tracer experiments, lost mass of the tracer is expected and associated with window of detection limitations or flow path geometries (Payn et al., 2009; Ward et al., 2023) implying there are flow paths and timescales longer than we can observe in reach-scale studies. Channel water balances similarly, associate net losses in stream flow to subsurface flows, but this loss is usually lumped with other potential losses like evapotranspiration (ET) and groundwater recharge at the reach scale (Payn et al., 2009). Underflow has also been used to explain experimental results, but as a hypothetical situation (e.g., Ward et al., 2016). Underflow also shows up in a variety of ways in models.

Some modeling efforts to understand feature-scale hyporheic exchange show underflow but treat the underflow as a by-product of the boundary conditions rather than the focus of the study (e.g., Gooseff et al, 2006 and Schmadel et al., 2017 to name a few). Other models have ignored underflow completely. Models like the Transient Storage Model, Advective Storage Path Model, and Advanced Terrestrial Simulator with Advection-Dispersion Equation with

Lagrangian Subgrids all model hyporheic exchange but explicitly ignore underflow and the longer residence times associated with underflow (Bencala & Walters, 1983; Painter et al., 2018; Worman et al., 2002). Other models have incorporated underflow to predict stream intermittency (e.g., Ward et al., 2018a) while others have noted the persistence of underflow in the subsurface compared to feature-scale exchange (e.g., "intermediate-scale flow paths" in Herzog et al., 2019). The inclusion of underflow is inconsistent across studies of river corridor exchange and thus, our understanding of its relative importance is not well understood.

Fundamental to our modern understanding of hydrology is the water balance, where the various sources and sinks of water to a control volume are quantified, balanced by the change in internal storage in a study system. Fluxes like stream flow, precipitation, and evapotranspiration are probably the most well understood with various methods to estimate their respective fluxes (Safeeq et al., 2021). Hillslope inputs have either been lumped in with stream flow or is estimated based on catchment area (Payn et al., 2009; Ward et al., 2018a). Changes in storage in the riparian zone of headwater mountain streams are minimal over annual cycles (e.g., Voltz et al., 2013). Thus, to close water balances, groundwater inflows or outflows are used to close the net gains or losses, respectively, in a water balance (Safeeq et al., 2021). The magnitude of underflow, despite being used to explain experimental results and being involved in models, is seldom measured or estimated. This does not mean that underflow should be neglected in water balance estimates though. In cases of stream intermittency, underflow can represent as mush of 100% of the down-valley flow in a river corridor, dominating surface flow (Ward et al., 2018a; Ward et al., 2020). Thus, underflow likely is a significant contributor to the water balance, but just how much is not well known.

Our overarching objective in this study is to empirically document underflow in a headwater mountain stream. Specifically, we ask:

1. How large is underflow compared to other fluxes known to be important at the reach and segment scales?
2. How different are interpretations of a channel water balance at the reach and segment scales?
3. To what extent do existing conceptual and mathematical models accurately reflect the spatial patterns in channel water balance and solute transport associated with underflow?

In this study, we conducted a series of multi-tracer studies in two bedrock-constrained segments of a headwater mountain stream at the H.J. Andrews Experimental Forest (Western Cascades, Oregon, USA). The bedrock constraints cause upwelling of all underflow to the surface, enabling an accurate estimate of total down-valley fluxes at multiple locations in the catchment. We calculate water balances for individual reaches within each segment and for the study segments. Finally, these water balances are compared to solute tracer studies that span each study segment to provide an independent estimate of reach- and segment-scale roles of the underflow.

5.3. Methods

5.3.1. Site Characterization

This study was conducted in Watershed 01 (WS01) at the H.J. Andrews Experimental Forest (HJA), in the Central Cascades of Oregon, (44.2332° N, 122.1762° W) in the summer of 2021 during baseflow conditions. WS01 is a second-order headwater stream draining about 96 ha at the outlet. Elevation ranges from 439 to 1027 m a.m.s.l (*Experimental Watersheds and Gauging Stations*, n.d.). The catchment is predominantly underlain by upper Oligocene / lower Miocene basaltic flows resulting in a narrow V-shaped valley and steep hillslopes (Swanson & James, 1975). An annual average of 2.1 m of rainfall occurs annually over the whole HJA forest and the average annual temperature is 9 °C (Ward et al., 2020). Additional site characterization is provided in a host of related studies including Anderson et al., (2005), Segura et al., (2019), Swanson and Jones (2002), Wondzell (2006), Ward et al., (2016), and Ward et al., (2018a).

This study focused on two segments along the stream in WS01, with the first downstream segment (lower segment) starting 48.7 m upstream from the gauge and extending 352.3 m upstream of the gauge. The second segment (upper segment) started at 367.7 m and ended at 540.9 m upstream of the gauge (Figure 4.1). Both segments were bounded by visible bedrock outcrops in the stream channel that forced all underflow into the surface stream, and match exactly the bedrock-bounded segments included in a host of prior studies of the basin (e.g., Herzog et al., 2019; Schmadel et al., 2017). A large bedrock step and extensive colluvial deposit are present between the gauge and the downstream bedrock outcrop in our study (0 to 48 m

upstream of the gauge). Each study segment was subdivided into five reaches, selected based on visible physical characteristics of the stream, including avoiding deep pools and visually identified surface transient storage (Jackson et al., 2013) to allow for complete mixing during solute tracer studies, instrumentation of locations that would remain perennially wet despite fluctuations in discharge (Ward et al., 2018a), and bounding of major morphologic features including locations where the stream splits into high- and low-flow channels and where the colluvium deposits visibly change in size. Reach length ranged from 12.7 m to 91.5 m with a median reach length of 49.1 m (Table 4.1).

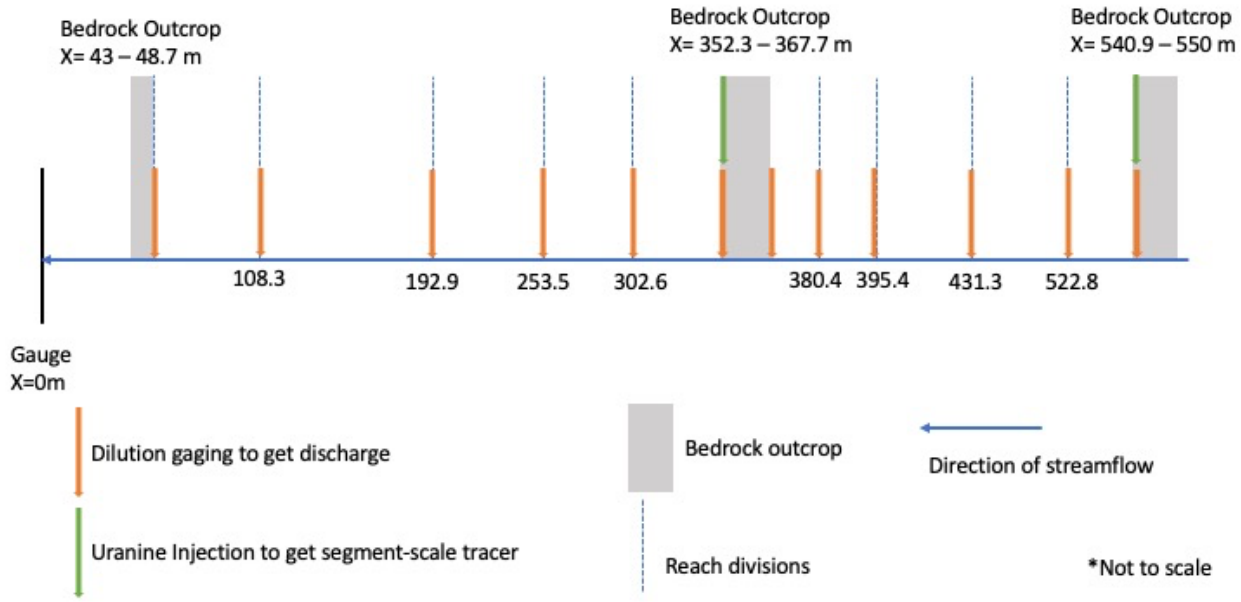

Figure 4.1: Diagram of experimental set up showing reach locations (blue dashed lines), bedrock outcrops (gray boxes), locations of dilution gaging (orange arrows), uranine injections (green arrows) along the stream of WS01.

Table 4.1: Reach lengths and locations in upper and lower segments bounded by bedrock in WS01. The downstream end of reach 5 is the upstream end of reach 4. The downstream end of reach 4 is the upstream end of reach 3 and so on.

Segment location	Segment Length (m)	Upstream end of reach (m from gauge)	Downstream end of reach (m from gauge)	Length of reach (m)	Reach #
Upper Segment	173.2	540.9	522.8	18.1	U1
		522.8	431.3	91.5	U2
		431.3	395.4	35.9	U3
		395.4	380.4	15	U4
		380.4	367.7	12.7	U5
Lower Segment	303.6	352.3	302.6	49.7	L1
		302.6	253.5	49.1	L2
		253.5	192.9	60.6	L3
		192.9	108.3	84.6	L4
		108.3	48.7	59.6	L5

5.3.2. Stream Solute Tracer Studies

5.3.2.1. Reach-scale dilution gauging

Dilution gauging was conducted at the up-and downstream end of each reach to calculate in-stream discharge (after Payn et al., 2009). Injections were conducted working from the downstream to the upstream end of each segment and repeated a total of three times during the study period (July 8 to August 27, 2021; Table 2 and Table 3). Electrical conductivity (EC) and stream temperature were recorded every 2 seconds using Onset HOBO U24 loggers. Electrical conductivity from solute tracers was converted to specific conductivity values to correct for

temperature, then to concentration (g/L) of salt (NaCl) using calibration curves developed by dissolving known masses of salt in stream water and after background-correcting the timeseries. Mixing lengths ranged from 3.8 m to 10.5 m and were determined on-site using visual inspection of discharge and morphology, following past studies in the basin (Ward et al., 2013; Ward, Zarnetske, et al., 2019). Overall, methods for choosing reach length, mixing, length, and calculating discharge into and out of each reach followed the methods of Payn et al., (2009), which have been successfully applied in multiple studies at this site (e.g., Ward et al., 2012; 2013; Ward, Zarnetske, et al., 2019). Tracer releases and mixing lengths were chosen to avoid locally unmixed regions directly within pools or downstream of identifiable inflows or along places where the channel splits. Mixing lengths were chosen to maximize turbulent self-contact within a representative volume of moving water, such as along an incised part of the channel (Payn et al., 2009). Discharge from the dilution gauging was calculated as

$$Q = \frac{M}{\int_0^t C(t)}$$

Where Q is discharge (m3/s), M is the injectate mass (g), C is the concentration (g/m3) and t is time (s).

Discharge in WS01 during baseflow recession develops a diurnal pattern (Bond et al., 2002; Voltz et al., 2013; Ward et al., 2018a). To account for changing time of day in dilution gauging, we estimated the time series of discharge at each location (QX(t); m3/s) as:

$$QX(t) = QG(t) * \frac{QX(obs)}{QG(obs)}$$

where QG(t) was the time series of discharge at the WS01 gauge, QX(obs) was the observed discharge at the dilution gauging location based on dilution gauging, and QG(obs) was the discharge at the gauge at the time when the tracer was initially released. We assumed that this approach includes synchrony of discharge time series along the study reach (e.g., maximum and minimum discharges occur at the same time of day) given past studies drawing this conclusion from the basin (Schmadel et al. 2017; Ward et al., 2018a), suggesting this. We also assumed that the ratio of discharge between the gauge and observation location does not change over the interpolation period, which is consistent with area proportional discharge estimates used in prior

modeling studies (Ward et al., 2020; Ward et al., 2018a) and consistent with empirical studies of discharge throughout the broader HJA basin (Ward, Zarnetske, et al., 2019). This approach was used to calculate the discharge time series at each dilution gauging location. Throughout the study, we used daily median discharge, calculated as the median value of QX(t) over a 24-hr period centered on the time of tracer release.

Table 4.2: Locations, dates, mixing length, and masses for reach-scale dilution gauging for the lower segment

Segment Location	Set #	Downstream end of reach location (m upstream from the gauge)	Date	Mixing Length (m)	Injection Mass (g)
Lower	1	48.7	7/8/21	4.3	63.518
		108.3	7/9/21	4.5	61.46
		192.9	7/12/21	6.2	57.152
		253.5	7/13/21	9.8	80.283
		302.6	7/14/21	5.2	91.88
		352.3	7/14/21	3.4	1004.15
Lower	2	48.7	7/19/21	4.3	57.544
		108.3	7/20/21	4.5	69.58
		192.9	7/21/21	6.2	54.367
		253.5	7/22/21	9.8	57.809
		302.6	7/23/21	5.2	60.266
		352.3	7/23/21	3.4	1064.531

Lower	3	48.7	7/25/21	4.3	50.343
		108.3	7/26/21	4.5	51.648
		192.9	7/27/21	6.2	69.873
		253.5	7/28/21	9.8	53.649
		302.6	7/29/21	5.2	53.566
		352.3	7/29/21	3.4	1034.34

Table 4.3: Locations, dates, and mixing length for reach-scale dilution gauging for the upper segment

Segment Location	**Set #**	**Downstream end of reach location (m upstream from the gauge)**	**Date**	**Mixing Length (m)**	**Injection Mass (g)**
Upper	1	367.7	8/8/21	4.1	65.384
		380.4	8/9/21	4.5	64.256
		395.4	8/10/21	3.8	53.505
		431.3	8/11/21	10.5	70.9
		522.8	8/12/21	8	70.761
		540.9	8/12/21	5.6	990.411
Upper	2	367.7	8/15/21	4.1	62.783
		380.4	8/16/21	4.5	55.745

		395.4	8/17/21	3.8	66.664
		431.3	8/18/21	10.5	78.045
		522.8	8/19/21	8	62.091
		540.9	8/19/21	5.6	891.349
Upper	3	367.7	8/23/21	4.1	74.295
		380.4	8/24/21	4.5	64.478
		395.4	8/25/21	3.8	60.347
		431.3	8/26/21	10.5	64.536
		522.8	8/27/21	8	63.307
		540.9	8/27/21	5.6	945.698

5.3.2.2. SEGMENT-SCALE URANINE TRACER RELEASE

A constant rate injection of uranine was conducted at the upstream end of each study segment. Durations for each release were approximately 6-hr. In-stream fluorescence sensors (Turner Cylops-7 Fluorescent Dye Sensor with PME C-FLUOR logger) were located at the boundaries of each study reach within the segments (two exceptions being attributable to a limited number of fluorometers available for the study). Sensors recorded fluorescence every 5 minutes and were converted to uranine concentrations using a calibration curve constructed by dissolving known masses of uranine in stream water. Stream discharge time series at each monitoring location were used to calculate the in-stream mass recovery of uranine at each location monitored. Injection rates of the uranine and water solution were recorded approximately every hour to ensure a constant rate of release (Table 4.5). The timing of uranine releases was such that they co-occurred with dilution gauging and salt tracer studies (Table 4.5).

The uranine data were background corrected and smoothed using a 2-hr moving window average to reduce noise in the sensor from light scattering.

As a solute tracer, uranine is known to undergo photolytic decay. To estimate the photolysis rate at our field site, a shallow dish of uranine dissolved in stream water was placed on the stream bank. Samples were collected from the dish and analyzed for concentration over a period of 11 days to measure the loss of uranine due to photolysis. We estimated a time-integrated photolysis rate for uranine in stream water by fitting a first-order decay model to the data, where t_{adv} is the advective time based on the transport of uranine peak concentrations along the segment, M0 is the mass of uranine at the upstream most logger, and M is the mass loss along each study reach.

$$M = M_0 * (1 - e^{-kt_{adv}})$$

Table 4.4: Details of the segment-scale dye tracer injections including injection location, time and date of injection, total injection time, and average injection rate.

Study segment	Location of injection (m)	Total injection time	Date of injection	Injection start time	Injection Rate
Lower	355.7	6hr 22 min	6/29/21	14:13	2.54 mL/s
Upper	546.5	6hr 22 min	8/3/21	13:45	2.70 mL/s

5.3.3. Reach- and Segment- Scale Water Balances

To construct a water balance for each reach and segment, we estimated the magnitudes of fluxes of water into and out of the river corridor, including in- and outflows in the streamflow and as underflow, exchanges with the atmosphere, and lateral inflows from hillslopes. To assess change in storage, we monitored water levels in the stream to estimate changes in the total depth of water stored in the riparian zone.

5.3.3.1. STREAM WATER INFLOWS AND OUTFLOWS

Surface inflows to and outflows from each study reach or segment were estimated from the dilution gauging detailed in the prior section, using the daily median discharge at each location as the basis for comparison. Stream discharges were calculated using dilution gauging, described in section 2.2. While uncertainty was not assessed for dilution gauging, past studies have estimated an 8.1% error (95% confidence interval; Schmadel et a. 2010) with detailed study at other sites At the HJ Andrews, an estimate of 10% error was invoked with no underlying basis by Ward et al. (2019).

5.3.3.2. UNDERFLOW INTO AND OUT OF EACH STUDY REACH

Underflow (Q_{sub}) was calculated using Darcy's Law:

$$Q_{sub} = -K * A * \frac{dh}{dl}$$

where K is hydraulic conductivity, dh/dl is the hydraulic gradient, and A is the cross-section area for flow. In our study, the hydraulic gradient was estimated based on the slope of the valley (after Ward et al., 2018a; Ward et al., 2020; Wondzell 2006). Cross-sectional area was estimated as the width of the valley multiplied by the expected depth of colluvium. Colluvium depth was estimated as d is depth of the colluvium (m), and w is the width of the valley (Ward et al., 2020):

$$d = 1 + 0.01 * w$$

Hydraulic conductivity in WS01 averaged 1.51e-04 m/s based on 7 falling head tests in WS01 (range 2.64e-05 m/s to 2.92e-04 m/s; Ward, Zarnetske, et al., 2019). Past studies have documented a range of hydraulic conductivities spanning 4.3×10^{-6} to 6.1×10^{-4} m/s based on falling head tests, using an average value of 7 x 10-5 m/s in modeling exercises (Kasahara &Wondzell, 2003), while Ward et al (2018) calibrated their model to a value of 5.62×10^{-6} m/s. Finally, underflow at the exposed bedrock bounding the study segments was assumed to be zero, consistent with modeling studies of the reach (Herzog et al., 2019; Schmadel et al., 2017; Ward et al., 2018b).

We used the slope, depth, width, and hydraulic conductivity values to estimate underflow in each study reach. We used the average hydraulic conductivity for our baseline estimate, and

include a range spanning the highest and lowest values reported by Ward, Zarnetske et al. (2019).

5.3.3.3. LATERAL INFLOW TO THE VALLEY BOTTOM FROM THE HILLSLOPES

Lateral inputs from the hillslope into the valley bottom were calculated as the inflow per unit drainage area calculated for the WS01 gauge multiplied by the difference in upslope accumulated area (UAA) at the upstream and downstream ends of the segments and reaches (Schmadel et al., 2017; Ward et al., 2020).

5.3.3.4. EVAPOTRANSPIRATION

Evapotranspiration (ET) data were obtained from OPENET (*OPENET*, 2023) using the area of WS01 that encompasses the study segments and the study period (Appendix Table 7.10). Riparian areas for OPENET were not extracted due to the limited resolution from the satellite imagery (i.e., it is unclear where the bounds of the riparian zone lay in the segment clipped). The ET fluxes in the river corridor were converted to a volumetric flow rate by multiplying them by the valley bottom area for each reach or segment. Estimates from OPENET were in the same range as prior publications documenting riparian ET in the watershed to be 1 mm/day to 3 mm/day (Bond et al., 2002; Moore et al., 2011).

5.3.3.5. PRECIPITATION

No precipitation was recorded at the nearby HJA Primary Meteorological Station (Daly et al., 2019) for the duration of the study period.

5.3.3.6. GROUNDWATER INFLOWS AND OUTFLOWS

No explicit estimates of losses to- nor gains from regional groundwater were made. Prior studies have consistently assumed the underlying intact bedrock to manifest as a no-flow boundary (Herzog et al., 2019; Schmadel et al., 2017; Ward et al., 2018a).

5.3.3.7. CHANGES IN STORAGE

Stage data from water level loggers were used to estimate the change in storage for the duration of the experiments. Stage data at the gage as well as along the two segments in this study were collected at 15-minute intervals for the duration of the experiments. Pressure transducers (Onset HOBO U20s) were placed in pools nearest the EC sensors to record stage along the study segments. Additionally, barometric pressure was recorded at the stream at the gauge, between the two segments, and at the top of the upper segment to correct pressure readings such that reflected actual changes in water level loggers. Because stage varies diurnally, a change in storage was estimated by fitting a linear regression to stage data through the study period. The total change in storage was estimated based on the total change in the linear regression during the study period.

5.3.3.8. ESTIMATION OF THE UNACCOUNTED-FOR INFLOWS ALONG STUDY SEGMENTS

Independently, we used the dilution gauging information to estimate the discharge that is unaccounted for in the stream at each observed location. We fit a linear regression between the up- and downstream discharges for each segment to represent the segment-scale change in surface water flow. The deviation of an observed streamflow from this linear relationship represents a volume of water known to have entered the study segment, but not present in the stream at a given location. The net volume of inflowing discharge that is unaccounted for represents a combination of underflow and the gross gains and losses of water to the river corridor (e.g., lateral inflows form hillslopes, outflows via ET).

5.4. RESULTS

5.4.1. SPATIAL VARIATION IN STREAM DISCHARGE

For all replicates in both reaches, stream discharge is highest on the bedrock at the upstream end of each study segment (Fig. 4.2). Stream discharge decreases mid-segment, where a larger proportion of flow travels down-valley as underflow, and then increases as downstream bedrock causes the flow to upwell and return to the surface (Fig. 4.2). In all but one case, we recorded a net loss of discharge between bedrock outcrops that bound the segments (-0.3843 to -3.1831 L/s in the downstream segment; -0.4250 to -1.4644 L/s in the upstream segment), and all

dilution gauging estimates were higher than the discharge recorded at the gauge when they were conducted.

The amount of discharge that is unaccounted for in the stream channel increases from zero at the upstream end of the study segment (where 100% of discharge is in the surface), then decreases to zero at the downstream end of each segment, where 100% of discharge returns to the surface stream (Figure 4.3). Within each segment, underflow ranges from 0.60% to 60.83% of the discharge entering the lower segment and 1.98% to 59.24% of the discharge entering the upper segment (Fig. 4.3). Underflow estimates using Darcy's Law within each study segment are in the same order of magnitude as unaccounted-for surface discharge using the mean hydraulic conductivity observed in the reach (see black diamonds in Fig. 4.3). In general, the discharge unaccounted for is larger than the magnitude estimated for underflow (i.e., the Darcy's Law estimates tend to be smaller than the magnitude of discharge that is not accounted for). However, we note that the magnitude of underflow is commonly within the range of possible underflows estimated using the full range of hydraulic conductivities observed in WS01 (whiskers in Fig. 4.3).

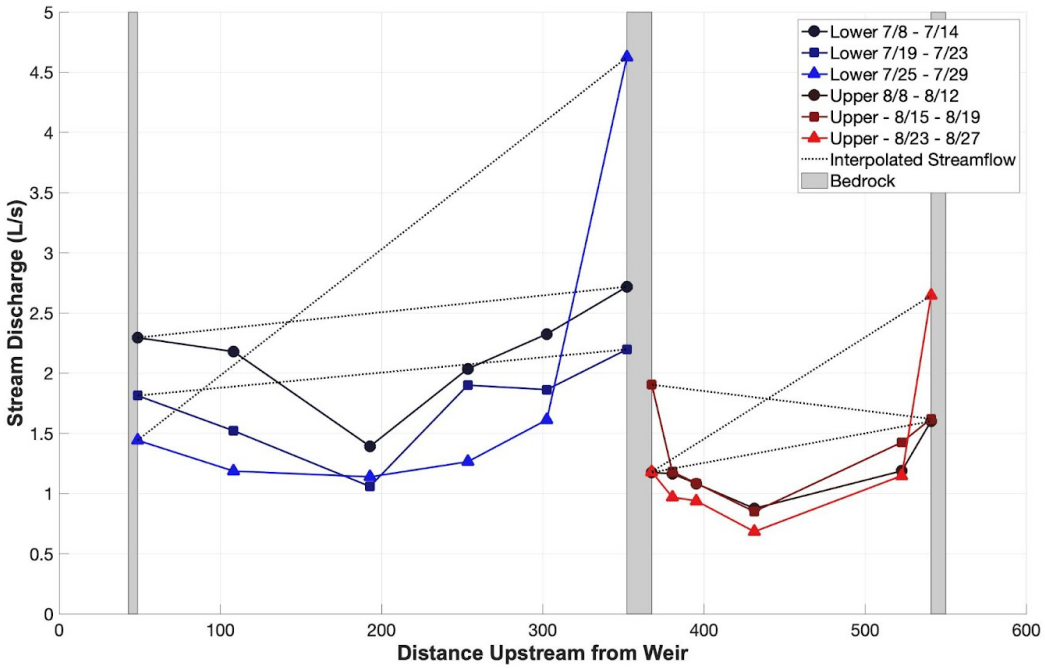

Figure 4.2: Measured discharge along both lower and upper segments. Bedrock sections are highlighted by the gray bars. Dashed lines are the interpolated values between the discharge measured over the outcropping bedrock.

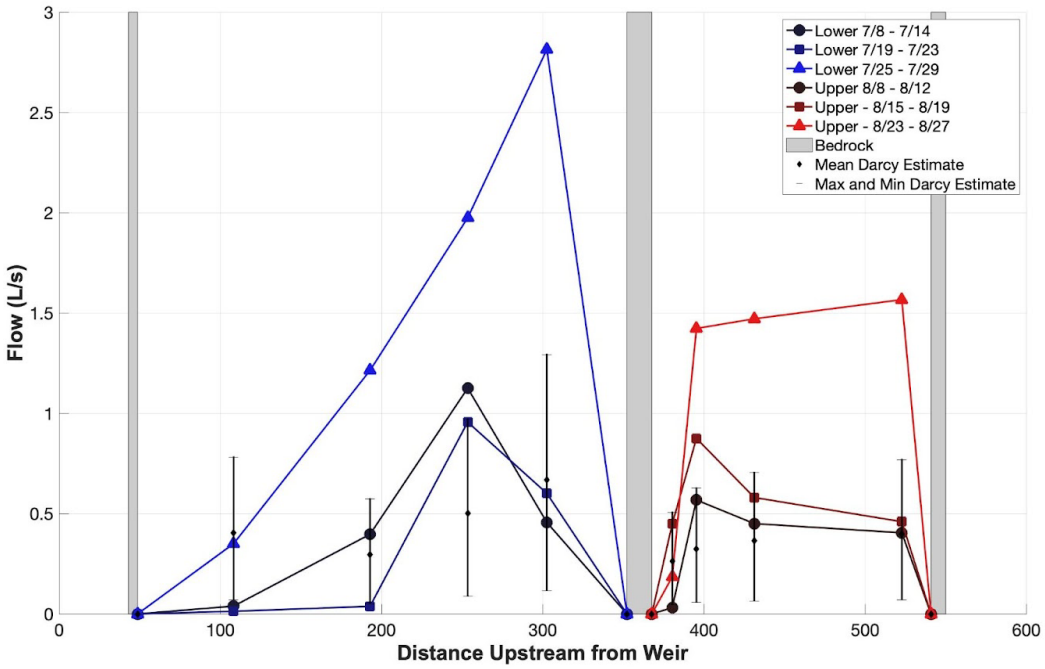

Figure 4.3: Difference from measured stream flows from interpolated values (colored lines). Subsurface flows estimated from Darcy's law are the vertical lines with minimum, mean, and maximum values of hydraulic conductivity.

Table 4.5: Change in surface discharge along each study reach and segment for the lower segment. All values reported in L/s. Negative values indicate net losses of stream flow.

Study dates	L1	L2	L3	L4	L5	Entire Segment
7/8 - 7/14	0.0169 (L/s)	-0.3988 (L/s)	-0.4421 (L/s)	0.9664 (L/s)	-0.5472 (L/s)	-0.4048 (L/s)
7/19 - 7/23	0.0751 (L/s)	-0.0693 (L/s)	-0.6372 (L/s)	0.6437 (L/s)	-0.3677 (L/s)	-0.3554 (L/s)
7/25 - 7/29	-2.5919 (L/s)	-0.4436 (L/s)	0.0930 (L/s)	0.2374 (L/s)	-0.3911 (L/s)	-3.0962 (L/s)

Table 4.6: Change in surface discharge along each study reach and segment for the upper segment. All values reported in L/s. Negative values indicate net losses of stream flow.

Study dates	U1	U2	U3	U4	U5	Entire Segment
8/8 - 8/12	-0.1562 (L/s)	-0.2417 (L/s)	0.2377 (L/s)	0.0987 (L/s)	-0.3991 (L/s)	-0.4606 (L/s)
8/15 - 8/19	0.0533 (L/s)	-0.5165 (L/s)	0.2541 (L/s)	0.1130 (L/s)	0.3177 (L/s)	0.2216 (L/s)
8/23 - 8/27	-1.2431 (L/s)	-0.3993 (L/s)	0.2844 (L/s)	0.0506 (L/s)	-0.1982 (L/s)	-1.5056 (L/s)

Table 4.7: Magnitude of unaccounted for discharge (L/s) at each location where discharge is measured along the study segment.

Lower	**48.7 m (bedrock)**	**108.3 m**	**192.9 m**	**253.5 m**	**302.6 m**	**352.3 m (bedrock)**
7/8 - 7/14	0	0.0392	0.3972	1.1257	0.4564	0
7/19 - 7/23	0	0.0133	0.0382	0.9565	0.6013	0
7/25 - 7/29	0	0.3508	1.2141	1.9750	2.8142	0
Upper	**367.7 m (bedrock)**	**380.4 m**	**395.4 m**	**431.3 m**	**522.8 m**	**540.9 m (bedrock)**
8/8 - 8/12	0	0.0318	0.5682	0.4505	0.4047	0
8/15 - 8/19	0	0.4493	0.8754	0.5806	0.4601	0
8/23 - 8/27	0	0.1852	1.4233	1.4717	1.5670	0

5.4.2. Reach- and Segment-Scale Water Balances for the River Corridor

5.4.2.1. Storage changes are minimal during the study period

The change in storage over the study period for each segment was zero (i.e., steady-state water balance). This is consistent with prior observations in the study system showing little change in storage during late summer baseflow (e.g., Voltz et al., 2013.). Our observations of the stream stage did not systematically decline during the study period (Appendix Table 7.9). For the linear regression model, a slope of decreasing water levels by 5.7×10^{-3} mm H2O/day was established, which is sufficiently small to account for less than 1 mm of stage change during the entirety of the experiment (7/2/21 to 8/31/21). This magnitude of change is smaller than the resolution of the water level loggers deployed in the field, so we take changes in storage to be zero during the study period.

5.4.2.2. Reach-scale water balances

Across all reach-scale water balances, surface stream flows are the predominant fluxes, accounting for 63.4 to 99.7% of inflows and 62.1 to 98.6% of outflows for reaches in lower segment (Figure 4.4). Reaches in the upper segment had similar results, with 66.6 to 99.5% of all inflows and 66.6 to 99.9% of all outflows being surface stream flows (Figure 4.5). The next largest fluxes at the reach scale are underflow into and out of each reach, ranging from 12.5 to 35.8% of inflows to and 12.6 to 35.0% of outflows from the lower segment, and 20.9 to 33.7% of inflows and 15.5 to 31.6% of outflows in the upper segment. These results do not include the 0% of inflows at the upstream bedrock nor outflows at the downstream bedrock.

Evapotranspiration from the riparian zone is estimated at 3.20 to 3.99 mm/d for the lower segment and 1.57 to 1.77 mm/d for the upper segment (*OPENET*, 2023). ET accounted for 1.01 to 2.80% of all outflows for reaches in the lower segment and 0.11 to 1.11% of all outflows from reaches in the upper segment. Lateral inflows from the hillslope represent 0.33 to 1.86% of inflows to reaches in the lower segment and 0.47 to 2.45% of inflows to reaches in the upper segment.

In all replicates, reaches L2 and U2 is losing water from upstream to downstream, and reaches L4 and U4 is gaining from upstream to downstream (Fig. 4.4, 4.5, right column). Gross losses dominate at the upstream two reaches for each segment, as downwelling stream water fills

the capacity of the subsurface to transport water as underflow. Gross gains dominate at the downstream reaches, showing upwelling from underflow. However, in every instance but one, the downstream most reach (L5 and U5) has net losses, indicating that the downstream bedrock outcrops do not necessarily force all underflow to return to the stream as expected (Fig 4.4, 4.5, right column).

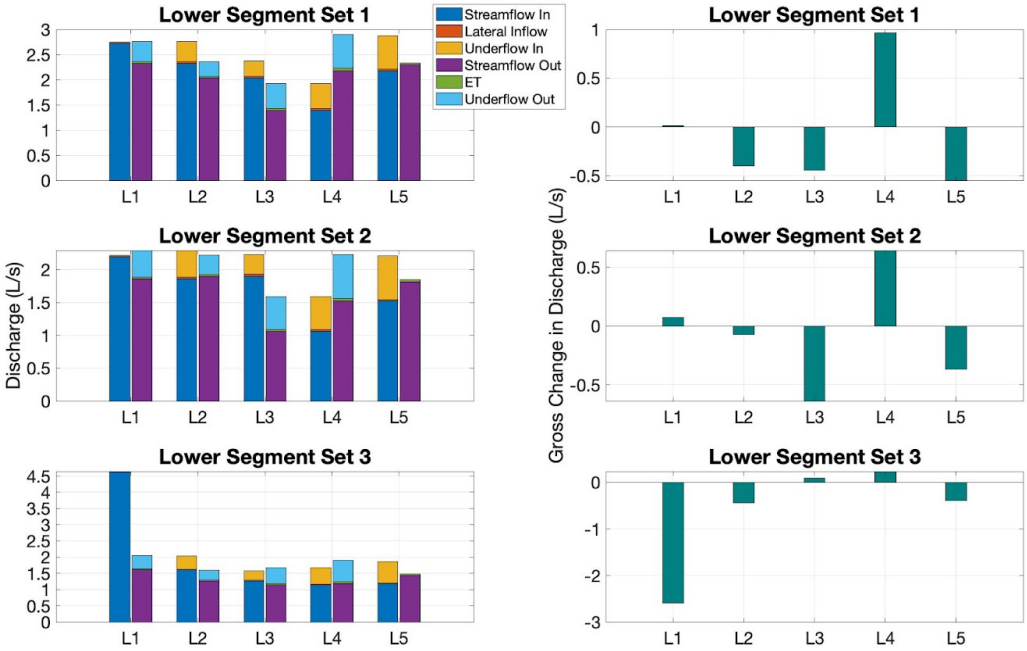

Figure 4.4: Water balance for each replicate study of the lower segments. The left column is the. Reach-scale water balance shows inflows (dark blue bars for stream flow, yellow for subsurface flows, and orange bars for hillslope inputs) and outflows (purple for stream flow, light blue for subsurface flows, and green bars for ET). The right column is gross gains and losses for each reach. Negative bars are losses from the stream and positive bars are gains to the stream.

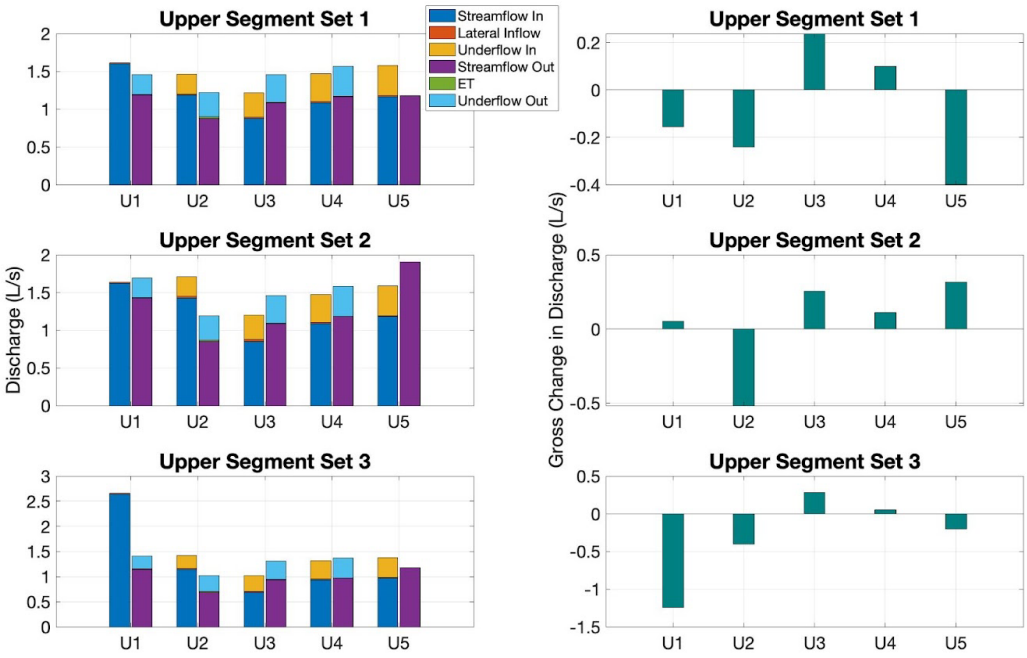

Figure 4.5: Water balance for each replicate study of the upper segments. The left column is the. Reach-scale water balance shows inflows (dark blue bars for stream flow, yellow for subsurface flows, and orange bars for hillslope inputs) and outflows (purple for stream flow, light blue for subsurface flows, and green bars for ET). The right column is gross gains and losses for each reach. Negative bars are losses from the stream and positive bars are gains to the stream.

5.4.2.3. SEGMENT-SCALE WATER BALANCES

At the segment scale, both segments are dominated by the net loss of water (2 of 3 replicates in the lower segment, 3 of 3 in the upstream segment, Figure 4.6). Because the segments are bounded by bedrock outcrops, underflow need not be estimated as an explicit inflow to nor outflow from the segment. Thus, surface stream inflows and outflows are the primary fluxes. Stream surface flows account for 94.7 to 98.0% of the total inflows for the lower segment and 93.9 to 97.3% of the total inflows for the upper segment. Stream surface flows account for 88.2 to 93.0% of total outflows for the lower segment and 96.9 to 97.2% of the total outflows for the upper segment. Lateral inflows for the segments represent 2.0 to 5.3% of the total inflows for the lower segment and 2.7 to 6.1% of the total inflows for the upper segment.

ET losses account for 7.0 to 12.0% of the total outflows for the lower segment, and 1.8 to 3.1% of the total outflows for the upper segment.

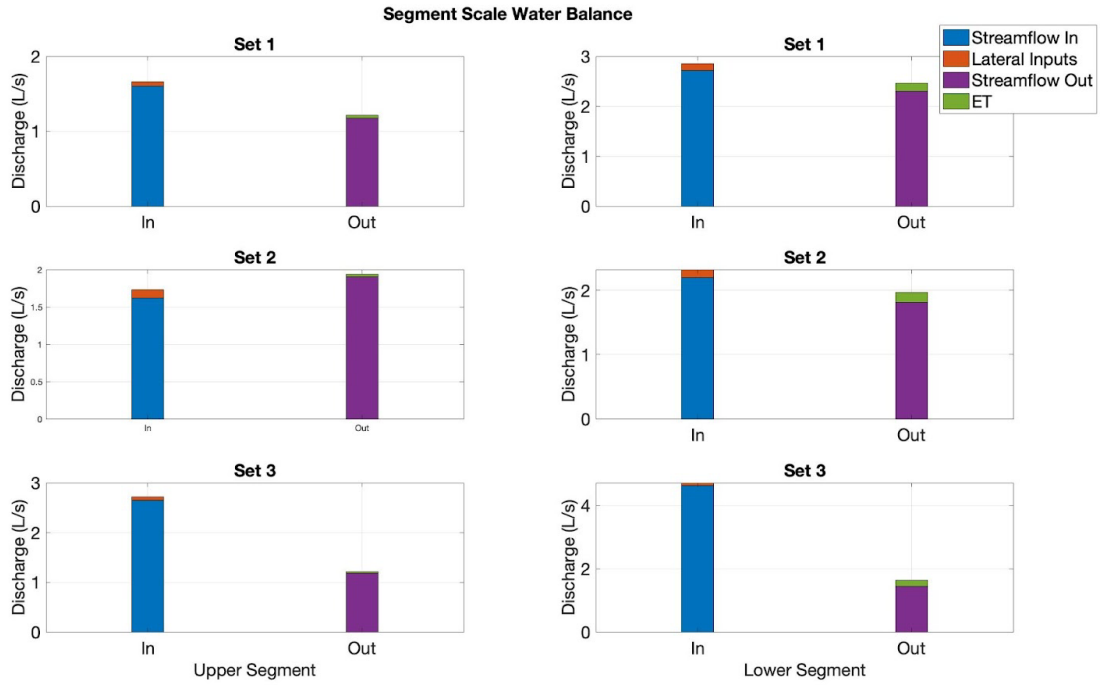

Figure 4.6: Segment-scale water balance for Upper Segment (left column) and lower segment (right column). Inflows are stream flow in (dark blue bar) and lateral hillslope inputs (orange). Outflows are stream flow leaving the segment (purple) and ET losses (green).

5.4.2.4. SEGMENT-SCALE URANINE MASS BALANCE

Consistent with segment-scale net losses in the channel water balance, uranine mass recovery also decreased along the study segments, even over the downstream bedrock portions where it was expected to return as underflow emerged in the stream (Figure 4.6). Less than 1% of the mass observed at the upstream-most reach of the lower segment was recovered at the downstream-most sensor. Similarly, only 8% of the mass observed at the upstream-most reach of the upper segment was recovered at the downstream-most sensor. The downstream sensors in each segment recorded uranine in the system for up to 18 days - until concentrations were too small to meaningfully measure - to limit the window of detection losses. Further, photolysis was

estimated based on the streamside experiment (k=0.159/day, consistent with Hixson et al., 2022). Losses due to photolysis, however, could not explain all mass loss (orange portion of bars in Figure 4.7). This suggests that there are outflows of water that began in the stream labeled with uranine that are not coming back at the segment scale over the outcropping bedrock as expected. This is an independent line of evidence showing the loss of water, consistent with the water balance results above.

Spatially, the uranine mass recovery shows a similar pattern to the channel water balances. In the upstream segment, there is evidence of return flow at the downstream end as mass recovery is greater at the downstream reaches (U4 and U5) than at the middle reach (U3). This indicates uranine mass downwelling at the head of the reach, traveling as underflow past U3, and returning to the stream channel (Figure 4.4). The elevated downstream recovery is not evident in the lower segment, but we acknowledge that photolysis may have been greater due to longer advective timescales and may have obscured any upwelling from underflow that occurred.

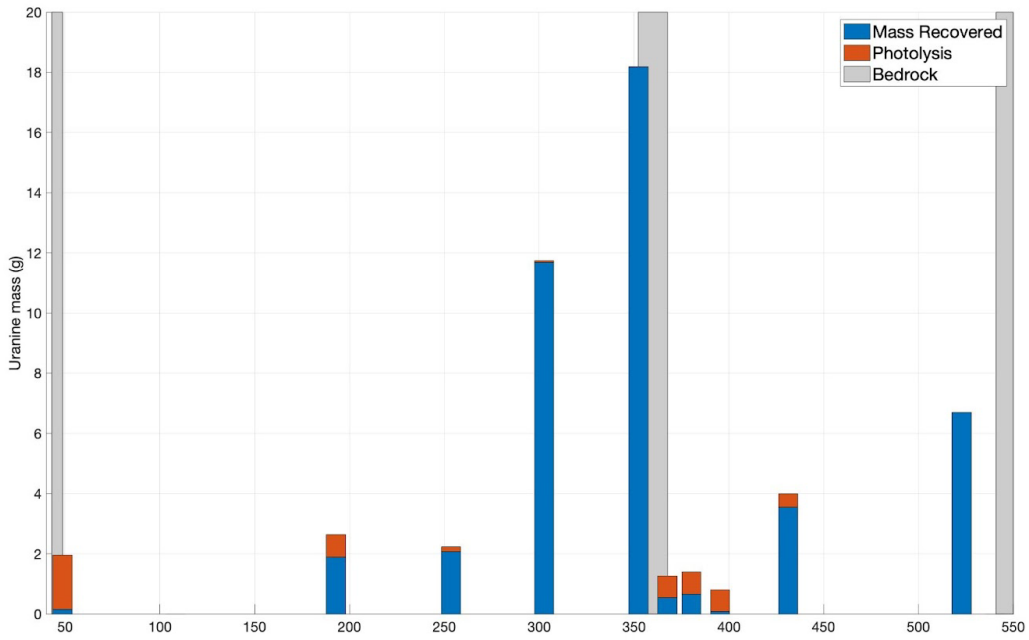

Figure 4.7: Uranine mass recovery for the upper and lower segments. Reach locations listed upstream to downstream going left to right. Blue bars are mass recovery of uranine measured by the in-stream sensors. Orange bars are mass lost to photolysis. Missing bars are due to a lack of sensors or incomplete recordings.

5.5. Discussion

5.5.1. Underflow Can Be Comparable in Magnitude to Streamflow in a Headwater Mountain Stream

While past studies have hypothesized the presence, magnitude, and impacts of underflow (e.g., Harvey et al., 2019; Herzog et al., 2019), we provide empirical, quantitative documentation of the magnitude and spatial patterns of underflow along a riparian corridor. Our estimates leveraged visible bedrock outcrops in the stream, allowing for quantification of total down-valley flow at multiple locations within a segment, which is a unique setting to enable these estimates. Ignoring fluxes from ET and lateral inputs, underflow can account for upwards of 66% of total flows using the method of interpolating streamflow across the segment (Fig. 4.3). Underflow accounted for as much as 35% of the total down-valley flow in our study when using Darcy's Law estimates (Fig. 4.4 and 4.5). Prior modeling studies, for example, have used Darcy's law to

estimate the magnitude of underflow (Ward et al., 2018a). We find those estimates are reasonable, with predicted magnitudes within the same order of magnitude as our empirical observations (e.g., dots and whiskers in Fig. 4.3). This range of potential subsurface flows is reasonable given the heterogeneity in hydraulic conductivity expected within the site (Ward, Zarnetske, et al., 2019), and suggests the Darcy's Law approximation is reasonable and a valid method for estimating underflow when interpolation of streamflow is not plausible (e.g., a stream segment not bounded by bedrock outcrops). We emphasize though that underflow should be explicitly accounted for in experiments and models given its contribution to the water balance.

5.5.2. Spatial Location and Scale of Study Change our Understanding of Transport and Water Balances in the River Corridor

The location of a study reach within a segment has the potential to alter our understanding of river corridor processes and underflow. For example, studying the upstream reaches (Reaches U1, L1, U2, and L2) of the segments for a study reach would result in net losses exceeding net gains (Figures 4.4 and 4.5), yielding an interpretation that the system is losing water and solutes to fates unknown. These studies along could not differentiate underflow (that would subsequently return to the stream) from transport along substantially longer flow paths. Moving further downstream, studying reaches in the middle of the segments (reach L3 and U3) can result in losses, gains, or practically no change at all in discharge or mass recovery. A balanced study with no losses or gains might be interpreted as complete mass recovery (unlikely), or at least offsetting gains from lateral inflows and losses to ET. However, our studies show that lateral inputs are nearly negligible, accounting for less than 3% of total inflows at the reach scale. Thus, interpretations would be incorrect. Moving to the downstream reaches within the segment, results would show predominantly gross gains to the stream. This could be interpreted as upwelling from an unknown source, but it would not be clear if this was hyporheic exchange, underflow, or regional groundwater inputs. In reality, gains represent a combination of intermediate and regional flow paths, but at the reach scale, we lack a way to differentiate them. Considering a single reach can result in differing interpretations of the processes occurring, depending on the location of that reach and how underflow is interacting with the stream and feature-scale hyporheic exchange.

Upstream losses 'fill' the down-valley capacity and upwell when a bedrock outcrop limits down-valley flow. While a water balance may be used to infer these flow paths, our uranine tracer further documents the coupling of upstream losses and downstream gains along study segments. Segment-scale studies can explain losses in a quantitative manner as either underflow or deeper groundwater losses, which normally would be unexplained by reach-scale studies. Streams do not just lose or just gain, and the combination of multiple reach studies into a segment highlights that streams are gaining and losing systematically (Covino & McGlynn, 2007; Payn et al., 2009). Additionally, the scale at which these processes are studied also can alter our understanding and interpretations (Becker et al., 2023). Studying a single reach means certain inputs and outputs cannot be separated, but by combining scales into a nested concept (multiple reaches in a segment), we can gain a more complete picture (Table 4.8).

Table 4.8: Conceptual table showing how process interpretations vary across space across a segment from observations of a single reach rather than in combination.

	Upstream reach	**Middle reach**	**Downstream reach**
Reach-scale tracer mass losses	High (loss to intermediate)	Moderate	Low (dominated by gains from intermediate)
Qunderflow	Low - downwelling is still occurring	Peak	Low - upwelling is limiting losses
Where do losses go?	"Longer than our study reach" but unclear if GW, ET, or underflow	"Longer than our study reach" but unclear if GW, ET, or underflow	Losses are limited, instead assumed gaining from hillslope is occurring

5.5.3. Possible Explanations of Water Losses

Contrary to long-held expectations that the study segments were net gaining, we found that the study segments were net losing in 5 of 6 replicates (Figure 4.6). Past studies have assumed net gains due to hillslope discharges to the valley bottom (Ward et al., 2018a) and accumulation of drainage area as a basis to estimate gains to the valley bottom (Schmadel et al.,

2017; Ward et al., 2018a). Even the visible presence of bedrock outcrops has been interpreted to infer that a continuous bedrock boundary must exist below the colluvium, which would prevent net losses to and gains from regional groundwater (Schmadel et al., 2017). Thus, the prevailing conceptual model is that of a gaining valley bottom, accumulating down-valley flow as the drainage area increases (e.g., Ward et al., 2018a; Ward et al., 2020). Still, other publications suggest that a large passive storage volume must exist within the basin to explain isotopic and solute transport signatures (e.g., Cain et al., 2019), but those also assume no net losses to regional groundwater. Contrary to these expectations, 100% of discharges measured on three distinct bedrock outcrops were larger than the downstream gauge, indicating that net loss continues from the study segment to the gauge. This is not only a baseflow phenomenon, as Ward et al. (2013) also found dilution gauging observations in the study segments to exceed gauge discharge during a storm response and recession spanning much higher discharges.

Additionally, while short spatial- and temporal-scale solute tracer studies may document patterns of gross gains and losses (e.g., salt tracers applied here and in Payn et al., 2009), we intentionally probed the intermediate flow paths with a segment-scale, constant rate injection of uranine. While we did document the presence of underflow, as evidenced by the increased mass recovery of uranine in downstream reaches for each segment (Figure 4.7), unexplained losses of uranine are too high to explain with photolysis alone. Thus, unexplained losses of uranine suggest the presence of another flux that is transporting water and uranine outside of the study segment. Examples of fluxes for losses might include net losses to a regional groundwater system or evapotranspiration of water by trees not in the valley bottom. Interaction with a large passive storage volumes (e.g., as described by Cain et al., 2019) with extended timescales could explain apparent losses of uranine without requiring a net change in fluxes along the segments. Indeed, other water balance studies suggest deeper storage and net groundwater export occur when water balances are not closed (Safeeq et al., 2021), and steep headwater catchments are more likely to lose water to aquifers than to gain water (Winter et al., 1999).

Finally, the potential for losses of water to regional groundwater systems is supported by the unit area discharge from WS01 (Figure 4.8). While this study was not a detailed basin-scale water balance, we did review discharge per unit watershed area at the gauge in WS01 compared to two paired watersheds (WS02, WS03) and a nearby headwater gauge (WS09). Given the comparable underlying geology, climate, weather, and land cover between the units, unit area

discharge may be expected to be comparable. However, we found significantly less discharge ($p<0.001$ for all pairwise Mann-Whitney-Wilcoxson U tests) from WS01. This comparison indicates that there is some additional process removing water from WS01 in comparison to the other basins. While we did not measure all components of a water balance across basins, we note that ET has not been sufficiently large to explain discharge differences in prior comparisons (e.g., Bond et al., 2002). Thus, we take this comparison as at least not disproving the potential that losses of water to regional groundwater are occurring in WS01.

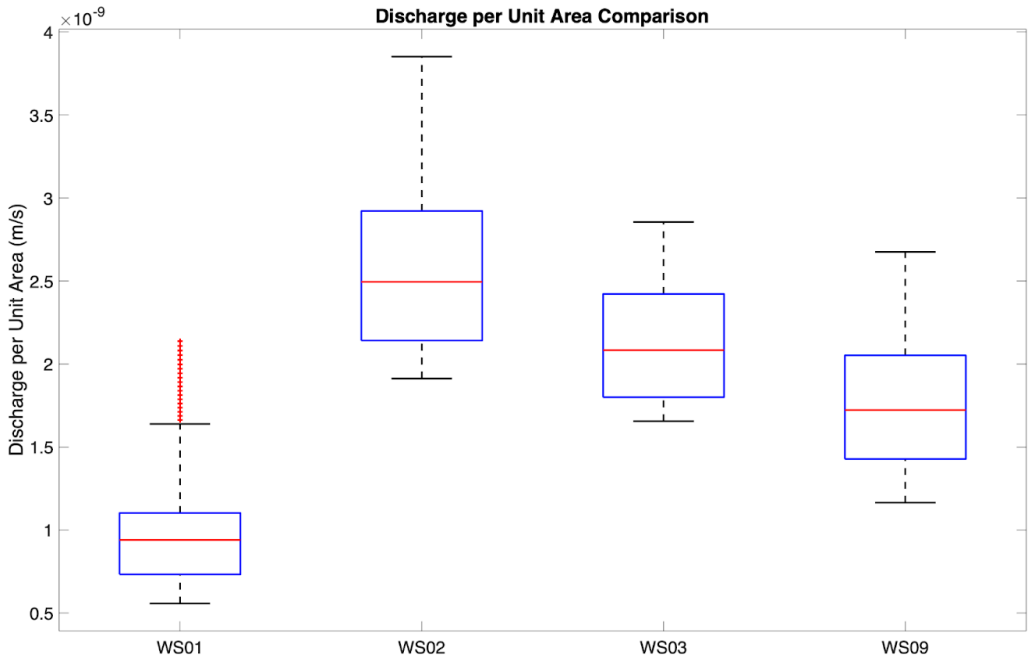

Figure 4.8: Comparison of discharge. Per unit area for WS01, WS02, WS03, and WS09 at the HJ Andrews during the time period this study was conducted. Data was gathered from the HJA Data portal (*Experimental Watersheds and Gauging Stations*, n.d.).

5.5.4. Updating Our Perceptual Model of WS01

Based on the results of this study, it is necessary to update the perceptual model of WS01 to more accurately include underflow and net losses from the river corridor. Here, we seek to integrate numerous prior studies – each of which has been conducted in WS01 with different goals and tools – to yield a coherent, multi-scale perceptual model. At the catchment scale, WS01 is broadly viewed as having a gaining stream (Harman et al., 2016; Schmadel et al., 2017; Ward, Payn, et al., 2013; Ward et al., 2018a). Of course, this must be true for at least the upper

portions of the basin, as streamflow is ultimately generated and a surface stream emerges. However, our observations consistently show unaccounted-for discharge in the river corridor in the lower segments of the basin. Thus, we synthesize the known stores and fluxes in the river corridor of WS01, detailing what is known and assumed. Our updated, multi-scale perceptual model for the catchment is presented in Tables 4.10-4.12, with each store and flux detailed in the subsequent paragraphs.

5.5.4.1. Relevance of evapotranspiration changes across scales

Past studies WS01 have assumed evapotranspiration in the valley bottom to be negligible and thus, ignored them in their perceptual models (e.g. Schmadel et al., 2017; Ward et al., 2018a). At the individual reach scale, ET accounted for less than 3% of total outflows in this study (section 3.2.2), but at the segment scale, ET starts to matter more, accounting for 7 to 12% of outflows in the lower segment. Still, during low-flow periods, ET is mostly negligible in its contribution to channel water balances (Bond et al., 2002). Additionally, stores of water used by ET on the hillslopes are likely disconnected from the storage that generates streamflow and recharges to groundwater (Brooks et al., 2010; Cain et al., 2019), but it is not clear how far down-slope this behavior holds. For example, trees on the lowest elevations of hillslopes may tap riparian water during the summer given their proximity. Additionally, depending on the season, ET losses can be from deep groundwater, the soil matrix in the hillslope, or mobile water from precipitation (Brooks et al., 2010). The impact of ET on catchment-scale water balances is highly scale-dependent (Thompson et al., 2011). For example, Safeeq et al., (2021) found that at a headwater catchment scale, ET makes up on average 43% of the water balance, and at the river basin scale, ET makes up an average of 18%. ET is a significant portion of the water balance that the reach and segment scales do not reflect in WS01. Additionally, in WS01, ET has been found to account for nearly 80% of precipitation during wet seasons (Bond et al., 2002). At the reach and segment scale, ET can be ignored but should not be ignored at the catchment scale.

5.5.4.2. Hillslope inputs are important at the catchment scale

Hillslope-derived fluxes enter the valley bottom as lateral inflows. These fluxes are commonly conceptualized as proportional to tributary hillslope area (e.g. Schmadel et al., 2017; Ward et al., 2020). Structural features in the valley bottom can affect how much water enters the

stream, which can change across scales (Jensco et al., 2009). However, we are not able to separate hillslope contributions to streams from other hydrologic exchanges, and thus contributions are usually assumed rather than measured/ documented (Bergstrom et al., 2016). Further, multiple studies suggest a strong relationship between catchment area and discharge (Burgers et al., 2014; de Vries et al., 1994 to name a few). Attempts to quantify hillslope contributions in WS01 revealed that soil properties, not topography were a stronger predictor of shallow soil water content in steep catchments (Jarecke et al., 2021). We also do not have a good way to link reach inputs to the watershed scale structure (Bergstrom et al., 2016).

Indeed, these lateral inflows are unknown and are often separated from models at smaller scales. Two-dimensional models have been used for nearly 20 years to study exchange in the hyporheic zone in WS01, yet all of them ignore lateral inputs (Anderson et al., 2006; Becker et al., 2023; Gooseff et al., 2005; Herzog et al., 2019; Schmadel et al., 2017). This study estimated lateral inputs to be minimal based on the water balance (Figures 4.4 – 4.6). However, a study done by Ward et al., (2018a) showed that watershed topology, including contributing areas, is a dominant control of water into the river corridor. As it stands, lateral inputs from the hillslope are likely important at the catchment scale but are less important at the reach scale.

5.5.4.3. STORAGE CHANGES IN SPACE AND TIME

At the reach and segment scales of this study, change in storage was treated as 0, given the minimal change of stage levels for the duration of the study. This is consistent with the relatively stable water levels observed in the riparian zone by Voltz et al., (2013). However, given the dramatic decrease in discharge throughout the water year in WS01 (2-3 orders of magnitude change during baseflow recession), increasing the temporal scale means a change in storage capacity cannot be treated as zero. The use of change in the storage being 0 has been widely applied across studies for water balance calculations that fail to identify sources of water that fluctuate substantially within the basin (Safeeq et. al 2021). However, this assumption results in an imbalance, even over decades (Han et al., 2020). Indeed, simulations in WS01 and beyond have shown that since the period of record, there has been a decline in flow connectivity and permanence (Ward et al., 2020). Methods to estimate the change in storage have involved measuring soil moisture at varying depths (e.g., Jarecke et al., 2021;, Safeeq et al, 2021), but the depths of the soil moisture probes are usually limited to the top 90 cm and then applied for the

volume of the catchment (Safeeq et al., 2021). Deeper storage is thus, not well understood. Assuming the change in storage to be 0 over time, or closing the water balance to estimate the change in storage can increase uncertainty in the contributions of the water model. Thus, changing storage should be considered in the perceptual model of WS01.

5.5.4.4. FLUXES INTO AND OUT OF THE RIVER CORRIDOR VARY ACROSS SPACE

The role of the riparian zone and river corridor itself varies depending upon the scale being considered. At the basin scale, the river corridor with the valley bottom broadly acts to collect and export water from the hillslopes (Bergstrom et al., 2016; Ward et al., 2018). Additionally, the river corridor lumps gains and losses and cannot be deciphered at the outlet of catchments (Bergstrom et al., 2016; Payn et al., 2009). At the segment scale, the river corridor appears to be net losing, during our study, lateral inflows were relatively small, and ET fluxes were insufficient to explain the losses. Therefore, the river corridor is net losing and recharging regional groundwater that is not explicitly known to exist, but does provide passive storage volumes required by lumped models (e.g., Cain et al. 2019), and explains the water balance not closing. At the reach scale, the river corridor may be either net losing or net gaining depending on the location of the study reach within the larger segment. The riparian zone and river corridor provide a different understanding of processes depending on the scale of study but are the means by which we measure fluxes and timescales.

5.5.4.5. KEEPING THE WATER BALANCE OPEN

Here, we link together the components of the water balance and the roles they play across scales to create a perceptual model for WS01 (Tables 4.9 - 4.11). We make the case for an open water balance as well given subsurface flows and deep groundwater losses cannot be directly measured, and fluxes such as ET, storage, and precipitation have uncertainty with them (Kampf et al., 2020). Additionally, the cause of the water losses in this study are unknown. Considering the fluxes listed below based on the scale of the study, and nesting the scales (multiple reaches within a segment, multiple segments within the catchment), we can have a more holistic view of the water balance in WS01 and better focus on areas less well understood (Kampf et al., 2020).

This perceptual model for WS01 (Tables 4.9 - 4.11) is presented across three scales: reach (Table 4.9), segment (Table 4.10), and catchment (Table 4.11), highlighting fluxes and stores (left column), what is needed to know about those fluxes and stores (middle column), and references for those (right column). As the scale changes, the fluxes that matter also change (e.g. lateral inputs are negligible at reach and segment scales but are not negligible at the catchment scale).

Table 4.9: Reach-scale perceptual model

Fluxes and Stores	**What we need to know**	**Reference**
Q_{stream} (Flux)	Intermittent during baseflow season – excess flow from the subsurface Diurnal fluctuations during baseflow	Ward et al., 2018
Q_{sub} (Flux)	Can be estimated with Darcy's law but changes spatially Use point measurements to parameterize across space Indirectly measured as loss from tracer studies	Kasahara and Wondzel 2003 Payn et al., 2009 This study; Schmadel et al., 2017, Kasahara and Wondzell 2003
Q_{lat} (Flux)	Negligible at reach scale Indirectly measured as gains from tracer studies	This study Payn et al., 2009
ET (Flux)	Minimal influence at the reach scale – does not explain losses ET decouples from influence on streamflow during baseflow season	This study Bond et all, 2002
GW (Store)	In the past has been ignored Could explain missing flows in water balance	Ward et al., 2018; Schmadel et al., 2017 This study

Table 4.10: Segment-scale perceptual model

Fluxes and Stores	What we need to know	Reference
Q_{stream} (Flux)	Intermittent during baseflow season – excess flow from the subsurface Diurnal fluctuations during baseflow	Ward et al., 2018
Q_{sub} (Flux)	Assumed 0 at outcropping bedrock Darcy's Law estimate is within range of flows but does not explain losses over bedrock	Payn et al. 2009; Ward et al., 2018; Schmadel et al., 2017 This study
Q_{lat} (Flux)	Has been ignored in models Storage for water during baseflow conditions	Schmadel et al., 2017 Segura et al., 2019
ET (Flux)	Minimal influence at the segment scale – does not explain losses ET decouples from influence on streamflow during baseflow season	This study Bond et al., 2002
GW (Store)	Previously ignored in models Modeled as gains to stream Could explain missing flows in water balance	Ward et al., 2018 Schmadel et al., 2017 This study

Table 4.11 Catchment-scale perceptual model

Fluxes and Stores	What we need to know	Reference
Q_{stream}	Measured at the gage and is an integrator of processes such as Q_{sub} Spatially and temporally intermittent	Ward et al., 2018 Ward et al., 2020
Q_{lat}	Related to Upslope Accumulated Area and streamflow at the gauge Storage for water during baseflow conditions	Ward et al., 2018 Segura et al., 2019
ET (Flux)	Riparian ET is disconnected from hillslope ET. Could explain missing flows in water balance Calculated at P/Q ratio	Cain et al., 2019 Safeeq et al., 2021
GW (Store)	Previously ignored or assumed to be unchanging Could explain losses that other fluxes do not Water to and from GW varies based on underlying geology	Ward et al., 2018 This Study Segura et al., 2019
P (Flux)	Point measurements at nearby stations and assumed to be homogenous across space	HJA data

5.6. CONCLUSIONS

This study set out to quantify underflow using the unique system in WS01 of the HJ Andrews Experimental Forest. First, we selected sites between outcropping bedrock sections, creating natural boundary conditions. The outcropping bedrock eliminated the possibility of underflow at the inlet and outlet of the study segments. Thus, underflow could be calculated as the difference between discharge at the inlet and discharge along the segment where underflow is possible. Underflow accounted for as much as ⅔ of total stream expected streamflow through the segments. Second, we calculated the water balance at the reach and segment scales for all sets in the lower and upper segments to understand spatial variation and how interpretations of the

fluxes change based on the location studied. We found that at the reach and segment scale, lateral hillslope inputs and fluxes from ET are minimal compared to streamflow and underflow, with underflow accounting for up to 35% of total flows using Darcy's Law. We also found that interpretations of processes change spatially through reaches. At upstream reaches of the segments, we consistently saw losses in the upstream reaches which could be interpreted as losses to groundwater or high ET. Middle reaches had minimal net gains or losses which could be interpreted as having minimal subsurface flows. Downstream reaches consistently saw gains in total flows which could be interpreted as the catchment having major lateral inputs. Combining multiple reach-scale studies at the segment scale provided a more comprehensive view of the fluxes occurring.

Surprisingly though, when performing a water balance, there were overall losses in discharge. A segment-scale tracer experiment with uranine dye also confirmed these losses. Furthermore, the losses continued beyond the individual segments to the stream gage recordings downstream. Based on different methods of recording losses, and the fact that the study catchment had lower discharge per unit area when compared to other catchments nearby, we believe that deep groundwater losses might be occurring, or that ET losses occur beyond the riparian zone. However, numerous studies at this site have all suggested that the bedrock is shallow (about three meters deep) and that it is intact based on the geology and onsite surveys (Schmadel et al., 2017; Swanson & James, 1975). Numerous models of WS01 have been built on the presumption that no deep groundwater losses occur, confounding the results of this study. Tying these results together, we believe that our perceptual model of WS01 needs to be updated to include known quantities of underflow and an open balance to account for these losses. While these findings are heavily site dependent, it is likely that the perceptual model regarding underflows, losses to GW, or the bottom of the watershed) and riparian zone influence are highly transferable.

5.7. ACKNOWLEDGMENTS

Thanks are to Steve Wondzell, Adam Ward, and Skuyler Herzog for your support both with the physical demands of conducting two 6-hour injections in a post-burnt watershed and for your thoughtfulness in making sure these experiments were successful given a global pandemic, a fire, and extreme heat.

Thanks also to the staff at the HJ Andrews for their support, lending helping hands, and use of facilities. Particular thanks to Mark Schulze and Greg Downing.

Data and facilities were provided by the H.J. Andrews Experimental Forest and Long Term Ecological Research (LTER) program, administered cooperatively by Oregon State University, the USDA Forest Service Pacific Northwest Research Station, and the Willamette National Forest. This material is based upon work supported by the National Science Foundation under the grant LTER8 DEB-2025755.

6. SYNTHESIS: UPSCALING HYPORHEIC STUDIES FROM REACH TO SEGMENT SCALE AND THE INCLUSION OF INTERMEDIATE-LENGTH FLOW PATHS

The field and modeling studies presented here demonstrate how interpretations of hyporheic exchange are often biased towards the reach scale, and that ignoring the segment scale and intermediate-length flow paths can greatly change our prediction of exchange when we upscale. Collectively, this work challenges earlier practices and assumptions for predicting hyporheic exchange and highlights how the prediction of exchange is skewed toward reach or feature scales.

As a hydrologic community, we want to collectively advance our understanding and predictive power of hydrologic processes. However, it is essential that we always make sure we are representing the processes we care about for the question we are trying to answer. As shown in Chapter 2, assumptions about representativeness and transferability can be propagated into future studies and models. Both the strategy used to select a study reach and the location of that reach determine if it is representative of the larger segment for hyporheic exchange. Typically, the biases towards the reach scale and shorter flow paths are unknown but have an influence on subsequent conceptual models. Thus, these unaccounted-for biases become embodied in our understanding of hyporheic zone processes. Accounting for within-site variability, selecting study reaches to match the processes we care about, and increasing the number of observations within and between sites will help improve the process understanding of hyporheic exchange and ensure that future studies are designed on valid assumptions.

Numerical models are valuable for expanding field studies across scales and conditions. However, these models often require a generalization of processes and arbitrary discretization that do not align with the current understanding of hyporheic exchange. It is known that underlying geology and geomorphology drive exchange from feature scale to catchment scale (Gooseff et al., 2006; Leopold et al., 1964; Payn et al., 2009). Models often move from reach to watershed scales ignoring the segment scale and underlying intermediate-length flow paths. The consequences of ignoring the segment-scale and intermediate-length flow paths are explored in Chapter 3. This study revealed that ignoring longer timescales generated by intermediate-length flow paths does impact our understanding of physical transport through the hyporheic zone. The difference in predicting mass transport between the baseline model and using truncated versions

of the transit time distribution was days for a 300 m segment. This difference would likely be compounded at a watershed scale. Using turnover points instead of arbitrary discretization did improve the prediction of transport compared to the baseline model, but only slightly, suggesting that how the reach is divided is less important than the truncation of transit times in the hyporheic zone. Finally, when coupled with a biogeochemical reaction, truncating times and discretization choices had little impact on predicting denitrification compared to the baseline model. Future work will be exploring reaction rates as they align with the timescales to understand what conditions for biogeochemical reactions matter most.

Intermediate-length flows in the subsurface might be important for physical transport as shown in the previous study. Despite the importance of these flows, our ability to directly measure and quantify these segment-scale flows is limited. Solute tracers assume mass loss is due to underflow (Payn et al., 2009), and models tend to artificially establish these intermediate-length flows in the subsurface, through boundary conditions (e.g., Herzog et al., 2019; Schmadel et al., 2017) but neither of these methods has the intention of quantifying the flows. Chapter 4 sought to quantify subsurface flows at both the reach scale and segment scale utilizing a unique setup existing in WS01 of the HJ Andrews Experimental Forest. Subsurface flows at both the reach and segment scales were successfully quantified using water balances of reach and stream segments between bedrock outcrops. Results revealed that subsurface flows account for up to one-third of all water fluxes in the channel water balance and that the segment scale provides a more complete picture of the water balance that a single reach study would not. Additionally, this study revealed that there were losses in stream discharge not explained by mass loss or subsurface flows, suggesting there was a loss to deeper groundwater not previously considered. Results from this study emphasize the importance of incorporating segment-scale characteristics, like intermediate-length flow paths, to provide a more complete understanding of hyporheic exchange processes.

This body of work highlights the importance of incorporating the segment scale in hyporheic studies when attempting to upscale to the catchment. Processes not normally seen at the reach scale like intermediate-length flow paths allow for a better representation of processes when we make predictions. It is important to note that these studies focused on steep cobble-bedded mountain streams draining headwater catchments. These results may not be directly applicable to other places or conditions, but the general concepts are relevant. Indeed, better

representation and better discretization are likely to change inferences and understanding in sites outside of the HJ Andrews. Also, having a better understanding of water balances across scales can improve process understanding, regardless of location. These findings also have relevance beyond the hyporheic zone and can expand into the whole river corridor. Overall, this work contributes to a growing body of literature on finding ways to effectively scale processes in the river corridor and advance our predictive power across spatial and temporal scales.

7. 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

Alexander, R. B., Smith, R. A., & Schwarz, G. E. (2000). Effect of stream channel size on the delivery of nitrogen to the Gulf of Mexico. *Nature*, *403*(6771), 758–761. https://doi.org/10.1038/35001562

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

Arrigoni, A. S., Poole, G. C., Mertes, L. A. K., O'Daniel, S. J., Woessner, W. W., & Thomas, S. A. (2008). Buffered, lagged, or cooled? Disentangling hyporheic influences on temperature cycles in stream channels. *Water Resources Research*, *44*(9), 1–13. https://doi.org/10.1029/2007WR006480

Becker, P. S., Ward, A. S., Herzog, S. P., & Wondzell, S. M. (2022). Data Files for Becker, P. S., A. S. Ward, S. P. Herzog, S. Wondzell (2022). Data Files for "Testing Hidden Assumptions of Representativeness in Reach-Scale Studies of Hyporheic Exchange." HydroShare. https://doi.org/https://doi.org/10.4211/hs.826fbc5bb04e4674b2df002d979b5390

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*(1), 1–16. https://doi.org/10.1029/2022WR032718

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

Bencala, K. E., Gooseff, M. N., & Kimball, B. A. (2011). Rethinking hyporheic flow and transient storage to advance understanding of stream-catchment connections. *Water Resources Research*, *47*(3), 1–9. https://doi.org/10.1029/2010WR010066

Bencala, K. E., & Walters, R. A. (1983). Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model. *Water Resources Research*, *19*(3), 718–724. https://doi.org/10.1029/WR019i003p00718

Bergstrom, A., Jensco, K., & McGlynn, B. (2016). Spatiotemporal processes that contribute to hydrologic exchange between hillslopes, valley bottoms, and streams. *Water Resources Research*, *52*(6), 4628–4645. https://doi.org/10.1111/j.1752-1688.1969.tb04897.x

Boano F., Harvey J.W., Marion A., Packman A.I., Revelli R.m Ridolfi L., W. A. (2014). Hyporheic flow and transport procesess: Mechanisms, models, and biogeochemical implications. *Reviews of Geophysics*, *52*(4), 603–679. https://doi.org/10.1029/88EO01108

Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., Wörman, A., & Boano F., Harvey J.W., Marion A., Packman A.I., Revelli R.m Ridolfi L., W. A. (2014). Hyophreic flow and transport processes: Machanisms, 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), 1–5. https://doi.org/10.1029/2008GL033554

Bond, B. J., Jones, J. A., Moore, G., Phillips, N., Post, D., Mcdonnell, J. J., & Wiley, J. (2002). The zone of vegetation influence on baseflow revealed by diel patterns of streamflow and vegetation water use in a headwater basin. 1677(October 2001), 1671–1677. https://doi.org/10.1002/hyp.5022

Boyer, E. W., Alexander, R. B., Parton, W. J., Li, C., Butterbach-Bahl, K., Donner, S. D., Skaggs, R. W., & Del Grosso, S. J. (2006). Modeling denitrification in terrestrial and

aquatic ecosystems at regional scales. *Ecological Applications*, *16*(6), 2123–2142. https://doi.org/10.1890/1051-0761(2006)016[2123:MDITAA]2.0.CO;2

Brooks, J. R., Barnard, H. R., Coulombe, R., & McDonnell, J. J. (2010). Ecohydrologic separation of water between trees and streams in a Mediterranean climate. *Nature Geoscience*, *3*(2), 100–104. https://doi.org/10.1038/ngeo722

Burgers, H. E. R., Schipper, A. M., & Hendriks, A. J. (2014). Size relationships of water discharge in rivers: Scaling of discharge with catchment area, main-stem length and precipitation. *Hydrological Processes*, *28*(23), 5769–5775. https://doi.org/10.1002/hyp.10087

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

Cain, M. R., Ward, A. S., & Hrachowitz, M. (2019). Ecohydrologic separation alters interpreted hydrologic stores and fluxes in a headwater mountain catchment. *Hydrological Processes*, *33*(20), 2658–2675. https://doi.org/10.1002/hyp.13518

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. (2009a). A model for lateral hyporheic flow based on valley slope and channel sinuosity. *Water Resources Research*. https://doi.org/10.1029/2008WR007442

Cardenas, M. B. (2009b). Stream-aquifer interactions and hyporheic exchange in gaining and losing sinuous streams. *Water Resources Research*, *45*(6), 1–13. https://doi.org/10.1029/2008WR007651

Castro, N. M., & Hornberger, G. M. (1991). Surface-Subsurface Water Interactions in an Alluviated Mountain Stream Channel. *Water Resources Research*, *27*(7), 1613–1621.

Condon, L. E., Markovich, K. H., Kelleher, C. A., McDonnell, J. J., Ferguson, G., & McIntosh, J. C. (2020). Where Is the Bottom of a Watershed? *Water Resources Research*, *56*(3). https://doi.org/10.1029/2019WR026010

Covino, T., McGlynn, B., & Mallard, J. (2011). Stream-groundwater exchange and hydrologic turnover at the network scale. *Water Resources Research*, *47*(12), 1–11. https://doi.org/10.1029/2011WR010942

Covino, T. P., & McGlynn, B. L. (2007). Stream gains and losses across a mountain-to-valley transition: Impacts on watershed hydrology and stream water chemistry. *Water Resources Research*, *43*(10), 1–14. https://doi.org/10.1029/2006WR005544

Daly, C., Schulze, M., & McKee, W. (2019). *Meteorological data from benchmark stations at the HJ Andrews Experimental Forest, 1957 to present.* Long-Term Ecological Research. Forest Science Data Bank, Corvallis, OR. https://doi.org/https://doi.org/10.6073/pasta/c021a2ebf1f91adf0ba3b5e53189c84f

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.

de Vries, H., Becker, T., & Eckhardt, B. (1994). *Power law distribution of discharge in ideal networks*. *30*(12), 3541–3543.

Drummond, J. D., Covino, T. P., Aubeneau, A. F., Leong, D., Patil, S., Schumer, R., & Packman, A. I. (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*(3), 1–11. https://doi.org/10.1029/2012JG002019

Elliott, H., & Brooks, N. H. (1997). Transfer of nonsorbing solutes to a streambed with bed forms : Theory permeable streambed and the overlying water of a stream or river is presented in this In a companion paper this issue] the results of experimental n + 00t Os u n + 00t Os q (x) = U •. 33(1), 123–136.

Experimental Watersheds and Gauging Stations. (n.d.). https://andrewsforest.oregonstate.edu/research/infrastructure/watersheds

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

Fischer, H., Kloep, F., Wilzcek, S., & Pusch, M. T. (2005). A river's liver - Microbial processes within the hyporheic zone of a large lowland river. *Biogeochemistry*, *76*(2), 349–371. https://doi.org/10.1007/s10533-005-6896-y

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.

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. (2014). A hydrogeomorphic river network model predicts where and why hyporheic exchange is important in large basins. *Geophysical Research Letters*, *41*(18). https://doi.org/10.1002/2014GL061099

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*, 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

Han, J., Yang, Y., Roderick, M. L., McVicar, T. R., Yang, D., Zhang, S., & Beck, H. E. (2020). Assessing the Steady-State Assumption in Water Balance Calculation Across Global Catchments. *Water Resources Research*, *56*(7), 1–16. https://doi.org/10.1029/2020WR027392

Harman, C. J., & Kim, M. (2014). An efficient tracer test for time-variable transit time distributions in periodic hydrodynamic systems. *Geophysical Research Letters*. https://doi.org/10.1002/2013GL058980

Harman, C. J., Ward, A. S., & Ball, A. (2016). How does reach-scale stream-hyporheic transport vary with discharge? Insights from rSAS analysis of sequential tracer injections in a headwater mountain stream. *Water Resources Research*. https://doi.org/10.1002/2016WR018832

Harvey, J., Gomez-Velez, J., Schmadel, N., Scott, D., Boyer, E., Alexander, R., Eng, K., Golden, H., Kettner, A., Konrad, C., Moore, R., Pizzuto, J., Schwarz, G., Soulsby, C., & Choi, J. (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. In *Water Resources Research* (Vol. 51, Issue 9). https://doi.org/10.1002/2015WR017617

Harvey, J. W., Wagner, B. J., & Bencala, K. E. (1996). Evaluating the Reliability of the Stream Tracer Approach to Characterize Stream-Subsurface Water Exchange. *Water Resources Research*, *32*(8), 2441–2451.

Herzog, S. P., Higgins, C. P., & McCray, J. E. (2016). Engineered Streambeds for Induced Hyporheic Flow: Enhanced Removal of Nutrients, Pathogens, and Metals from Urban Streams. *Journal of Environmental Engineering*, *142*(1), 1–10. https://doi.org/10.1061/(ASCE)EE.1943-7870.0001012.

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*, 1–17. https://doi.org/10.1029/2019WR025763

Hixson, J. L., Ward, A. S., McConville, M. B., & Remucal, C. K. (2022). Release Timing and Duration Control the Fate of Photolytic Compounds in Stream-Hyporheic Systems. *Water Resources Research*, *58*(11). https://doi.org/10.1029/2022WR032567

Jackson, T. R., Haggerty, R., and Apte, S. V. (2013). A fluid-mechanics based classification scheme for surface transient storage in riverine environments: quantitatively separating surface from hyporheic transient storage. *Hydrology and Earth System Sciences, 17,* 2747-2779. www.hydrol-earth-syst-sci.net/17/2747/2013/ doi:10.5194/hess-17-2747-2013

Jan, A., Coon, E. T., & Painter, S. L. (2021). Toward more mechanistic representations of biogeochemical processes in river networks: Implementation and demonstration of a multiscale model. *Environmental Modelling and Software*, *145*(August), 105166. https://doi.org/10.1016/j.envsoft.2021.105166

Jarecke, K. M., Bladon, K. D., & Wondzell, S. M. (2021). The Influence of Local and Nonlocal Factors on Soil Water Content in a Steep Forested Catchment. *Water Resources Research*, *57*(5), 1–21. https://doi.org/10.1029/2020WR028343

Jencso, K. G., McGlynn, B. L., Gooseff, M. N., Wondzell, S. M., Bencala, K. E., & Marshall, L. A. (2009). Hydrologic connectivity between landscapes and streams: Transferring reach- and plot-scale understanding to the catchment scale. *Water Resources Research*, *45*(4), 1–16. https://doi.org/10.1029/2008WR007225

Kampf, S. K., Burges, S. J., Hammond, J. C., Bhaskar, A., Covino, T. P., Eurich, A., Harrison, H., Lefsky, M., Martin, C., McGrath, D., Puntenney-Desmond, K., & Willi, K. (2020).

The Case for an Open Water Balance: Re-envisioning Network Design and Data Analysis for a Complex, Uncertain World. *Water Resources Research*, *56*(6), 1–19. https://doi.org/10.1029/2019WR026699

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

Kiel, B. A., & Bayani Cardenas, M. (2014). Lateral hyporheic exchange throughout the Mississippi River network. *Nature Geoscience*, *7*(6), 413–417. https://doi.org/10.1038/ngeo2157

Larkin, R. G., & Sharp, J. M. (1992). On the relationship between river-basin geomorphology, aquifer hydraulics, and ground-water flow direction in alluvial aquifers. *Geological Society of America Bulletin*, *104*(12), 1608–1620. https://doi.org/10.1130/0016-7606(1992)104<1608:OTRBRB>2.3.CO;2

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. *Hydrol. Earth Syst. Sci. Discuss*, *May*, 1–37.

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

Moore, G. W., Jones, J. A., & Bond, B. J. (2011). How soil moisture mediates the influence of transpiration on streamflow at hourly to interannual scales in a forested catchment. https://doi.org/10.1002/hyp.8095

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

OPENET. (2023).

Orghidan, T. (1959). A new habitat of subsurface waters: The hyporheic biotope. *Fundamental and Applied Limnology*, *176*(4), 291–302. https://doi.org/10.1127/1863-9135/2010/0176-0291

Painter, S. L. (2018). Multiscale Framework for Modeling Multicomponent Reactive Transport in Stream Corridors. *Water Resources Research*, *54*(10), 7216–7230. https://doi.org/10.1029/2018WR022831

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). https://doi.org/10.1029/2008WR007644

Peter, K. T., Herzog, S. P., Tian, Z., Wu, C., McCray, J. E., Lynch, K., & Kolodziej, E. (2018). Evaluating emerging organic contaminant removal in an engineered hyporheic zone using high resolution mass spectrometry. *Water Research*, *150*, 140–152.

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

Rathore, S. S., Jan, A., Coon, E. T., & Painter, S. L. (2021). On the Reliability of Parameter Inferences in a Multiscale Model for Transport in Stream Corridors. *Water Resources Research*, *57*(5), 1–21. https://doi.org/10.1029/2020WR028908

Runkel, R. L., & Broshears, R. E. (1991). One-dimensional transport with inflow and storage (OTIS): A solute transport model for small streams. *CADSWES-Center for Advanced Decision Support for Water and Environmental Systems - Department of Civil Engineering - University of Colorado.*, 91. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:One-Dimensional+Transport+with+Inflow+and+Storage+(+OTIS+):+A+Solute+Transport+Model+for+Small+Streams#0

Safeeq, M., Bart, R. R., Pelak, N. F., Singh, C. K., Dralle, D. N., Hartsough, P., & Wagenbrenner, J. W. (2021). How realistic are water-balance closure assumptions? A demonstration from the southern sierra critical zone observatory and kings river experimental watersheds. *Hydrological Processes*, *35*(5), 1–16. https://doi.org/10.1002/hyp.14199

Schmadel, N. M., Ward, A. S., Lowry, C. S., & Malzone, J. M. (2016). Hyporheic exchange controlled by dynamic hydrologic boundary conditions. *Geophysical Research Letters*. 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

Soulsby, C., Malcolm, I. A., Tetzlaff, D., & Youngson, A. F. (2009). Seasonal and inter-annual variability in hyporheic water quality revealed by continuous monitoring in a salmon spawning stream. *River Research and Applications*, *25*(10), 1304–1319. https://doi.org/10.1002/rra.1241

Stanford, J. A., & Ward, J. V. (1988). Stanford and Ward - The hyporheic habitat of river ecosystems. *Nature*, *335*, 64–66.

Stewart, M. K., Morgenstern, U., & McDonnell, J. J. (2010). Truncation of stream residence time: How the use of stable isotopes has skewed our concept of streamwater age and origin. *Hydrological Processes*, *24*(12), 1646–1659. https://doi.org/10.1002/hyp.7576

Stonedahl, S. H., Harvey, J. W., Detty, J., Aubeneau, A., & Packman, A. I. (2012). Physical controls and predictability of stream hyporheic flow evaluated with a multiscale model. *Water Resources Research*, *48*(10), 1–15. https://doi.org/10.1029/2011WR011582

Stonedahl, S. H., Harvey, J. W., & Packman, A. I. (2013). Interactions between hyporheic flow produced by stream meanders, bars, and dunes. *Water Resources Research*, *49*(9), 5450–5461. https://doi.org/10.1002/wrcr.20400

Stonedahl, S. H., Harvey, J. W., Wörman, A., Salehin, M., & Packman, A. I. (2010). A multiscale model for integrating hyporheic exchange from ripples to meanders. *Water Resources Research*, *46*(12), 1–14. https://doi.org/10.1029/2009WR008865

Swanson, F. J., & James, M. E. (1975). Geology and geomorphology of the H.J. Andrews Experimental Forest, Western Cascades, Oregon. In *Res. Pap. PNW-188* (p. 14 p.).

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*, 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

Thompson, S. E., Harman, C. J., Troch, P. A., Brooks, P. D., & Sivapalan, M. (2011). Spatial scale dependence of ecohydrologically mediated water balance partitioning: A synthesis framework for catchment ecohydrology. *Water Resources Research*, *47*(5), 1–20. https://doi.org/10.1029/2010WR009998

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

Toth, J. (1963). A Theoretical Analysis of Groundwater Flow in Small Drainage Basins 1 of phe low order stream and having similar t he outlet of lowest impounded body of a relatively. *Journal of Geophysical Research*, *68*(16), 4795–4812. https://doi.org/10.1029/JZ068i016p04795

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). The River Continuum Concept. *Canadian Journal of Fisheries and Aquatice Sciences*, *37*(1).

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

Wagner, B. J., & Harvey, J. W. (1997). Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies. *Water Resources Research*, *33*(7), 1731–1741. https://doi.org/10.1029/97WR01067

Ward, A. S. (2016). The evolution and state of interdisciplinary hyporheic research. *Wiley Interdisciplinary Reviews: Water*, *3*(1), 83–103. https://doi.org/10.1002/wat2.1120

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*. 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., Gooseff, M. N., Voltz, T. J., Fitzgerald, M., Singha, K., & Zarnetske, J. P. (2013). 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. https://doi.org/10.1002/wrcr.20434

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., Payn, R. A., Gooseff, M. N., Mcglynn, B. L., Bencala, K. E., Kelleher, C. A., Wondzell, S. M., & Wagener, T. (2013). Variations in surface water-ground water interactions along a headwater mountain stream : Comparisons between transient storage

and water balance analyses. *Water Resources Research*, *49*, 3359–3374. https://doi.org/10.1002/wrcr.20148

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). 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., Gooseff, M. N., Covino, T., Herzog, S., McGlynn, B., & Payn, R. A. (2023). Breaking the Window of Detection: Using Multi-Scale Solute Tracer Studies to Assess Mass Recovery at the Detection Limit. *Water Resources Research*, *59*(3), 1–16. https://doi.org/10.1029/2022WR032736

Ward, A. S., Wondzell, S. M., Schmadel, N. M., & Herzog, S. P. (2020). Climate Change Causes River Network Contraction and Disconnection in the H.J. Andrews Experimental Forest, Oregon, USA. *Frontiers in Water*, *2*(April), 1–10. https://doi.org/10.3389/frwa.2020.00007

Ward, A. S., Wondzell, S. M., Schmadel, N. M., Herzog, S., Zarnetske, J. P., Baranov, V., Blaen, P. J., Brekenfeld, N., Chu, R., Derelle, R., Drummond, J., Fleckenstein, J. H., Garayburu-Caruso, V., Graham, E., Hannah, D., Harman, C. J., Hixson, J., Knapp, J. L. A., Krause, S., … Wisnoski, N. I. (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., Derelle, R., Drummond, J., Fleckenstein, J. H., Garayburu-Caruso, V., Graham, E., Hannah, D., Harman, C. J., Herzog, S., Hixson, J., Knapp, J. L. A., Krause, S., Kurz, M. J., Lewandowski, J., … Wondzell, S. M. (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

Winter, T. C., Harvey, J. W., Lehn Franke, O., & Alley, W. M. (1999). *Ground water and surface water a single resource* (Issue USGS Circular 1139). https://doi.org/https://doi.org/10.3133/cir1139

Wlostowski, A. N., Gooseff, M. N., Bowden, W. B., & Wollheim, W. M. (2017). Stream tracer breakthrough curve decomposition into mass fractions: A simple framework to analyze and compare conservative solute transport processes. *Limnology and Oceanography: Methods*, *15*(2), 140–153. https://doi.org/10.1002/lom3.10148

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*. https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.12135-9

Wondzell, S. M., LaNier, J., & Haggerty, R. (2009). Evaluation of alternative groundwater flow models for simulating hyporheic exchange in a small mountain stream. *Journal of Hydrology*, *364*(1–2), 142–151. https://doi.org/10.1016/j.jhydrol.2008.10.011

Worman, A., Packman, A. I., Johansson, H., & Jonsson, K. (2002). Effect of flow-induced exchange in hyporheic zones on longitudinal transport of solutes in streams and rivers. *Water Resources Research, 31*(1), 1-15. https://doi.org/10.1029/2001WR000769

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: Biogeosciences*, *116*(1), 1–12. 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

8. APPENDIX

8.1. CHAPTER 2 SUPPLEMENTARY INFORMATION

This SI includes a model schematic of the COMSOL model setup and moving windows, as well as tables including information regarding the study reaches used for the analysis, COMSOL model element qualities, inputs used to calculate *RSF*, and resulting p-values and percent error values for hyporheic exchange metrics. SI is also available: http://www.hydroshare.org/resource/826fbc5bb04e4674b2df002d979b5390

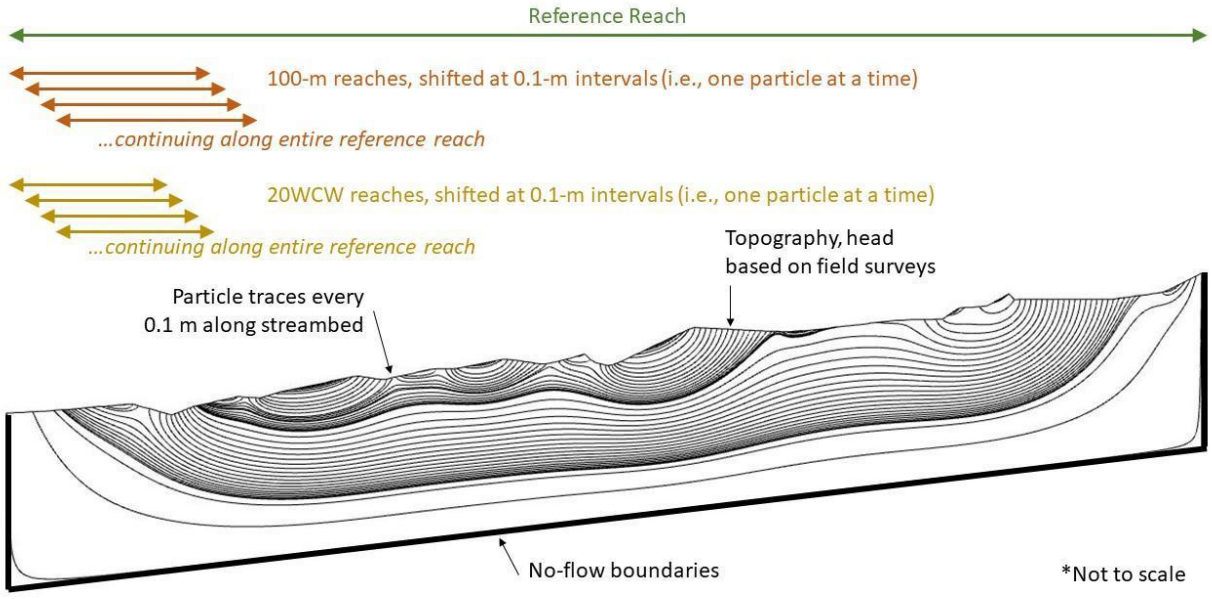

Figure 7.1: Model schematic showing the particle traces, no-flow boundaries, and sliding windows of 20 WCW and 100 m, based on (Herzog et al., 2019; Schmadel et al., 2017).

Table 7.1: Characteristic summaries of study reaches in the HJ Andrews

Reach	Team and Survey year	Reference Source	Reach Length (m)	Drainage Area (m^2)	Landform Type	20WCW length (m)*
Cold Creek*	2014, Ward et al.	(Ward et al., 2019)	246.7	582416.7	Sardine	88.4
Unnamed Creek*	2014, Ward et al.	(Ward et al., 2019)	256.5	1136325	Earth Flow	34.8
WS01*	2014, Ward et al.	(Ward et al., 2019)	537.2	760335.4	Little Butte	19

WS03*	2014, Ward et al.	(Ward et al., 2019)	541.6	784494.9	Little Butte	18.4
REU06	2019, Becker et al.		15.5	1339456	Little Butte	
Syn155	2019, Becker et al.		16.2	811838.9	Little Butte	
Syn161	2019, Becker et al.		16.8	56705.98	Little Butte	
REU04	2019, Becker et al.		19.5	474845	Little Butte	
Syn044	2019, Becker et al.		21.7	1019843	Sardine	
Syn085	2019, Becker et al.		21.9	300315.9	Sardine	
REU08	2019, Becker et al.		23	719965.8	Little Butte	
Syn053	2019, Becker et al.		23.2	270883.4	Earth Flow	
Syn082	2019, Becker et al.		23.8	348233.9	Little Butte	
REU03	2019, Becker et al.		26.1	796870.7	Earth Flow	
Syn060	2019, Becker et al.		26.7	105644.8	Earth Flow	
Syn071	2019, Becker et al.		30	1014683	Little Butte	
Syn058	2019, Becker et al.		35	670661.7	Earth Flow	
Syn064	2019, Becker et al.		36.8	153736.8	Earth Flow	
Syn043	2019, Becker et al.		38.4	86425	Sardine	
Syn074	2019, Becker et al.		43.5	499902.7	Little Butte	
Syn059	2019, Becker et al.		43.8	368894.2	Earth Flow	
LO241	2004, Anderson et al.	(Anderson et al., 2005)	54.7	992520.8	Little Butte	
LO214	2004, Anderson et al.	(Anderson et al., 2005)	61.8	847543.7	Earth Flow	

LO224	2004, Anderson et al.	(Anderson et al., 2005)	89.9	2064717	Earth Flow	
LO282	2004, Anderson et al.	(Anderson et al., 2005)	95.8	474845	Sardine	
LO395	2004, Anderson et al.	(Anderson et al., 2005)	104.5	962094.6	Sardine	
LO334	2004, Anderson et al.	(Anderson et al., 2005)	131.3	5721441	Sardine	
LO348	2004, Anderson et al.	(Anderson et al., 2005)	141.7	3747207	Sardine	
LO428	2004, Anderson et al.	(Anderson et al., 2005)	199.1	30892709	Little Butte	
LO356	2004, Anderson et al.	(Anderson et al., 2005)	202.2	16776008	Sardine	
LO403	2004, Anderson et al.	(Anderson et al., 2005)	267.2	61886662	Little Butte	
LO407	2004, Anderson et al.	(Anderson et al., 2005)	278.4	58835434	Little Butte	
LO416	2004, Anderson et al.	(Anderson et al., 2005)	400	52596433	Little Butte	

Table 7.2 Mesh element qualities

Reach	**Number of Elements**	**Minimum quality**	**Average Quality**
Cold Creek	281908	0.4942	0.9295
Unnamed Creek	304912	0.4148	0.9303
WS01	434,484	0.4999	0.928
WS03	499454	0.1218	0.927
REU06	12889	0.6182	0.9254
Syn155	17840	0.6399	0.926

Syn161	18720	0.5712	0.926
REU04	16448	0.5512	0.9222
Syn044	18188	0.499	0.9232
Syn085	22082	0.5182	0.9104
REU08	19271	0.6005	0.921
Syn053	19106	05145	0.9233
Syn082	20180	0.5782	0.9168
REU03	21614	0.6283	0.9249
Syn060	22859	0.4846	0.9173
Syn071	25414	0.5668	0.9202
Syn058	29132	0.5827	0.9234
Syn064	32348	0.5605	0.9233
Syn043	32641	0.6277	0.9247
Syn074	38430	0.5492	0.9224
Syn059	24335	0.5671	0.916
LO241	45365	0.6298	0.9211
LO214	51893	0.6305	0.92
LO224	74274	0.5474	0.9255
LO282	77831	0.5438	0.9217
LO395	116671	0.545	0.9297
LO334	148925	0.5621	0.9313
LO348	159058	0.5378	0.9303
LO428	273744	0.5381	0.9336
LO356	225740	0.5458	0.9278
LO403	365210	0.5323	0.9341
LO407	380403	0.5349	0.9367
LO416	558654	0.5395	0.9342

Table 7.3: Inputs for Reaction Significance Factors (RSF)

Watershed	Discharge, Q (m^3/s)	Channel width, w (m)	Reach length, L_c (m) Segment	Reach length, Lc (m) 20 WCW
Cold Creek	0.016	4.42	246.7	88.4
Unnamed Creek	0.002	1.74	256.5	34.8
WS01	0.002	0.95	537.2	19
WS03	0.001	0.92	541.6	18.4

Explanation of other variables

s is the residence time in the storage zone. Transit times from each particle was used for this

ris the intrinsic reaction timescale in the storage zone (s), fixed at 10 hour following Harvey et al., 2018.

Ls is the river turnover length calculated as Ls=Qqs*w where Q is discharge and qsis hydrologic exchange flux (m s^{-1}), extracted from particle traces.

Kruskal- Wallis p-values for TTD and RSF

Percent of p-values from Kruskal-Wallis pairwise tests that are less than 0.10, 0.05, and 0.01 for each reach length (20WCW and 100m) for Cold Creek, Unnamed Creek, Watershed 01, and Watershed 03 compared to the reference segment. (ie they are significantly different).

Table 7.4: p_{kw} for transit time distributions (TTD)

	p_{kw} **<0.1**		p_{kw} **< 0.05**		p_{kw} **<0.01**	
	20WCW	**100 m**	**20WCW**	**100 m**	**20WCW**	**100 m**
Cold (20WCW = 88m)	14.8	12.8	8.1	5.7	2.5	2.0
Unnamed (20WCW=38.4 m)	78.0	78.7	70.3	77.6	55.8	73.2
WS01 (20WCW=19 m)	54.7	27.3	45.0	17.1	30.3	6.3
WS03 (20WCW= 18.4 m)	54.2	52.6	45.7	41.9	33.0	32.2

Table 7.5: p_{kw} for Reaction Signifcance Factor (RSF)

	p_{kw} <0.1		p_{kw} < 0.05		p_{kw} <0.01	
	20WCW	**100 m**	**20WCW**	**100 m**	**20WCW**	**100 m**
Cold (20WCW = 88m)	100	100	100	100	99.4	98.2
Unnamed (20WCW=38.4 m)	100	93.9	98.7	89.5	87.8	75.6
WS01 (20WCW=19 m)	100	100	100	100	100	100
WS03 (20WCW= 18.4 m)	88.8	90.7	86.1	90.0	80.6	87.7

Table 7.6: Percent difference for Exchange Flux (Q_{hef}) and percent upwelling (P_{up})

	WS	**<1%**	**≥1 %<5%**	**≥5 <10%**	**≥ 10 <50%**
Q_{hef} **(20 WCW)**	**Cold**	3.6	23.5	45.1	27.8
	Unnamed	3.2	8.7	6.5	78.8
	WS01	3.6	15.8	13.9	64
	WS03	1.8	8.6	7.9	42.5
Q_{hef} **(100 m)**	**Cold**	10.7	36.8	24.9	27.5
	Unnamed	4.8	2.4	7.5	85.3
	WS01	3.0	26.3	31.7	39.0
	WS03	7.4	33.8	22.4	36.3
P_{up} **(20 WCW)**	**Cold**	30.4	59.1	10.5	0
	Unnamed	6.5	24.1	27.2	42.2
	WS01	6.3	24.0	22.4	42.6
	WS03	2.0	10.0	9.7	42.5
P_{up} **(100 m)**	**Cold**	25.7	68.0	6.3	0

	Unnamed	9.8	48.2	31.7	10.2
	WS01	9.6	23.7	26.9	39.8
	WS03	4.2	22.2	47.9	25.7

8.2. Chapter 3 Supplementary Information

Table 7.7: PFLOTRAN Inputs

Species	**Concentration**
DOM	700.0 e-6 (μM)
O_2 (aq)	0.21 G O_2 (g)
H+	6.0 (P)
HCO_3-	400e-6 G CO_2 (g)
NO_3-	1.0e-15 (μM
N_2 (aq)	0.78 G N_2 (g)

8.2.1. Reaction Equations in Input File

```
DOM aerobic respiration
  MICROBIAL_REACTION
    REACTION 1.0e+00 DOM1  + 1.0e+00 O2(aq)  -> 1.0e+00 HCO3- + 1.0 H+
    RATE_CONSTANT       1.0e-06
    MONOD
      SPECIES_NAME       O2(aq)
      HALF_SATURATION_CONSTANT 6.0e-06
      THRESHOLD_CONCENTRATION 1.1e-20
    /
    MONOD
      SPECIES_NAME       DOM1
      HALF_SATURATION_CONSTANT 4.5e-05
```

```
    THRESHOLD_CONCENTRATION 1.1e-20

DOM denitrification
  MICROBIAL_REACTION
   REACTION 1.0e+00 DOM1  + 0.8e+00 NO3- + 0.8 H+  -> 1.0 HCO3- + 1.0 H+  + 0.4
N2(aq)
   RATE_CONSTANT       1.6e-07
   MONOD
    SPECIES_NAME       NO3-
    HALF_SATURATION_CONSTANT 5.0e-01 #make this big to drive to zero and get first
order kinetics
    THRESHOLD_CONCENTRATION 1.1e-20
   /
   MONOD
    SPECIES_NAME       DOM1
    HALF_SATURATION_CONSTANT 4.5e-05
    THRESHOLD_CONCENTRATION 1.e-20
   /
   INHIBITION
    SPECIES_NAME       O2(aq)
    TYPE MONOD
    INHIBITION_CONSTANT 3.0e-7
```

8.3. CHAPTER 4 SUPPLEMENTARY INFORMATION

Table 7.8: Stage data information including location and correlation coefficients with the water stage level at the weir. Small p-value suggests strong correlation

Lower Segment Locations (m)	Correlation coefficient, p-value	Upper Segment Locations (m)	Correlation coefficient, p-value
47.4	0.1843, 0	382.7	0.1204, 0
94.1	0.1136, 0	396.2	0.2213, 0
196.3	0.1516, 0	440.0	0.1569, 0
245.8	0.2070, 0	484.8	0.1570, 0
306.7	0.1561, 0	515.9	-0.2044, 0
352.3	0.1111	539.7	-0.2631, 0

Table 7.9: Linear regression model coefficient and p-values of stage data

Segment	Intercept	Slope (mm H2O/day)	p-value
Lower	-2908.8	0.004	p <<1e-3
Upper	5009.2	-0.006	p << 1e-3

Table 7.10 OpenET data

Date	ET (mm/day)
7/8/21	4.345
7/9/21	5.453
7/10/21	5.614
7/11/21	6.296
7/12/21	5.738
7/13/21	5.949
7/14/21	5.731
7/19/21	5.416
7/20/21	4.956
7/21/21	4.116

7/22/21	5.164
7/23/21	5.474
7/25/21	5.17
7/26/21	7.574
7/27/21	3.556
7/28/21	7.91
7/29/21	7.147
8/8/21	3.507
8/9/21	5.26
8/10/21	5.415
8/11/21	5.237
8/15/21	5.163
8/16/21	5.339
8/17/21	2.67
8/18/21	4.709
8/19/21	4.525
8/23/21	4.7
8/24/21	4.843
8/25/21	4.262
8/26/21	3.365
8/27/21	4.479

Shape file coordinates that ET data was gathered for:

-122.25916431751105, 44.20758321187398, -122.25873815536036, 44.206210278251056, -122.25704820128159, 44.20582545506822, -122.2569908507285, 44.20492995925534, -122.25611580885018, 44.20445226829652, -122.25456695363394, 44.2038488513005, -122.25122548959918, 44.20527782828669, -122.2507505945396, 44.206180277342085, -122.24978139245651, 44.20720915446756,-122.25223759916847, 44.20754383495265, -122.25509374315155, 44.20815609691281, -122.25737963744906, 44.20825161288477