Photocopy and Use Authorization

In presenting this dissertation in partial fulfillment of the requirements for an advanced degree at Idaho State University, I agree that the Library shall make it freely available for inspection. I further state that permission for extensive copying of my dissertation for scholarly purposes may be granted by the Dean of the Graduate School, Dean of my academic division, or by the University Librarian. It is understood that any copying or publication of this dissertation for financial gain shall not be allowed without my written permission.

Signature_______________________________________________________

Date__________________________________________________________
Quantifying Patterns and Processes of Intermittent Stream Biogeochemistry

and of Introductory STEM Classroom Behavior

by Ruth B. MacNeille

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctorate of Arts

in the Department of Biological Sciences

Idaho State University

Fall 2020
Copyright (2020) Ruth B. MacNeill
To the Graduate Faculty:

The members of the committee appointed to examine the dissertation of Ruth MacNeille find it satisfactory and recommend that it be accepted.

______________________________
Kathleen Lohse,
Major Advisor

______________________________
Colden Baxter,
Committee Member

______________________________
Rosemary Smith,
Committee Member

______________________________
Keith Reinhardt,
Committee Member

______________________________
Sarah Godsey,
Graduate Faculty Representative
Dedication

My pursuit of a Doctor of Arts was inspired by Dr. Micha Miller and Dr. Erin Naegle of Columbia College, CA, and by my students outdoors and in the classroom. My desire to become the best science teacher I could be has motivated this journey. I dedicate this work to all of the above, to my parents and family, as well as to my own hard work, which has brought me further than I could have imagined.
Acknowledgements

First, I would like to acknowledge and thank my major advisor, Dr. Kitty Lohse and the rest of my committee members: Dr. Colden Baxter, Dr. Sarah Godsey, Dr. Rosemary Smith, and Dr. Keith Reinhardt. Each has been supportive of my growth over the past five years in many ways and each has been highly influential to my research and professional development. I would also like to thank Dr. Micha Miller and Dr. Erin Naegle for being a major support with regards to mentorship in teaching and teaching careers.

I would also like to acknowledge Idaho State University for awarding me the Doctorate of Arts (D.A.) fellowship, the ISU Department of Biological Sciences and Geosciences, and the College of Science and Engineering (CoSE) for supporting graduate students. This research was funded by NSF RC CZO Cooperative Agreement #EAR 1331872, USDA-ARS, and the Idaho State University Doctorate of Arts fellowship in biology.

Part 1: Special thanks to the United State Department of Agriculture Agricultural Research Service (USDA ARS) Northwest Watershed Research Center personnel for logistical and infrastructure support in addition to the Lohse Lab and Idaho State University. Additional acknowledgments include Dr. Dewayne Derryberry, Dr. Bruce Finney, Susan Parsons, Emma McCorkle, Kayla Glossner, and Anna Radke. Special dedication to the memory of Dr. Wayne “Doc” Minshall. Additional thanks to Kate Behn for cocreating semivariogram R-code.

Part 2: Special thanks to Dr. Anna Grinath, Dr. Karen Appleby, and Dr. Barbra Frank for contributions and specifically the D.A. program.

Thank you to the Biology Graduate Student Association (BGSA) for all you do for the graduate experience at ISU.

Lastly, I would like to thank and acknowledge the emotional support of my family and friends without whom such an endeavor would not have been possible.
# Table of Contents

List of Figures ................................................................................................. ix

List of Tables .................................................................................................... xi

Abstract ........................................................................................................... xii

Chapter I: General Introduction to the Research .............................................. 1

  1.1 Structure and Process ............................................................................... 1

  1.2 Patterns and Processes at Multiple Scales ............................................... 4

  1.3 Research Overview .................................................................................. 6

  1.4 Works Cited ............................................................................................. 11

Chapter II: Influence of Drying and Wildfire on Longitudinal Chemistry Patterns and Processes of Intermittent Streams ........................................... 18

  Abstract ........................................................................................................ 18

  2.1 Introduction ............................................................................................. 19

  2.2 Methods ................................................................................................... 28

  2.3 Results ...................................................................................................... 42

  2.4 Discussion ............................................................................................... 59

  2.5 Conclusions ............................................................................................ 71

  2.6 Works Cited ............................................................................................ 74

  2.7 Supplementary Materials ....................................................................... 84
Chapter III: Spatial Complexity Theory to Practice: Quantifying Fine-Grain Biogeochemical Heterogeneity of Drying and Wildfire Patterns of Intermittent Streams

Abstract ................................................................. 91

3.1 Introduction .......................................................... 92

3.2 Methods .................................................................... 102

3.3 Results ...................................................................... 114

3.4 Discussion ............................................................... 130

3.5 Conclusions .............................................................. 144

3.6 Works Cited .............................................................. 146

3.7 Supplementary Materials ........................................... 157

3.8 Appendices A-C ......................................................... 165

Chapter IV: What Teaching Techniques Are Currently Used at a Midsized Public University in Introductory STEM Classrooms?

Abstract ................................................................. 173

4.1 Study Motivation ........................................................ 174

4.2 Theoretical Framework .............................................. 177

4.3 Prior Research .......................................................... 179

4.4 Research Design ....................................................... 189

4.5 Results ................................................................. 196
4.6 Discussion ............................................................................................................... 206

4.7 Conclusion .............................................................................................................. 217

4.8 Works Cited ............................................................................................................ 219

4.9 Supplementary Materials ...................................................................................... 228

4.10 Appendices A-D .................................................................................................... 235
List of Figures

Figure 2.1 2016 Study Design at Reynolds Creek Experimental Watershed Critical Zone Observatory.................................................................32
Figure 2.2 2016 Weir Discharge, Temp (°C), pH, and DO (mg/L).......................43
Figure 2.3 Initial Cluster Analysis.................................................................46
Figure 2.4 2016 Longitudinal Analyte Patterns...............................................48
Figure 2.5 Analyte Differences from April to June...........................................50
Figure 2.6 In-Stream Evaporative Processes and Groundwater Analyses...........52
Figure 2.7 Autochthonous or Allochthonous Organic Carbon Sourcing................58
Figure 3.1 Hypothesized Semivariograms....................................................101
Figure 3.2 2016 and 2017 Sampling and Study Area ....................................104
Figure 3.3 Workflow Diagram.......................................................................108
Figure 3.4 Discharge of the Unburned JD and Burned MC Measured at the Weir ....116
Figure 3.5 2016 and 2017 Longitudinal Concentration Plots............................118
Figure 3.6 Linear Model Versus GAM Model Fits and Residual Plots................119
Figure 3.7 2016 DIC Semivariograms for April and June.................................124
Figure 3.8 2017 Semivariograms for DIC, DOC, TN, and PO₄³⁻ ........................130
Figure 3.9 Revised Conceptual Semivariograms.............................................142
Figure 4.1 Active Learning Gains at Multiple Scales......................................181
Figure 4.2 Percentage of Total Aggregated Observations of Student and Instructor Behaviors.................................................................199
Figure 4.3 Student and Instructor COPUS Behaviors as a Percentage of the Total Class Time...............................................................201
Figure 4.4 Most Time Spent by Students and Instructors at Idaho State University Compared to Universities Nationally…………………………..202

Figure 4.5 Distribution of ISU Instructors’ Self-Reported Practices…………………….204

Figure 4.6 Correlations of Instructor Self-Assessment (TPI) Scores and Anonymous Student Surveys………………………………………………….206
List of Tables

Table 2.1 Study Aims.................................................................26
Table 2.2 Characteristics of the Unburned JD and Burned MC Sub-Catchments...........30
Table 2.3 Stable Water Isotope Analysis.........................................................53
Table 3.1 AIC Values Based on the GAM Fit for the Data and Normalized by the Sample Size (N)........................................................................122
Table 3.2 Spatial Statistical Parameters for DIC, Cl\textsuperscript{−}, DOC, TN, and PO\textsubscript{4}\textsuperscript{3−} ..........126
Table 4.1 Groupings of COPUS Classroom Observation Behaviors within an ICAP Framework..................................................................................................193
Table 4.2 Percentages of Classroom Observations of Classroom Behaviors for Students and Instructors and Percentage of Time Spent Measured by COPUS.............197
Table 4.3 Differences and Similarities Synthesized from Instructors’ Self-Reported Practices........................................................................................................205
Quantifying Patterns and Processes of Intermittent Stream Biogeochemistry and of Introductory STEM Classroom Behavior

Dissertation Abstract—Idaho State University (2020)

Changing environmental conditions and disruptions are altering how streams and science classrooms are structured and how they function. Although stream spatial complexity has been theoretically emphasized, quantifying heterogeneity remains limited, especially in intermittent streams. In the Intermountain West United States, streams are experiencing earlier drying and the region is experiencing elevated wildfire conditions with climate change. Similarly, classroom active learning structures have been shown to benefit student learning, yet lecture dominates higher education practices. This dissertation explores the shared theme of linking theory and practice in two research areas: (a) biogeochemical patterns and processes in intermittent streams, and (b) introductory college science classroom behavior. In Part 1, I conducted sampling and quantified biogeochemical patterns and processes at fine-grain intervals (10 to 50 m) in two streams in southwestern Idaho, United States, one unburned and one recently burned, as they dried. In contrast to expectations, streams showed weak evidence of in-stream evapoconcentration with drying. However, streams showed groundwater contributions increased with drying, which was more apparent in the burned stream. Organic carbon sourcing was highly variable in both streams. Then I quantified patterns using semivariograms and showed heterogeneity increased for dissolved inorganic and organic carbon but decreased for other nutrients with drying. Small solute semivariogram ranges detected (>150 m) may indicate that processing in intermittent headwater streams occurs at smaller scales than larger or perennial streams, and longer ranges in burned stream
chemistry may support the telescoping ecosystem model, indicating disturbance like fire may increase patch size.

In Part 2, I conducted observations of student and instructor behaviors in introductory science classrooms, instructor self-assessments, and anonymous student surveys to assess the extent of active learning implementation. Observed classroom behaviors were largely less active for instructors (71%) and students (81%) and mirrored national data. Little (>1%) student time was spent on core science skills such as predicting and presenting. Instructors, who reported using more active learning spent more time on the processes of science, and their students did more group work. Future teaching professional development recommendations include supporting instructors to facilitate active student behaviors and conducting student evaluations.

Key Words: intermittent streams, biogeochemistry, evaporative processes, evapoconcentration, groundwater, allochthonous, autochthonous, stable water isotopes, heterogeneity, semivariogram, science education, active learning, ICAP framework, undergraduate STEM, teaching professional development, classroom behaviors, COPUS, classroom observations.
Chapter I: General Introduction to the Research

1.1 Structure and Process

The relationship between structure and function is central to stream ecology (Frissell et al., 1986; Fisher et al., 1998a; Montgomery, 1999; Poole, 2002) and science education (Wieman et al., 2010; Petersen and Gorman, 2014; Baepler et al., 2016; Kranzfelder et al., 2020). Streambed structure, for example, influences stream ecosystem processes such as nutrient cycling (Fisher et al., 1998a; Malard et al., 2002; Fisher et al., 2004). Stream geomorphology influences complex flow paths for water exchange (Ward, 1989; Fisher and Welter, 2006) that include vertical flow exchange in hyporheic zones (Valett et al., 1994; Fisher et al., 1998a; Gomez et al., 2017), lateral flow exchange with riparian (Dwived et al., 2017), hillslopes (Boyer et al., 1997), and sandbar parafluvial zones (Fisher et al., 1998b; Gomez et al., 2017). Such zones can create anaerobic environments connected by water and alter nitrogen cycling (Jones et al., 1995; Malard et al., 2002; McClain et al., 2003; Fisher et al., 2004; Harms and Grimm, 2008; Bernhardt et al., 2017). Therefore, the amount of stream water connecting these structures to each other impacts the complexity of biogeochemical patterns and processes (Fisher et al., 2004; von Schiller et al., 2008; Gomez et al., 2017). Despite these observations and growing theoretical emphasis on the importance of spatial complexity on biogeochemistry (Fisher et al., 1998a; Pool, 2002; Fisher et al., 2004), few studies have evaluated how stream structure complexity influences biogeochemical patterns and processes (Cooper et al., 1997).

Disturbances such as flooding or wildfire can influence patterns and processes in streams (Fisher et al., 1998b), as well as seasonal phenomena such as stream
intermittency, or drying by altering the water flow (Fisher and Welter, 2006). For example, fire can increase the amount of stream water by reducing evapotranspiration from the water budget (Atchley et al., 2018; Wine and Cadol, 2016). Stream drying patterns have been shown to impact nutrient retention (von Schiller et al., 2008) and biofilm life cycles and phosphorous processing (Timoner et al., 2010). However, the ecological and hydrological study of stream intermittency has only recently emerged as a robust subfield to which many disciplines including biogeochemistry contribute (Busch et al., 2019), but there are few conceptual models that incorporate this phenomenon (Costigan et al., 2016; Allen et al., 2020). This occurs despite the fact that intermittent streams represent at least 60% of streams in the United States (Nadeau and Rains, 2007; Datry et al., 2017) and make up ~80% the streams in the western United States (Levick et al., 2008). Both stream intermittency (Döll and Schmied, 2012) and wildfire will likely increase under current climate change scenarios in regions such as the Great Basin ecosystem (Klos et al., 2014; Abatzoglou and Williams, 2016; Parks et al., 2016). Both phenomena can cause heterogeneity at the meter scale with consequences for an entire stream network (Costigan et al., 2016).

Similar to stream complexity, classroom structure is complex and can be physical, curricular, or social. Patterns of information or idea flow can influence student learning, identity, and sense of belonging. When information flows primarily from teacher to student (i.e., lecture, answering questions, etc.), students behave passively by listening and receiving information (Chi and Wylie, 2014). Classroom architecture often facilitates and reinforces the position of the instructor as the main focus (Petersen and Gorman, 2014; Baepler et al., 2016). However, when students are asked to discuss a prompt with
other students, the students play a more active role to develop their own understanding, sense-making, and learning (Resnick et al., 2015; Grinath and Southerland, 2019; Kranzfelder et al., 2020). Information flows in more complex patterns, and when students are generating their own ideas, deeper learning processes are promoted (Chi, 2009; O’Conner et al., 2015; Kranzfelder et al., 2020). Classroom structure is reflected in patterns and processes of classroom behavior (Kranzfelder et al., 2020), and measuring this behavior can help describe and consider the structure with regard to teaching effectiveness and student learning.

*Active learning* is an umbrella term used to describe a multitude of teaching practices that seek to put students at the center of generating information (Chi, 2009; Chi and Wylie, 2014) and promote student interactions in classrooms (Kranzfelder et al., 2020). Active learning in science, technology, engineering, and mathematics (STEM) classrooms has increased student learning (Freeman et al., 2014), reduced the achievement gap (Strimaitis et al., 2017; Canning et al., 2018; Theobald et al., 2020), increased student satisfaction (Ueckert et al., 2011), and increased student retention (Canning et al., 2018). Despite overwhelming evidence supporting the use of active learning in classrooms, lecture has prevailed as the dominant teaching strategy in higher education (Burrowes, 2003; Tanner and Allen, 2006; Derting and Ebert-May, 2010; Ueckert et al., 2011; Weimer, 2013). This may be in part because faculty who learn to teach by observing mentors tend to reproduce these modes of teaching, though many other models and techniques are possible (Brownell and Tanner, 2012). While exposing teachers to different pedagogical techniques can change their conceptual models of teaching and impact practice (Katz et al., 2010; Wilson, 2013), the dominant paradigms
have consequences for the teaching practices implemented and can make change difficult (Brownell and Tanner, 2012).

1.2 Patterns and Processes at Multiple Scales

Patterns and processes in streams and science education occur at multiple organizational, spatial, and time scales. Spatial patterns in streams occur at habitat to catchment to network scales (Frissell et al., 1986; Poole, 2002; Fisher and Welter, 2006). To move between multiple scales, extrapolation of processes at smaller scales is common in ecology (Fisher et al., 2004). However, it is possible to communicate directly between fine-grain and landscape scale patterns with a study scope that accommodates measurements at both scales (Cooper et al., 1997; Turner and Chapin, 2006; Fisher et al., 2004; McGuire et al., 2014). Stream researchers often use predetermined parameters defined by the physical environment such as habitat to investigate patterns and processes (Frissell et al., 1986; Fisher et al., 1998a, Montgomery, 1999). While there may be important process domains that occur with geomorphology (Montgomery, 1999), rarely do research designs allow patterns and processes in streams to emerge without preset parameters (Cooper et al., 1997; Scown et al., 2016). Stream chemistry data are particularly well poised to provide continuous spatial measurements (McGuire et al., 2014; Scown et al., 2016); continuous measurements are useful for moving between scales and also allow for patterns to emerge (Cooper et al., 1997; Peterson et al., 2006; Scown et al., 2016). With a few exceptions, stream studies do not often quantify the changing patterns at scales required to observe and describe these changes to structure (Cooper et al., 1997; McGuire et al., 2014; Scown et al., 2016). In part, this has been due to difficulty in both sampling in a spatially continuous manner at a fine enough grain over
a large enough extent (Dent and Grimm et al., 1999) and applying spatial statistics (Cooper et al., 1997; Peterson et al., 2006). For ecologists, this design and approach counters traditional parametric statistics that seek to control and diminish variability between sites, not exploit the variability (Turner and Chapin, 2006). Quantifying stream drying patterns has an added difficulty of describing patterns as the stream structure changes with drying.

Scales in science education impact student learning, which is shaped by curriculum. Curriculum on smaller scales can be defined as a particular lesson, or by a particular course like introductory biology, and at larger scales, curriculum can be defined as an entire degree program (Cobb et al., 2003; Porter and Rossner, 2006; Wieman et al., 2010; Bradforth and Miller 2015; Matz et al., 2018). All scales contribute to experiences that form neurological connections in a student’s brain; from a structural perspective, forming neurological connections defines the learning process (Zull, 2011). At small scales, the sequence of student-instructor interaction influences students’ cognitive processes (Greeno, 2015). For instance, a common question/response sequence scenario occurs when an instructor asks the class a question, a student responds, and then the instructor evaluates or elaborates on a student’s response without further student elaboration. In this sequence, the students’ response role is complete after answering the initial prompt, which the instructor evaluates so that the instructor is cognitively active longer and at a deeper level (Crowe et al., 2008) than the student. This sequence suggests students’ contributions stop after an initial response and they then cognitively disengage from further developing or justifying their ideas (Greeno, 2015). Changes to this interaction sequence and positioning of the instructor as the provider of information can
result in a deeper learning process (Greeno, 2015). To identify these patterns, observing classroom behavior of students and instructors can be a tool for gaining insight into what learning processes are occurring. Patterns can either emerge through qualitative methods or be tested using more quantitative methods (Creswell and Clark, 2011).

Unlike stream processes, education researchers often find it more difficult to extrapolate between scales in education, given the contextual nature of using humans as a study subject (Merriam, 2009; O’Conner et al., 2015). However, similar to scales in stream patterns, educational scales also communicate among each other, especially because institutional support plays a large role in structuring whether instructors feel encouraged to engage to teach (Fanghanel, 2007; Wieman et al., 2010; Matz et al., 2018). Institutional support is key because people, be they students or instructors, are externally motivated by evaluation (Wieman et al., 2010; Bradforth et al., 2015). If instructors are not incentivized through promotion and tenure or other recognition to professionally develop their teaching skills, they will orient their efforts toward those merits that are considered (Wieman et al., 2010; Bradforth et al., 2015). Observing classroom behavior of both instructors and students can help to elucidate patterns and processes that are occurring and identify areas of support in the classroom on an institutional level.

1.3 Research Overview

Streams and classrooms share a similar disconnect between theory and practice; this research addresses this gap in each. Theoretically, stream ecologists have recognized the significance of spatial complexity on ecosystem processes (Frissell et al., 1986; Fisher et al., 1998a; Montgomery, 1999; Poole, 2002; Thorp et al., 2006; Winemiller et al., 2010); however, the tools to quantify these patterns have not often been employed
(Cooper et al., 1998; Scown et al., 2016), especially in intermittent streams (Costigan et al., 2016). Similarly, in classrooms, research on the use of active learning has been supported by gains in students learning (Freeman et al., 2014), satisfaction (Ueckert et al., 2011), and equity (Strimaitis et al., 2017; Canning et al., 2018; Theobald et al., 2020). Yet, the implementation of active learning teaching techniques lags behind traditional lecture in higher education (Prince et al., 2004; Waldrop, 2015). Studying patterns and processes of both intermittent stream biogeochemistry and classroom behavior offers important insights to both fields individually and also highlights surprising links between the two subjects.

This dissertation examined patterns and processes in two parts; Part 1 focuses on intermittent stream biogeochemistry, and Part 2 focuses on classroom behavior across introductory STEM courses. In Part 1, I sought to investigate the biogeochemistry patterns and processes that change with stream drying and following fire. I used a temporally repeated high-spatial scope to investigate changes in two streams, one unburned and one recently burned in southwestern Idaho, United States. In Chapter 2, I asked what patterns and process shifts were observed as streams dried. Specifically, I tested for shifts in three hypothesized processes influencing biogeochemical patterns: evaporative processes, groundwater influence, and organic carbon sources. My results suggested that in-stream evaporative processes did shift substantially with seasonal changes, but that groundwater influence increased with stream drying. I detected previously unobserved longitudinal patterns, especially with regard to fine-grain variability of organic carbon sourcing. Furthermore, I was able to make inferences about the somewhat muted effects of fire and stream drying on these processes. The
spatiotemporal biogeochemical patterns that I observed with stream drying were complex and led me to think deeply about how to quantify and describe them.

While recognized as theoretically important, quantifying spatiotemporal patterns have not often been applied to continuous biogeochemical measurements (Turner and Chapin, 2006; Scown et al., 2016). In Chapter 3, I asked how heterogeneity and patch size changed as streams dried. I hypothesized that heterogeneity would increase at low flows and that following fire, stream drying would result in greater biogeochemical heterogeneity. Building on the data set for Chapter 1, I added another time stamp at low flow with higher density sampling in order to quantify spatial dependency. I applied geostatistical tools and used a time series approach to capture and describe biogeochemical spatial complexity as the streams dried. I described a method for quantifying spatial patterns in nonbranching headwater streams as they dried. My results demonstrated that heterogeneity increased for dissolved carbon with stream drying but decreased for other nutrients, and suggested that fire may impact patch dynamics at low flow. Collectively, my results may be consistent with predictions set forth by the telescoping ecosystem model (TEM; Fisher et al., 1998b) with regard to biogeochemical patch size following fire. At low flow, solute patch size was larger in the burned stream than the unburned stream, which did not support my initial hypothesis.

In Part 2, I transition to education research investigating active learning in introductory STEM classrooms. Similar to the traditional oversight of stream drying as a ubiquitous process in intermittent streams, active learning has been promoted conceptually but not in practice (Prince et al., 2004; Waldrop, 2015). In Chapter 4, I assess common teaching practices employed in introductory STEM courses at Idaho State
University (ISU). Utilizing external observations, instructor self-assessment, and anonymous student surveys, I investigated teaching strategies and practices in introductory STEM courses. I quantified patterns of classroom behavior of both instructors and students throughout the semester and then qualitative instructor self-reported practices. These results suggested that, similar to national patterns, most of the time students displayed passive behavior like listening and asking questions while instructors promoted these behaviors by lecturing and writing on the board. Behaviors that were rarely observed were predicting and students presenting, which both reflect nationally identified biology education goals (AAAS, 2009). Students self-reported that they felt more engaged when they also asked more questions. My data also provided suggestions for overcoming barriers to using active learning techniques and identified specific areas for teaching professional development (TPD) support and curriculum development. In addition, my results provided a baseline of data to measure the success of future teaching, curriculum, or institutional interventions.

Streams and STEM classrooms share a powerful opportunity to examine patterns at multiple scales that challenge the dominant paradigm in their respective fields. A unifying theme between the two projects is a disconnect between theory and practice, and each offered an opportunity to reflect on patterns able to be examined based on measurement grain and extent. In addition, each takes the perspective that inputs (chemical inputs in the stream or informational inputs into students’ brains) are not passively transported by their respective vessels but are actively transformed, changing through time and space. Thus, both projects seek to identify and map patterns at large and small scales. The results of these research projects provide foundational patterns for
systems that may experience shifts and thereby provide baseline data, methodological approaches, and recommendations for future research.
1.4 Works Cited


Part I

Chapter II: Influence of Drying and Wildfire on Longitudinal Chemistry Patterns and Processes of Intermittent Streams

Abstract

Stream drying and wildfire are projected to increase with climate change in the western United States, and both are likely to impact stream chemistry patterns and processes. To investigate drying and wildfire effects on stream chemistry (carbon, nutrients, anions, cations, and isotopes), we examined seasonal drying in two intermittent streams in southwestern Idaho, one stream that was unburned and one that burned eight months prior to our study period. During the seasonal recession following snowmelt, we hypothesized that spatiotemporal patterns of stream chemistry would change due to increased evaporation, groundwater dominance, and autochthonous carbon production. With increased nutrients and reduced canopy cover, we expected greater shifts in the burned stream. To capture spatial chemistry patterns, we sampled surface water for a suite of analytes along the length of each stream with a high spatial scope (50-meter sampling along ~2500 meters). To capture temporal variation, we sampled each stream in April (higher flow), May, and June (lower flow) in 2016. Seasonal patterns and processes influencing stream chemistry were generally similar in both streams, but some were amplified in the burned stream. Mean dissolved inorganic carbon (DIC) concentrations increased with drying by 22% in the unburned and by 300% in the burned stream. In contrast, mean total nitrogen (TN) concentrations decreased in both streams, with a 16% TN decrease in the unburned stream and a 500% TN decrease (mostly nitrate) in the burned stream. Contrary to expectations, dissolved organic carbon (DOC) concentrations
varied more in space than in time. In addition, we found the streams did not become more evaporative relative to the Local Meteoric Water Line (LMWL) and we found weak evidence for evapoconcentration with drying. However, consistent with our expectations, strontium-DIC ratios indicated stream water shifted towards groundwater-dominance, especially in the burned stream. Fluorescence and absorbance measurements showed considerable spatial variation in DOC sourcing each month in both streams, and mean values suggested a temporal shift from allochthonous towards autochthonous carbon sources in the burned stream. Our findings suggest that the effects of fire may magnify some chemistry patterns but not the biophysical controls that we tested with stream drying.

2.1 Introduction

Intermittent streams, those streams experiencing periods of disconnected surface water flow (Larned et al., 2014), currently constitute ~30% of the total river length and discharge of the world river network and about half of the United States (US) network (Datry et al. 2014). Despite the ubiquity of stream drying, perennial streams have historically dominated our understanding of stream chemistry patterns. Indeed, study and understanding of intermittent streams have lagged behind perennial ones (Costigan et al., 2016; Allen et al., 2020; Busch et al., 2020). For example, traditional stream chemistry studies have often focused on temporally intensive measurements at the outlet of a catchment and assumed that downstream patterns are representative of the upstream segment (e.g., Fisher and Likens, 1973; Schiff and Aravena, 1990; Boyer et al., 1997). However, this assumption may not be valid and the approach is unlikely to capture the spatiotemporal variability of intermittent streams (Godsey and Kirchner, 2014; Costigan
et al., 2016). To address this assumption, the scope of intermittent stream studies, which is determined by the extent (area or time-period over which measurements are conducted) and the grain (frequency of measurement through space or time) need to be adjusted (Schneider, 2001; Fausch et al., 2002). Indeed, high spatial-scope studies adopting a spatially continuous sampling approach (fine-grain) carried out over larger stream segment lengths (greater extent) have revealed previously undetected spatial patterns along the length of perennial streams and stream networks (Fausch et al., 2002; Likens and Buso, 2006; McGuire et al., 2014). Stream intermittence is characterized by contractions and disconnections that occur with drying at the scale of meters with consequences for stream structure at segment to network scales (Stanley et al., 1997; Dent and Grimm, 1999; Zimmer et al., 2013; Hale and Godsey, 2019; Jensen et al., 2019). Thus, temporally repeated high-spatial scope investigations seem necessary to understand patterns and mechanisms associated with stream intermittence.

Stream intermittency is projected to increase in extent, frequency, and duration as climate changes (Datry et al., 2014), with potentially important consequences for stream chemistry. In the western US, 80% of streams experience drying (USGS hydrographical database, 20061), and by 2050, intermittent streams and ephemeral streams (those that only experience event-based flow) are projected to increase by 5-7% (Döll and Schmied, 2012). In basins with moderate relief and elevation, such as the Snake, Great Salt Lake, and Oregon Closed Basins, snowmelt-fed streams are particularly susceptible to future stream drying as these areas may experience 100% loss of snow by 2050 (Klos et al., 2014). In most of the western US, projections include earlier snowmelt (Stewart et al.,

---

2004; Abatzoglou et al., 2014). Although watershed geomorphology and vegetation determine hydrologic responses to precipitation shifts (Tague and Grant, 2009; Godsey et al., 2014), potential climate change impacts to streams include early-season drying (Datry et al., 2014; Jaeger et al., 2014; Tennant et al., 2015), lower base flows, and increased groundwater sourcing (Tague and Grant, 2009; Godsey et al., 2014). Stream surface water chemistry patterns will likely reflect and respond to these anticipated changes, but the patterns and processes dominating these responses are complex because of potential feedbacks.

Coincident with changing snowmelt and low-flow patterns, wildfire is also increasing in its frequency and spatial extent in the western US (Abatzoglou and Williams, 2016; Parks et al., 2016), which alters land cover and influences stream chemistry (Hauer and Spencer, 1998; Mast et al., 2016). In particular, fire combusts vegetation and biotic soil components and releases mineral forms of nutrients and ions into soils (Murphy et al., 2006; Rau et al., 2007) that can be transported into streams (Spencer et al., 2003; Mast et al., 2016). Fire impacts to the water budget of a catchment can be substantial, especially in forested systems (Kinoshita and Hogue, 2015; Wine and Cadol, 2016; Atchley et al., 2018). Reduced terrestrial vegetation can lower evapotranspiration (ET) rates (Poon and Kinoshita, 2018), which can increase stream flow (Kinoshita and Hogue, 2015; Costigan et al., 2016; Wine and Cadol, 2016; Atchley et al., 2018), increase sediment yields (Subiza and Brand, 2018; Vega et al., 2020), and decrease terrestrial carbon inputs (Bixby et al., 2015; Cooper et al., 2015). Moreover, fire opens up the stream canopy increasing radiation, temperature, and wind (Naiman and Sedell, 1980; Poole and Berman, 2001; Bixby et al., 2015), which can increase
evaporation (Poole and Berman, 2001; Maheu et al., 2014) and in-stream primary production (Davis et al., 2013; Rugenski and Minshall, 2014; Cooper et al., 2015). To date, most studies have focused on forested watersheds with less attention paid to seasonally snow-dominated mountain areas where sagebrush steppe vegetation often predominates. In these ecosystems, fire regimes are shifting with the spread of invasive Bromus tectorum (cheatgrass; Bradley, 2009; Bradley et al., 2016) and increasing fire frequency and fuel availability (Link et al., 2006, Abatzoglou and Kolden, 2011). The few studies conducted in these ecosystems have rapid recovery of ET losses due to increased grass and herbaceous cover with only marginal and short-term effects on streamflow (Flerchinger et al. 2016, Fellows et al. 2018). Because stream drying and fire are accelerating, there is a critical need to understand how both phenomena will affect spatiotemporal patterns of stream surface water chemistry in these contexts.

Natural drivers of stream drying and stream chemistry patterns include increases in surface water evaporation (Brooks and Lemon, 2007; Gallo et al., 2012) or ET losses (Poon and Kinoshita, 2018; Warix, 2020) and/or changes in subsurface connectivity (Brooks et al., 2015; Costigan et al., 2016); fire may have consequences for these processes as well (Kinoshita and Hogue, 2015; Wine and Cadol, 2016; Atchley et al., 2018; Poon and Kinoshita, 2018). For example, increased in-stream evaporation may cause evapoconcentration, or an increased concentration of solutes as the amount of water decreases as found in hot desert streams at low flows (Brooks and Lemon, 2007; Gallo et al., 2012). Under low-flow conditions, evapoconcentration may occur in patches of open canopy stream that experience particularly high radiation, and contribute to overall patchiness in stream solute patterns. Fire may promote such processes by opening
stream canopies (Cooper et al., 2015), which can increase in-stream evaporative processes by increasing radiation and wind (Maheu et al., 2014) or increase streamflow by reducing evapotranspiration (Kinoshita and Hogue, 2015; Costigan et al., 2016; Wine and Cadol, 2016; Atchley et al., 2018; Poon and Kinoshita, 2018). Alternatively, patchiness may arise from subsurface processes like changing patterns of hillslope connectivity and water residence times that influence groundwater inputs of water and solutes (Zimmer and McGlynn, 2017; Dohman, 2019; van Meeryeld et al., 2019). Dynamic dissolved organic carbon (DOC) patterns in streams have been linked to springtime upland snowmelt and lateral flushing in alpine catchments of the Rocky Mountains (Boyer et al., 1997). However, the flushing process in the headwaters at Reynolds Creek Critical Zone Observatory (RC CZO) located in southwest Idaho was found to be slightly more complex (Radke et al., 2019) than Boyer et al.’s (1997) lateral flushing paradigm. At RC CZO, geophysical and hydrochemical evidence pointed to vertical and then lateral springtime flushing. The prior year’s soil-water and DOC flushed vertically to the saprolite during snowmelt and then laterally along this interface to the stream. DOC in the stream was primarily allochthonous, presumably sourced from DOC that had accumulated in the soil matrix during the dry summer rather than from in-stream processes (Radke et al., 2019).

As streams dry, connectivity to hillslope carbon sources may shift and in-stream primary production may also be patchy. The resulting stream chemistry reflects the relative magnitudes of these processes and whether these changes occur synchronously or not. Both autochthonous (in-stream) and allochthonous (terrestrially produced) sources of carbon may shift as hydrologic and biologic inputs change seasonally, with stream size,
or following fire (Minshall et al., 1989; Cooper et al., 2015). For example, high DOC concentrations can result from autochthonous carbon production during low flows in desert streams in Arizona (Jones et al., 1996; Brooks and Lemon, 2007). Similarly, in warm, humid systems in Tennessee, low flows and open canopy in autumn are associated with high DOC concentrations driven by increased autochthonous processing (Mulholland and Hill, 1997). Immediately following fire, stream algal blooms can result from elevated nutrient levels and open canopy that increase sunlight while also reducing the amount of available allochthonous carbon (Cooper et al., 2015). The chemical signatures of the organic carbon in the dissolved organic matter (DOM) can characterize organic carbon as autochthonous or allochthonous in origin. Fluorescence Index (FI) is a metric of spectral properties of fulvic acids component of DOM and absorbance coefficient (a_{254}[m^{-1}]) is a metric of carbon aromaticity; both tools help to evaluate autochthonous or allochthonous sourcing (McKnight et al., 2001; Brooks and Lemon, 2007; Inamdar et al., 2012). The connectivity of streams to the terrestrial surroundings may impact the variability of DOM sourcing. We expect that smaller headwater catchments may be more hydrologically connected to proximal terrestrial carbon (Hornberger et al., 1994; Brooks et al., 1999) and thus, more impacted by terrestrial processes (Creed et al., 2015). However, hillslope connections may also vary more in low-flow conditions and across the seasonal streamflow recession following snowmelt.

In this exploratory study, we measured spatial patterns of biogeochemistry in two mountainous, intermittent headwater streams as they dried. We sampled along the length of each stream, hereafter referred to as the longitudinal sampling or patterns. Sampling occurred following snowmelt through mid-summer in both streams: one unburned stream
and one recently burned stream. We used a spatially continuous sampling approach over 2.5 km repeated over three months during the growing season to assess changes in solute sourcing patterns and processes. We asked three questions: (a) How does surface stream chemistry vary longitudinally and temporally in intermittent streams as they begin to dry? (b) What are potential driving processes that can explain these patterns? and (c) How do these patterns and processes change following fire? We hypothesized stream chemistry would vary spatially downstream as a function of different physical and biological processes, that spatial patterns would shift temporally as a function of seasonal drying, and that these shifts would be more pronounced following fire. Specifically, we hypothesized that carbon concentrations would increase owing to in-stream evapoconcentration, groundwater connectivity, and in-stream primary production (Table 2.1). On the other hand, we hypothesized that nutrient concentrations would decrease owing to increased uptake with drying, with higher nitrogen losses following fire in the burned stream owing to an open canopy and high nitrogen availability (Table 2.1). To test our hypotheses, we performed an initial cluster analysis to evaluate similarity and dissimilarity amongst chemical properties including solute concentrations, water isotopes, and spectral indices. We then evaluated spatial and temporal patterns and tested hypotheses about associated processes using a combination of these properties that emerged as relevant to our questions.
<table>
<thead>
<tr>
<th>Study aims</th>
<th>Expectations as streams dry (April to June) for unburned and burned streams</th>
<th>Rationale</th>
</tr>
</thead>
</table>
| **Spatial** | Analytes will group into those that behave more conservatively and those that behave more reactively. Distinct longitudinal patterns will be observed, and these will differ between groups of analytes.  
*Approach:* solute concentrations, isotope, and spectral indices collected at high-spatial scope (50-m intervals over 2,500 m). Initial cluster analysis. | Source/sink dynamics will become more pronounced due to less surface water connectivity and increased local in-stream production and/or unique chemical signature of deeper groundwater sources with distinct flow path chemistry; this will be greater following fire when nutrients are high and uptake/biological production can be highly heterogeneous in the stream. |
| **Temporal** | Changes in the longitudinal spatial patterns and increases in conservative concentrations; decreases in bio-reactive solutes with greater shifts in burned stream.  
*Approach:* repeated solute concentration measurements across three months (April, May, and June) as stream shifts from snowmelt to low flow. Initial cluster analysis. | Higher conservative solutes concentrations during lower water flow due to decreased snowmelt and increased groundwater sourcing. Lower observed bio-reactive solute concentrations due to in-stream biotic uptake. |

*Study Aims.*
<table>
<thead>
<tr>
<th>Processes and mechanisms tested</th>
<th>Physical hydrological</th>
<th>Biological</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-stream evaporation, evapoconcentration, dilution</td>
<td>Autochthonous or allochthonous carbon sourcing</td>
</tr>
<tr>
<td><strong>Temporal shifts</strong></td>
<td>Temporal shifts with heavier stable water isotopes later in the season and increased evidence of in-stream evapoconcentration, greater shifts in burned stream. <em>Analyses</em>: stable water isotopes slopes, source/sink chloride dynamics.</td>
<td>As growing season progresses and water velocity flows, nutrients will be more labile and in-stream production will contribute.</td>
</tr>
<tr>
<td><strong>As ambient air temperature</strong></td>
<td>As ambient air temperature increases, evaporative processes will increase including evapoconcentration, especially in burned stream where an open canopy offers no sun or wind protection.</td>
<td>As snow melt diminishes, water will shift from shallow subsurface water to deeper groundwater.</td>
</tr>
</tbody>
</table>

Notes: Outlines the study framework and aims (left two columns under “Study aims” in bold outline), expectations, and rationale for stated explanations. The approach and method employed to observe patterns and test processes are described under each expectation.
2.2 Methods

2.2.1 Experimental Design

2.2.1.2 Study Site. Our study took place at the Reynolds Creek Experimental Watershed (RCEW) and RC CZO, a 239 km² watershed located southwest of Boise, ID, USA. The USDA Agricultural Research Service (ARS) established RCEW in 1960 as an experimental watershed representative of the Intermountain West region in the United States (Marks et al., 2011), and the ARS has monitored long-term precipitation and stream discharge trends (Seyfried et al., 2000, 2018). The RCEW extends over a steep climatic gradient with mean annual precipitation varying from 250 to 1100 mm/yr and mean annual temperatures from 5.5 °C to 11°C. This climatic variability is driven by the nearly 1000 m elevation range. At lower elevations, rain is the dominant form of precipitation in the RCEW whereas snow is dominant at the highest elevations (Nayak et al., 2010, Kormos et al., 2014). Peak stream discharges across the watershed are driven by snowmelt patterns (Pierson et al., 2001). Vegetation includes Wyoming sagebrush steppe in the lower elevations, transitioning to mountain big sagebrush (Artemisia tridentata), antelope bitterbrush (Purshia tridentata), rabbitbrush (Chrysothamnus viscidiflorus), western juniper (Juniperus occidentalis), aspen (Populus tremulodes) and coniferous forest (mostly Douglas fir (Pseudotsuga menziesii)), at higher elevations (Seyfried et al., 2018). The site is underlain by Cretaceous granites and volcanic rocks such as the Miocene Salmon Creek Volcanics (McIntyre, 1972).

To evaluate spatial and temporal variation in stream chemistry, we studied two headwater streams within the RCEW known to be intermittent, Johnston Draw (basin
area 1.83 km$^2$) and Murphy Creek (basin area 1.32 km$^2$; Seyfried et al., 2000; Pierson et al., 2001; Patton et al., 2018). Johnston Draw was ideal for this study owing to the availability of streamflow data measured at a dropbox v-notch 90° weir (Seyfried et al., 2000; Godsey et al., 2018). We opportunistically studied Murphy Creek to evaluate possible extremes in spatial and temporal variation in stream chemistry following wildfire. Specifically, the Soda Fire burned 68 km$^2$ of RCEW in August 2015, including Murphy Creek (Vega et al., 2020). Briefly, the fire was classified as moderate severity in the study area and consumed nearly all above-ground live vegetation resulting in more than 60% bare ground (bare soil, ash, and rock), with high sediment delivery to the stream (Vega et al., 2020). In October 2015, a dropbox v-notch 90° weir with pressure transducer for discharge measurements was re-activated in Murphy Creek. Hereafter we refer to Johnston Draw as “unburned JD” and Murphy Creek as “burned MC". Unburned JD and burned MC are comparable in basin area, discharge, elevation, mean annual precipitation (MAP), and aspect (Seyfried et al., 2000; Table 2.2). Annual water yields for 2016 were lower than the long-term average for both watersheds (Glossner, 2019). Though some studies find increased water yield after fire due to loss of vegetation (Atchley et al., 2018), in burned MC, the 2016 water year represented the third lowest streamflow year out of 12 years (1968-1977 and 2016-2017) of RC CZO data (Vega et al., 2020). The major differences between the two catchments are recent fire history and lithology (Table 2.2). The lithology of unburned JD is predominantly granodiorite with minor lithologic discontinuities including a combination of quartz latite and rhyolite flows underlying the high plateau at the top of the catchment, and a small olivine-rich basalt flow that extends into the outflow (McIntyre, 1972). Basalt and Salmon Creek
Volcanics, underlies the burned MC. Unburned JD is described in more detail by Godsey et al. (2018) and Patton et al. (2018, 2019) and burned MC is described in more detail in Seyfried et al. (2000), Pierson et al. (2001), and Vega et al. (2020).

Table 2.2

*Characteristics of the Unburned JD and Burned MC Sub-Catchments.*

<table>
<thead>
<tr>
<th></th>
<th>Unburned JD</th>
<th>Burned MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drainage area (km²)</td>
<td>1.8</td>
<td>1.32</td>
</tr>
<tr>
<td>Streamflow direction</td>
<td>East with upper ~500 m south</td>
<td>East</td>
</tr>
<tr>
<td>Mean annual discharge (Q, m³/s)</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>Weir type</td>
<td>Dropbox v-notch 90°</td>
<td>Dropbox v-notch 90°</td>
</tr>
<tr>
<td>Elevation (m)</td>
<td>1490-1850</td>
<td>1383-1822</td>
</tr>
<tr>
<td>MAP (mm/year)</td>
<td>550</td>
<td>500-600</td>
</tr>
<tr>
<td>Lithology</td>
<td>Granodiorite</td>
<td>Salmon Creek Volcanics and basalt</td>
</tr>
<tr>
<td>2015 Fire</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Primary vegetation</td>
<td>Mountain big sagebrush, Juniper, alder, aspen, bunch grasses</td>
<td>(pre fire) Mountain big sagebrush, antelope bitterbrush, rabbitbrush, bunch grasses</td>
</tr>
</tbody>
</table>
2.2.1.2 Sampling Design. To investigate spatiotemporal chemistry patterns in unburned JD and burned MC, we used a high-spatial scope approach: temporally repeated high-spatial density measurements at the km extent. In April, May, and June 2016, we sampled each stream at 50 m intervals (the grain of our study) over ~2,500 m distance (the extent of our study; Figure 2.1). Each 50 m site is referred to as a reach site (blue dots in Figure 2.1) and the reach sites span over ~2500 m distance, the extent, which we refer to as the stream segment (Figure 2.1). In each stream segment, sampling started at the weir and continued to the uppermost observed surface flows (Figure 2.1) to include the entire sub-catchment. We refer to the suite of 50 m reach sites collectively as the stream longitudinal profile, and the spatial patterns identified using these sites as longitudinal patterns. Water samples were collected in pools at the stream thalweg following Dent and Grimm (1999).
Figure 2.1

2016 Study Design at Reynolds Creek Experimental Watershed Critical Zone Observatory.

Notes: (A) Reynolds Creek Critical Zone Observatory (RC CZO) in southwestern, Idaho, USA (yellow dot). Outlined in turquoise are two headwater sub-catchments, unburned JD and burned MC, which are nested within the RC CZO, the perimeter of which is outlined in yellow. (B) shows the study scope with the unburned JD (lower) and burned MC (upper) longitudinal sampling grain of 50 m interval, or stream reach sites (blue circles), spanning ~2500 m extent. Longitudinal sampling sites start at a weir (green star) and extend to the uppermost reach with observed surface flow in each stream (red “x”), as identified in the field in April 2016. Colors indicate lithology based on McIntyre (1972).

2.2.2 Biogeochemical Characterization of Intermittent Streams

2.2.2.1 Physical and Chemical Field Properties. We characterized each stream reach site by measuring a suite of in situ properties including temperature (°C), pH, dissolved oxygen (DO mg/L), and estimated canopy cover. Water temperature (°C) and
pH were measured at each stream reach site using an Oakton pH 110 Series probe (Vernon Hills, IL) calibrated with 4, 7 and 10 pH standard solutions; DO was similarly measured at each reach site with a YSI dissolved oxygen probe (Burlington, VT). Canopy cover was visually estimated in July at half of the stream sites using upward-oriented fisheye lens photos at pool height. Maximum canopy coverage was estimated at alternating sites from photos taken during full leaf-out. Canopy coverage was categorized based on the proportion of the photo covered by foliage (0-25%, 25-50%, 50-75%, and 75-100%). Discharge (m3/s) was calculated from continuous stage measurements using a stage-discharge rating curve only at the outlet weir of each stream segment (Figure 2.1, B). We acknowledge that direct measures of discharge and associated water balance at stream reach site would have augmented our understanding of stream dynamics and local water balance, but this was beyond this study. Accurately measuring discharge throughout a stream network at low flows is methodologically challenging and time-consuming. Instead, we focus here on concentration patterns and use indirect and environmental tracer approaches described below to understand inputs to and outputs from each reach site.

2.2.2 Surface Water Collection. To establish biogeochemical patterns in each stream, we collected surface water samples in the field and then filtered and analyzed them in the lab for the concentrations of a suite of chemical constituents. In the field, samples were collected in 250 mL amber high-density polyethylene bottles (HDPE) bottles, which had been rinsed three times and pre-leached in 18.2 MOhm distilled (DI) water. Bottles were then rinsed three times in the field with the stream sample and then filled to eliminate headspace. Samples were transported by backpack out of the steep
terrain (~35-60 total per stream per sample period), and the water samples were kept refrigerated (4 °C) until filtered and analyzed. For carbon and total nitrogen (TN) analysis, water samples were filtered within 72 hours (typically within 24 hours) in the lab using vacuum filtration through pre-combusted 0.7µm Whatman glass fiber filters (GFF) into DI- and sample-rinsed 60 ml amber HDPE bottles. Amber HDPE bottles were used instead of glass owing to transport hazards (Sanderman et al., 2009). The remaining sample was syringe-filtered through a 0.45 µm Puradisc nylon filter into four 60 ml clear HDPE bottles for analysis of nutrients, anions, cations, isotopes, and fluorescence index. Samples were collected moving from downstream to upstream to minimize disturbance.

2.2.2.3 Laboratory Analysis. Stream chemistry was analyzed for a suite of chemical constituents, including dissolved inorganic carbon (DIC), DOC, TN, nutrients including ammonium-N, nitrate (NO₃-N), orthophosphate (PO₄³⁻-P), anions including chloride (Cl⁻) and sulfate (SO₄²⁻), and cation concentrations including base cations and rare earth elements such as strontium (Sr). In addition, we analyzed stream water for stable water isotopes (δ¹⁸O and δ²H), fluorescence index (FI), and absorbance coefficient (a₂₅⁴ [m⁻¹]). FI and a₂₅⁴ methods are discussed in the DOC sourcing section. DIC, DOC and TN concentrations were measured on a Shimadzu TOC-V CSH (Columbia, MA, USA) equipped with an ASI-V autosampler and TNM-1 chemiluminescence detector for TN. Errors <2-3% were accepted for concentrations >1 mg C/L. High DIC concentrations (~10-30 mg DIC/L as C) resulted in difficulty using the non-purgeable organic carbon method (NPOC). Thus, we calculated DOC as the difference between TC and DIC and propagated the errors associated with this calculation (~0.3-0.9 mg/L). Cross validation of our DOC values with the Perdrial Environmental Biogeochemistry Lab at the
University of Vermont (Burlington, VT) showed reasonable agreement (n = 140, r = 0.65). Owing to the higher error associated with DOC by difference, we approached DOC concentrations with some caution but retained them because DOC and spectral characteristics showed similarly variable patterns, and DOC provided an interesting contrast to other nutrients and DIC. Nutrients were measured on a Westco Discrete Analyzer (Unity Scientific, Brookfield, CT, USA), an automated chemical spectrophotometer. We accepted a <10% error for NO$_3^-$-N and PO$_4^{3-}$-P concentrations <1 mg/L and a 20% error for NH$_4^+$-N <0.10 mg/L. For most of the analyses, we utilize TN because both NO$_3^-$ and NH$_4^+$ were below detection limit, with the exception of early season samples and those from burned MC. Anions Cl$^-$ and SO$_4^{2-}$ were analyzed on a Dionex ion chromatograph (ICS-5000, Sunnyvale, CA, USA) with AS18 column 4 X 250 mm, and we accepted <10% error for sample <1 mg/L and 2-3% for samples > 1 mg/L. Cations were measured by the Center for Archaeology, Materials and Applied Spectroscopy (CAMAS) lab (Idaho State University, Pocatello, ID) on a Thermo X-II 283 series Inductively Coupled Plasma Mass Spectrometer (ICP-MS) equipped with a Cetac 240-position 284 liquid autosampler (ThermoFisher Scientific). Dilutions were 1:10 sample to de-ionized water. Some cations were below the detection limit of 10 ppb. We reliably measured Sr, Ba, Na, Mg, Al, Si, K, Ca, Sc, Ti, Fe, and Zn for all three months. Lastly, stable water isotope samples were analyzed at ISU/CAMAS Stable Isotope Laboratory on a Thermo Scientific, High Temperature Conversion Elemental Analyzer (TC-EA) interfaced to a Delta V Advantage mass spectrometer. Precision for both δ$^{18}$O and δ$^2$H was better than ±0.2‰ and ±2.00‰ respectively.
Lastly, autochthonous and allochthonous DOC sourcing was investigated by analyzing the spectral characteristics of DOM in the lab, specifically Fluorescence Index (FI) and absorbance. Completed by the Perdrial Environmental Biogeochemistry Lab at the University of Vermont (Burlington, VT), these characteristics were measured using the Aqualog Fluorescence and Absorbance Spectrometer (Horiba, Irvine, CA, USA). The excitation (EX) wavelength range spanned from 250-600 nm (increment 3nm) and emission (EM) ranged from 212-619 nm (increment 3.34 nm). All excitation emission matrices (EEMs) were blank-subtracted (nanopure water, resistivity 18MΩ cm⁻¹), corrected for inner filter effects, and Raman-normalized (Ohno, 2002; Miller et al., 2010). All samples were diluted to absorbance values below 0.3 and we computed relevant indices, including the FI (calculated as the intensity at Emission 470 nm divided by the intensity at Emission 520 nm for Excitation at 370 nm (Cory & McKnight, 2005)). Absorbance (a) at 254 nm can also be used as a direct measure of aromaticity of DOM. We report the absorption coefficient (a₂₅₄ [m⁻¹]) which is calculated independently of DOC concentrations and using Equation 2.1 (Inamdar et al., 2012; Green and Blough, 1994) below:

Equation 2.1: \[ a_{254} [m^{-1}] = (UV \text{ absorbance at 254 nm}) \times 2.303 \times 100. \]

Even though specific UV absorbance (SUVA₂₅₄) is the most commonly used indicator for DOM aromaticity (Weishaar et al., 2003), we refrained from its use due to the high error in DOC measurements and report \( a_{254} [m^{-1}] \) instead.
2.2.3 Spatiotemporal Pattern Analyses

2.2.3.1 Initial Clustering of Water Properties and Biogeochemistry. To assess possible common synchronous or asynchronous drivers of biogeochemistry, we initially identified chemical variables within each stream that were strongly and weakly related over time using cluster analysis. We explored highly correlated groupings among chemical variables described above using cluster variable analysis for each month in order to understand a given variable’s relationship with other variables over time. The loaded variables were the chemical properties and analyte concentrations measured in the field and lab including temperature, pH, DO, carbon (DIC and DOC), anions (Cl⁻), nutrients (TN and PO₄³⁻), all cations measurable across the three months (Sr, Ba, Na, Mg, Al, Si, K, Ca, Sc, Ti, Fe, and Zn), and stable water isotopes (δ¹⁸O and δ²H). The cluster variable analyses were conducted in JMP (version 14.2) which first grouped variables into cluster components determined by first and second principal components and then ranked variables within cluster components based on the strength determined by a cluster algorithm (SAS Institute Inc., 2017). This algorithm iteratively clusters variables based on a combination of strength of the principal component and relative eigenvalue strength. The order in which the analytes and chemical properties are reported within each cluster is based on the highest Pearson correlation coefficient (r) within the cluster. The results of these analyses informed our choice of analytes used for longitudinal pattern reporting and testing our specific process hypotheses. We chose analytes that juxtapose each other by selecting both highly ranked variables within clusters of strongly temporally correlated variables as well as analytes that were not strongly clustered. We did this in order to represent a spectrum of more conservative solutes such as Cl⁻ to more biological
reactive ones such as TN in each stream. The above criteria supported reporting longitudinal patterns for DIC, Cl\(^-\), Sr, DOC, TN, and PO\(_4^{3-}\). However, we assessed temporal patterns from April to June in all chemical properties and solutes as well.

2.2.3.2 Longitudinal Spatial Patterns and Temporal Tendencies. To explore the biogeochemical patterns both along the length of the stream and monthly changes as the stream dried, we plotted the longitudinal patterns of all solutes for each month, but focused on those identified above in the cluster analysis: DIC, Cl\(^-\), Sr, DOC, TN, and PO\(_4^{3-}\). These analytes also represent more conservative analytes and more reactive analytes as defined by Baker and Webster (2017). Stream data collected longitudinally from non-perennial streams presented challenges. First, samples collected 50 m apart in a stream are not statistically independent, second as streams dried April to June, it was impossible to sample in some locations and stream sample sizes changed, and third underlying longitudinal patterns complicated typical methods of quantifying spatial heterogeneity (details in statistical challenges of stream drying data discussion section). Overall temporal comparisons of chemical properties, solutes, and cations from high to low flow were analyzed using mean differences and propagated errors from April to June within each stream. We determined differences to be substantial when they were larger than errors. As detailed above (Table 2.1), we expected to observe longitudinal and temporal similarities in both streams from April to June and predicted higher concentrations of carbon with evapoconcentration, lower concentrations of nutrients as in-stream primary production seasonally increased, and increasing cation concentrations as both streams became increasingly groundwater-dominated. However, we expected
differences between the two streams like higher TN due to fire and higher cation loss due to volcanic lithology in burned MC.

2.2.4 Explaining Patterns: Testing Hypothesized Processes

2.2.4.1 In-Stream Evaporative Processes. The potential evaporative signature of stream water was assessed by plotting stable water isotopes $\delta^{18}$O and $\delta^2$H against the Global Meteoric Water Line (GMWL) as well as the Local Meteoric Water Line (LWML; $\delta^2$H=7.1($\delta^{18}$O)-6.3), with data derived by Tappa et al. (2016). We compared monthly $\delta^{18}$O and $\delta^2$H data within each stream to the LMWL and calculated the monthly best-fit slopes based on simple linear regression and the LMWL-intercepts. To test for statistical differences between months and streams, we ran t-tests on the slopes of the regressed isotopic relationships. We hypothesized that the streams would undergo evaporation with drying in this semi-arid system (Fritz and Clark, 1997), and thus, that streams’ isotopic data would fall below the LMWL (Gat, 1996). In addition, we predicted both streams would exhibit an increasingly evaporative signature as the streams dried, with a stronger shift in burned MC.

We evaluated the role of in-stream evapoconcentration in explaining chemistry patterns associated with stream drying by comparing the behavior of $\delta^{18}$O to Cl$^-$ between upstream and downstream sites. We calculated predicted downstream values of Cl$^-$ concentration ($[Cl^-]$) based on ratios of $\delta^{18}$O from upstream (subscript “U” in equation 2.2) and downstream (subscript “D”) samples and the observed upstream Cl$^-$ ($[Cl^U]$) values (Mulholland and Hill, 1997; Gallo et al., 2012):

$$[Cl^-] = \frac{[Cl^U][\delta^{18}O_D]}{[\delta^{18}O_U]}$$
Because both $\delta^{18}$O and Cl$^-$ are expected to be conservative tracers (Baker and Webster, 2017), equation 2 should lead to accurate predictions of Cl$^-$ and each downstream site unless evapoconcentration or flushing were dominant processes. Therefore, we compared the predicted [$Cl_{D}$] and observed Cl$^-$ concentrations for each downstream site. If evapoconcentration were driving patterns in the water samples, we would expect to observe differences in the predicted versus observed Cl$^-$ concentrations. To explain any differences, we would need to parse out the role of flushing versus evapoconcentration (Gallo et al., 2012).

**2.2.4.2 Groundwater influence.** We used water isotopes and the Sr/DIC ratio to evaluate the potential influence of groundwater as streams dried. First, we calculated unburned JD’s and burned MC’s isotopic intercepts with the LMWL (Tappa et al., 2016) using each monthly regression (Fritz and Clark, 1997) and evaluated temporal shifts. We then compared the stream’s LMWL-intercepts to the mean rain, snow, and groundwater samples collected by Radke et al. (2019) in a nearby but higher elevation watershed within RC CZO (Reynolds Mountain East ~2100 m elevation). In this way, isotopic means of rain ($\delta^2$H -66.5 ± 12.7 and $\delta^{18}$O -8.4 ± 2.4, mean ± SE), snow ($\delta^2$H -129 ± 10.2 and $\delta^{18}$O -16.8 ± 1.3, mean ± SE) and groundwater samples ($\delta^2$H -121 ± 1.04 and $\delta^{18}$O -16.4 ± 0.18, mean ± SE) were assumed to be reasonable endmembers for our sites as well. As the streams dried, we expected the stream-LMWL intercepts to shift towards the groundwater signature later in the season if groundwater made up a larger component of the streams’ surface water.

We also plotted Sr against DIC for each month because a previous study showed Sr to be a good indicator of groundwater at the RC CZO (Radke et al., 2019). We
calculated RC CZO groundwater and precipitation (rain and snow) averages from samples collected by Radke et al. (2019) in late summer (August 2017), and used these values as endmembers in our analysis. Sr averaged 142.74 ± 8.77 ppb (mean ± SE) and DIC averaged 21.20 ± 0.55 mg C/L (mean ± SE). Rain and snow water samples had average values of 1.37 ± 1.63 ppb (mean ± SE) and 1.54 ± 1.80 mg C/L (mean ± SE) for Sr and DIC, respectively (Radke et al., 2019). We compared the upstream/downstream Sr-DIC relationship by comparing the uppermost 1,000 m sites with those in the lowest 1,000 m. We expected to observe a shift towards the groundwater ratio reflected in the late summer Sr/DIC ratios in unburned JD and burned MC that would be consistent with the shift in isotopic signatures.

**2.2.4.3 Autochthonous or allochthonous DOC sourcing.** Allochthonous carbon can be spectrally distinguished from autochthonous due to the structural carbon components in terrestrial vegetation, like lignin, which are also more aromatic. FI and \( a_{254} \) measure these spectral components and help classify carbon sources as allochthonous or autochthonous (McKnight et al., 2001; Inamdar et al., 2012). We compared FI to published ranges of these values associated with more autochthonous or allochthonous sources (Inamdar et al., 2012). Based on these values, we interpreted FI values between 1.2-1.5 as reflecting an allochthonous signature whereas FI above 1.7 reflects an autochthonous/microbial signature (McKnight et al., 2001). Furthermore, we report the variability using coefficient of variation (CV) because FI had no underlying longitudinal patterns in either stream (see discussion section, “2.4.3.2. Statistical Challenges of Stream Drying Data”). However, in our calculations, we controlled for
sample size in each stream by including only those sites that had water flowing all three months so as not to bias any one month.

Aromatic and humic DOM is a product of terrestrial vegetation so that a higher aromatic $a_{254}$ signature reflects a more allochthonous DOM with structural carbon. Thus, the relationship between FI and $a_{254}$ is inverted, such that lower $a_{254}$ result from less aromatic DOM, more autochthonous/microbial derived carbon and higher $a_{254}$ result from more aromatic, more allochthonous derived carbon (Inamdar et al., 2012). Due to increased in-stream primary production as both streams dried, we expected to observe FI to increase reflecting shifts from more allochthonous to autochthonous carbon, and $a_{254}$ to decrease reflecting diminished aromatic properties with a more pronounced shift in burned MC.

2.3 Results

2.3.1 Biogeochemical characterization of intermittent streams

2.3.1.1 Stream hydrologic conditions. Seasonal drying in both intermittent headwater streams was similar, with weir discharges (Q) decreasing in April, May, and June 2016 (Figure 2.2, A). Drying occurred in both streams between April and June. For unburned JD, all of the April sampling sites (n=57) had surface flow, 89% persisted in May (n=51), and only 65% sustained surface flows in June (n=37). Drying occurred in reaches interspersed throughout the unburned JD stream. Similarly, all of the reach sites in burned MC were flowing in April (n=59), but only 73% in May (n=43), and 64% sustained flow in June (n=38). Unlike in unburned JD, drying in burned MC occurred from the top of the stream to the bottom as though the entire stream contracted
longitudinally (though subsequent work showed this pattern did not persist later in the summer; Warix, 2020). Thus, by June, 35% of unburned JD’s and 36% of burned MC’s reach sites had dried, with the highest proportion of sites drying from May to June in unburned JD stream and from April to May in burned MC stream. Over the sample period, unburned JD stream discharge exhibited a slightly greater relative change (April Q ~ 0.022 m$^3$/s to June Q <0.005 m$^3$/s) than burned MC (April Q ~ 0.017 m$^3$/s to June Q ~0.005 m$^3$/s; Figure 2.2, A). Burned MC also exhibited flashier stream flow as spikes in discharge, presumably responding to event-based precipitation or snowmelt inputs (Figure 2.2, A) that did not occur as strongly in unburned JD.

Figure 2.2

2016 Weir Discharge, Temp (ºC), pH, and DO (mg/L).

Notes: 2016 discharge (A) of the unburned JD (blue) and burned MC (orange) measured at the weir (most downstream sample point); symbols reflect the discharge on the sample
dates. Longitudinal patterns of (B and C) temperature (°C) and estimated canopy cover (% cover in grey bars), (D and E) pH (-log [H+]), and (F and G) dissolved oxygen (DO mg/L).

2.3.1.2 Physical and chemical field properties. The streams were similar in physical and chemical properties measured in the field but as expected, canopy cover differed between the unburned stream and burned stream. Stream temperature showed strong longitudinal and increasing temporal shifts in both streams, whereas pH and DO were more variable longitudinally and temporally (Figure 2.2, B-G and Supplementary Table 2.1). The two streams differed in canopy cover in that 44% of unburned JD had >25% canopy cover during full leaf-out whereas 100% of sites sampled in burned MC had <25% canopy cover, and most sites fell very close to 0% canopy cover even in July 2016 (Figure 2.2, B, C). Although some burned MC sites had >0% canopy cover in July, the post-fire vegetation differed from the vegetation in the unburned JD. The riparian vegetation of burned MC in July was dominated by fast-growing, non-woody shrubs and small, young willows because all vegetation, namely sagebrush and bunch grasses were charred to the ground (Vega et al., 2020). This contrasted with the overhanging riparian alder, willow, and juniper in unburned JD.

2.3.2 Biogeochemical pattern analysis

2.3.2.1 Initial clustering of water properties and biogeochemistry.
Correlations between stream analytes varied between streams and these relationships changed as each stream dried. We identified groups of solutes that behaved similarly or dissimilarly for each stream at different moments in the season. For example, in unburned JD, we observed positive correlations between Sr, Mg, Ca, K, Na, Ba, Fe, DOC, temperature (Temp), DIC, δ18O, TN, Cl⁻, Sc, and Si in April. Ti, Al, DO, pH, PO4³⁻, and
Zn were negatively correlated with these analytes and chemical properties, but positively correlated with each other (Figure 2.3, A). By June, Mg, Fe, Cl, K, Temp, TN, δ¹⁸O, Sc, and Si were positively correlated with each other, but shifted to being negatively correlated with other analytes or chemical properties. From April to June, shifts from negatively to positively correlated analytes were observed between Ti, DO, pH, and PO₄³⁻ and other analytes or properties in unburned JD (Figure 2.3, E). We observed similar (though not identical) patterns within burned MC (Figure 2.3, B, D, and E). Strongly correlated (either positively or negatively) analytes were placed into clusters for each month and stream (Supplementary Table 2.3). Based on these results, we selected representative analytes to illustrate different clusters and behaviors for the longitudinal analysis, including DIC, DOC, TN, PO₄³⁻, Sr and Cl⁻.
Figure 2.3

Initial Cluster Analysis.

Notes: Cluster analysis shows correlations between analytes by rows for April (panels A and B), May (C and D), and June (E and F) for unburned JD (left column) and burned
MC (right column). Analytes include: temperature, or Temp, (°C), DO (mg/L), pH, DIC (mg C/L), Cl$^-$ (mg Cl/L), DOC (mg C/L), TN (mg N/L), PO$_4^{3-}$ (mg P/L), $\delta^{18}$O, and cations (ppb): Sr, Ba, Na, Mg, Al, Si, K, Ca, Sc, Ti, Fe, and Zn. Analytes are labeled across the top of each x-axis and are listed in the same order along the y-axis in each panel. Analyte order is based on cluster analysis groupings where analytes within each cluster are more strongly correlated (higher Pearson correlation coefficient or r-value) to each other than to other groups of analytes. The r-values are colored by saturation where the higher the r-value, the more saturated the color, and grey represents a low r-value and low correlation. Red indicates positive correlations and blue negative correlations.

### 2.3.2.2 Longitudinal spatial patterns

Common longitudinal patterns were evident in both unburned JD and burned MC streams (Figure 2.4) for many of the more conservative analytes (e.g., DIC, Sr, and Cl$^-$) than the biologically reactive ones (e.g., DOC, TN, and PO$_4^{3-}$). As expected, mean concentrations generally increased for DIC, Sr, and Cl$^-$ downstream (Figure 2.4, A-F). The exception to this pattern were consistently high concentrations at the top of unburned JD (relative to downstream concentrations) at reach sites that were south-facing and meadow-like. Concentrations of more reactive analytes, DOC, TN, and PO$_4^{3-}$, tended to decrease downstream in unburned JD whereas DOC and PO$_4^{3-}$ were relatively invariant longitudinally in the burned MC stream. TN was exceptional in that it varied in the uppermost portions of both streams with patterns that changed each month of the sampling campaign.
Notes: Longitudinal patterns of (A and B) dissolved inorganic carbon (DIC mg C/L), (C and D) strontium (Sr ppb), (E and F) chloride (Cl⁻ mg Cl/L), (G and H) dissolved organic carbon (DOC mg C/L), (I) Total Nitrogen (TN) mg N/L, and (K) orthophosphate (PO₄³⁻) mg P/L.
carbon (DOC mg C/L), (I and J) total nitrogen (TN mg N/L), (K and L) phosphate (PO$_4^{3-}$ mg P/L) for unburned JD (blue, left column) and burned MC (orange, right column) streams. Values from April are represented by circles, May sites are diamonds, and June sites are squares along the length of the stream where 0 m is the weir and ~2,500 m is the upper stream reach.

2.3.2.3 Temporal shifts. As expected, mean differences in concentrations between April and June showed that TN concentrations substantially decreased over the growing season in each stream, and by an order of magnitude in burned MC (Figure 2.5, B). Mean TN concentration decreased by >80% from April to June (Figure 2.5, B) in burned MC stream (1.14 ± 0.03 to 0.19 ± 0.02 mg N/L, mean ± SE); TN was composed mostly of NO$_3^-$; even in April, NH$_4^+$ was largely below our instrument detection limit. We observed a similar but less pronounced decrease in TN in unburned JD (Figure 2.5, A), where a 39% decrease occurred over the same months (0.36 ± 0.04 to 0.22 ± 0.03 mg N/L, mean ± SE). Unlike TN, PO$_4^{3-}$ decreased only slightly in the burned MC stream over the growing season and did not exhibit a measurable change in unburned JD (Figure 2.5, A, B and Supplementary Table 2.3). In contrast to our expectations, DOC exhibited no clear temporal shift from April to June in unburned JD and slightly decreased in burned MC (Figure 2.5).

In both streams, we observed similar temporal patterns in the average differences from April to June for base cations and other solutes (Figure 2.5, C, D). Like DIC and Cl$^-$, most cations increased from April to June in both streams as indicated by negative difference values between April and June mean concentrations (Figure 2.5, C, D, and Supplementary Table 2.4). However, in unburned JD, some cation concentrations decreased as indicated by positive difference values as was the case for Si, Fe, and Ba. As expected, burned MC exhibited three to ten times larger cation differences than unburned
JD for several cations including Na, Mg, Si, K, and Ca (Figure 2.5 C, D, and Supplementary Table 2.4).

Figure 2.5

*Analyte Differences from April to June.*

Notes: Mean solute differences from April to June are shown for (A) unburned JD and (B) burned MC for temperature, or Temp, (°C), pH (-log [H+]), DO (mg/L), DIC (mg C/L), DOC (mg C/L), TN (mg N/L), PO₄³⁻(mg P/L), Cl⁻(mg Cl/L), DOC (mg C/L), TN (mg N/L), PO₄³⁻(mg P/L). Mean cation differences (ppb) from April to June are shown for unburned JD (C) and burned MC (D). Error bars show propagated error and asterisks indicate where differences are greater than the propagated error. Positive differences indicate decreasing in concentration from April to June and negative differences indicate increasing concentration from April to June (arrows for reference). Note pH measures are representative of log scale, so small changes are not equivalent to raw concentrations of other solutes. Additionally, please note the difference in axes scales between panel (C) and (D).
2.3.3 Explaining patterns: testing hypothesized processes

2.3.3.1 Evaporative processes. Isotopic signatures for both streams showed that evaporation was occurring, which we expected. However, counter to our predictions, we did not find evidence of temporal shifts towards increasing evaporation with stream drying in either stream. The regression slopes of the $\delta^{18}$O-$\delta^{2}$H relationships for all months for both unburned JD and burned MC fell below the GMWL and LMWL (Tappa et al., 2016), indicating an evaporative signature in these two streams compared to local precipitation (Figure 6, A, B). In the unburned JD, we observed a lower-sloped evaporative trend with drying as April’s slope ($\pm$ SE) was 2.7 ± 0.16, May’s was 3.6 ± 0.27, and June’s slope 4.7 ± 0.30 (Table 2.3). All months within unburned JD were statistically different so that April differed from May ($t=-5.2$, $p=<0.0001$, $df=103$), May differed from June ($t=-4.4$, $p=3.7E-05$, $df=81$), and April from June ($t=-14.14$, $p=<0.0001$, $df=84$). In the burned MC stream, April’s slope ($\pm$ SE) was 4.3 ± 0.30, May’s was 3.0 ± 0.26, and June’s was 4.5 ± 0.23 (Table 2.3). In contrast to unburned JD, burned MC slopes for April and June did not statistically differ from each other ($t=-0.51$, $p=0.61$, $df=91$), although both differed from May ($t=-2.6$, $p=0.01$, $df=95$ and $t=-5.1$, $p=<0.0001$, $df=76$, respectively) which had the most evaporative signature over the study period.

Between the two streams, slopes were only significantly different from each other in April ($t=-10.24$, $p=<0.0001$, $df=108$) and not in May ($t=1.0$, $p=0.32$, $df=89$) or June ($t=0.6$, $p=0.56$, $df=70$). We generally did not observe longitudinal patterns of evaporation (Figure 2.6, A, B), with the exception of the uppermost portion of unburned JD. There, the aspect and vegetative cover of the stream was distinct from downstream reaches:
these south-facing meadow sites had a strong evaporative signature in April (Figure 2.6, A) and had dried by May.

Figure 2.6

In-Stream Evaporative Processes and Groundwater Analyses.

Notes: Stable water isotopes ($\delta^{18}$O and $\delta^2$H) for sites in both unburned JD (A) and burned MC (B) in April (light circles), May (mid-tone diamonds), and June (dark squares). The global meteoric water line (GMWL) is shown in solid black and the local meteoric water line (LMWL) is the dashed red line ($\delta^2$H=7.1($\delta^{18}$O)-6.3; Tappa et al., 2016). Mean rain (pink diamond), mean snow (yellow triangle), and mean well values (green triangle) samples were collected by Radke et al. (2019) at a higher elevation site in Reynolds Creek watershed. Source-sink dynamics of measured and predicted Cl$^+$ concentrations calculated using $\delta^{18}$O upstream and downstream site ratios for the unburned JD (C) and

---

**A**

Unburned JD

- April
- May
- June
- Meadow
- Well
- Rain
- Snow
- GWML
- LMWL

**B**

Burned MC

- April
- May
- June

**C**

Stable water isotopes ($\delta^{18}$O and $\delta^2$H) for sites in both unburned JD (A) and burned MC (B) in April (light circles), May (mid-tone diamonds), and June (dark squares). The global meteoric water line (GMWL) is shown in solid black and the local meteoric water line (LMWL) is the dashed red line ($\delta^2$H=7.1($\delta^{18}$O)-6.3; Tappa et al., 2016). Mean rain (pink diamond), mean snow (yellow triangle), and mean well values (green triangle) samples were collected by Radke et al. (2019) at a higher elevation site in Reynolds Creek watershed. Source-sink dynamics of measured and predicted Cl$^+$ concentrations calculated using $\delta^{18}$O upstream and downstream site ratios for the unburned JD (C) and

---

**E**

Stable water isotopes ($\delta^{18}$O and $\delta^2$H) for sites in both unburned JD (A) and burned MC (B) in April (light circles), May (mid-tone diamonds), and June (dark squares). The global meteoric water line (GMWL) is shown in solid black and the local meteoric water line (LMWL) is the dashed red line ($\delta^2$H=7.1($\delta^{18}$O)-6.3; Tappa et al., 2016). Mean rain (pink diamond), mean snow (yellow triangle), and mean well values (green triangle) samples were collected by Radke et al. (2019) at a higher elevation site in Reynolds Creek watershed. Source-sink dynamics of measured and predicted Cl$^+$ concentrations calculated using $\delta^{18}$O upstream and downstream site ratios for the unburned JD (C) and

---

**F**

Stable water isotopes ($\delta^{18}$O and $\delta^2$H) for sites in both unburned JD (A) and burned MC (B) in April (light circles), May (mid-tone diamonds), and June (dark squares). The global meteoric water line (GMWL) is shown in solid black and the local meteoric water line (LMWL) is the dashed red line ($\delta^2$H=7.1($\delta^{18}$O)-6.3; Tappa et al., 2016). Mean rain (pink diamond), mean snow (yellow triangle), and mean well values (green triangle) samples were collected by Radke et al. (2019) at a higher elevation site in Reynolds Creek watershed. Source-sink dynamics of measured and predicted Cl$^+$ concentrations calculated using $\delta^{18}$O upstream and downstream site ratios for the unburned JD (C) and
the burned MC (D). Predicted downstream Cl\textsuperscript{-} concentrations are compared to measured downstream Cl\textsuperscript{-} concentrations. Red solid line represents identical measured and predicted concentrations and the dashed grey lines enclose the prediction interval. Cl\textsuperscript{-} values above the red line indicate Cl\textsuperscript{-} concentrations greater than expected, or sources (green), and Cl\textsuperscript{-} concentration below the red line indicate Cl\textsuperscript{-} values that are less than expected, or sinks (orange). Sr and DIC relationships for the unburned JD (E) and the burned MC (F). Mean precipitation (both snow and rain) and mean groundwater endmembers from Radke et al. (2019) are represented by pink diamonds and green triangles, respectively. Grey bars on endmembers show the SE. Uppermost reach sites (>2000 m) of each stream are outlined in red for each month water was present.

Table 2.3

**Stable Water Isotope Analysis.**

<table>
<thead>
<tr>
<th></th>
<th>Isotope Analyses</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Slope</td>
<td>± SE</td>
<td>LMWL Intercept (per-mil)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unburned JD</td>
<td>April 2.7</td>
<td>0.16</td>
<td>-111.47</td>
</tr>
<tr>
<td></td>
<td>May 3.6</td>
<td>0.27</td>
<td>-111.09</td>
</tr>
<tr>
<td></td>
<td>June 4.7</td>
<td>0.30</td>
<td>-114.04</td>
</tr>
<tr>
<td>Burned MC</td>
<td>April 4.3</td>
<td>0.30</td>
<td>-122.36</td>
</tr>
<tr>
<td></td>
<td>May 3.0</td>
<td>0.26</td>
<td>-117.04</td>
</tr>
<tr>
<td></td>
<td>June 4.5</td>
<td>0.23</td>
<td>-121.73</td>
</tr>
</tbody>
</table>

Notes: The slope ± standard error and the LWML-intercept extracted from the data presented in Figure 2.6 (A and B).

**2.3.3.2 In-stream evapoconcentration or dilution influences on solute concentrations.** To investigate potential effects of evaporation or dilution on the concentration of solutes as streams dried, we evaluated source-sink dynamics of chloride compared to δ\textsuperscript{18}O. In both streams, most measured Cl\textsuperscript{-} fell within the error of the predicted Cl\textsuperscript{-} based on δ\textsuperscript{18}O, indicating the two analytes behaved conservatively, with no major new gains or losses of water. Given this finding, we did not further investigate
spatial or temporal deviations due to either evapoconcentration or dilution (Figure 2.6, C, D). Although we hypothesized that in-stream evapoconcentration would be a driver of increased analyte concentrations, we found few sites in either stream that showed concentrations outside of the δ\textsuperscript{18}O-based prediction intervals. Additionally, counter to our predictions, we observed no differences between streams despite their difference in canopy cover.

2.3.3.2 Groundwater Influence. The isotopic signature of both streams showed surface water sources to be mixed contributions of snowmelt and rain: each stream’s isotope values fell between the snow and rain mean values reported by Radke et al. (2019; Figure 2.6, A, B). Based on intercepts, water in both streams had dominant contributions from snow/groundwater; water sources for unburned JD reflected a larger rain contribution as indicated by isotopic-intercepts with the LMWL (~ -111 δ\textsuperscript{2}H per mil) whereas the LMWL-intercept for burned MC stream (~ -120 δ\textsuperscript{2}H per mil) reflected a larger snow and/or groundwater contribution (Figure 2.6, A, B). Because the groundwater and snow isotopic signatures were similar, we could not distinguish between the two sources with isotopic tracers. Counter to our expectations, no clear temporal shifts in water sources emerged in the isotope data in either stream (Table 2.3).

However, temporal shifts in groundwater-precipitation mixing were apparent by pairing the isotopic patterns with observed shifts in DIC and other analyte concentrations. For example, we observed that DIC and Sr were strongly correlated (Figure 2.3). When compared to the average precipitation (rain and snow) and late summer groundwater Sr/DIC endmembers, stream water was sourced from precipitation in April and shifted towards groundwater sourcing in June in each stream (Figure 2.6, E, F). In contrast to
isotopes of water, rain and snow precipitation had similar DIC and Sr signatures and these signatures were substantially different from groundwater DIC and Sr signatures (Radke et al., 2019). We likely did not capture an additional endmember in unburned JD in the upper portion of the catchment (Figure 2.6, E) whereas most of the variation in the burned MC was captured by two endmembers, potentially with exception of a few downstream sites (Figure 2.6, F). The uppermost reach sites (>2000 m) of each stream (flowing in April and May) had different signatures; the signatures from the upstream reaches in unburned JD fell close to groundwater and burned MC upstream reach sites fell close to a precipitation signature (Figure 2.6, E, F and Supplementary Figure 2.1).

2.3.3.3 Organic Carbon Sourcing: Autochthonous and Allochthonous processes. DOM sources and spectral characteristics varied along the profile of each stream during each month; FI varied by similar amounts in both streams whereas $a_{254}$ [m$^{-1}$] varied more in unburned JD (Figure 2.7). Many FI values fell between the 1.5 allochthonous derived threshold and the 1.7 autochthonous/microbially derived threshold (Figure 2.7, A, B), indicating mixed carbon sources in both streams. Unburned JD had a slight but surprising longitudinal FI pattern in April (mean 1.59 ± 0.01 SE) when we observed a more autochthonous/microbial FI signature downstream, while the upper reach of the stream had a more allochthonous signature (Figure 2.7, A, circles). May generally had an allochthonous signature (mean 1.49 ± 0.003 SE) as did June (mean 1.50 ± 0.007 SE), and this similarity between seasons was counter to our predictions. Interestingly, while May and June mean FI values were similar, June exhibited greater variation (CV 2.8) than May (CV 1.2), with points spanning the thresholds between autochthonous and allochthonous sources in both months (Supplementary Table 2.5).
In burned MC (Figure 2.7, B), we observed a mixed autochthonous and allochthonous FI signature in April (mean 1.55 ± 0.005 SE), and a similar but more varied FI signature in May (mean 1.50 ± 0.009 SE). June’s FI varied along the stream profile with an overall downstream increase towards autochthonous/microbial signature (mean 1.62 ± 0.02 SE). This tendency matched our expectations of increasing in-stream productivity, which we expected would be more pronounced in the burned stream. In addition, burned MC’s higher mid-summer FI variation was reflected in the relatively high June FI CV (7.5) compared to that of April (1.0) and May (3.5; Supplementary Table 2.5). Despite the overall autochthonous tendency in June, there were also sites for which the FI value fell below or close to the allochthonous threshold (1.5, Figure 2.7, B, squares). We did not expect to observe this high variation.

We expected that as streams dried, DOC aromaticity (high in terrestrially derived carbon) would be decreasing due to increased in-stream productivity, especially in burned MC, and thus, we expected decreases in $a_{254}$ values. Overall, unburned JD had higher $a_{254}$ values that were more longitudinally and seasonally variable than burned MC (Figure 2.7, C, D and Supplementary Table 2.5). We expected higher overall aromaticity in unburned JD. Though we expected to see higher FI values coupled with low $a_{254}$ values, we did not observe these relationships consistently. In April, unburned JD had higher $a_{254}$ values (mean 44.5 ± 1.4 SE) coupled with higher FI values (Figure 2.7, E). The unburned JD maintained higher $a_{254}$ values in May (mean 48.1 ± 3.0 SE) and dropped in June (mean 38.5 ± 2.9 SE); this decrease in aromaticity suggests less terrestrial carbon and followed our expectations. Compared to unburned JD, burned MC $a_{254}$ values shifted less from April (mean 26.2 ± 0.48 SE) to May (mean 27.2 ± 0.54 SE) and June (mean 29.1 ± 0.53
SE, Figure 2.7, D). The differences in a254 temporal patterns between the streams were illustrated by the relationship of FI and a254 (Figure 2.7, E, F) where the spread of FI values was greater in the burned MC, but the spread of a254 was greater in the unburned JD. We hypothesized that increased in-stream productivity would result in shifts of increased FI and decreased a254 values. However, the spatial patterns we observed in both streams were more longitudinally varied and had less clear seasonal shifts than expected.
Figure 2.7

Autochthonous or Allochthonous Organic Carbon Sourcing.

Notes: (A) and (B) Fluorescence index (FI), (C) and (D) absorbance coefficient ($a_{254}$ $[\text{m}^{-1}]$), and (E) and (F) the relationship of FI to $a_{254}$ for stream profile of unburned JD (blue, left) and burned MC (orange, right); shape, shade, and color of months are described in...
Figure 2.2. Grey shaded boxes (A and B) indicate the 1.5-1.2 allochthonous-derived threshold and the 1.7-2.0 autochthonous/microbial-derived threshold.

2.4 Discussion

Stream intermittency and fire regimes are changing in the Intermountain West and cold montane shrubland ecosystems that characterize the northern Rocky Mountains and Great Basin are especially vulnerable to these changes. We hypothesized stream chemistry patterns would change seasonally and longitudinally in response to both drying and fire, and that process shifts would explain these biogeochemical patterns. Specifically, we hypothesized that with drying, in-stream evaporative processes, increased groundwater influence, and shifts in DOC sourcing from allochthonous to autochthonous would contribute to differences in these longitudinal stream chemistry patterns. In a recently burned watershed, we expected the impacts of these processes would be amplified.

2.4.1 Stream Drying: Nonlinear Spatiotemporal Biogeochemical Patterns

Both streams exhibited nonlinear longitudinal patterns in chemistry that changed with stream drying. Spatially variable and distinct patterns among analytes were revealed by this study’s high-scope sampling approach and breadth of measured chemical constituents. For instance, we identified potential springs/deeper groundwater inputs where temperature or solute concentrations remained stable despite seasonal shifts at other locations longitudinally along the streams (e.g. around ~1500 m in unburned JD and ~ 2000 m in burned MC). In contrast to our expectations, we found similar longitudinal patterns and processes driving these patterns in both streams. This provides some evidence for emerging regional behavior in the subsurface processes (i.e.
groundwater contributions) driving stream chemistry. Our findings collectively point to the power of high-scope sampling to capture both the temporal and spatial variability in intermittent stream processes, and they contribute empirical data on how spatiotemporal patterns and processes may change with stream drying in headwater streams like these.

2.4.1.1 Streams Similarities: DIC Dominance with Drying. Given the differences in fire history and lithology, the similarity of chemistry patterns and shifting processes between these two streams was surprising. In both unburned JD and burned MC, DIC and many cations exhibited clear longitudinal patterns that increased in mean concentration with stream drying. Though we did not directly analyze concentration-discharge relationships longitudinally at each 50 m reach site, chloride source/sink dynamics did not suggest strong spatial inputs or outputs in any month. Monthly increases in DIC and other conservative solute concentrations were consistent with dilution behavior of concentration-discharge observed over the water year at the weir of each stream (Glossner, 2019). While conservative solutes in larger streams often display chemostatic behavior (Godsey et al., 2009), our data is consistent with chemodynamic concentration-discharge relationships documented in smaller catchments (Hunsaker and Johnson, 2017) and intermittent streams (Bernal et al., 2019). Instead of evapoconcentration or dilution, our findings point to deeper groundwater sources driving increased concentrations as the streams dried. Our findings are consistent work by Wlostowski et al. (in review), whose findings describe loamy and sand-rich soils and relatively shallow bedrock in unburned JD resulting in more baseflow dominated streamflow with less subsurface water storage compared to CZOs with clay-rich soils. Distinct from other baseflow-dominated streams, groundwater contributed to low flow in
our streams, not high flow as observed in the Jemez River Basin Critical Zone observatory in New Mexico (McIntosh et al., 2017). Our results build off of those reported by Radke et al. (2019) at a higher elevation RC CZO catchment during snowmelt, where deeper flow paths rather than shallow subsurface flow paths contributed to stream DOC as observed by Boyer et al. (1997). Deeper flow paths also contribute to increasing ion concentration patterns at low flow in Dry Creek Experimental Watershed near Boise, ID (McNamara et al., 2005). Our findings, along with these earlier works suggest a regional occurrence of limited near-surface storage capacity due to loamy soils and the importance of deeper groundwater for stream water.

In contrast to DIC, more biologically reactive solutes TN, DOC, and PO$_4^{3-}$ either decreased or were relatively invariant with stream drying. Mean TN concentrations decreased throughout the growing season, and DOC showed relatively high variability and no clear longitudinal pattern or large seasonal shift in either stream. With respect to TN, we expected temporal decreases because nitrogen is biologically reactive and often subject to uptake, especially during low flows and low concentrations (Moatar et al., 2017). On the other hand, many studies of stream carbon dynamics have found clear seasonal patterns in DOC (Hornberger et al., 1994; Boyer et al., 1997; Mulholland and Hill, 1997; McGuire et al., 2014; Hale and Godsey, 2019). However, we did not observe a strong DOC pattern with drying in either stream. Rather, we found strong temporal shifts in DIC, and DIC concentrations that were 2 to 3 fold higher than DOC in unburned JD and burned MC. Thus, DIC, rather than DOC, appeared to dominate stream carbon dynamics at our sites.
Nonlinear spatial patterns in longitudinal stream chemistry spatial patterns have been observed in other studies that employed high spatial-scope measurements, but few have examined these longitudinal patterns in the context of stream drying or across such a broad suite of biologically reactive and conservative analytes. In one such example, within snowmelt-driven, deciduous forested streams at Hubbard Brook, NH, Likens and Buso (2006) sampled at 100 m intervals over a network and observed nonlinear spatial patterns in a wide suite of constituents including DOC, DIC, K, Na, SO$_4^{2-}$, NO$_3^-$, Sc, dissolved Si, and Ca (also see McGuire et al., 2014; Zimmer and Lautz, 2014). Similarly, Dent and Grimm (1999) found that NO$_3^-$-N and soluble reactive phosphorus (SRP or PO$_4^{3-}$) varied longitudinally in the hot desert at Sycamore Creek, AZ. Moreover, based on 25 m intervals over 10 km of stream, they observed that NO$_3^-$-N decreased and SRP increased with succession following monsoonal flooding. These findings agree with our observations of decreasing TN concentrations and relatively invariant (unburned JD) or slightly decreasing PO$_4^{3-}$ (burned MC) with drying.

2.4.1.2 Stream Differences: Lithology and Fire Impacts. We found differences between the two streams we studied in cation concentrations, the proportion of open canopy, and magnitude of seasonal decrease in TN, organic carbon and water sourcing. Lithological differences likely account for higher cation concentrations in volcanic burned MC compared to granitic unburned JD stream (Meybeck, 1987; Ibarra et al., 2016). It is also likely that fire impacts increased cation concentrations through ash and sediment inputs and/or mineralization of these inputs, especially during snow melt. At the watershed scale, sediment losses were high during winter and snowmelt in the year following fire, a low flow year, and sustained in 2017 (>450 compared to 20 g m$^{-2}$ yr$^{-1}$
mean sediment yield in Johnston). Sediment losses were low during the drying summer months (<5% of annual sediment yield) indicating low contributions of particulates during this study period (Glossner, 2019); mineralization of these ash-derived products warrants further study.

Elevated nitrogen levels in April and May in burned MC decreased by >80% in June. These TN results supported our hypothesis that nitrogen levels would be elevated immediately following wildfire but that decreases in TN would be observed as burned MC dried due to potentially increased nutrient uptake. Other studies such as Murphy et al. (2006) and Rau et al. (2007) have documented that there is often a lag in nitrate uptake following fire due to the time required to re-establish uptake by denitrifying bacteria populations and vegetation. In some cases, high nitrogen levels have been documented in streams for years following fire (Hauer and Spencer, 1998; Mast et al., 2016), potentially related to low DOC from terrestrial carbon losses (Rodríguez-Cardona et al., 2020).

However, TN levels in burned MC decreased relatively quickly to concentrations similar to those of unburned JD by June. Finally, DOM sourcing followed different temporal tendencies in the two streams: burned MC shifted as expected from more allochthonous to autochthonous whereas unburned JD showed more mixed contributions. These findings will be discussed further below in the section “Statistical Challenges of Stream Drying Data.”

Fire can have a large effect on water budget and cause increases in stream flow as catchment ET diminishes following vegetation losses (Kinoshita and Hogue, 2015; Atchley et al., 2018), but we did not find strong evidence of this influencing stream chemistry drying patterns. This in part may be because burned MC experienced a
relatively low water year in 2016 (Vega et al., 2020; Glossner, 2019) and rapid growth of grasses and herbaceous was observed. Rapid fire recovery (within a year) has been observed previously at RC CZO with regards to gross ecosystem production (Fellows et al., 2018) and ET (Flerchinger et al., 2016), with statistically marginal short-term increases to stream flow (Flerchinger et al., 2016). Moreover, ET rates have been shown to be relatively low in sagebrush in RC CZO (Sharma et al., 2019, 2020). Isotopic LMWL-intercepts indicated differing water sources between the streams and suggested unburned JD was more rainfed than burned MC, which was fed more by groundwater/snow. Given the vegetation differences, it is possible unburned JD had more shallow subsurface storage of rainwater than burned MC. However, as vegetation regrew in burned MC, there was no clear LMWL-intercept shift from April to June suggesting changes in soil water storage capacity. Alternatively, water source differences may have reflected pre-fire differences and further study would be required to parse this out. As mentioned, we observed stream chemistry responses to fire in some instances like elevated TN, but we did not detect increased stream flow. These observations are consistent with the weak evidence weak evidence of monthly changes in evapoconcentration or dilution processes and burned MC’s lack of shifts in water isotopic water sources or evaporative slopes. In addition to the similarity of stream patterns indicating the importance of deeper subsurface processes on stream chemistry, the similarity may indicate muted effects of fire on stream chemistry in the burned watershed.
2.4.2 Processes Explaining Spatiotemporal Patterns.

Our findings highlight the utility of stream chemistry to elucidate both hydrological processes (evaporation or dilution and water sourcing) and the in-stream biological processes (potential autochthonous carbon-sourcing) that contribute to spatiotemporal patterns as streams dry seasonally and following wildfire. A strength of our data was that we were able to evaluate surface-groundwater interactions, which has been identified as a key research priority to understanding intermittent stream controls (Costigan et al., 2016). Our inferences are only made possible by the high spatial scope sampling approach we employed.

2.4.2.1 Evaluating Hydrological Processes: Shift Towards Groundwater with Drying. By quantifying evaporation and groundwater influence, we could distinguish between chemistry drying patterns driven by surface processes and chemistry drying patterns driven by subsurface hydrologic processes in these streams. Isotopic slopes showed both streams were evaporative; however, we observed no temporal shifts towards increased in-stream evaporation with drying. Contrary to our expectations and despite differences in canopy cover, both streams showed similar evaporative trends in water isotopes with no statistical difference in slope between the two streams in May or June. Furthermore, we found only weak evidence of in-stream evapoconcentration or dilution on the spatial stream chemistry patterns in either stream during any month. In contrast, Sr-DIC ratios provided strong evidence that stream surface water shifted towards a deeper groundwater signature with drying. These low-flow groundwater contributions differ from the lateral input dominance of snowmelt described by Boyer et al. (1997). The Sr-DIC relationships varied longitudinally and were better captured by precipitation and
groundwater endmembers in burned MC than unburned JD. In April, the Sr-DIC signature of the uppermost portion of unburned JD led us to reason that either that these sites are spring-fed or that our sampling missed the shift from precipitation to groundwater sourcing. In general, increased contact time between water and soil/bedrock may explain elevated cation and DIC concentrations in groundwater (Fritz and Clark, 1997; Ritcher and Billings, 2015; Olshansky et al., 2019), each of which increased with drying in both streams. Although isotope LMWL-intercepts did not show strong shifts in water sourcing between months as Sr-DIC ratios did, the isotopic signatures of snow and groundwater were indistinguishable. This suggests two potential scenarios. First, as snowmelt recharges deeper groundwater, snowmelt increases its solute concentrations. Alternatively, shallow subsurface evaporation could cause evapoconcentration (Radke et al., 2019). Quantifying subsurface processes such as those that would affect the residence times of water (Brooks et al., 2015), carbon in groundwater, or potential evapoconcentration in soils (Radke et al. 2019) may be important for understanding stream dynamics at RC CZO and warrants further regional study.

2.4.2.2 Evaluating Biological Processes: Carbon Sourcing is Longitudinally Variable. Rather than exhibiting distinct seasonal shifts over time, biologically active solutes, such as organic carbon, showed a mixed signature of allochthonous/autochthonous sourcing and composition in both streams each month. We found DOC sourcing (FI) and aromatic properties ($a_{254}$) varied in both streams, but burned MC was less aromatic overall ($a_{254}$). Longitudinally, unburned JD had mixed sourcing (FI values) and a wide spanning aromatic signature ($a_{254}$) even in June, though there was a tendency towards being less aromatic. Although unburned JD FI observations
were similar to summer 2017 samples from higher elevation streams at RC CZO (Radke et al., 2019), they were counter to our expectations. Canopy cover in the unburned JD was heterogeneous: leaf out may have decreased sunlight availability to the stream, limiting productivity and contributing more allochthonous material throughout the summer. Burned MC more closely matched our expectations of increased in-stream productivity with a more autochthonous FI in June and an overall less aromatic signature, but FI was nevertheless unexpectedly variable. More in-stream productivity may have been supported by high (~20 fold more) nitrogen availability in the early season and an open canopy from April to June, combined with the generally reduced vegetation as allochthonous carbon sources post fire. While there were temporal differences between the streams, the high variability that characterized DOM sourcing and spectral characteristics reflected a mixed organic carbon signature in both streams. This may be typical of small-area catchments that are often hydraulically connected to surrounding terrestrial uplands (Creed et al., 2015). Nevertheless, burned MC showed clearer, and potentially less complex patterns of increasing groundwater influence and autochthonous DOC. Though it is unclear how characteristic these patterns may have been before the fire in burned MC, it is possible the fire’s reduction of canopy cover and increase of TN may have contributed to more pronounced temporal shifts towards autochthonous DOC.

2.4.3 Study Approach: Pros and Cons

2.4.3.1 Power of High Spatial-Scope Sampling and Integrated Methodology. Collectively, our findings have contributed to recognizing and understanding stream chemistry patterns that cannot be addressed with traditional sampling approaches that
rely on a small number of spatially discrete locations, such as stream outlets (e.g., Fisher and Likens, 1973; Boyer et al., 1997, Mulholland and Hill, 1997; Hood et al., 2006; Jaffé et al., 2008, etc.). For instance, Jaffé et al. (2008) found FI to vary from 1.15 to 1.75 at the landscape to biome scale. We found comparable variation both longitudinally down the 2.5 km stream segments at a 50 m grain and temporally across months we studied. This finding suggests that traditional coarse-grain spatial sampling of DOM sources 1) may not capture heterogeneity of the streams above or below a sampling point and 2) may lead to inappropriate inferences about a given stream network. It is unclear if network convergence would occur downstream in DOM sourcing, as has been observed in DOC concentrations (Asano et al., 2009; Hale and Godsey, 2019), and if so, at what scales (Creed et al., 2015). Our findings provide evidence to support the Creed et al. (2015) hypothesis that DOM sourcing dynamics within a single catchment may be more complex than thought and may have consequences for the chemical composition and quality of organic matter being transported downstream.

Stream chemistry studies have traditionally either focused on hydrological processes (Brooks et al., 2015) or on in-stream biological processes (e.g., Minshall et al., 1989; Dent and Grimm, 1999), and study scope determines at what scale(s) patterns can be detected or processes can be assessed (Schneider, 2001; Fausch et al., 2002). In this observational study, our aim was to explore both hydrological and biological processes that may drive stream chemistry patterns using integrated methodology at a high spatial scope as streams dry. Taking this spatial approach required that we balance trade-offs between distance covered (extent) and sampling interval (spatial grain); we did not study a large stream network, but focused on the headwater stream segment, where we
expected to see the most pronounced drying and biogeochemical patterns and variation. Although we also used repeated monthly sampling to investigate temporal patterns, we focused our efforts at the beginning of the drying season with moderate temporal resolution (monthly) over a limited temporal extent (three months at the beginning of one drying season).

Due to their small size, headwater streams are highly variable with regards to stream chemistry (Zimmer et al., 2013; Creed et al., 2015; Hunsaker and Johnson, 2017), and thus they are likely to be more susceptible to seasonal and spatial regime shifts such as increased stream intermittency and fire, both promoted by climate change. The particulars of these shifts in headwaters are important to understand in various biomes for study design as well as water quality monitoring (Abbott et al., 2018). Future work might include spatially distributed streamflow measurements to accompany stream chemistry observations because it is possible that the local water balance differs seasonally and longitudinally, affecting in-stream evapoconcentration/dilution dynamics (e.g., Glossner, 2019; Warix, 2020). Lastly, we balanced sampling intensity and number of streams studied. Although our inferences are limited by the fact that we studied only two streams, the similarities between the two suggest some of the patterns we observed may be expected throughout the RC CZO and region. Our analysis of a wide range of chemical constituents at 50 m reach intervals complemented the work of Hale and Godsey (2019) who measured DOC at 200 m intervals at the network scale in southeastern Idaho, USA, showing that additional analytes can improve process interpretation. We suggest that exploratory studies, such as this one, are important to tease apart which processes are
effectively uniform throughout a stream reach, and which are likely to be spatially heterogeneous and seasonally dynamic.

**2.4.3.2 Statistical Challenges of Stream Drying Data.** Additional challenges existed applying typical statistical approaches used to describe perennial stream chemistry patterns to spatiotemporal drying patterns. It was particularly challenging to describe temporal patterns, compare the longitudinal spatial variation between months as streams dried, and to quantify the spatial patterns of a given month. Much of this difficulty stemmed from the loss of data points as streams dried, impacting the extent and resolution and effectively reducing the sample size each month and leaving few perfect statistical approaches. Some authors have reported maximums and minimums to capture seasonal differences within a network (Likens and Buso, 2006; Brooks and Lemon, 2007; Zimmer et al., 2013), but this limits comparisons between variables, particularly of different magnitudes or different units of measure. Coefficient of variation (CV; calculated by standard deviation, SD, divided by the mean) is a unitless tool for quantifying variation and is therefore appealing to compare different variables. As streams dry, CV has been used to quantify temporal patterns (Likens and Buso, 2006), spatial patterns at varying scales (Dent et al., 2001), and spatiotemporal stream chemistry patterns (Hale and Godsey, 2019) to compare across varying monthly means and sample sizes. However, CV is sensitive to changes in the mean, as Dent and Grimm (1999) point out, as well as the sample size which makes CV potentially problematic for drying streams. For example, in our data if April and June solute concentrations have the same SD (representing spatial variation) but June has a higher mean, June will have a lower CV. This could reflect the loss of sampling locations rather than changes in spatial
variability of the sites that remained wet in June. Dent and Grimm (1999) acknowledged the limitation and emphasized CV only in specific instances.

We also faced difficulties in quantifying spatial patterns in our streams due to underlying, nonlinear longitudinal patterns, which changed with stream drying. Brooks and Lemon (2007) addressed this issue by reporting means and SD for an upper, middle, and lower region of the San Pedro River, AZ. To control for differences in the mean, sample size and underlying longitudinal patterns, we calculated CV in instances for which underlying longitudinal patterns did not change with drying, which occurred only for FI. More sophisticated approaches have included constructing semivariograms to quantify spatial patterns (McGuire et al., 2014) and spatiotemporal patterns (Dent and Grimm, 1999). The data from our study required an in-depth approach to detecting and quantifying spatiotemporal patterns which was beyond the scope of this paper, and we more thoroughly addressed quantifying patterns in a separate, forthcoming publication. Continued development of robust statistical methods and conceptual models that incorporate drying (changes in sample size) is merited. This is a vital step to developing more appropriate approaches and tools to understand patterns and mechanisms of stream drying (Jensen et al., 2019).

2.5 Conclusions

Understanding intermittent stream ecosystem structure and processes requires a shift in stream conceptual models to include stream drying (Allen et al., 2020) and testing controls on drying like surface water-groundwater interactions (Costigan et al., 2016). Exploring patterns using a high-spatial scope approach, biogeochemical data enabled us
to test several hypotheses about processes associated with stream drying and following fire. We found headwater streams had distinct chemical patterns that shifted longitudinally with seasonal drying and showed similar patterns despite fire history differences. In this study, an unburned and a burned stream exhibited similar temporal patterns with stream drying and similar shifts in processes. We argue that detecting shifts in the processes that influenced these patterns required measurements at a fine grain over an intermediate extent. DIC dominated carbon dynamics in both streams, with increasing concentrations with drying. We observed similar drying shifts in increasing Sr and Cl\(^{-}\), relatively invariant DOC, decreasing TN, and relatively invariant PO\(_4^{3-}\). Cations generally increased from April to June with more substantial losses in burned MC potentially due to volcanic bedrock. The most striking difference in temporal patterns between the unburned JD and burned MC was the dramatic loss of TN from April to June in the burned MC. As the streams dried, surface water chemistry patterns could be influenced by surface evaporative processes, dilution, and/or from subsurface processes with shallow and deep groundwater influence. During stream drying, the snow-fed, semi-arid mountainous catchments we studied were both characterized as evaporative streams, but with weak evidence for in-stream evapoconcentration or dilution as either stream dried. There were not substantial in-stream evaporative differences despite canopy cover differences between unburned JD and burned MC. Instead, stronger evidence for deeper groundwater processes influencing stream chemistry patterns with longitudinal distinctions were observed in these streams, with more clear chemistry shifts with drying in burned MC. The similarities between streams may suggest that subsurface structure mutes the effects of fire and lithology differences at RC CZO. Organic carbon sourcing
was mixed and highly longitudinally varied in both streams, but with an overall shift towards autochthonous in burned MC. These findings contribute to understanding streams in the sagebrush steppe by exploring spatiotemporal stream chemistry patterns and processes shifts as stream drying and following wildfire. Such empirical data are critical to the development of more sophisticated stream models, which are especially needed as climate changes across the western U.S.
2.6 Works Cited


2.7 Supplementary Materials

Supplementary Figure 2.1

*Longitudinal Ground Water Patterns.*

![Diagram showing longitudinal ground water patterns](image)

Notes: Sr and DIC relationships for unburned JD (A) and burned MC (B); shade and color are the same as described in Figure 2.6. Mean precipitation (rain and snow, pink diamonds) and groundwater (green triangles) endmembers are described in Figure 2.6. The streams are split into three longitudinal sub-segment portions upstream of the weir: upper (> 2000 m above the weir; dark purple outline), mid-stream (1000 m to 2000 m above the weir; mid-tone purple outline), and lower (within 1000 m upstream of the weir; light purple outline) to reflect observed downstream groupings in longitudinal variation. The uppermost surface flow in April is indicated by a red “x” and the downstream weir is indicated by a green star in the watershed map and box colors correspond with graphs.
Supplementary Table 2.1

*Field chemical characteristics.*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>April-June</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temp °C</strong></td>
<td>Mean ± SE n</td>
<td>Mean ± SE n</td>
<td>Mean ± SE n</td>
<td>Mean ± Error</td>
</tr>
<tr>
<td>Unburned JD</td>
<td>9.50 0.604 57</td>
<td>11.83 0.668 51</td>
<td>13.90 0.772 38</td>
<td>-3.87 0.94</td>
</tr>
<tr>
<td><strong>pH</strong></td>
<td>7.00 0.031 57</td>
<td>6.84 0.057 51</td>
<td>6.81 0.087 38</td>
<td>0.25 0.076</td>
</tr>
<tr>
<td><strong>DO mg/L</strong></td>
<td>11.31 0.166 57</td>
<td>10.53 0.100 51</td>
<td>9.06 0.410 38</td>
<td>2.44 0.40</td>
</tr>
<tr>
<td><strong>Temp °C</strong></td>
<td>Mean ± SE n</td>
<td>Mean ± SE n</td>
<td>Mean ± SE n</td>
<td>Mean ± Error</td>
</tr>
<tr>
<td>Burned MC</td>
<td>10.20 0.264 60</td>
<td>11.83 0.668 51</td>
<td>16.11 0.58 38</td>
<td>-5.92 0.64</td>
</tr>
<tr>
<td><strong>pH</strong></td>
<td>7.28 0.076 60</td>
<td>6.84 0.057 51</td>
<td>7.47 0.074 38</td>
<td>-0.19 0.11</td>
</tr>
<tr>
<td><strong>DO mg/L</strong></td>
<td>11.18 0.113 60</td>
<td>10.53 0.097 36</td>
<td>10.38 0.153 38</td>
<td>0.82 0.19</td>
</tr>
</tbody>
</table>

Notes: Standard error (SE), sample size (n), mean difference from April-June, and mean difference propagated error for unburned JD and burned MC in April, May, and June.
Supplementary Table 2.2

Solute Summary Table.

<table>
<thead>
<tr>
<th>Analyte mg/L</th>
<th>April Mean ± SE n</th>
<th>May Mean ± SE n</th>
<th>June Mean ± SE n</th>
<th>April-June Mean ± Error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unburned JD</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>6.50 ± 0.21 56</td>
<td>9.23 ± 0.47 51</td>
<td>11.49 ± 0.50 35</td>
<td>-3.14 ± 0.87</td>
</tr>
<tr>
<td>Cl⁻</td>
<td>1.15 ± 0.030 54</td>
<td>1.16 ± 0.036 50</td>
<td>1.34 ± 0.036 35</td>
<td>-0.2 ± 0.047</td>
</tr>
<tr>
<td>DOC</td>
<td>4.75 ± 0.53 53</td>
<td>5.01 ± 0.28 52</td>
<td>4.63 ± 0.29 35</td>
<td>0.09 ± 0.35</td>
</tr>
<tr>
<td>TN</td>
<td>0.36 ± 0.037 56</td>
<td>0.27 ± 0.024 50</td>
<td>0.22 ± 0.03 37</td>
<td>0.142 ± 0.048</td>
</tr>
<tr>
<td>PO₄³⁻</td>
<td>0.037 ± 0.007 54</td>
<td>0.037 ± 0.004 49</td>
<td>0.042 ± 0.006 37</td>
<td>-0.0043 ± 0.0097</td>
</tr>
<tr>
<td><strong>Burned MC</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td>5.42 ± 0.30 57</td>
<td>10.45 ± 0.47 41</td>
<td>15.67 ± 0.68 36</td>
<td>-7.21 ± 0.73</td>
</tr>
<tr>
<td>Cl⁻</td>
<td>0.825 ± 0.037 55</td>
<td>1.11 ± 0.020 39</td>
<td>1.47 ± 0.025 36</td>
<td>-0.64 ± 0.045</td>
</tr>
<tr>
<td>DOC</td>
<td>3.00 ± 0.12 51</td>
<td>2.83 ± 0.11 41</td>
<td>3.33 ± 0.16 39</td>
<td>-0.33 ± 0.20</td>
</tr>
<tr>
<td>TN</td>
<td>1.14 ± 0.032 58</td>
<td>0.915 ± 0.023 42</td>
<td>0.19 ± 0.02 38</td>
<td>0.95 ± 0.037</td>
</tr>
<tr>
<td>PO₄³⁻</td>
<td>0.094 ± 0.006 58</td>
<td>0.116 ± 0.004 42</td>
<td>0.13 ± 0.004 38</td>
<td>-0.037 ± 0.0071</td>
</tr>
</tbody>
</table>

Notes: Solutes standard error (SE), sample size (n), mean difference from April-June, and mean difference propagated error for unburned JD and burned MC in April, May, and June.
Supplementary Table 2.3

*Cations Summary Table.*

<table>
<thead>
<tr>
<th>Cation</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>April-June</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SE</td>
<td>Mean ± SE</td>
<td>Mean ± SE</td>
<td>Mean ± Error</td>
</tr>
<tr>
<td>Sr</td>
<td>58.26 ± 2.9</td>
<td>73.42 ± 5.6</td>
<td>74.85 ± 4.5</td>
<td>-16.59 ± 5.35</td>
</tr>
<tr>
<td>Na</td>
<td>6507 ± 231</td>
<td>9906 ± 317</td>
<td>7634 ± 329</td>
<td>-1127.41 ± 402.0</td>
</tr>
<tr>
<td>Ba</td>
<td>162 ± 12.2</td>
<td>164 ± 16.02</td>
<td>304 ± 29.8</td>
<td>-142.09 ± 32.23</td>
</tr>
<tr>
<td>Mg</td>
<td>1768 ± 82.77</td>
<td>2793 ± 159</td>
<td>2292 ± 139</td>
<td>-524.08 ± 162.0</td>
</tr>
<tr>
<td>Al</td>
<td>130 ± 10.95</td>
<td>241 ± 12.8</td>
<td>35.17 ± 7.6</td>
<td>95.21 ± 13.33</td>
</tr>
<tr>
<td>Si</td>
<td>17657 ± 224</td>
<td>19572 ± 212</td>
<td>16156 ± 308</td>
<td>1501.35 ± 381.60</td>
</tr>
<tr>
<td>K</td>
<td>1577 ± 99.7</td>
<td>2033 ± 128</td>
<td>1411 ± 102</td>
<td>166.32 ± 142.65</td>
</tr>
<tr>
<td>Ca</td>
<td>5960 ± 255</td>
<td>9597 ± 529</td>
<td>7681 ± 452</td>
<td>-1720.6 ± 519.35</td>
</tr>
<tr>
<td>Sc</td>
<td>6.58 ± 0.087</td>
<td>6.14 ± 0.136</td>
<td>5.10 ± 0.154</td>
<td>0.48 ± 0.17</td>
</tr>
<tr>
<td>Ti</td>
<td>10.22 ± 0.527</td>
<td>11.22 ± 0.653</td>
<td>6.16 ± 0.419</td>
<td>4.06 ± 0.67</td>
</tr>
<tr>
<td>Fe</td>
<td>158 ± 10.29</td>
<td>153 ± 15.0</td>
<td>118 ± 33.1</td>
<td>40.7 ± 34.67</td>
</tr>
<tr>
<td>Zn</td>
<td>6.41 ± 0.574</td>
<td>72.0 ± 25.58</td>
<td>11.7 ± 0.756</td>
<td>-7.3 ± 0.95</td>
</tr>
<tr>
<td>Sr</td>
<td>30.42 ± 2.3</td>
<td>71.19 ± 5.0</td>
<td>108.48 ± 5.9</td>
<td>-76.8 ± 6.30</td>
</tr>
<tr>
<td>Na</td>
<td>3525 ± 212</td>
<td>5541 ± 241</td>
<td>7556 ± 222</td>
<td>-3908.32 ± 299.64</td>
</tr>
<tr>
<td>Ba</td>
<td>42.21 ± 2.03</td>
<td>59.2 ± 2.15</td>
<td>109 ± 3.37</td>
<td>-66.5 ± 3.96</td>
</tr>
<tr>
<td>Mg</td>
<td>1090 ± 86.1</td>
<td>2401 ± 179</td>
<td>4100 ± 242</td>
<td>-2971.33 ± 256.59</td>
</tr>
<tr>
<td>Al</td>
<td>15.72 ± 1.15</td>
<td>21.66 ± 3.19</td>
<td>18.08 ± 4.15</td>
<td>-2.35 ± 4.31</td>
</tr>
<tr>
<td>Si</td>
<td>13568 ± 921</td>
<td>18952 ± 687</td>
<td>22839 ± 293</td>
<td>-9270.39 ± 967.38</td>
</tr>
<tr>
<td>K</td>
<td>2097 ± 137</td>
<td>2832 ± 131</td>
<td>3393 ± 124</td>
<td>-1223.45 ± 181.48</td>
</tr>
<tr>
<td>Ca</td>
<td>3519 ± 244</td>
<td>7167 ± 474</td>
<td>10728 ± 514</td>
<td>-7086.19 ± 565.97</td>
</tr>
<tr>
<td>Sc</td>
<td>4.07 ± 0.254</td>
<td>5.91 ± 0.168</td>
<td>7.13 ± 0.16</td>
<td>-3.06 ± 0.30</td>
</tr>
<tr>
<td>Ti</td>
<td>3.27 ± 0.231</td>
<td>5.93 ± 0.306</td>
<td>7.20 ± 0.40</td>
<td>-3.84 ± 0.46</td>
</tr>
<tr>
<td>Fe</td>
<td>34.32 ± 2.27</td>
<td>64.02 ± 2.94</td>
<td>86.05 ± 3.11</td>
<td>-50.81 ± 3.84</td>
</tr>
<tr>
<td>Zn</td>
<td>4.82 ± 0.675</td>
<td>4.87 ± 0.403</td>
<td>8.49 ± 2.00</td>
<td>-1.75 ± 0.86</td>
</tr>
</tbody>
</table>

Notes: Cations standard error (SE), sample size (n), mean difference from April-June, and mean difference propagated error for unburned JD and burned MC in April, May, and June.
Supplementary Table 2.4

*Cluster analysis of analytes and parameters for unburned JD and burned MC during each month.*

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Unburned JD</th>
<th>Burned MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>April</td>
<td>May</td>
</tr>
<tr>
<td>1</td>
<td>Sr</td>
<td>Mg</td>
</tr>
<tr>
<td></td>
<td>Ca</td>
<td>Na</td>
</tr>
<tr>
<td></td>
<td>Na</td>
<td>Sr</td>
</tr>
<tr>
<td></td>
<td>Ba</td>
<td>Cl⁻</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>Fe</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>Ti</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>Al</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>Ba</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>δ¹⁸O</td>
</tr>
<tr>
<td>2</td>
<td>Ti</td>
<td>pH</td>
</tr>
<tr>
<td></td>
<td>Al</td>
<td>Fe</td>
</tr>
<tr>
<td></td>
<td>Fe</td>
<td>PO₄³⁻</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
### Cluster 3

<table>
<thead>
<tr>
<th>DO</th>
<th>pH</th>
<th>DOC</th>
<th>Temp</th>
<th>DOC</th>
<th>pH</th>
<th>Temp</th>
<th>DOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

### Unburned JD

<table>
<thead>
<tr>
<th>Cluster</th>
<th>April</th>
<th>May</th>
<th>June</th>
<th>April</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Temp</td>
<td>TN</td>
<td>Temp</td>
<td>Temp</td>
<td>TN</td>
<td>Sc</td>
</tr>
<tr>
<td></td>
<td>DIC</td>
<td>Temp</td>
<td>TN</td>
<td>Zn</td>
<td>δ¹⁸O</td>
<td>TN</td>
</tr>
<tr>
<td></td>
<td>δ¹⁸O</td>
<td>Zn</td>
<td>δ¹⁸O</td>
<td>Zn</td>
<td>--</td>
<td>Sc</td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td>--</td>
<td>TN</td>
<td>--</td>
<td>--</td>
<td>Si</td>
</tr>
<tr>
<td></td>
<td>Cl⁻</td>
<td>--</td>
<td>Cl⁻</td>
<td>--</td>
<td>--</td>
<td>Si</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>Sc</td>
</tr>
</tbody>
</table>

### Burned MC

<table>
<thead>
<tr>
<th>Cluster</th>
<th>April</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Sc</td>
<td>Si</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Si</td>
<td>Si</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Si</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cluster</th>
<th>April</th>
<th>May</th>
<th>June</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>PO₄³⁻</td>
<td>Zn</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Supplementary Table 2.5

*FI and a254 Summary Table.*

<table>
<thead>
<tr>
<th></th>
<th>Unburned JD</th>
<th>Burned MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (n)</td>
<td>SE</td>
</tr>
<tr>
<td>April</td>
<td>1.59 (n=38)</td>
<td>0.01</td>
</tr>
<tr>
<td>May</td>
<td>1.49 (n=50)</td>
<td>0.003</td>
</tr>
<tr>
<td>June</td>
<td>1.50 (n=37)</td>
<td>0.007</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Unburned JD</th>
<th>Burned MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (n)</td>
<td>SE</td>
</tr>
<tr>
<td>April</td>
<td>44.5 (n=38)</td>
<td>1.4</td>
</tr>
<tr>
<td>May</td>
<td>48.1 (n=50)</td>
<td>3.0</td>
</tr>
<tr>
<td>June</td>
<td>38.5 (n=37)</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Notes: Summary of monthly mean, sample size (n), standard error (SE), and coefficient of variation (CV), and CV sample size for fluorescence index (FI) values for each stream in April, May, and June. CV was calculated using sites with surface water for all three months to prevent biasing any month with more stream flow and only for FI because there were no underlying longitudinal patterns.
Chapter III: Spatial Complexity Theory to Practice: Quantifying Fine-Grain Biogeochemical Heterogeneity of Drying and Wildfire Patterns of Intermittent Streams

Abstract

The spatial structure of intermittent streams changes as stream flow seasonally expands and contracts and also likely respond to disturbances like wildfire; stream biogeochemistry patterns likely reflect these changes in structure. While the importance of spatial complexity has been theorized extensively, few studies have quantified this complexity and scales at which patches occur or change in streams. We sought to understand and quantify spatial complexity of biogeochemical patterns in two headwater streams, one unburned and one recently burned, as they experienced drying following snowmelt in southwest Idaho, United States. To describe spatial complexity and patch size in surface water, we employed a high-spatial-scope sampling of biogeochemistry. We sampled each stream at 50-m intervals over a ~2,500-m segment. To quantify how patterns changed, we repeated sampling as the streams dried in April, May, and June 2016. Then, to better define patch size, we sampled the streams again at 10-, 25-, and 50-m intervals during low flow in July 2017. We hypothesized that spatial complexity and heterogeneity would increase and patch size would decrease with stream drying, and more so in the burned stream. We found that statistical methods that relied on central tendency, such as coefficient of variation (CV), did not consistently describe the variation in our data, and instead our data lent itself to geostatistical methods. We applied a time series modeling approach to remove underlying nonlinear longitudinal patterns in the data, and then used the detrended residuals to calculate the standard deviation (SD) and
build semivariograms. Consistent with our hypotheses, relative Akaike Information
Criterion (AIC) values of the underlying trends increased in model complexity. An
increasing SD and semivariogram sills for dissolved carbon indeed showed increased
heterogeneity with drying. However, contrary to our hypotheses, nutrient heterogeneity
decreased with drying in both streams. Semivariogram ranges suggested that the burned
stream had larger patch sizes, which supports the telescoping ecosystem model (TEM),
which predicts that following disturbance, ecosystems expand and increase nutrient travel
distance in streams. Utilizing spatially continuous biogeochemistry measurements
enabled us to identify spatial patterns that define ecosystem processes at previously
undetected scales.

3.1 Introduction

Stream spatial complexity, or heterogeneity, is reflected in water chemistry
patterns (Poole, 2002; Kirschner et al., 2004; Peterson et al., 2006) and has long been
recognized by ecologists as important to stream function (Townsend, 1989; Fisher et al.,
1998a; Montgomery, 1999; Fisher et al., 2004; Winemiller et al., 2010). Intermittent
streams, those that undergo seasonal drying, have pronounced structural complexity,
which changes temporally as flow expands and contracts (Stanley et al., 1997; Larned et
al., 2010; Costigan et al., 2016; Gomez et al., 2017). Yet fundamental questions about the
impact of low-flow conditions on heterogeneity and ecosystem processes remain
unanswered. In particular, how does drying alter spatiotemporal heterogeneity in stream
chemistry? Answering these questions is particularly pressing in the western United
States (Levick et al., 2008; Datry et al., 2014) where many intermittent streams are fed by
snowmelt, and many of these regions are projected to experience loss in snow cover and increased stream drying (Döll and Schmid, 2012) as well as increased fire frequency (Abatzoglou and Williams, 2016; Parks et al., 2016) with climate change. Comparing higher flow to lower flow patterns may elucidate seasonal shifts in intermittent stream structure following snowmelt.

Ecologists describe stream heterogeneity using a hierarchical framework in which patterns occur at multiple, nested scales (Frissell et al., 1986; Fisher et al., 1998a), but scales proposed within hierarchies are predetermined by geomorphology, vegetation, or habitat parameters (Frissell et al., 1986; Fisher et al., 2004; Montgomery, 1999). Units within these parameters are known as patches (Poole, 2002), which describe distinct, relatively homogenous areas compared to their surroundings that contribute to overall heterogeneity in streams (Fisher and Welter, 2006; Winemiller et al., 2010). From an ecosystem perspective (e.g., Pringle et al., 1988; Townsend, 1989; Winemiller et al., 2010), biogeochemical patches could describe areas with similar concentrations reflecting certain biological processes like nitrogen cycling (Dent and Grimm, 1999; Dent et al., 2001; Fisher et al., 2004, Fisher and Welter, 2006). At small scales, stream biogeochemical patches theoretically correspond with habitat subunits (Fisher et al., 1998a). At larger scales, process domains (Montgomery, 1999) and functional process zones (FPZ; Thorp et al., 2006) theoretically describe stream processes defined by patterns of geomorphology and flow. However, biogeochemical patterns may exist at additional, not easily predetermined scales. For instance, McGuire et al. (2014) quantified spatial patterns of biogeochemistry in perennial streams and defined nested patterns across 32 tributaries in Hubbard Brook, NH. Patch size and pattern was distinct among
chemical constituents, and some constituents exhibited patches nested within larger domains while others did not (McGuire et al., 2014). Thus, assuming the scales at which biogeochemical patterns occur may have consequences for pattern detection. Despite the theoretical emphasis on spatial complexity, aside from a few studies (e.g., Dent and Grimm, 1999; McGuire et al., 2014; Isaak et al., 2010), heterogeneity in streams has remained relatively unquantified. Using spatial and geostatistical tools can quantify patch sizes that emerge from high-spatial-scope data (Cooper et al., 1997; Peterson et al., 2006; Turner and Chapin, 2006; Scown et al., 2016).

Streams’ spatial structures vary through time with seasonal factors such as water flow (Fisher et al., 2004) or disturbance (Lake, 2000) like burn history. In a review of conceptual stream models, Allen et al. (2020) identified the telescoping ecosystem model (TEM; Fisher et al., 1998b) as one of a few conceptual stream frameworks that poses hypotheses about how changing flow could impact streams’ structure and function. Indeed, TEM predicts that stream spatial structure will expand in response to disturbances, particularly monsoonal flooding. This is because high, scouring flows can decrease habitat complexity and disrupt populations, like algal communities, which results in increased downstream nutrient transport that would otherwise be cycled/retained via uptake (Fisher et al., 1998b; Fisher et al., 2004). Findings by Dent and Grimm (1999) supported predictions of TEM, showing nitrogen and phosphorus patch size decreased as a stream progressed through succession recovering following monsoonal flooding. Disturbance and successional recovery in arid desert streams, which make up the framework of TEM (Fisher et al., 1998b; Fisher et al., 2004), are similar but distinct phenomena from seasonal stream drying following snowmelt. During stream
drying, as surface water disconnects and contracts, heterogeneity might increase (Dent et al., 2001; Fisher et al., 2004; Brooks and Lemon, 2007; Larned et al., 2010; von Schiller et al., 2010; Zimmer et al., 2013). Patch size might decrease during lower flow if longer water residence times cause increased biotic and abiotic interactions in the streambed (Dent and Grimm, 1999; Moatar et al., 2017). Shifts in groundwater exchanges or the relative influence of groundwater exchanges during low flow (Valett et al., 1994; Fisher et al., 1998a; Zimmer et al., 2013; MacNeill et al., 2020) may also decrease patch size. Alternatively, water chemistry may be more homogeneous because sources of streamflow may be better mixed at low flows if snowmelt or rain contributions to the stream diminish and deeper groundwater alone sustains baseflow (Segura et al., 2019).

Predicting how TEM applies to patch size following fire is challenging. Depending on prefire ecosystem conditions and burn severity, fires reduce terrestrial vegetation, which can a decrease evapotranspiration (Poon and Kinoshita, 2018), reorganize vegetation communities (Fellows et al., 2018), increase stream discharge (particularly in forested catchments; Kinoshita and Hogue, 2015; Wine and Cadol, 2016; Atchley et al., 2018), elevate nitrogen (Mast et al., 2016), and decrease dissolved organic carbon (DOC; Bixby et al., 2015; Cooper et al., 2015; Rodríguez-Cardona et al., 2020). However, fire can also cause increased autochthonous carbon production (Davis et al., 2013; Rugenski and Minshall, 2014; Cooper et al., 2015). Increased stream water velocity could decrease biologic uptake and extend nutrient travel distance (Fisher et al., 1998b), resulting in increased patch size similar to monsoon effects (Dent and Grimm, 1999). On the other hand, if nutrients are elevated following fire, algae communities could recover quickly, decreasing nutrient uptake time and increasing patchiness. Analyzing spatial
patterns following stream drying and fire may yield insight regarding the applicability of various theories on heterogeneity and TEM hypotheses.

To detect heterogeneity at various scales and over time, study scope must be considered, which is determined by the grain, or sample interval through time or space, and the extent, or total area or amount of time measured (Cooper et al., 1997; Gustafson, 1998; Fausch et al., 2002). Ecologists and hydrologists have argued that investigating landscape-scale stream processes, like drying and fire, requires a spatial scope of examination of intermediate (hundreds of meters to tens of kilometers) extents measured at a fine grain (tens to hundreds of meters), which requires using more spatially continuous sampling than has been traditional for water chemistry studies (e.g., Dent and Grimm, 1999; Schneider, 2001; Fausch et al., 2002; Poole, 2002; Kirschner et al., 2004; Likens and Buso, 2006; Brooks and Lemon, 2007; Gallo et al., 2012; Zimmer et al., 2013; McGuire et al., 2014). Biogeochemical studies of an intermediate spatial scope include McGuire et al. (2014), who sampled at a grain of 100 m; Zimmer et al. (2013), who sampled at grain of 50 m; and Dent and Grimm (1999), who sampled at a grain of 25 m. Although logistically difficult (Dent and Grimm, 1999; McGuire et al., 2014; Jensen et al., 2019), such an approach is appropriate for investigating longitudinal stream chemistry patterns that change with stream drying and following fire because each are processes that have meter-scaled heterogeneity with consequences for watershed-scaled water chemistry (MacNeille et al., 2020). Spatially continuous and repeated fine-grain measurements have revealed stream drying patterns including hysteresis (Jensen et al., 2019), diel evapotranspiration (Warix, 2020), strong correlations with dissolved organic carbon (DOC) concentrations (Hale and Godsey, 2019), variable nitrogen groundwater
inputs (Dent and Grimm, 1999), variation in groundwater contributions of ions (Zimmer et al., 2013), and highly variable organic carbon sourcing (MacNeille et al., 2020). However, intermittent stream patterns analyzed at a repeated high spatial scope pose both an opportunity for rich description and a statistical challenge as well (Cooper et al., 1997).

Traditional statistical techniques include those that rely on the central tendency of the data such as mean, median, minimums/maximums, or coefficient of variation (CV; Brooks and Lemon, 2007; Asano et al., 2009; Gomez et al., 2009; von Schiller et al., 2010; Zimmer et al., 2013; Hale and Godsey, 2019). However, the central tendency of data is limited in its ability to describe the patterns or the complex variation along a stream profile, which include possible nonlinear underlying patterns in a given stream segment, the spatial relationships of fine-grain measurements (Peterson et al., 2006), or the loss of data points as streams dry (MacNeille et al., 2020), and changing means (Dent and Grimm, 1999). These limitations are especially problematic if the objective is to understand the changing complexity of spatial patterns. Instead, geostatistical tools that use spatial statistics offer a more appropriate approach. Yet, with the exception of McGuire et al. (2014) and Dent and Grimm (1999), few examples exist of applying geostatistics to biogeochemical studies such that patterns can emerge from field data.

Applying terrestrially developed geospatial techniques to streams is nuanced, and tools developed for a time series approach can be applied to questions of spatial patterns in streams (Cooper et al., 1997; Ganio et al., 2005; Peterson et al., 2006; Scown et al., 2016). Specifically, fine-grain measurements violate sample independence and display autocorrelation, which is especially true in streams where water moves between locations.
in a downstream direction (Cooper et al., 1997; Peterson et al., 2006). Thus, basic assumptions of parametric statistics are violated. In stream networks, scholars face the challenge of having a branched spatial structure where neighboring measurements are connected in structurally complex ways (either Euclidean distance, flow-connected, or flow unconnected; Peterson et al., 2006; Ver Hoef et al., 2006; Isaak et al., 2014), so models must accommodate these nuances. Predictive models have been developed that use modified kriging to describe large watershed spatial patterns (e.g., Isaak et al., 2010; Money et al., 2011; Isaak et al., 2014). For example, Isaak et al. (2010) were able to apply network statistics to model thermal patterns and patches that were suitable for rainbow trout (*Oncorhynchus mykiss*) and bull trout (*Salvelinus confluentus*). While useful, this method applied species-specific thermal parameters to examine patch size (Isaak et al., 2010). These predetermined parameters can be juxtaposed to more exploratory approaches that allow patterns to emerge from the statistical relationships between field-collected, empirical, high-spatial-scope data themselves (Ganio et al., 2005; Thorp et al., 2006; Scown et al., 2016). Stream chemistry is particularly well suited to study emergent patterns because of its ever-present nature (compared to organisms or habitat). Still, heterogeneity in nonbranching headwater stream segments, which are considered highly variable (Datry et al., 2014), have remained largely unquantified.

Semivariograms are geostatistical tools that quantify spatial relationships between data and produce parameters to describe the heterogeneity. Used in the process of kriging, semivariograms use the residuals of modeled data and average variability (as calculated by semivariance, the response variable) at a given separation distance between points (binned, independent variable). The averaged variance at a given distance builds
the empirical semivariogram graph and fits a semivariogram model. The resulting parameters are (Figure 3.1,C) (a) a sill, an asymptote that occurs at a separation distance at which points are statistically independent and represents overall variation as semivariance; (b) a range, or ascending leg of the model, occurs at separation distances at which points are autocorrelated; and (c) the nugget, which occurs when at separation distance zero, a semivariance greater than zero is calculated (resulting in not defining a distance at which semivariance is zero) and is interpreted as either not sampling at fine enough grains to determine the autocorrelated relationship or measuring a variable where spatial relationships do drive the patterns. Semivariograms have been used to describe the geomorphology of streambeds using high resolution imagery (Chen et al., 2019; Scholl et al., 2020), but less frequently the chemistry patterns that reflect biophysical processes (e.g., Dent and Grimm, 1999; McGuire et al., 2014). One strength of spatial statistical approaches is that patterns can emerge from spatially continuous data (Cooper et al., 1997; Scown et al., 2016).

In this study, we explored biogeochemical patterns at a high-spatial-scope measurement in surface water of two headwater, intermittent streams, one unburned and one burned, as they dried. We built upon results from MacNeille et al. (2020) that showed longitudinal trends in stream chemistry with drying and fire apparently controlled by shifts in groundwater contributions and showing highly variable allochthonous and autochthonous carbon contributions as streams dried. Here, we asked the following questions: How does the heterogeneity of biogeochemical patterns change in surface water as headwater streams dry and following fire? What is the patch size in intermittent headwater streams for biogeochemical analytes, and does it vary with analyte? How do
spatial patterns and patch sizes vary in surface water as headwater streams dry immediately following fire?

To investigate these questions, we quantified patterns from stream chemistry data collected at 50-m intervals over ~2,500 m in April, May, and June 2016. In addition, we collected samples at a more fine grain (10-, 25-, and 50-m intervals) during low flow in July 2017, which allowed patterns to emerge using a number of statistical approaches. These included CV, standard deviation (SD) of modeled data, and semivariograms. Due to possible increases in patchiness associated with drying, we hypothesized that with drying, heterogeneity would increase and that the patch sizes for each constituent would decrease. Specifically, with drying, we expected increases in SD of modeled data and a higher semivariogram sill, shorter semivariogram range, and larger nugget (Figure 3.1,C). We also expected each analyte to exhibit a distinct patch size, as defined by the semivariogram range. Finally, we hypothesized greater heterogeneity would occur following wildfire as the burned stream dried, and that this would diminish 2 years postfire. Thus, we expected the low flow sill to decrease from 2016 to 2017 in the burned stream (Figure 3.1,E).
Hypothesized Semivariograms.

Notes: (A) and (B) represent two hypothetical watersheds where variations in stream chemistry concentrations are depicted by colored circles along the stream. (A) depicts more surface connectivity and a more homogeneous spatial pattern. (B) depicts a stream experiencing surface water fragmentation due to drying where the spatial pattern is more heterogeneous. (C) Semivariograms have three parameters: (1) sill, or overall semivariance ($\gamma$), separation distances (h; binned distances) at which points are statistically independent; (2) range, the ascending part of the graph that indicate the separation distances at which points are statistically dependent on each other; (3) nugget, at a separation distance of zero, the semivariance is not zero due to either too coarse sampling, error of measurements, or something other than space explains the relationship at that scale. We hypothesized that with stream drying, streams become more heterogeneous. We expected to observe a higher sill, shorter range, and larger nugget. (D) At low flow, we hypothesized that immediately following fire the stream (orange) will be more heterogeneous than after ~2 years of recovery (dark red). Separation distances are depicted by the brown (shorter separation distance) and red (longer separation distance) brackets and arrow pointing to where in the semivariogram the separation distances might be averaged. (E) We expected to observe a lower sill, longer range, and smaller nugget ~2 years after fire.
3.2 Methods

3.2.1 Study Approach

3.2.1.1 Study Sites. Our study took place in two subwatersheds nested within the Reynolds Creek Experimental Watershed and Critical Zone Observatory (RC CZO), located in southwest Idaho, United States. The RC CZO was established by the USDA Agricultural Research Service (ARS) in 1960 as a representative watershed of the Intermountain West region in the United States (Marks et al., 2011). With a nearly 1,000-m elevation gradient, RC CZO climate varies from a mean annual precipitation (MAP) of 250 to 1,100 mm/yr and mean annual temperatures (MAT) of 5.5 °C to 11°C. Precipitation at lower elevations is primarily rain and at highest elevations is primarily snow (Nayak et al., 2010; Kormos et al., 2014). Snowmelt drives peak discharge in streams across the watershed (Pierson et al., 2001). RC CZO vegetation transitions from Wyoming sagebrush steppe at lower elevations to mountain sagebrush (Artemisia tridentata), western juniper (Juniperus occidentalis), aspen (Populus tremuloides), and coniferous forest mostly comprising Douglas fir (Pseudotsuga menziesii), at higher elevations (Seyfried et al., 2018). Riparian vegetation at midelevations specifically includes alder and willow.

To evaluate stream biogeochemical spatiotemporal pattern complexity, we studied two intermittent, headwater streams, Johnston Draw Creek and Murphy Creek (Patton et al., 2018; Pierson et al., 2001; Seyfried et al., 2000). The two streams were similar in that each had a dropbox v-notch 90° weir at the outlet where streamflow data were collected (Seyfried et al., 2000; Godsey et al., 2018), and they were comparable in basin area, discharge, elevation, MAP, MAT, and aspect (Seyfried et al., 2000; MacNeille et al.,
2020). The streams differed in their burn history and their lithology. In August 2015, a wildfire classified as moderate within the study area, the Soda Fire, burned 68 km² of the RC CZO, including Murphy Creek, and resulted in an estimated 60% bare ground in this catchment (Vega et al., 2020). In contrast, wildfire had not been reported in Johnston Draw during the previous 60 years that the ARS monitored the watershed. Hereafter we refer to Johnston Draw as “unburned JD” and Murphy Creek as “burned MC.” The lithology of unburned JD is predominantly granodiorite with some quartz latite and rhyolite in the upper portions of the catchment and basalt flow lower in the outflow (McIntyre, 1979), whereas the lithology of burned MC is mostly Salmon Creek Volcanics, which are basaltic. Unburned JD is described in more detail by Godsey et al. (2018) and Patton et al. (2018, 2019), and burned MC is described in more detail in Seyfried et al. (2000), Pierson et al. (2001), and Vega et al. (2020).

3.2.1.2 Sampling Design. To investigate spatiotemporal stream biogeochemistry patterns, we continuously sampled surface water moving upstream along the length of each stream and repeated this at several time points in 2016 and 2017 as streams dried following snowmelt. Building upon previous studies (Dent and Grimm 1999; Zimmer et al., 2013, McGuire et al., 2014), we sampled at varying densities. Initially we sampled both streams at 50-m intervals in April, May, and June 2016, moving upstream from the outlet weir (Figure 3.2). We refer to the fine-grain sampling sites as 50-m sites, and collectively as the longitudinal profile, making up longitudinal patterns of each stream. We define the overall stream extent or distance (~2,500 m in April to ~2,000 m in June) as a stream segment (as described in MacNeille et al., 2020). We repeated this sampling
in July 2017 when discharge was comparable to June 2016. We sampled at 10-, 25-, and 50-m intervals (Figure 3.2) and samples were collected until a sample size for each density was as close to \( n = 50 \) as possible. We sampled this way in order to satisfy requirements for robust spatial statistical analysis. All water samples were collected at the stream thalweg as suggested by Dent and Grimm (1999).

Figure 3.2

2016 and 2017 Sampling at Reynolds Creek Experimental Watershed Critical Zone Observatory.

Notes: (A) Reynolds Creek Critical Zone Observatory (RC CZO) in southwest, Idaho, USA. Samples were collected until either 50 samples were collected (as in the 10-m case) or the stream had no surface water (as in the 25-m and 50-m cases). Longitudinal sampling sites start at a weir (yellow star) and extend to the uppermost reach with
observed surface flow in each stream (green “x”), as identified in the field in April 2016. Color indicates elevation relief.

3.2.1.3 Stream Water Collection. To characterize the streams, in situ physical and chemical parameters including temperature (°C), dissolved oxygen (DO mg/L), and pH were collected along with samples at each 50-m site. Stream pH and temperature were measured using an Oakton pH 110 Series probe (Vernon Hills, IL) calibrated with 4, 7, and 10 pH standard solutions. DO was measured with a YSI Dissolved Oxygen probe. Sampling of field sites occurred throughout the stream each year, but in 2017 sample intervals differed either due to time limitations or shallow surface area. For instance, we measured temperature at every 2017 reach site but DO and pH only at every 50-m reach site; surface water samples were collected at every possible interval. We did not measure discharge at each 50-m site because low volume prevented reliable direct discharge measurements, and the alternative, dilution gauging, would have influenced the chemical composition of our samples. Instead, we report the discharge collected at the respective stream weirs.

We collected surface water in 18.2 MΩm distilled (DI)-rinsed and -leached 250-mL amber bottles by starting at the outlet weir and moving in the upstream direction. At each sampling location, the bottles were rinsed three times with the stream sample and then filled to eliminate any remaining headspace. Samples were carried out of the field by foot (~60–35 bottles) and refrigerated (4°C) until they were filtered within 72 hours. Filtering for dissolved organic carbon (DOC), and total dissolved nitrogen (TN) occurred by vacuum through pre-combusted 0.7µm Whatman glass fiber filters (GFF) into pre-sample rinsed 60 ml amber high-density polyethylene bottles (HDPE) bottles. The remaining sample was used for dissolved inorganic carbon (DIC) and anion (Cl⁻) analysis
through a hand syringe-filtered through a 0.45 µm nylon filter into 60 ml clear high-density polyethylene bottles (HDPE) bottles.

**3.2.1.4 Biogeochemical Laboratory Analyses.** In both 2016 and 2017, we analyzed surface water for DIC, DOC, anions (chloride or Cl\(^{-}\) and sulfate or SO\(_4\)^{2-}\), TN, and nutrient (ammonium-N, nitrate-N [NO\(_3\^-\)-N], hereafter nitrate, and orthophosphate-P [PO\(_4^{3-}\)-P], hereafter phosphate or PO\(_4^{3-}\)) concentrations. Additional constituents were analyzed in 2016 and are presented in MacNeille et al. (2020) but are not included here because they were not reanalyzed during the 2017 sampling. In brief, MacNeille et al. (2020) showed that conservative solutes increased in concentration with drying and groundwater inputs and were not subject to changes in evaporation with drying.

In the lab, DIC, DOC and TN concentrations were measured on a Shimadzu (Tokyo, Japan) TOC-V 266 CSH, which was equipped with an ASI-V autosampler and TNM-1 chemiluminescence detector for TN. For concentrations >1 mg C/L on measurements, <2%–3% error was accepted. High values of DIC (~10–30 mg DIC/L as C) resulted in incomplete and variable removal of DIC from the nonpurgeable organic method (NPOC). Instead, DOC was calculated by total carbon (TC) measured minus the measured DIC and propagated error. In a cross-validation analysis, our DOC values (TC-DIC) and NPOC values from the Perdrial Environmental Biogeochemistry Lab at University of Vermont (Burlington, VT) agreed reasonably (n = 140, r = 0.65). NO\(_3^-\), NH\(_4^+\), and PO\(_4^{3-}\) were measured on an automated chemical spectrophotometer Westco Discrete Analyzer (Unity Scientific, Brookfield, CT). We accepted a <10% error for NO\(_3^-\)N, and PO\(_4^{3-}\)P concentrations <1 mg/L and a 20% error for NH\(_4^+\)N <0.10 mg/L.
However, with the exception of April 2016 samples from burned MC collected closest to the fire and during high snowmelt, both NO\textsubscript{3}\textsuperscript{-} and NH\textsubscript{4}\textsuperscript{+} were below or at instrument detection limit, so for most statistical analysis we utilized TN. Anions were measured using ion chromatography on a ThermoFisher Scientific Dionex (Sunnyvale, CA) ICS-5000; we accepted <10% error for sample <1 mg/L and 2%–3% for samples >1 mg/L.

We used DIC, Cl\textsuperscript{-}, DOC, TN, and PO\textsubscript{4}\textsuperscript{3-} in the spatial analyses we report here. We chose these constituents based on the results of multivariate analyses reported in MacNeille et al. (2020).

### 3.2.2 Quantifying Spatiotemporal Patterns with Stream Drying

As per Cooper et al. (1997), we used multiple approaches to quantify spatial complexity in stream chemistry. We first explored traditional methods and calculated the mean, SD, and CV, which emphasized the central tendency of the data to describe changes. We reported means (Supplementary Figure 3.1) but had difficulty utilizing CV to describe variation. CV was sensitive to (a) underlying, nonlinear longitudinal patterns that changed as sites dried, and (b) increasing concentration means with drying (Supplementary Figure 3.1). Given these issues, it was not clear how to best approach the data when calculating CV. We calculated CV using four different approaches that included using all the 2016 data or using only 50-m sites in 2016 that could be compared across all three months (i.e., holding sample size constant with 50-m sites that had water present all three months) and a moving average (per Asano et al., 2009) to address underlying patterns or not (Supplementary Figure 3.1). CV results were not consistent (Supplementary Figure 3.1), so instead we evaluated spatial patterns using a time series
approach to detrend nonlinear longitudinal data, evaluated changes in spatial complexity over time, and used these more robust approaches to interpret spatiotemporal patterns.

3.2.2.1 Spatial Statistical Analyses Overview. To evaluate spatial dependence in stream chemistry, we followed a workflow displayed in Figure 3.3. We first fit nonlinear models to the longitudinal patterns to detrend the data (Figure 3.3,1). We then compared model complexity across the 2016 months for each analyte (Figure 3.3,2). Using residuals (Figure 3.3), we then calculated SD (Figure 3.3,4a) and built semivariograms (Figure 3.3,4b).

Figure 3.3
Workflow Diagram.

Notes: Diagram proceeds left to right (1-4) for the spatial statistical method we employed and our predictions (red text) for a hypothetical data set during low flow (open circles) compared to high flow (solid circles). We (1) modeled the data to detrend using the (generalized additive model) GAM smoother to detrend and then (2) tested this using the AIC/n methods described in Section 3.2.2.2. The residuals (3) of the detrended data were then used to calculate the (4a) SD of the modeled data (SDGAM) and constructed
semivariograms (4b). Each dot on the semivariograms (4b) was labeled with the number of pairs that are averaged into the separation distance bin.

3.2.2.2 Detrending Nonlinear Longitudinal Patterns Using a Time Series Approach. Nonlinear longitudinal patterns underlaid our stream data, were constituent specific, and changed with drying (MacNeille et al., 2020). To quantify the spatial relationship between reach sites within the streams, we removed the underlying downstream trends. Visual inspection of the longitudinal stream profiles indicated that a linear model was inappropriate for our data. Instead, knowing we wanted to quantify the autocorrelation (statistical dependence) between fine-grain data points, we used a time series approach of applying smoother to model and detrend the spatial patterns (Cooper et al., 1997). We evaluated smoother options including generalized additive model (GAM), lowess, and smooth.spline models from which we could extract Akaike Information Criterion (AIC; Akaike, 1974) values. AIC is a model selection tool that optimizes model fit to the data by balancing the best fit (lowest error) with the least complex fit (lowest degrees of freedom). The lowest AIC value represents the most parsimonious model fit. Being able to extract AIC was important because we utilized it as a relative value to quantify and compare model complexity (Section 3.2.2.3).

Of the smoothers, we determined the GAM smoother to be the most appropriate for two reasons. First, GAM automatically calculates AIC and uses the optimal model (lowess and smooth.spline required individually calculated AIC values). Second, an analysis of a simulated data set similar in behavior and size to our data, demonstrated GAM consistently modeled the data most similarly to the known data function (Supplementary Figure 3.2). Specifically, GAM is a generalized linear model (GLM) in which the predicted linear model is a sum of localized functions. We used the R package
mgcv using the function gam.fit (R version 1.1.456; Wood, 2011). The GAM model was determined by the convergence of a generalized cross validation (GCV) estimation and un-biased risk estimator (UBRE) model selection criterion. The optimal GAM model was chosen by using penalized likelihood maximization, and GAM specifically uses penalized iteratively reweighted least squares (P-IRLS; Wood, 2000). We reported the sum of squared errors (SSE) of linear model (SSE$_{LM}$) versus an SSE of the GAM model (SSE$_{GAM}$) to assess which was a more appropriate model (Supplementary Table 3.1). Based on those results, we proceeded using the GAM smoother to first detrend our data each month in order to analyze the temporal and spatial patterns explicitly.

3.2.2.3 Temporal Shifts: Quantifying Changes in Longitudinal Patterns. In order to evaluate how the longitudinal stream profile trends changed over time, we quantified the GAM model for each 2016 month (April, May, and June) and each constituent and compared between months. To accomplish this, we first calculated a complexity measure for each stream month by fitting the same nonparametric GAM smoother in R to each stream and month for a given constituent (Equation 3.1):

Equation 3.1: $C = N - df\text{-resid}$

where $C$ term is complexity, or number of parameters being used, $N$ is sample size, and $df\text{-resid}$ is the regression degrees of freedom.

Thus, the GAM smoother changed curvature to fit the data each month, and we then quantified qualities of the smoother to compare between the months of each stream as depicted hypothetically in the workflow diagram (Figure 3.2). Two ways that smoothness of fit was quantified were (a) the simplicity or complexity of the curve fitted, measured in
terms of degrees of freedom, and (b) how closely the data followed or “fit” the curve, measured in the size of the sum of the squared residuals, which is also SSE. Overall, a smoother, less complex curve would have smaller values for both of these measures and underlying patterns would be unique to each constituent. To compare a constituent’s smoother fit across the three months in each stream, we used AIC (Akaike, 1973). In this context, we used AIC as a tool to measure smoothness of fit to compare between months of a given constituent, by rationally balancing two criteria, rather than as a model selection tool. However, as streams dried, each month had different numbers of observations, so AIC/n was used to normalize by sample size (Equation 3.2). We ran two trials of these analyses for comparison, one using all the data for each month where sample size changed between months, and one holding sample size constant using only data from sites with water sampled each month. Lastly, given the small sample sizes of drier months, we explored the use of AIC corrected, but found the results did not change our interpretation of temporal patterns, and, therefore we judged that AIC/n was the more straightforward approach. The following Equation 3.2 normalizes AIC for sample size:

\[
\text{Equation 3.2: } \frac{AIC_{GAM}}{N} = \frac{\log(SSE)}{N} + \frac{2C}{N}
\]

where SSE is sums of squared error (calculated by residuals), C is a term for complexity generated by the GAM fit, and N is sample size.

By quantifying the relative change of the GAM fit using AIC/n, SSE/n, and C/n, we were able to represent the temporal changes in terms of complexity of GAM smoother fit for a given constituent as each stream dried from April to June.

We wanted to calculate the spatial relationships independent of the underlying patterns and to use as many points as possible each month to capture the complexity and
effect on patterns as reach sites dried. The subsequent analyses could be compared independent of data loss with drying and were not sensitive to changing mean as was CV; thus, we did not continue to exclude points that only experienced one or two months of surface water present as we examined spatial relationships.

3.2.2.4 Quantifying Spatial Heterogeneity: SD of GAM fit and Semivariograms. After detrending the stream data as described (Section 3.2.2.2), we quantified the spatial heterogeneity of each stream by calculating SD_{GAM} and semivariance using all the data points of solute concentrations for each month. SD_{GAM} was calculated in R (version 1.1.456) after GAM fit was applied to data, the SD of the residuals was taken and compared to semivariogram results. As recommended by Cooper et al. (1997), we thought it important to use more than one analysis to quantify spatial patterns, as semivariograms are sensitive to parameters set by the researcher (Peterson et al., 2006).

Semivariograms were built by using the average semivariance (Equation 3.3) of autocorrelated pairs for a given distance \( h \), or bin size:

Equation 3.3: \( \gamma(h) = \frac{1}{2|N(h)|} \sum_{N(h)} (z_i - z_j)^2 \)

where \( \gamma \) is the semivariance of \( h \), a given separation distance, \( N \) is sample size, and \( z \) is the concentration at a given site, \( i \) or \( j \), of \( h \) separation distance between \( i \) and \( j \).

Our sampling method assumed the single-point measurement at the thalweg of the stream was representative of the stream cross section at that location. However, semivariograms are built using \((x, y)\) coordinates, so we generated a random \( x \) (0–0.1 m)
and used the actual stream distance from downstream to upstream (0 m located at the weir) to build \((x, y)\) coordinates. Importantly, stream distance considered points all to be possible neighboring pairs. Given that we did not need to consider stream network branching (Peterson et al., 2006), these distances were appropriate. Semivariogram construction is sensitive to several parameters including the number of pairs, bin size, and number of bins (Webster and Oliver, 1992; Cooper et al., 1997), and small sample sizes cause uncertainty by influencing each of these parameters (Cooper et al., 1997; McGuire et al., 2014). The semivariogram is built using average semivariance for a given separation distance and requires about 50 pairs contributing to the average to be statistically robust (Webster and Oliver, 1992; McGuire et al., 2014). We constructed semivariograms in R using the automap package function autofitVariogram (R version 1.1.456; Hiemstra et al., 2008), which estimated the sill, range, and nugget from the data by using fit.variogram function in the gstat package of R (Gräler et al., 2016; Pebesma, 2004). We used a minimum of 50 pairs per bin (Cooper et al., 1997; McGuire et al., 2014), and we held the spherical model constant as has been standard (Dent and Grimm, 1999; Cressie, 2005; Peterson et al., 2006) to generate a “best fit” semivariogram sill (\(\text{Sill}_{BF}\)). We explored the parameters of the smallest SSE semivariogram model and compared to the spherical semivariogram model (Supplementary Table 3.2). We found that almost always the sills were the same and that spherical was the most common optimal semivariogram model when considering the 2016 and 2017 data (Supplementary Table 3.2). However, in 2016 we had smaller sample samples owing to drying and our design, so we conducted a sensitivity analysis to determine the error in our semivariogram sills running the spherical model for each data set twenty times, each with
one random point removed. We referred to this as leave-one-out (LOO) sensitivity. Thus, for 2016 months we reported the best fit sill and the range of LOO sills ($\text{Sill}_{\text{LOO}}$) possible based on our sensitivity analysis, which we refer to as the sensitivity margin. Though sample sizes for the 2017 data were more robust, we conducted the LOO sensitivity analysis for reference for sensitivity with 2016 data.

3.3 Results

3.3.1 Hydrological Conditions

Stream drying and decreasing discharge from early season to later season (Figure 3.4) characterized the hydroclimatic conditions of both intermittent streams in 2016. Patterns in 2016 are detailed in MacNeille et al. (2020) whereas patterns in 2017 are reported in this study and comparisons to 2016 are emphasized here. In April 2016, both streams had surface water at 100% of the 50-meter sites ($n = 57$ for unburned JD and $n = 59$ for burned MC) with all water present in all sites over the sample area including the weir site. Weir discharges were $\sim 0.02$ m$^3$/s for the unburned JD and $\sim 0.017$ m$^3$/s for the burned MC. By June 2016, the unburned JD experienced slightly less drying with 65% of 50-m sites ($n = 37$) with surface water present but slightly greater decrease in weir discharge ($<0.005$ m$^3$/s) relative to the burned MC, where 64% of 50-m sites ($n = 38$) had surface water present and weir discharge of $\sim 0.005$ m$^3$/s. In 2016, the unburned JD experienced greatest drying from May to June (89% in May to 65% in June of 50-m sites had surface water) whereas more of the burned MC dried earlier from April to May (73% of 50-m sites had surface water in May).
Both streams were drier in July 2017 than in June 2016 when compared to the 50-m site baseline measured in April 2016; however, the difference in drying was greater in unburned JD. Compared to April 2016 (n = 57), unburned JD had surface water present for 39% of 50-m sites (n = 22) in July 2017 (covering ~1,300-m extent) compared to 65% of 50-m sites (n = 37) in June (covering ~2,100-m extent). Compared to April 2016 (n = 59), burned MC had a more similar number of 50-m sites with surface water present in July 2017 relative to June 2016. In July 2017, burned MC had surface water present for 56% of 50-m sites (n = 33 over ~1,800-m extent) compared to 64% of 50-m sites (n = 38 over ~2,000-m extent) in June 2016. Though both streams had more surface water present at the 50-m grain and over a larger extent in June 2016 than July 2017, with larger differences in unburned JD, the streams had similar discharge patterns in June 2016 and July 2017.

Weir discharge patterns were comparable between streams for each late season sample period. Despite the percentage water present patterns in July 2017 for each stream, weir discharge was slightly higher for unburned JD in July 2017 than it had been in June 2016 (Figure 3.4,A). June 2016 weir discharge in unburned JD was lower (<0.005 m$^3$/s) than burned MC (~0.005 m$^3$/s). However, in July 2017 weir discharge was slightly higher for unburned JD and lower in burned MC, all of which were <0.005 m$^3$/s.

In situ field measurements showed increases in temperature in each stream in July 2017 (mean ± SE 15.2 ± 0.14 in unburned JD and 20.6 ± 0.35 in burned MC) compared to June 2016 (mean ± SE 14.0 ± 0.8 in unburned JD and 16.1 ± 0.6 in burned MC). On the other hand, pH was more comparable in unburned JD in June 2016 and July 2017 (mean ± SE 7.4 ± 0.07 in June 2016 and 6.8 ± 0.05 in July 2017), but in burned MC pH
was higher in June 2016 than July 2017 (mean ± SE 6.8 ± 0.09 in June 2016 and 7.1 ± 0.04 in July 2017). DO (mg/L) was also higher on average for both streams in June 2016 than July 2017 (mean ± SE 9.1 ± 0.4 in June 2016 and 8.5 ± 0.3 in July 2017 in unburned JD; 10.4 ± 0.1 in June 2016 and 7.6 ± 0.3 in July 2017 in burned MC) compared to April 2016. DO showed relative variability longitudinally in July 2017 (Figure 3.4).

Figure 3.4

*Discharge of the Unburned JD and Burned MC Measured at the Weir.*

Notes: 2016 (lighter shade) and 2017 (darker shade) discharge of the (A) unburned JD (blue) and (B) burned MC (orange) measured at the weir (most downstream sample point); bars in A and B reflect the sample dates and symbols are placed at the discharge at those sample dates. Longitudinal patterns of (B and C) temperature (°C), (D and E) pH (-log [H+]), and (F and G) dissolved oxygen (DO mg/L). Sample sizes (n) indicated in legend are the number of sites with water present, but number of sites with a given field measurement differed due to differing strategies using the instruments (i.e., measuring DO at 50 m in 2017, not at 10, 25, and 50 m), limitations due to shallow surface water, or field instrument malfunctions.
3.3.2 Solute Patterns with Drying

Distinct longitudinal patterns were exhibited in each stream with stream drying in 2016 as reported by MacNeille et al. (2020), and generally these patterns were repeated in July 2017 (Figure 3.5) in both streams. DIC concentrations were greater in 2017 for both streams than June 2016, as was Cl for unburned JD. In unburned JD the uppermost stream reaches (~1,800 m to 2,500 m in Figure 3.4) and downstream near the weir (0 m Figure 3.5) were associated with high constituent concentrations in 2016. The uppermost segment of the unburned JD stream was characterized by meadow grasses riparian vegetation and south-facing aspect compared to larger shrubs and trees (alder, juniper, sagebrush) downstream. These unburned JD sites were dry in July 2017 and therefore were not resampled in the subsequent year.

DOC, TN, and PO₄³⁻ behaved similarly in both streams across time, with some 2017 exceptions. DOC had relatively invariant temporal patterns for both streams in 2016, but had lower mean concentrations in July 2017 in unburned JD (Table 3.2). TN mean concentrations had a decreasing trend through the 2016 growing season; however, the burned MC stream had a sixfold decrease and July 2017 concentrations were similarly low (June 2016 0.19 ± 0.037 and July 2017 0.20 ± 0.005). PO₄³⁻, again similar between streams, had slightly increasing mean concentrations with stream drying from April to June 2016 and then again increased with increased stream drying in 2017.
Figure 3.5

2016 and 2017 Longitudinal Concentration Plots.

Notes: Samples were collected at 50-m intervals in 2016 and at 50-, 25-, and 10-m intervals in July 2017. Sample sizes between sample periods are indicated in the legend next to each month symbol and include the weir.

3.3.3 Quantifying Biogeochemical Spatiotemporal Patterns with Stream Drying

3.3.3.1 Detrending Nonlinear Longitudinal Stream Patterns. General additive model (GAM) fits removed underlying stream patterns and resulted in better fits than...
linear models as measured by sum squared error (SSE). For example, using the June DIC fit comparison for each stream, we observed that the linear fit was not appropriate, and linear model residuals confirmed this with remaining secondary patterns in residual plots (Figure 3.6). Furthermore, the linear model SSE for DIC was greater than the SSE of the GAM model by about sevenfold for unburned JD and twofold for burned MC. Similarly, higher $SSE_{\text{linear}}$ values were found for all constituents in both streams for nearly all months with some DOC expectations (Supplementary Table 3.1). We proceeded using the GAM fit for consistency.

Figure 3.6

*Linear Model Versus GAM Model Fits and Residual Plots.*

Notes: The left graphs illustrate the difference between using a linear fit (dashed line) and the GAM smoother fit (solid line) for June DIC concentrations in the (A) unburned JD (blue) and (D) burned MC (red) stream. Sums squared error (SSE) for the GAM
(SSE$_{\text{GAM}}$) and the linear model (SSE$_{\text{linear}}$) are listed for each stream. Residual plots for each the GAM model (A1 and B1) and the linear model (B1 and B2) are to the right.

### 3.3.3.2 Temporal Shifts: Quantifying Changes in Longitudinal Patterns.

Comparing longitudinal model complexity using AIC showed that relative model complexity increased for conservative analytes with stream drying but was more variable for biologically reactive analytes and in burned MC. Generally, changes in AIC/n values indicated increases in the complexity term (Equation 3.1 and 3.2), but there were exceptions such as unburned JD DOC in June (Supplementary Table 3.3 and Supplementary Figure 3.3). In unburned JD, DIC and Cl$^-$ increased in model complexity (indicated by increasing AIC values, Table 3.1, A, C and Supplementary Table 3.3 and Supplementary Figure 3.3) from April to June, while DOC AIC values increased slightly in JD unburned but decreased in unburned MC. Differences between the streams were calculated for TN and PO$_4^{3-}$, which increased in burned MC, but in unburned JD relative model complexity comparisons differed by the data set. TN and PO$_4^{3-}$ increased only when 50-m interval sites that dried were excluded in unburned JD (Table 3.1, C), but when using all the data (Table 3.1, A) to compare relative model complexity, TN and PO$_4^{3-}$ decreased in value from April to June in unburned JD.

In burned MC, constituents were more variable. DIC and DOC decreased with stream drying from April compared to June while Cl$^-$, TN, and PO$_4^{3-}$ increased. Some analyte models had the highest AIC value and complexity in May (Table 3.1 and Supplementary Table 3.2). This may be explained by the combined effects of increasing model complexity with drying and still having water present in upstream sites (which dried in June but had high solute concentrations) which each contributed to the overall model complexity for the longitudinal pattern. Such was the case for PO$_4^{3-}$ patterns in
both streams and the unburned JD’s DOC and the burned MC’s DIC patterns (Table 3.1). We did not compare 2017 model complexity to 2016 because we did not observe the longitudinal patterns throughout the drying season for an appropriate comparison. In addition, due to the loss of sites that may have influenced pattern complexity, to quantify the longitudinal patterns temporally, we explored model complexity across 2016 months by calculating AIC using only sites that had surface water present in April, May, and June (Table 3.1, C, D)
Table 3.1

AIC Values Based on the GAM Fit for the Data and Normalized by the Sample Size (N).

<table>
<thead>
<tr>
<th></th>
<th>(A) Unburned JD</th>
<th>(B) Burned MC</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIC</td>
<td>Cl⁻</td>
<td>DOC</td>
<td>TN</td>
<td>PO₄³⁻</td>
<td>DIC</td>
<td>Cl⁻</td>
<td>DOC</td>
<td>TN</td>
<td>PO₄³⁻</td>
<td>DIC</td>
<td>Cl⁻</td>
<td>DOC</td>
<td>TN</td>
<td>PO₄³⁻</td>
<td>DIC</td>
<td>Cl⁻</td>
<td>DOC</td>
<td>TN</td>
<td>PO₄³⁻</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>0.26</td>
<td>0.20</td>
<td>0.41</td>
<td>0.27</td>
<td>0.27</td>
<td>0.34</td>
<td>0.17</td>
<td>0.32</td>
<td>0.26</td>
<td>0.05</td>
<td>0.45</td>
<td>0.12</td>
<td>0.16</td>
<td>0.36</td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td>0.44</td>
<td>0.30</td>
<td>0.45</td>
<td>0.31</td>
<td>0.28</td>
<td>0.31</td>
<td>0.41</td>
<td>0.20</td>
<td>0.40</td>
<td>0.13</td>
<td>0.31</td>
<td>0.41</td>
<td>0.20</td>
<td>0.40</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June</td>
<td>0.59</td>
<td>0.51</td>
<td>0.43</td>
<td>0.15</td>
<td>0.14</td>
<td>0.31</td>
<td>0.41</td>
<td>0.20</td>
<td>0.40</td>
<td>0.13</td>
<td>0.31</td>
<td>0.41</td>
<td>0.20</td>
<td>0.40</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Only sites with surface water in all three months

|        | (C) Unburned JD | (D) Burned MC |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |        |
|--------|----------------|---------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|        | DIC | Cl⁻ | DOC | TN | PO₄³⁻| DIC | Cl⁻ | DOC | TN | PO₄³⁻| DIC | Cl⁻ | DOC | TN | PO₄³⁻| DIC | Cl⁻ | DOC | TN | PO₄³⁻|
| April  | 0.43| 0.21| 0.58| 0.40| 0.18 | 0.31| 0.12| 0.40| 0.28| 0.02 | 0.51| 0.13| 0.22| 0.40| 0.26 |
| May    | 0.54| 0.35| 0.63| 0.41| 0.15 | 0.51| 0.13| 0.22| 0.40| 0.26 | 0.32| 0.51| 0.23| 0.39| 0.14 |
| June   | 0.62| 0.51| 0.48| 0.43| 0.37 | 0.32| 0.51| 0.23| 0.39| 0.14 |

Notes: (A) and (B) are AIC values calculated using each month’s entire data set and normalized by N. For (C) and (D) (labeled “Only sites with surface water in all 2016 months”), the AIC values calculated using only sites that had surface water sampled in all three months in order to compare among months without the impact of stream drying (losing sites between months).
3.3.3.3 Quantifying Spatial Heterogeneity: SD of the GAM fit and Semivariograms. SD\textsubscript{GAM} and semivariograms showed alignment for all constituents in all months. SD\textsubscript{GAM} and semivariograms showed overall variance increased for dissolved carbon with stream drying and decreased for other analytes. Generally, the sensitivity of semivariogram sills increased with smaller sample sizes as streams dried in June compared to April. We used DIC semivariograms to illustrate the best fit sill and LOO sensitivity (Figure 3.7), which we report for all analytes (Table 3.2).

In the case of DIC, overall semivariance increased with stream drying in both streams in 2016 (Figure 3.7). In unburned JD, DIC semivariance (Figure 3.8, A; Table 3.2) increased fourfold from April (0.21 best fit sill with 0.17–0.22 sensitivity margin) to June (0.74 best fit sill with 0.50–1.7 sensitivity margin), and in burned MC (Table 3.2) semivariance increased 40-fold from April (0.035 best fit sill with 0.03–0.04 sensitivity margin) to June (1.6 best fit sill with 1.5–1.7 sensitivity margin; Figure 3.7B). DIC semivariogram patterns were confirmed by SD\textsubscript{GAM}, which rose from 0.44 to 0.79 from April to June 2016 in unburned JD and from 0.035 to 1.2 from April to June for burned MC.
Figure 3.7

2016 DIC Semivariograms for April and June.

![Semivariograms for April and June](image)

Notes: Semivariograms for the unburned JD stream (blue; A) and for burned MC stream (red; B) for 2016 April and June show the average semivariance ($\gamma$(mg C/L)$^2$) at a given separation distance (m) between points throughout the stream. Solid lines show the best fit semivariogram and shaded area show the leave one out (LOO) sensitivity analysis. The range of sill semivariance is based on the sensitivity analysis and is labeled near the best fit sill for 2016. Small open circles show the average semivariogram calculated using a number of pairs, which are labeled. Note the “full nugget” effect in each case.

In July 2017, DIC overall semivariance was relatively equivalent to June 2016 in unburned JD as seen in overlap in June 2016 0.5–1.7 LOO sensitivity margin and 0.60 best fit June 2017 (Table 3.2). In contrast, burned MC DIC semivariance decreased in July 2017 (0.87) by nearly twofold compared to June 2016 with no overlap (1.6 best fit and LOO sensitivity margin 1.5–1.7), and instead, burned MC 2017 July DIC semivariance overlapped with the LOO sensitivity margin of unburned JD (Figure 3.7, B and Table 3.2). Similar to DIC semivariance, SD$_{GAM}$ were equally similar as June 2016 SD$_{GAM}$ values were 70% of July 2017 values in both streams.
DOC semivariance exhibited interannual patterns opposite of DIC in each stream. 2017 DOC was lower than 2016 DOC in unburned JD by 50% (1.2 in 2016 compared to 0.58 in 2017) and higher in burned MC by 70% (0.93 in 2016 compared to 1.3 in 2017). DOC semivariance had less overlap within the streams between 2016 and 2017 when considering LOO sensitivity margins. However, burned MC 2017 best fit (1.3) slightly overlapped unburned JD June 2016 (LOO sensitivity margin 1.1–1.3; Table 3.2). We generally found the larger sample sizes in April had smaller LOO sensitivity such that the sensitivity of sill results was smaller than it was in June, when sample sizes were smaller (Table 3.2). DOC SD_GAM values were very similar to semivariogram sills.
Table 3.2

*Spatial Statistical Parameters for DIC, Cl\(^-\), DOC, TN, and PO\(_4\)\(^3^-\).*

<table>
<thead>
<tr>
<th></th>
<th>Unburned JD</th>
<th>Burned MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIC</td>
<td>DIC</td>
</tr>
<tr>
<td></td>
<td>SD(<em>{GAM}) Sill(</em>{BF}) Sill(_{LOO}) Range (m) n</td>
<td>SD(<em>{GAM}) Sill(</em>{BF}) Sill(_{LOO}) Range (m) n</td>
</tr>
<tr>
<td>April</td>
<td>0.44 0.21 0.17–0.22 - 53</td>
<td>0.17 0.036 0.033–0.037 - 58</td>
</tr>
<tr>
<td>May</td>
<td>0.62 0.53 0.15–0.56 - 50</td>
<td>0.18 0.037 0.031–0.041 - 42</td>
</tr>
<tr>
<td>June</td>
<td>0.79 0.74 0.50–1.7 - 36</td>
<td>1.2 1.6 1.5–1.7 - 38</td>
</tr>
<tr>
<td>July</td>
<td>1.1 0.60 0.58–0.62 76 79</td>
<td>0.86 0.87 0.83–88 81 91</td>
</tr>
<tr>
<td></td>
<td>Cl(^-)</td>
<td>Cl(^-)</td>
</tr>
<tr>
<td></td>
<td>SD(<em>{GAM}) Sill(</em>{BF}) Sill(_{LOO}) Range (m) n</td>
<td>SD(<em>{GAM}) Sill(</em>{BF}) Sill(_{LOO}) Range (m) n</td>
</tr>
<tr>
<td>April</td>
<td>0.16 0.02 0.021–0.025 - 53</td>
<td>0.15 0.03 0.026–0.042 - 55</td>
</tr>
<tr>
<td>May</td>
<td>0.02 0.005 0.003–0.005 - 48</td>
<td>0.05 0.003 0.002–0.003 - 39</td>
</tr>
<tr>
<td>June</td>
<td>0.10 0.009 0.004–1.01 - 35</td>
<td>0.05 0.003 0.002–0.003 - 37</td>
</tr>
<tr>
<td>Month</td>
<td>DOC</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>July</td>
<td>0.30</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>DOC</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>SD_{GAM}</td>
<td>Sill_{BF}</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>0.70</td>
</tr>
<tr>
<td>May</td>
<td>0.71</td>
<td>0.54</td>
</tr>
<tr>
<td>June</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>July</td>
<td>0.78</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>TN</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td>SD_{GAM}</td>
<td>Sill_{BF}</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.0015</td>
</tr>
<tr>
<td>May</td>
<td>0.03</td>
<td>0.001</td>
</tr>
<tr>
<td>June</td>
<td>0.12</td>
<td>0.05</td>
</tr>
<tr>
<td>July</td>
<td>0.02</td>
<td>0.00042</td>
</tr>
<tr>
<td></td>
<td>PO_{4}⁻³-</td>
<td></td>
</tr>
<tr>
<td>April</td>
<td></td>
<td></td>
</tr>
<tr>
<td>May</td>
<td></td>
<td></td>
</tr>
<tr>
<td>June</td>
<td></td>
<td></td>
</tr>
<tr>
<td>July</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SD\textsubscript{GAM}</td>
<td>Sill\textsubscript{BF}</td>
</tr>
<tr>
<td>------</td>
<td>----------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>April</td>
<td>0.43</td>
<td>0.002</td>
</tr>
<tr>
<td>May</td>
<td>0.02</td>
<td>0.0003</td>
</tr>
<tr>
<td>June</td>
<td>0.03</td>
<td>0.003</td>
</tr>
<tr>
<td>July</td>
<td>0.01 7E-05</td>
<td>6.9E-05–8.0E-05</td>
</tr>
</tbody>
</table>

Notes: Results from the SD of the modeled data (SD\textsubscript{GAM}) and semivariogram parameters using a spherical semivariance model: the best fit sill (Sill\textsubscript{BF}), leave-one-out sensitivity analysis spread of modeled sills (Sill\textsubscript{LOO}), the range (m) for July 2017 data, and the sample size used for each analysis.
While dissolved carbon as both DIC and DOC overall semivariance (sill) and SD$_{GAM}$ increased from April to June 2016 in both streams, Cl$^-$, TN, and PO$_4^{3-}$ overall semivariance and SD$_{GAM}$ were variable between the streams (Table 3.2). For instance, Cl$^-$ decreased from April to June in unburned JD (0.02 to 0.01, respectively) and in burned MC (Sill$_{BF}$ April 0.03 to June 0.002). Similarly, SD$_{GAM}$ in unburned JD decreased from 0.16 to 0.10 and from 0.15 to 0.05 in burned MC. These variance patterns were observed for PO$_4^{3-}$ as well. TN differed in that semivariance and SD$_{GAM}$ increased for unburned JD and were relatively invariant for burned MC. In July 2017, overall semivariance and SD$_{GAM}$ increased in both streams compared to June 2016 for Cl$^-$. However, TN and PO$_4^{3-}$ both decreased in unburned JD, and for burned MC, TN increased while PO$_4^{3-}$ exhibited the same variability (Table 3.2).

In 2017 our sampling strategy allowed us to determine well-defined range distances or patch size (as per Dent and Grimm, 1999; McGuire et al., 2014) for various solutes (Figure 3.8 and Table 3.2). DIC ranges were similar between streams (76 and 81 m for unburned JD and burned MC); however, Cl$^-$ and biologically reactive solutes DOC, TN, and PO$_4^{3-}$ all differed (Table 3.2). Cl$^-$ differed between the two streams as the unburned JD range was 152 m, while burned MC was 39 m. DOC, TN, and PO$_4^{3-}$ each had a shorter range in unburned JD (9.5, 23, and 33 m, respectively; Table 3.2) compared to burned MC (124, 61, 69 m, respectively; Table 3.2). Given the similarities between DOC sills, we were surprised by the difference in DOC ranges.
Figure 3.8

2017 Semivariograms for DIC, DOC, TN, and PO$_4^{3-}$.

![Diagram showing semivariograms for DIC, DOC, TN, and PO$_4^{3-}$](image)

Notes: Semivariograms for DIC, DOC, TN, and PO$_4^{3-}$ for unburned JD (top row, blue) and burned MC (bottom row, red). Open circles are the empirical semivariogram being modeled using the spherical model.

3.4 Discussion

Despite the ubiquity of stream drying, especially in headwater streams, and increasing wildfire risk across the Western United States, the impacts of these processes on stream spatial structure have not been well quantified. We collected repeated high-spatial scope data as streams dried following snowmelt, in the headwaters of an unburned and a recently burned stream. Using a combined time series approach and geostatistical tools, our findings partially supported our hypotheses. Collectively our results indicated that during lower flow, model complexity did indeed increase for most analytes, while spatial heterogeneity increased for dissolved carbon, but not for nutrients. Ranges or patch size were longer in burned MC than unburned JD, which did not support our hypotheses, but may have implications for TEM hypotheses (Fisher et al., 1998b). First, we discuss the results including the importance of detrending nonlinear longitudinal
patterns, the temporal changes in the complexity of underlying longitudinal patterns with drying, and the changes in heterogeneity with stream drying and following fire. We then revisit our conceptual diagram and discuss the sensitivity of our results, along with future research needs.

3.4.1 Quantifying Spatiotemporal Biogeochemical Patterns with Stream Drying: A Spatial Statistical Analysis

Unlike others who have used CV to describe heterogeneity (e.g., Asano et al., 2009; Gomez et al., 2009; von Schiller et al., 2011; Hale and Godsey, 2019), we found CV to be an unsatisfactory measurement to describe drying heterogeneity patterns due to inconsistency with calculation approaches taken for stream drying data sets (Supplementary Figure 3.1 and 3.8 Appendices A-C) and between spatial statistical methods. Congruence between SDGAM and semivariance suggested that SDGAM was an appropriate method to describe the spatial variation, or heterogeneity, of our data. With an appropriately robust 2017 data set, semivariograms allowed patch size to emerge by describing the distance at which our samples no longer displayed spatial dependence, or homogeneity.

3.4.1.1 Nonlinear Longitudinal Stream Patterns: Detrending. Longitudinal stream patterns were nonlinear and our data were best modeled, or detrended, using a GAM fit; adopting this time series approach allowed us to isolate the spatial relationships between reach sites as each stream dried. Consistently lower SSEGAM compared to the SSELM demonstrated the nonlinear nature of longitudinal patterns. Though many studies indicate high spatial variability in small catchments (Asano et al., 2009; Abbott et al.,
2018; Hale and Godsey, 2019), few have reported on the nonlinearity of a nonbranching stream segment that we observed in our results. Others before us examining stream chemistry patterns (Gustafson, 1998; Dent and Grimm, 1999; Ganio et al., 2005; Zimmer et al., 2013; Hale and Godsey, 2019) have either had linear longitudinal data, assumed a linear fit, or did not report how or if detrending was appropriate; in our experience, geostatistical programs that provide spatial analysis tools do not allow for a complete exploration of modeling the data appropriately. As our data showed, a linear fit to the data would lump underlying patterns into the spatial heterogeneity by including them in residuals (Figure 3.6).

Detrending the data to isolate spatial patterns helped us to address longitudinal differences that may have occurred due to changes in water flux either spatially or temporally. Increases in stream water can occur following fire due to vegetation losses (Atchley et al., 2018), but in sagebrush steppe ecosystems, these increases can be marginal (Wilcox, 2002; Flerchinger et al., 2016). Inability to measure discharge at every 50-m site limited our direct assessment of potential water flux influencing chemistry concentrations down the streams. However, MacNeill et al. (2020) found little evidence in either of these two study streams using predicted/measured Cl\textsuperscript{-} longitudinal or temporal patterns being driven by changes in water flux. Nevertheless, using a smoother to model the data further reduced the potential impacts water flux influence could have had on spatial relationships among 50-m sites because the SD\textsubscript{GAM} and semivariogram analysis utilized residuals around the model. In this way, detrending data by adopting a time series approach (Cooper et al., 1997) was a strength in our method that allowed us to evaluate the relative complexity between models and examine the temporal changes of
the longitudinal chemistry patterns. Then we were able to move on and quantify the spatial relationships between 50-m sites separately from the influence of underlying longitudinal patterns.

3.4.1.2 Spatial Complexity Increased with Stream Drying. In both streams, our hypotheses were somewhat supported, as the underlying longitudinal patterns increased in complexity for some analytes with drying, while heterogeneity within each stream increased for dissolved carbon and decreased for nutrients with drying. The underlying models generally increased in complexity, even when normalized by changing sample size (AIC/n). Exceptions to this were in unburned JD TN and PO₄³⁻, which each showed different model results depending on the data set used to model the data, and DIC and DOC model complexity in burned MC, which decreased. In part, these deviations may be explained by the biological reactivity and cycling of TN, PO₄³⁻, and DOC (MacNeillie et al., 2020). Specifically, TN and PO₄³⁻ AIC/n values in unburned JD decreased from April to June when all data were used but increased when the analysis included only the 33 50-m sites with surface water present over all three months. 50-m sites in unburned JD lost due to surface water not present included the uppermost stream sites that dried beginning in May. These reaches were in a meadow with distinctively high concentrations relative to the downstream sites. In addition, MacNeillie et al. (2020) found DOC sourcing to be variable longitudinally, so perhaps DOC behavior was less predictable than other solutes.

Our inferences are based on comparisons of AIC/n values, which provided a relative measure of change as streams dried, largely based on the complexity term (C), but not statistically significant differences between months’ longitudinal changes. Detrending the data enabled us to analyze the longitudinal trends separately from the
spatial relationships between measured 50-m sites because our interpretations of heterogeneity and patch size are based on analyses of the residuals.

While our hypothesis that longitudinal complexity would increase with drying was largely supported, our hypothesis that analytes would increase in heterogeneity was only supported by dissolved carbon spatiotemporal patterns. From April to June, dissolved carbon increased in $SD_{GAM}$ and semivariance, and therefore heterogeneity. We previously reported in MacNeille et al. (2020) that increasing DIC concentrations had a strong relationship with cation concentrations associated with increasing deeper groundwater inputs in both study streams. Increases in the $SD_{GAM}$ and semivariance suggest that though these inputs may be more representative of a deep groundwater sourcing signature, when the stream is less connected during low flow, the surface water patterns are more heterogeneous. Increased DIC heterogeneity corroborate the findings of Zimmer et al. (2013), who observed increased variability of conservative solutes in groundwater at low flow.

Compared to DIC, DOC concentrations were relatively longitudinally variable, which was reflected in higher overall semivariance. In addition, DOC had high $SD_{GAM}$ and semivariance despite less variant and lower mean concentrations, which further emphasizes the value of quantifying spatial trends using $SD_{GAM}$ and semivariance because information is not easily gleaned without spatial analyses. DOC’s increasing semivariance patterns corroborate earlier findings in these streams indicating that carbon spectral indices, fluorescence index (FI), and absorbance coefficient ($abs_{254}[m^{-1}]$) were surprisingly varied longitudinally at 50-m intervals (MacNeille et al., 2020). Lastly,
unburned JD and burned MC DIC similarities between overall semivariance (SillBF) may have indicated some recovery of burned MC two years postfire.

In contrast to DIC and DOC, nutrients TN and PO4³⁻ heterogeneity decreased with the exception of 2016 TN patterns in unburned JD. These TN and PO4³⁻ patterns may be a result of more direct reflection of biological uptake and cycling from in-stream heterotrophic and autotrophic organisms. TN and PO4³⁻ uptake is subject to seasonal variation as uptake increases with temperature (Mulholland and Hill, 1997) and decreased discharge (Peterson et al., 2001). As concentrations decreased, if biological demand was higher or increased to higher levels than nutrient availability, perhaps heterogeneity would decrease due to the limited excess supply of TN or PO4³⁻. Indeed, as nitrogen concentrations demonstrated, the majority of nitrogen was organic as inorganic levels of NH4⁺-N and NO₃⁻-N were not detectable expect in burned MC in April, indicating there was little nitrogen that remained in the inorganic form. We were surprised to find that Cl⁻ heterogeneity decreased with stream drying in both streams, because we expected it to behave as DIC did due to its relatively conservative behavior (MacNeille et al., 2020).

3.4.2 Biogeochemical Patch Size and Previously Undetected Functional Process Zones

“Patches” in prior stream literature are not often examined using a spatially continuous approach that allows them to empirically emerge (Cooper et al., 1997; Ganio et al., 2005; Scown et al., 2016). However, semivariograms defined patterns based on spatial relationships of biogeochemistry, not on other predetermined parameters such as habitat. We equated patch size to range distance, distances in which constituents displayed statistical dependence, as others have done before us (Cooper et al., 1997; Dent
and Grimm, 1999; McGuire et al., 2014); thus, we used the 2017 semivariogram ranges to quantify stream patches during low flow. Our results suggest that patch size differed between conservative and bioreactive analytes, with bioreactive solutes patches showing stronger similarity, and that patch sizes differed between streams, with DIC being exceptionally similar. Though we expected to find differences in range values between constituents, especially between conservative and biologically reactive solutes, we were surprised by this DIC range similarity. DIC and Cl−, despite both having a strong relationship to groundwater sources with stream drying (MacNeill et al., 2020), differed in range distances from each other within each stream. On the other hand, DOC, TN, and PO₄³⁻ range distances in unburned JD were shorter than DIC. Though we calculated a longer DOC range in burned MC, TN or PO₄³⁻ had similarly short ranges to one another.

The patch sizes we calculated may be related to nutrient travel distances and have implications for nutrient dynamics. The distance for which a nutrient molecule travels before biological uptake occurs, the uptake length, is calculated by the water velocity, water depth, and nutrient concentration divided by the areal uptake rate (Newbold et al., 1981; Fisher et al., 2004; Baker and Webster, 2017; Tank et al., 2017). Though we have not measured nutrient uptake, our observations indicate water velocity and depth decreased and TN and PO₄³⁻ concentration also decreased in both streams from April to June. Biological uptake would likely increase with the growing season (Mulholland and Hill, 1997), but if we controlled for nutrient uptake rate and assumed it stayed the same as it was during higher flow conditions (higher numerator with higher water velocity, depth, and nutrient concentration), nutrient travel distance would decrease with drying. If solute concentration increased substantially, like DIC did, we would instead expect travel
longer distances relative to those solutes with decreasing concentrations. This may explain why we saw that patch size for TN and PO$_4^{3-}$ was shorter than DIC. Indeed, following flooding in Sycamore Creek, Martí et al. (1997) found decreases in nitrogen uptake lengths that matched Dent and Grimm’s (1999) decreasing nitrogen semivariogram ranges in the same stream. It also seems likely that there is a relationship between the patch size for these nutrients that interact and enable autotrophic and heterotrophic activity in streams. For instance, Rodriguez-Cardona et al. (2020) found that stream nitrogen levels in burned catchments remained elevated after DOC sources were combusted following fire. Stoichiometrically, we observed less DOC relative to phosphorous concentrations in both streams, but these ratios were reduced to one fourth in burned MC compared to unburned JD. Further research would be required to investigate relationships between nutrient range distance and spiraling travel distance in the study streams’ context.

The fine-scale patch sizes we detected were scaled to our study scope (Kirschner et al., 2004; Peterson et al., 2006) and may represent process domains and FPZ at scales smaller than previously detected in larger or perennial or larger streams. Because drying occurred on the meter scale, we used a sampling grain as fine as 10 m to define patterns. We calculated ranges all were under 152 m and most under 100 m, which represented smaller patches than defined by other stream biogeochemistry research. In addition, given that 2017 ranges were robustly defined with averaged semivariogram at small separation distances, our study scope has potentially enabled us to define patches at the smallest distances at which constituents exhibit statistical relationships. These patterns reflect the study scope and provide evidence that supports the high variability of headwater streams.
that have been reported by others (Peterson et al., 2001; Asano et al., 2009; Zimmer et al., 2013; Abbott et al., 2018; Hale and Godsey, 2019). In a perennial stream network, McGuire et al. (2014) found hierarchical patterns that were scaled to a larger extent (over 32 tributaries, many of which were about 2 km) and a coarser grain (100 m) than our study; McGuire et al. (2014) reported patches of 500 to >3,000 m. Like McGuire et al. (2014) found, we expected the smallest patches we defined were nested within larger patches associated with larger process domains. Instead of perennial streams, we sought to define patches in drying streams at scales more similar to those measured by Dent and Grimm (1999) over the 10-km extent of Sycamore Creek at a 25-m grain. Dent and Grimm (1999) found NO$_3^-$ patch size decreased from ~2,000 m to ~500 m and PO$_4^{3-}$ from ~2,500 m to ~1,000 m as the stream recovered following flooding. We had hoped to test range distance directly as the stream dried April to June, but given the inability to determine 2016 ranges at our study scope, results remained inconclusive as a direct measure. Based on our results, we suspect similar spatiotemporal patterns occur with seasonal drying as do following monsoonal flooding (Dent and Grimm, 1999) but that shifts are more subtle.

The patch lengths that emerged from the semivariogram ranges in burned MC were longer compared to the unburned JD, which may support TEM predictions (Fisher et al., 1998b). The TEM predicts longer nutrient travel times due to less overall nutrient cycling following a disturbance like fire. Due to lacking prefire data for burned MC, we cannot infer causality. However, six months after wildfire, burned MC had less overall variation (lower sill) in April for dissolved carbon than unburned JD, though 2016 range could not be robustly defined. Past research has shown similarities in the 2016 patterns
between these streams in regard to concentration patterns and processes associated with drying (MacNeill et al., 2020). However, quantifying the fine-scale spatial patterns between the streams suggests longer patch sizes in burned MC and, thus, process domains and FPZs. Informed by the TEM predictions for nutrient spiraling, these findings supported TEM, as unburned JD likely has more habitat complexity formed by woody debris, for instance, that we observed lacking in burned MC after the fire.

3.4.3 Conceptual Model: Headwater Streams as They Dry

We revised our hypothesized semivariogram conceptual model to illustrate how our findings suggest biogeochemical heterogeneity and spatial patterns change as streams dry and following fire, and these results may have implications for the ecosystem processes of headwater streams (Figure 3.9). For instance, our results defined ranges at smaller distances than most previous work and suggest that carbon increases in heterogeneity while nutrient heterogeneity may decrease. Dissolved carbon best supported our initial hypothesis that heterogeneity would increase with stream drying. DIC and DOC heterogeneity may increase for related but different reasons. DIC concentrations increased with drying, associated with increased groundwater inputs (MacNeill et al., 2020). Zimmer et al. (2013) visually showed that groundwater ions (as well as maximums and minimums) have high concentration variation during low flow, and our DIC results revealed similar increased heterogeneity with more groundwater dominance in these RC CZO watersheds.

During stream drying, TN and PO$_4^{3-}$ behaviors were fairly coupled by the measures of increased longitudinal complexity, decreased heterogeneity, and similar
patch size. This coupling of TN and PO$_4^{3-}$ in the patterns, especially the patch size, may suggest that these patches correspond to uptake rate, but not uptake efficiency (von Schiller et al., 2008). Dent and Grimm (1999) reported similar behaviors between nitrogen and phosphorous as both patch sizes decreased with succession. However, in other intermittent headwater streams of Mediterranean environments, the demand of NH$_4^+$ and PO$_4^{3-}$ were decoupled (von Schiller et al., 2008). Thus, though heterogeneity decreased for each TN and PO$_4^{3-}$, the processes driving low heterogeneity may be different during low flow. TN biologic demand may increase, while biologic demand PO$_4^{3-}$ may be lower relative to its availability during low flow (von Schiller et al., 2008), especially in ecosystems where TN is more limiting (Martí et al., 1997); such is the case at RC CZO (Schwabedissen et al., 2017). Indeed, the stoichiometry suggested that PO$_4^{3-}$ was on average in greater abundance than TN (most of which was organic nitrogen); June in burned MC was the most extreme case of limited nitrogen compared to PO$_4^{3-}$ availability. This later season lower N:P ratio may have occurred because of higher PO$_4^{3-}$ availability from increased sediment transport to the stream following fire (Vega et al., 2020), while higher N:P ratios in unburned JD may have been due to relatively less PO$_4^{3-}$ availability due to physical retention by woody debris that were absent in burned MC.

Our results suggest streams shift toward storing rather than transporting when going dry. Hale and Godsey (2019) found that at small scales (<1.46 km$^2$), intermittency was a stronger predictor of DOC concentrations than catchment area. Although we did not observe strong temporal differences using mean DOC concentration values with drying, if intermittency is a strong predictor of DOC concentrations, it may also be a strong predictor of DOC spatial relationships. Similarly, Lohse et al. (2020) found that
decay rates increased with more water present and that daily water presence accounted for the majority of the variation. As already discussed, nutrient heterogeneity decreased, and this may be because uptake increases during the growing season and low concentrations in our streams may cause nutrients to be retained or cycled quickly (Peterson and Grimm, 1992; Martí et al., 1997). In these ways, our results may suggest biologically reactive analytes are being retained within the stream either physically or as biomass during low flow compared to high flow. Similarly, during fire recovery, as displayed by larger patch sizes in burned MC, streams may transport more materials (Rodriguez-Cardona et al., 2020) compared to unburned streams. Understanding nutrient dynamics during the time of drying is critical, because this may be the time of highest productivity like biofilms (Timoner et al., 2010). In Mediterranean ecosystems, the period of drying indicates major shifts in the epilithic (streambed surface) life cycle and high biological processing rates (Timoner et al., 2010). Similar patterns may occur at RC CZO, which warrants further research.
Figure 3.9

_Revised Conceptual Semivariograms._

Notes: Conceptual semivariograms for (A) dissolved carbon, (B) nutrients TN and PO$_4^{3-}$, and (C) fire effects. In (A) and (B), the dotted line denotes low flow semivariance and the solid line denotes high flow. In (C), the denotes the unburned stream (blue dotted line) relative to the burned stream (orange dotted line).

### 3.4.4 Ranges Are Dynamic and Sensitive: Sample Design and Modeling in Future Research

Ranges are dynamic because patches measured via fine-grain, high-spatial-scope approaches are sensitive to the parameters set within the semivariogram. Study scope impacts pattern detection, but sample design also impacts the small-scale spatial relationships defined by the semivariogram range. We, like others before us (Dent and Grimm, 1999; McGuire et al., 2014; Hale and Godsey, 2019), first used consistent sampling intervals in 2016 throughout the stream segment. In this exploratory study, we found that 50-m intervals were too long to capture autocorrelation, or the semivariogram range, which occurred at a grain finer than 50 m given our overall study extent (~2,500 m). Had we increased the extent of our sampling and used a coarser sampling grain, larger patches likely would have also emerged. Given that our objective was to quantify patterns in the headwater where drying occurred readily and variation is thought to be
high (Creed et al., 2015), we did not adjust our study extent, but instead in 2017 we
adjusted the measurement grain. We used a finer grain of 10 and 25 m and were able to
capture the semivariogram ranges for the solutes we tested. However, we employed a
sampling design that sampled 10 intervals at the most downstream portion of our stream
segment. Alternatively, we could have employed a nested sampling design whereby 10-m
intervals were scattered in clusters throughout the entire segment. Such a design might
better estimate this relationship, which could include change between different reaches
throughout, but the range would be less well defined at any given reach. Given our
design, we were able to robustly define solute ranges for the downstream portion of each
stream. Though these results have informed our conceptual model for spatial patterns
with stream drying and following fire, further research would be required to determine
how representative these ranges are throughout these and other semiarid headwater
streams.

In addition, semivariogram analyses required many choices that could have
influenced the range output we observed. These included how pairs at given separation
distances were binned (Peterson et al., 2006) and which semivariogram model fit was
used to estimate the parameters from the empirical semivariogram (Gorsich and Genton,
2000). In each case we chose conservative approaches. First, we set the bins to have a
minimum of 50 pairs per bin. Second, spherical, gaussian, and exponential models are
comparable in shape (Gorsich and Genton, 2000), but we held the spherical model
consistent as others before us have done to compare parameters across solutes (Peterson
et al., 2006), even when the spherical model was not identified with lowest SSE.
Nevertheless, especially for small data sets, these models can yield different ranges for
the same data (Gorsich and Genton, 2000), which we found to be the case with our data. Gorsich and Genton (2000) suggested a method of using the derivative of each model to determine the best fit. Whether it is best to hold the semivariogram model constant as we have done or use either the lowest SSE for all data sets or the method Gorsich and Genton (2000) suggest could be investigated further for data sets like ours in the future.

3.5 Conclusions

Quantifying spatial patterns in intermittent streams is challenging for a number of reasons, including that these methods require a great deal of time, expertise, and judgement. Because of this, while theorized in the literature, techniques for quantifying patterns from continuous spatial measurements have not often been applied. However, as stream intermittency and wildfires increase, how these conditions will impact stream structure, especially as it relates to ecosystem processes, are of increasing importance. A strength and distinctive contribution of our study is that we used biogeochemistry to quantify spatial heterogeneity and patch size as two headwater streams dried, one recently burned. We collected high-spatial-scope data repeated at several time points so that patterns could emerge with drying and at low flow conditions. Adopting a time series approach and then applying geostatistical methods, we were able to define heterogeneity and patch size (>160 m) potentially of the smallest process domain detectable at a landscape extent during low flow. First, we found nonlinear longitudinal patterns that required detrending to remove. As we hypothesized, we found that underlying patterns’ model complexity increased as both streams dried. We then used residuals of the modeled data to isolate the spatial relationships between fine-grain reach sites and analyzed
heterogeneity using the SD of the detrended data and building semivariograms. Results from both of these approaches suggested dissolved carbon heterogeneity increased with drying, supporting our hypothesis. Contrary to our hypothesis, however, nutrient heterogeneity decreased with drying in both streams. Patch sizes suggested that the burned stream had larger patch sizes and supported the TEM hypothesis, which predicted that following disturbance, there would be an ecosystem expansion and increased nutrient travel distance in streams. While the effects of wildfire fit into conceptual models of ecosystem processes and spatial complexity such as TEM, stream drying as a seasonal phenomenon may not.

We feel the tools we employed, especially using the SD of the detrended data, have been described in stream ecology literature but have remained largely unapplied for quantifying patterns in streams at our study scope. We found measuring biogeochemistry in a spatially continuous manner and using spatial statistical tools to be especially appropriate for quantifying spatial variation. Furthermore, applying these tools to biogeochemistry enabled patterns to emerge and emphasized changing spatial complexity as streams dried. Given the increasing structural complexity we found associated specifically with headwater intermittency, future research quantifying potential shifts in nutrient spiraling with repeated measurements and processing by autotrophic and heterotrophic microorganism communities like biofilms would build on our findings. Spatial complexity could also be examined during different phases of the drying and rewetting phenomenon.
3.6 Works Cited


3.7 Supplementary Materials

Supplementary Figure 3.1

CV Calculation Methods (A) and Results (B).

A. Traditional methods: Coefficient of variation (CV) and moving mean average. To account for underlying longitudinal patterns, we calculated the average CV of a 5-site moving window (MW) CV, similar to methods presented by Asano et al. (2009). We conducted this analysis using all the data (“all data MW”) and using data from sites that had surface water present all three months (“3 surface water MW”). The moving window approach compensated for longitudinal patterns along the profile, however, shared the drawbacks of either changing sample size or excluding data for stream sites that dried. Ultimately, like others before us (Dent and Grimm, 1999), we emphasized trends more than specific CV and predicted CV would increase as streams dried if heterogeneity was increasing. We also continued our analysis of spatiotemporal heterogeneity using a time series approach to spatial statistical analyses that included modeling the stream patterns and using residuals to calculate standard deviation (SD\(_{GAM}\)) and build semivariograms as described by Cooper et al. (1997).
B.

Notes: A) DIC (A-B) and TN (C-D) CVs and mean concentrations (mg C/L) for unburned JD (A, C; left, blue) and burned MC (B, D; right, orange) for April, May, and June 2016. CVs are calculated using four approaches: all the data (dashed line, lighter color), all the data using a 5-window moving average (dotted line, lighter color), only sites that had surface water all 3 months (dashed line, dark), and only sites that had surface water all 3 months using a 5-window moving average (solid line, dark). Means are calculated using all the data (solid black circles) and for those sites that had surface water present all three months (open circles). E and F are tables providing the numerical values using all the data (right 5 columns) and using sites with surface water present all three months (left 4 columns) of mean, SD, CV, CV MA (moving average), and sample size (n).
Supplementary Figure 3.2

**Smoothen Selection.**

In these simulations we know what the true model is, so we can make some useful comparisons.

<table>
<thead>
<tr>
<th></th>
<th>Gam</th>
<th>Lowess</th>
<th>Smooth.spline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complexity, AIC</td>
<td>Complexity, AIC</td>
<td>Complexity, AIC</td>
</tr>
<tr>
<td>Simulation 1</td>
<td>7.90, 248.7</td>
<td>2, 250.94</td>
<td>17.5, 235.41</td>
</tr>
<tr>
<td>Simulation 2</td>
<td>8.18, 82.88</td>
<td>8.75, 87.57</td>
<td>8.25, 83.15</td>
</tr>
</tbody>
</table>

Based, at least on these two simulations we know a few things
- Smooth.spline occasionally produces total nonsense, C = 17.5 was an absurd fit.
- Lowess produced too smooth a fit for the first one.
- Gam, at least for these examples, does much better. For example, in sim 2 it produced the simplest model and the lowest AIC. In sim 1 it was the only one that didn’t produce nonsense.
- Gam is easy to implement, as it finds the optimal model for our study.

A) Simulated data function:  

B) GAM: closest fit to known function

C) Lowess: too simple,  
(low complexity, high error

D) Smooth.spline: too complicated, underfit  
overfit (high complexity, low error

A) Simulated function  

B) GAM and Simulated

C) Lowess and Simulated  

D) Smooth.spline and
Supplementary Table 3.1

**SSE of Linear (SSE\textsubscript{LM}) and GAM (SSE\textsubscript{GAM}) Fits.**

<table>
<thead>
<tr>
<th></th>
<th>Unburned JD</th>
<th>Burned MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIC</td>
<td>DIC</td>
</tr>
<tr>
<td><strong>SSE\textsubscript{LM}</strong></td>
<td><strong>SSE\textsubscript{GAM}</strong></td>
<td><strong>SSE\textsubscript{LM}</strong></td>
</tr>
<tr>
<td>April</td>
<td>106</td>
<td>10</td>
</tr>
<tr>
<td>May</td>
<td>554</td>
<td>19</td>
</tr>
<tr>
<td>June</td>
<td>145</td>
<td>22</td>
</tr>
<tr>
<td>July</td>
<td>290</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td><strong>Cl\textsuperscript{-}</strong></td>
<td><strong>Cl\textsuperscript{-}</strong></td>
</tr>
<tr>
<td><strong>SSE\textsubscript{LM}</strong></td>
<td><strong>SSE\textsubscript{GAM}</strong></td>
<td><strong>SSE\textsubscript{LM}</strong></td>
</tr>
<tr>
<td>April</td>
<td>1.7</td>
<td>1.3</td>
</tr>
<tr>
<td>May</td>
<td>0.031</td>
<td>0.012</td>
</tr>
<tr>
<td>June</td>
<td>1.2</td>
<td>0.32</td>
</tr>
<tr>
<td>July</td>
<td>7.5</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td><strong>DOC</strong></td>
<td><strong>DOC</strong></td>
</tr>
<tr>
<td><strong>SSE\textsubscript{LM}</strong></td>
<td><strong>SSE\textsubscript{GAM}</strong></td>
<td><strong>SSE\textsubscript{LM}</strong></td>
</tr>
<tr>
<td>April</td>
<td>60.8</td>
<td>33.7</td>
</tr>
<tr>
<td>May</td>
<td>82.7</td>
<td>23.7</td>
</tr>
<tr>
<td>June</td>
<td>71.4</td>
<td>39.0</td>
</tr>
<tr>
<td>July</td>
<td>105</td>
<td>46.2</td>
</tr>
<tr>
<td></td>
<td><strong>TN</strong></td>
<td><strong>TN</strong></td>
</tr>
<tr>
<td><strong>SSE\textsubscript{LM}</strong></td>
<td><strong>SSE\textsubscript{GAM}</strong></td>
<td><strong>SSE\textsubscript{LM}</strong></td>
</tr>
<tr>
<td>April</td>
<td>2.1</td>
<td>0.05</td>
</tr>
<tr>
<td>May</td>
<td>0.64</td>
<td>0.05</td>
</tr>
<tr>
<td>June</td>
<td>0.65</td>
<td>0.54</td>
</tr>
<tr>
<td>July</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td><strong>PO\textsubscript{4}^{3-}</strong></td>
<td><strong>PO\textsubscript{4}^{3-}</strong></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>$\text{SSE}_{LM}$</th>
<th>$\text{SSE}_{GAM}$</th>
<th>$\text{SSE}_{LM}$</th>
<th>$\text{SSE}_{GAM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>0.16</td>
<td>0.09</td>
<td>0.1</td>
<td>0.09</td>
</tr>
<tr>
<td>May</td>
<td>0.031</td>
<td>0.013</td>
<td>0.28</td>
<td>0.03</td>
</tr>
<tr>
<td>June</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
<td>0.17</td>
</tr>
<tr>
<td>July</td>
<td>0.03</td>
<td>0.008</td>
<td>0.06</td>
<td>0.04</td>
</tr>
</tbody>
</table>
### Supplementary Table 3.2

#### Semivariogram Model Parameters Comparison.

<table>
<thead>
<tr>
<th></th>
<th>Unburned ID</th>
<th>Burned MC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIC</td>
<td>DIC</td>
</tr>
</tbody>
</table>
|                | Sill<sub>b</sub> | Range<sub>b</sub> | Sill<sub>spherical</sub> | Range<sub>spherical</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | Model<sub>lowest SSE</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | SSE<sub>lowest SSE</sub>
| April          | 0.12        | 3.2E-05   | 0.12        | 3.2E-05   | 0.036      | 3.1E-07   | Sph     | 0.036      | 3.1E-07   | 1.1E-07   |
| May            | 0.53        | 6.0E-06   | 0.43        | 0.0002    | 0.037      | 9.1E-07   | Hole    | 0.040      | 3.8E-07   | 3.8E-07   |
| June           | 0.73        | 5.1E-05   | 0.73        | 1.1E-05   | 1.6         | 9.1E-07   | Gau     | 1.6         | 3.4E-05   | 3.4E-05   |
| July/2017      | 0.60        | 7.7E-02   | 0.64        | 3.1        | 1.8E-03    | 8.3E-02   | 0.002   | 8.9         | 3.2E-02   | 0.0002    |
|                | CI          | CI        |
|                | Sill<sub>b</sub> | Range<sub>b</sub> | Sill<sub>spherical</sub> | Range<sub>spherical</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | Model<sub>lowest SSE</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | SSE<sub>lowest SSE</sub>
| April          | 0.02        | 1.7E-07   | 0.02        | 1.7E-07   | 0.03        | 7.0E-07   | Hole    | 0.03        | 9.2E-08   | 9.2E-08   |
| May            | 0.0004      | 3.8E-11   | 0.0003      | 2.4E-11   | 0.003      | 4.6E-11   | Gau     | 0.003      | 3.9E-11   | 3.9E-11   |
| June           | 0.009       | 3.2E-09   | 0.009       | 3.2E-09   | 0.003      | 5.0E-09   | Hole    | 0.003      | 4.0E-09   | 4.0E-09   |
| July/2017      | 0.10        | 1.4E-04   | 0.10        | 1.4E-04   | 0.04        | 2.1E-06   | Sph     | 0.04        | 2.1E-06   | 2.1E-06   |
|                | DOC         | DOC       |
|                | Sill<sub>b</sub> | Range<sub>b</sub> | Sill<sub>spherical</sub> | Range<sub>spherical</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | Model<sub>lowest SSE</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | SSE<sub>lowest SSE</sub>
| April          | 0.70        | 7.8E-04   | 0.8         | 1.9E-03   | 0.56        | 8.1E-05   | Hole    | 0.5         | 1.9E-05   | 1.9E-05   |
| May            | 0.54        | 7.5E-05   | 0.54        | 7.5E-05   | 0.48        | 4.2E-05   | Hole    | 0.52        | 1.8E-05   | 1.8E-05   |
| June           | 1.1         | 3.7E-05   | 1.1         | 3.7E-05   | 0.92        | 2.1E-05   | Sph     | 0.92        | 2.1E-05   | 2.1E-05   |
| July/2017      | 0.58        | 0.001     | 0.58        | 0.001     | 1.3         | 1.2E-04   | 0.002   | 1.3         | 1.2E-04   | 1.2E-04   |
|                | TN          | TN        |
|                | Sill<sub>b</sub> | Range<sub>b</sub> | Sill<sub>spherical</sub> | Range<sub>spherical</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | Model<sub>lowest SSE</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | SSE<sub>lowest SSE</sub>
| April          | 0.0015      | 5.2E-09   | 0.0016      | 9.2E-10   | 0.003      | 6.3E-10   | Hole    | 0.002      | 6.2E-10   | 6.2E-10   |
| May            | 0.001       | 2.5E-10   | 0.001       | 2.5E-10   | 0.003      | 3.1E-10   | Sph     | 0.001      | 3.1E-10   | 3.1E-10   |
| June           | 0.05        | 2.0E-08   | 8.2         | 1.5E-09   | 0.0007     | 3.6E-10   | Pen     | 0.0007     | 2.8E-10   | 2.8E-10   |
| July/2017      | 4.1E-04     | 9.4E-10   | 4.5E-04     | 5.1E-10   | 0.0025     | 2.1E-08   | Gau     | 0.0026     | 1.8E-08   | 1.8E-08   |
|                | PO<sub>4</sub> | PO<sub>4</sub> |
|                | Sill<sub>b</sub> | Range<sub>b</sub> | Sill<sub>spherical</sub> | Range<sub>spherical</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | Model<sub>lowest SSE</sub> | Sill<sub>lowest SSE</sub> | Range<sub>lowest SSE</sub> | SSE<sub>lowest SSE</sub>
| April          | 0.002       | 1.8E-10   | 0.002       | 1.8E-10   | 0.0017     | 2.3E-10   | Hole    | 0.0017     | 1.3E-10   | 1.3E-10   |
| May            | 0.0003      | 4.5E-11   | 0.0003      | 2.4E-11   | 0.0009     | 3.1E-10   | Gau     | 0.0009     | 3.1E-10   | 3.1E-10   |
| June           | 0.001       | 7.2E-10   | 0.011       | 1.4E-12   | 0.005      | 1.6E-11   | Gau     | 0.005      | 9.3E-12   | 9.3E-12   |
| July/2017      | 8.0E-9      | 6.3E-11   | 8E-05       | 3.5E-11   | 0.1E-04    | 1.4E-09   | Sph     | 0.1E-04    | 1.4E-09   | 1.4E-09   |

Notes: Compares parameters for the lowest sums squared errors (SSE) semivariogram model with the spherical semivariogram model for DIC, Cl<sup>-</sup>, DOC, TN, and PO<sub>4</sub><sup>3-</sup> for the unburned stream (left) and the burned stream (right). 2016 best fit (not leave-one-out sensitivity analysis models) and 2017 sills (Sill<sub>b</sub>spherical and Sill<sub>lowest SSE</sub>) are labeled for comparisons. Ranges for 2017 spherical (Range<sub>b</sub>spherical) and lowest SSE (Range<sub>lowest SSE</sub>) semivariogram models were also compared.
Supplementary Table 3.3

**AIC, SSE, and C Terms Summary Tables.**

<table>
<thead>
<tr>
<th></th>
<th>DIC</th>
<th>CI</th>
<th>DOC</th>
<th>TN</th>
<th>PO$_2^{\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC/N</td>
<td>SSE/N</td>
<td>C/N</td>
<td>AIC/N</td>
<td>SSE/N</td>
</tr>
<tr>
<td>April</td>
<td>0.26</td>
<td>0.19</td>
<td>0.11</td>
<td>0.2</td>
<td>0.01</td>
</tr>
<tr>
<td>May</td>
<td>0.44</td>
<td>0.38</td>
<td>0.19</td>
<td>0.3</td>
<td>0.0002</td>
</tr>
<tr>
<td>June</td>
<td>0.59</td>
<td>0.61</td>
<td>0.25</td>
<td>0.51</td>
<td>0.009</td>
</tr>
</tbody>
</table>

**Burned**

<table>
<thead>
<tr>
<th></th>
<th>DIC</th>
<th>CI</th>
<th>DOC</th>
<th>TN</th>
<th>PO$_2^{\infty}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC/N</td>
<td>SSE/N</td>
<td>C/N</td>
<td>AIC/N</td>
<td>SSE/N</td>
</tr>
<tr>
<td>April</td>
<td>0.34</td>
<td>0.03</td>
<td>0.17</td>
<td>0.17</td>
<td>0.02</td>
</tr>
<tr>
<td>May</td>
<td>0.45</td>
<td>0.22</td>
<td>0.09</td>
<td>0.12</td>
<td>0.003</td>
</tr>
<tr>
<td>June</td>
<td>0.31</td>
<td>1.4</td>
<td>0.10</td>
<td>0.41</td>
<td>0.002</td>
</tr>
</tbody>
</table>

**Unburned**

*only 50 m interval sites with surface water all three months*

<table>
<thead>
<tr>
<th></th>
<th>DIC N=33</th>
<th>CI N=35</th>
<th>DOC N=34</th>
<th>TN N=36</th>
<th>PO$_2^{\infty}$ N=33</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC/N</td>
<td>SSE/N</td>
<td>C/N</td>
<td>AIC/N</td>
<td>SSE/N</td>
</tr>
<tr>
<td>April</td>
<td>0.43</td>
<td>0.30</td>
<td>0.21</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>May</td>
<td>0.54</td>
<td>0.076</td>
<td>0.26</td>
<td>0.35</td>
<td>0.002</td>
</tr>
<tr>
<td>June</td>
<td>0.62</td>
<td>0.43</td>
<td>0.27</td>
<td>0.51</td>
<td>0.009</td>
</tr>
</tbody>
</table>

**Burned**

*only 50 m interval sites with surface water all three months*

<table>
<thead>
<tr>
<th></th>
<th>DIC N=36</th>
<th>CI N=30</th>
<th>DOC N=31</th>
<th>TN N=37</th>
<th>PO$_2^{\infty}$ N=36</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AIC/N</td>
<td>SSE/N</td>
<td>C/N</td>
<td>AIC/N</td>
<td>SSE/N</td>
</tr>
<tr>
<td>April</td>
<td>0.31</td>
<td>1.2</td>
<td>0.15</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>May</td>
<td>0.51</td>
<td>0.04</td>
<td>0.25</td>
<td>0.13</td>
<td>0.002</td>
</tr>
<tr>
<td>June</td>
<td>0.32</td>
<td>1.6</td>
<td>0.10</td>
<td>0.51</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Notes: The AIC/N, sums squared error/N (SSE/N), and the complexity term/N (C/N) where N is sample size for all the data (top half of the table) and for the sample size controlled between months to only sites with water all three months.
Supplementary Figure 3.3

*AIC, SSE, and C terms summary graphs.*

**A.**

![AIC/N, SSE/N, and C/N graphs](image)

**B.**

![AIC, SSE, and C terms as calculated in AIC/N equation](image)

Notes: Graph of A) the AIC/N (blue), C/N (grey), and SSE/N (orange) terms and B) the AIC/N (blue), (log10SSE)/N (grey), and 2C/N (orange) terms.
3.8 Appendices A-C

3.8 Appendix A

Methods for Calculating CV.

**Traditional methods: Coefficient of variation (CV) and moving mean**

**average.** CV has been used by past researchers to describe stream spatial heterogeneity over time (e.g. Dent and Grimm, 1999; Asano et al., 2009; Hale and Godsey, 2019). Defined as the standard deviation divided by the sample mean, our data posed difficulty for using CV because 1) the sample size and mean changes as streams dry and 2) the streams appeared to have underlying longitudinal patterns at the sampling interval. Thus, we used four approaches to calculating monthly CV for each analyte (DIC, Cl⁻, DOC, TN, PO₄³⁻) in each stream for 2016 and 2017, which each had benefits and draw-backs. The first was to use all the data for each month in order to capture all the variation as the stream dried (called “all data”). However, this approach compared CV across changing sample sizes and means (April has a larger sample size/covers a larger extent than June and changing mean), to which CV is sensitive. The second approach addressed changing sample size by holding sample size constant between months as streams dried (called “all 3 months”). To do this, we excluded data for sites that did not have surface water all three months (i.e. were dry). While this method allows for direct comparisons of CV between months, it excluded important data that represented the spatiotemporal change in the system we intended to quantify. Furthermore, neither of the above approaches addressed potential underlying longitudinal patterns, which add a layer to the complexity of analyzing spatial heterogeneity.
To account for underlying longitudinal patterns, we calculated the average CV of a 5-site moving window (MW) CV, similar to methods presented by Asano et al. (2009). We conducted this analysis using all the data (“all data MW”) and using data from sites that had surface water present all three months (“3 surface water MW”). The moving window approach compensated for longitudinal patterns along the profile, however, shared the drawbacks of either changing sample size or excluding data for stream sites that dried. Ultimately, like others before us (Dent and Grimm, 1999), we emphasized trends more than specific CV and predicted CV would increase as streams dried if heterogeneity was increasing. We also continued our analysis of spatiotemporal heterogeneity using a time series approach to spatial statistical analyses that included modeling the stream patterns and using residuals to calculate standard deviation (SD\text{GAM}) and build semivariograms as described by Cooper et al. (1997).
3.8 Appendix B

*Results for Calculating CV.*

**Method 1: CV and moving mean analyses.** Quantifying the spatiotemporal patterns using CV gave mixed results across approaches and analytes for each stream and was sensitive to changes in means relative to standard deviation. In some cases, CV using four approaches yielded different variation between the same month using different approaches, and importantly different temporal trends in variation as streams dried (Supplementary Figure 3.1, 3.2). For instance, CV was highest in almost all cases for April and May when all the data was used to calculate CV, and for most analytes, using the moving window approach to calculate CV resulted in lower overall CV for each month (Supplementary Table 3.1). For DIC in each stream, CV temporal trends as streams dry differ between approaches of using all the data or holding sample size constant (Supplementary Figure 3.1, 3.2). For DIC and TN in both streams, CV is increased, especially when all the data is used for the calculation, as mean values decrease. CV was impacted by standard deviation and mean changing disproportionately, for instance in the case of burned MC TN. Standard deviation decreased from April to June by almost 50% but mean concentration decreased by >80%, thus CV increased substantially though this change was driven mostly by a larger decrease in mean concentration than in standard deviation. Later, we will show that the resulting drying trends of a given analyte calculated using CV differed in several examples from those calculated using AIC comparisons and semivariance.
Figure below demonstrate the difference in CV calculations using four different approaches and the means (open and closed) dots. For various methods depicted in the legend.
The table below gives the calculations for the graphs above showing CV calculations using four methods.

<table>
<thead>
<tr>
<th></th>
<th>Unburned JD</th>
<th></th>
<th>Burned MC</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DIC mg C/L</td>
<td></td>
<td>DIC mg C/L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>CV N=33</td>
<td>Mean ± SD</td>
<td>CV N=33</td>
</tr>
<tr>
<td></td>
<td>CV N=33</td>
<td></td>
<td>CV N=34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n</td>
<td></td>
<td>n</td>
<td></td>
</tr>
</tbody>
</table>

| April            | 7.67 ± 0.88 | 63.92      | 13.84     | 5.83 ± 0.38  | 13.38      | 5.54      |
| May              | 9.16 ± 3.25 | 56.59      | 7.66      | 8.35 ± 2.09  | 26.94      | 8.33      |
| June             | 13.15 ± 6.00| 27.83      | 9.46      | 11.42 ± 3.60 | 25.45      | 10.67     |

|                  | CI          |              | CI        |              |
|                  | Mean ± SD   | CV N=33      | Mean ± SD | CV N=33      |
|                  | CV N=35     |              | CV N=33   |              |
|                  | n           |              | n         |              |

| April            | 1.54 ± 0.25 | 18.11      | 11.85     | 1.07 ± 0.16  | 15.11      | 11.34     |
| May              | 1.15 ± 0.25 | 22.90      | 0.50      | 1.09 ± 0.15  | 14.00      | 5.86      |
| June             | 1.34 ± 0.22 | 19.72      | 7.44      | 1.34 ± 0.22  | 15.21      | 8.48      |

| DOC              |             |              |           |              |
|                  | Mean ± SD   | CV N=33      | Mean ± SD | CV N=34      |
|                  | CV N=34     |              | CV N=34   |              |
|                  | n           |              | n         |              |

| April            | 4.78 ± 1.56 | 32.64      | 23.32     | 4.68 ± 0.80  | 19.32      | 13.77     |
| May              | 5.91 ± 1.94 | 30.88      | 11.04     | 5.66 ± 1.20  | 20.64      | 12.26     |
| June             | 4.65 ± 1.60 | 35.44      | 23.33     | 4.51 ± 1.11  | 33.44      | 19.44     |

| TN               |             |              |           |              |
|                  | Mean ± SD   | CV N=33      | Mean ± SD | CV N=34      |
|                  | CV N=34     |              | CV N=34   |              |
|                  | n           |              | n         |              |

| April            | 1.25 ± 0.27 | 75.77      | 9.27      | 0.23 ± 0.03  | 12.84      | 14.00     |
| May              | 1.24 ± 0.15 | 62.44      | 9.44      | 0.19 ± 0.09  | 14.00      | 15.30     |
| June             | 1.21 ± 0.15 | 75.11      | 11.91     | 0.18 ± 0.05  | 25.77      | 22.23     |

| PO₄³⁻            |             |              |           |              |
|                  | Mean ± SD   | CV N=33      | Mean ± SD | CV N=34      |
|                  | CV N=34     |              | CV N=34   |              |
|                  | n           |              | n         |              |

| April            | 0.189 ± 0.006 | 50.66  | 54.00 | 58 | 0.08 ± 0.02 | 120.13 | 63.97 |
| May              | 0.021 ± 0.002 | 53.97  | 54.88 | 52.52 | 0.12 ± 0.020 | 56.57 | 51.42 |
| June             | 0.042 ± 0.040 | 53.75  | 53.57 | 52.89 | 0.13 ± 0.025 | 50.37 | 50.28 |
3.8 Appendix C

Discussion for Calculating CV.

Challenges with coefficient of variation. Although a common practice to use CV to describe variability in streams (e.g., Dent and Grimm, 1999; Asano et al., 2009; Hale and Godsey, 2019), we found that CV posed significant challenges to quantify spatiotemporal heterogeneity and often CV results did not agree with SD\textsubscript{GAM} or semivariogram results. Similar to challenges described by Cooper et al. (1997) in quantifying spatial patterns in perennial streams generally, quantifying spatiotemporal patterns for the two streams as they dried posed two major difficulties. First changing means and sample size as streams dried, and the second was longitudinal patterns, which also changed with stream drying, that underlay the spatial relationships between fine-spatial scale sites within the streams (Figure 3.4). For example, as burned MC dried in 2016, mean DIC concentration increased 3-fold while standard deviation increased 2-fold but overall CV decreased. This was also clearly exhibited by TN in burned MC with drying from April to June 2016 for both data scenarios using all the data and when sample size was held constant as mean TN concentration decreased 6-fold while standard deviation decreased by 2-fold and CV substantially increased (Supplementary Figure 3.1). Asano et al. (2009) demonstrate similar relationships where increased catchment area results in increased mean solute concentration and lower CV. In Asano et al.’s (2009) view, larger lower spatial variation. However, our findings highlight the limitations of using CV to assess spatial heterogeneity because SD\textsubscript{GAM} and semivariogram results show DIC semivariance (sill) to increase with drying in both streams. Thus, although Dent and Grim (1999) de-emphasized CV values and focused on
large changes, in our data inferences about increased CV indicating increased heterogeneity in a given month would be different using the two CV versus spatial statistics.

In addition to not agreeing with spatial statistical analyses, our results demonstrated that in most analyte cases, CV and, importantly, the interpretation of temporal patterns based on CV differed depending on which of the four approaches presented were used to calculate CV (Supplementary Figure 3.1, 2; Supplementary Table 3.1). For instance, in both streams, DIC CV increased from April to June (2016) when all the data was used to calculate CV compared to when the sample size was held constant between months ("3 month surface flow"; Supplementary Figure 3.1). For instance, in unburned JD where high solute concentrations were measured at the highest elevation sites, which differed from the bottom sites in aspect and vegetation (Figure 3.4). Once these sites were lost as the stream dried, the CV was impacted as standard deviation decreased when using all the data with no moving window for both DIC and TN. Therefore, this change in CV may have reflected changes in the underlying longitudinal pattern of the streams, not the spatial heterogeneity between fine-grain 50 m reach sites and the loss of sites with each sampling period. Under drying conditions, when using CV, we found it was unclear whether it was more appropriate to retain data and compare CVs of changing sample size or to force constant sample size, but exclude data. The different results found by the different CV approaches would result in different interpretations of the patterns occurring with stream drying and that alone disqualified CV as an appropriate tool to describe variation in our data. In the two ways outlined above, our results supported our expectation that CV would be an unsatisfactory analysis method.
Instead of relying on CV, we relied on spatial statistical analyses to evaluate and quantify heterogeneity.
Part 2

Chapter IV: What Teaching Techniques Are Currently Used at a Midsized Public University in Introductory STEM Classrooms?

Abstract

*Active learning* is an umbrella term for student-centered teaching techniques that engage students in the learning process, also described as being generative in the construction of knowledge. Despite a rich body of scholarship providing evidence that active learning improves learning and increases equity in introductory science courses, traditional lecturing remains the dominant teaching method used and modeled in undergraduate classrooms. Lack of faculty teacher training and limited institutional support or incentives are among some barriers. However, two recent surveys both show that Idaho State University (ISU) faculty and former students would like improved instructional practices. In this study, we investigated instructional practices used by triangulating external observer of classroom behavior (Classroom Observational Protocol for Undergraduate STEM or COPUS; Smith et al., 2014), instructor self-assessment (Teaching Practices Inventory or TPI; Wieman and Gilbert, 2014), and anonymous student survey data. We asked the following questions: What are the patterns of student and instructor behavior in introductory science, technology, engineering, and mathematics (STEM) courses at a midsized state university? What practices do instructors implement (or not implement) in these courses? What instructional practices can we identify to inform future teaching professional development (TPD)? We analyzed observational behavior data using the interactive, constructive, active, and passive framework developed by Chi and Wylie (2014) and categorized student and instructor
behaviors into “more” or “less” generative behaviors. Overall at ISU, the most frequent behaviors were less generative, and between instructors and students, instructors displayed generative behaviors more of the time than did students. The least frequent behaviors observed were specific to science practices including predicting and communication. Overall, the majority of instructors reported no formal teaching training and did not collect evaluative data from students throughout the semester. Finally, we found that of the highest TPI scores, instructors reported less lecturing, more time spent on the process of science, and more student questions; these practices contrasted with the lowest TPI scores. This project identifies gaps in ISU professional development and provides specific, locally derived suggestions to improved future faculty teaching professional development. ISU has an opportunity to increase evidence-based teaching practices and improve student satisfaction in introductory science courses.

4.1 Study Motivation

Instruction is a major component of any undergraduate student’s course experience. *Active learning* refers broadly to education reform that places students’ own construction of knowledge at the forefront of teaching (Weimer, 2013). In science, technology, engineering, and mathematics (STEM) fields, active learning environments offer students opportunities to engage in scientific thinking that faculty identify as important in the development of future scientists (AAAS, 2009). Many science-related skills that science educators seek to teach in introductory science courses, such as making predictions, evaluating evidence, or communicating with other scientists, require students to engage in a generative learning process. Furthermore, in a meta-analysis of over 200
classrooms, Freeman et al. (2014) found that active learning instructional techniques reduced failure rates in introductory undergraduate STEM courses by 55% overall. Active learning also has implications for diversity in STEM as underrepresented undergraduate student retention increased in active learning introductory biology courses (Canning et al., 2018).

Introductory and nonmajors science courses are critical to promoting science literacy and may be the first and sometimes only interaction students have with science (Jin and Bierma, 2013). Instructional quality in introductory courses is arguably the most important for student retention in STEM fields (Gasiewski et al., 2012) or exposure to science. In this way, these courses help to address science literacy, the ability of the general public to understand science (Krajcik and Sutherland, 2010). Introductory courses have the potential to spark interest or extinguish it for students experiencing science for the first time academically (Gasiewski et al., 2012). Though these courses can be overlooked as meaningful teaching opportunities by instructors, they are the gateway to any more specialized science fields and build the foundation for advanced classes (Gasiewski et al., 2012). Thus, active learning techniques are important to integrate into introductory curriculum.

However, despite education researchers’ widespread agreement that active learning increases student learning and the growing body of literature outlining and promoting these teaching practices (Brawner et al., 2002; Handelsman et al., 2004; Prince, 2004; Knight and Wood, 2005; Michael, 2006; Wood, 2009; Freeman et al., 2011; Weimer, 2013; Freeman et al., 2014; McGuire and McGuire, 2015; Harackiewicz et al., 2016; Strimaitis et al., 2017; Theobald et al., 2020), active learning remains
underpracticed by STEM faculty (Burrowes, 2003; Tanner and Allen, 2006; AAAS, 2009; Wood, 2009; Ueckert et al., 2011; Tanner, 2013; Freeman et al., 2014; Stains et al., 2018). Both broad and situation-specific barriers may account for continued use of traditional lecture, and teasing barriers apart in any one instance will require localized analyses and support. Establishing a baseline of classroom activity is a first step that can provide the measurement basis and direction for change.

At a midsized state university, Idaho State University (ISU), classroom behavior and the use of active learning remains poorly documented across STEM. However, both instructors and students desire improved instruction. In a 2018 ISU instructor survey, 90% identified active learning as the top teaching professional development (TPD) desire (ISU Program for Instructional Effectiveness, Fall 2018). Simultaneously, ISU conducted a student attrition survey that identified instructional quality as a top institutional reason that students stopped enrollment (42% in 2019; ISU Academic Affairs Office, 2019). In order to investigate both instructor and student concerns, we assessed classroom behaviors that characterized introductory STEM courses. Thus, we focused on characterizing common instructor behaviors and common student behaviors using classroom observation protocols to evaluate the use of active learning in these environments. If teaching support for implementing more active learning environments is offered in the future, we will be able to use this study to track growth toward this means. These results can also contribute to developing more effective TPD strategies and curriculum.
4.2 Theoretical Framework

*Active learning* is a widely recognized term for a broad array of teaching techniques that aim to offer opportunities for constructing knowledge. The term is problematic because it is poorly defined and unconstrained in its use (Prince, 2004). However, we use this term because it is commonly recognized outside of education researchers. Supporting student-generated knowledge is at the core of active learning. Thus, we refer to the interactive, constructive, active, and passive (ICAP) learning framework (Chi and Wylie, 2014) to help define generative student learning environments.

The ICAP framework helps to constrain active learning by structuring types of learning activities hierarchically as related to the depth of understanding (Chi and Wylie, 2014). *Interactive* (I) modes of engagement are defined as student-to-student talk, specifically dialogue between students to collaboratively elaborate, diagram, or refine ideas that students have generated (Chi and Wylie, 2014). Within the collaborative process, Chi and Wylie (2014) point out that students have to justify and defend ideas through dialogue, which further facilitates deep learning. *Constructive* (C) modes of engagement are generative, such as when students self-reflect, explain, or compare and contrast knowledge on their own (Chi and Wylie, 2014). *Active* (A) engagement is defined as being kinesthetic in nature, where students are physically manipulating materials or taking notes (Chi and Wylie, 2014). Lastly, *passive* (P) engagement is listening, reading, or observing. In the ICAP framework, interactive learning activities result in greater levels of student understanding such that I>C>A>P (Chi and Wylie, 2014). In our analyses, following the ICAP framework as outlined by Chi and Wylie
(2014), we grouped ICAP into more generative and less generative categories in order to understand STEM classroom behaviors. In our categorization, interactive and constructive behaviors were grouped as more generative and active and passive behaviors were grouped as less generative. Specific behavior groupings are discussed in Section 4.2.1, Table 4.3.

Under the ICAP framework, student interaction is placed at the highest level of engagement and deepest learning outcomes. ICAP has roots in learning and cognitive science and pulls from constructivist theories that learning occurs when students are cognitively engaged in producing their own ways of knowing (constructing), or reconstructing models through dialogue (co-constructing; Chi, 2009). Placing student interaction at the core of learning is echoed in other pedagogical frameworks as well, such as ambitious science learning (Windschitl et al., 2012), which is also built around student discourse. Though Chi (2009) points out that not all student interactions can be considered equal in being constructive for both parties, student-student interactions make this most possible. Indeed, this emphasis has been confirmed by many studies where increased student interaction has led to measurable student learning gains (Smith et al., 2009; Tanner, 2009; Ueckert et al., 2011; Tanner et al., 2013 Lund et al., 2015; Baily et al., 2018; Warfa et al., 2018). For instance, Baily et al. (2018) examined students who conducted reciprocal peer tutoring (RPT) in which students switched roles between a question asker and a question answerer in required study sessions around quiz material. Students who participated in the RPT interactions averaged ~4% higher on all exam scores than their counterparts who did not engage in these interactions. Warfa et al. (2018) analyzed collaborative student dialogues in a chemistry course and observed
groups establishing their own criteria to justify chemical processes and norms for depicting these processes. As Chi and Wylie (2014) point out, student interactions promote deeper student learning because students cognitively work through problems together.

4.3 Prior Research

4.3.1 STEM Diversity and Inclusion through Active Learning

Evidence shows that active learning environments are more equitable for underrepresented groups and may increase diversity in STEM (Tanner, 2013; Freeman, 2014; Bradforth and Miller, 2015; Strimaitis et al., 2017; Harackiewicz et al., 2016; Theobald et al., 2020). First, active learning courses have reduced failing rates (Freeman et al., 2014); this may result in increasing diversity if students most likely to struggle in courses can be predicted by demographic factors like race, gender, first-generation student, and so on. Indeed, Theobald et al. (2020) found that active learning class activities reduced the achievement gap observed between majority and underrepresented undergraduates by 33%, and that higher intensity of in-class activities were positively related to this reduction. In addition, the achievement gap was reduced in active learning high school biology curriculum that offered a wider range of hands-on science activities (Strimaitis et al., 2017) and when undergraduates created informal, frequent essays relating science content to their lives (Canning et al., 2018).
4.3.2 Evidenced-Based Success in Undergraduate Education at Multiple Scales

Three general focal points—instructional practices, student learning, and curriculum development—broadly encompass the lenses through which aspects of science education research occur (Talanquer, 2014; Reeves et al., 2016). Evidence for student success associated with active learning has been well documented and associated with each of these research focal points at multiple scales, or levels, of curriculum (Porter and Rossner, 2006; Wieman et al., 2010; Bradforth et al., 2015; Matz et al., 2018). Curriculum is implemented through specific activities in classrooms, course structure, and departmental structure of students’ four-year course of study (Figure 4.1). The implementation of active learning at small and large scales has shown gains in both student learning and satisfaction (Armbruster et al., 2009; Cleveland et al., 2017).
Notes: Active learning at three levels of implementation and examples of evidence showing learning gains at each. (A) depicts institutional levels where Derting and Ebert-May (2010) conducted a longitudinal study on student learning gains over the course of their undergraduate careers in a revised teaching curriculum (yellow represented the original curriculum and blue represented the revised) in major areas of biology. Learning gains were measured by the biology field subject exam, and results were broken down by subject including cell, molecular, organism, population/ ecology/evolution, and test totals. Course-level (B) active learning gains illustrated by the results of Freeman et al. (2014) found greater than 10% decreased failure rates in a meta-analysis of active learning STEM classes (n = 225). Instructional level (C) learning gains were reported by Baily et al. (2018) from a regularly assigned reciprocal peer-tutoring session by students who significantly tested higher on class exams than their counterparts who were not assigned the activity.

4.3.2.1 Institutional Structure. Institutional restructuring of students’ four-year course of study has allowed for the assessment of long-term learning gains of student-centered teaching. Derting and Ebert-May (2010) conducted a study at Michigan State
University (MSU) through which the biology introduction series courses were restructured around an inquiry-based curriculum. Students who went through the introductory inquiry curriculum performed higher at the end of their course of study on standardized tests and adopted mindsets that more closely resembled those of professional scientists compared to those who did not (Derting and Ebert-May, 2010). While these results were compelling, it is important to note that institutional support played a crucial role in executing active learning on this departmental scale (Matz et al., 2018). After the initial MSU biology series course revamp, a three-dimensional learning framework designed to improve students’ science practices, mastery of cross-cutting concepts (those that are transdisciplinary), and understanding of core disciplinary ideas was implemented and researched across introductory biology, chemistry, and physics courses (Matz et al., 2018). However, institutional differences in support such as the biology department program coordinators and the history of supporting innovative pedagogy led to varied implementation success between departments, where more implementation resulted in higher student learning gains within the various departments (Matz et al., 2018). Similarly, Wieman et al. (2010) identified department-level support as being the most important to widespread change such as an entire program of study. It is at this level that goals for students can be developed as institutional learning goals for science students and at this level support such as an education-specific role can be filled as a resource for faculty (Wieman et al., 2010).

4.3.2.2 Course Structure. At the course level, adopting active learning curriculum frameworks that structure large introductory undergraduate STEM courses has been shown to produce significant improvements among majority nonmajor students
(Ueckert et al., 2011) and poorly performing students (Ueckert et al., 2011; Freeman et al., 2014). Ueckert et al. (2011) saw a decrease in failing grades over 6 years and improved attitude toward biology after students participated in a large (900 students total per semester), redesigned active-learning introductory biology course. The course was restructured to include an electronic student response system (clicker system), pair-share discussion, web quizzes (offering immediate feedback), scaffolded (incremental) reading assignments, and group presentations in the lectures. The study reports a decrease in D/F/W grades from 34% to 25% and an increase in A/B grades from 40% to 51% following the redesign (Ueckert et al., 2011). Importantly, students self-reported increased confidence and content relevance, as well as decreased feelings that biology is difficult to understand (Ueckert et al., 2011).

A flipped classroom is a course-level active learning technique that helps to manage cognitive load (Abeysekera and Dawson, 2015) and has also seen learning gains compared to the lecture-based version of the same course (Barral et al., 2018). In flipped classrooms, students engage with class content outside of class through lecture videos and use class time for collaborative activities (Strayer, 2011). Compared to a nonflipped general biology course, Barral et al. (2018) found that students preformed higher on exams across question difficulties, though most pronounced at lower difficulty. The collective results from these studies have been supported by many others across small and large scales (Prince, 2004; Knight and Wood, 2005; Michael, 2006; Armbruster et al., 2009; Derting and Ebert-May, 2010; Freeman et al., 2014; Strimaitis et al., 2017; Canning et al., 2018; Theobald et al., 2020).
4.3.2.3 Instructional Structure. Individual instructors can facilitate specific activities that promote active learning even in large classes, which often promote or require interaction and discussion between students. One such example is “think-pair-share,” which asks students to explain material to each other during lectures and has been empirically shown to result in learning gains (Coleman et al., 1997; Smith et al., 2009; McGuire and McGuire, 2015). For instance, Smith et al. (2009) measured learning gains after the use of think-pair-share throughout 70-minute lectures, during which individual students applied peer-discussed concepts to novel scenarios. These increases were measured even when no student in the discussion group “knew the answer” (Smith et al., 2009). The act of explanation is one of the best learning tools (Tanner, 2009; McGuire and McGuire, 2015). Think-pair-share offers students a chance to engage with material in a way that resembles inquiry (inherent to practicing science) because they have to think before receiving “the right answer” from an instructor (Tanner and Allen, 2006). This is one of many techniques that has demonstrated learning success and could be implemented by faculty without restructuring a syllabus or with departmental support.

Despite the evidence that active learning results in learning gains, especially for underrepresented students, active learning is not common in STEM courses. In a study across 24 doctorate-granting institutions and 541 instructors observed, lecturing remains the most common behavior in classrooms by instructors and listening remains the most common student behavior (Stains et al., 2018). Stains et al. (2018) found that 80% of STEM courses across disciplines and course levels were “didactic,” or largely just lecture based, while 27% were “interactive,” using some active learning strategies. Only 8% of courses observed incorporated student-centered learning into the entire course structure.
Similar results have been found at other institutions using similar measurement tools (e.g., Smith et al., 2014; Lund et al., 2015). So, what are potential barriers to implementing active learning strategies more widely?

4.3.3 Barriers to Active Learning

Critical barriers to more widespread active learning implementation in higher education include the lack of formal teacher training and TPD with an established culture of “teaching as one was taught” and of financial support of incentives, logistics, and infrastructure. Often in higher education, teaching is “passed down” through an apprenticeship of observation (Borg, 2004); teaching is often a modeled skill set, not intentionally, thoughtfully developed through evidence-based techniques (Wilson and Cole, 1991; Feixas and Euler, 2012). Thus, difficulty arises in changing a skill set that is considered to be passively acquired. However, this does not mean instructors have no desire for change (Knight et al., 2006; Southerland et al., 2011). A study across STEM disciplines of science lecturers’ instructional knowledge found that while many lecturers emphasized the importance of content knowledge, most demonstrated little pedagogical knowledge yet had high recognition of its usefulness (Fraser, 2016). In contrast, Auerbach and Andrews (2018) found that instructors who were identified as using active learning instruction also had a deep level of pedagogical knowledge. Experts and novices were recruited from teaching specific networks, and expert instructors were distinguished by three qualities described in Auerbach et al. (2018): (a) more than 4 years of experience implementing active learning, (b) effectiveness of student learning from pre- to postsemester using validated measurements, and (c) evidence of adjusting teaching based
on systematic critical reflective practice comparing student learning data to intended learning objectives (Auerbach et al., 2018). In addition, experts had engaged in more than 40 hours of TPD and many had presented at TPD trainings (Auerbach et al., 2018). While there are known issues with self-reporting (May-Ebert et al., 2011), Auerbach and Andrews (2018) found that instructors who self-identified as active learning instructors showed high consistency with each other in analyzing pedagogical techniques observed in classroom instruction. Thus, the barrier of training in teaching may be a result of lacking institutional support and culture around teaching professional development and generally low valuing of teaching within academic research communities (Brownell and Tanner, 2012; Wieman et al., 2010; Fraser, 2016; Matz et al., 2018).

Another common barrier is often identified in the logistics of running active learning activities, either due to large class size or classroom infrastructure (Whiteside et al., 2010; Petersen and Gorman, 2014). Instructors most interested in overcoming these barriers should be targeted for TPD, and those advocating for improved infrastructure are those who express pedagogical discontentment (Southerland et al., 2011). *Pedagogical discontentment* describes those instructors who feel frustrated because they recognize that their teaching goals or philosophy are misaligned with their practice (Southerland et al., 2011). These instructors are primed for self-reflection and motivated for change (Southerland et al., 2011).

**4.3.3 Elucidating Patterns of Instruction and Learning in Classrooms**

To best support teaching practices that facilitate generative learning in classrooms, assessing the existing patterns of instruction in classrooms is a crucial first
Patterns can be considered across time (one lesson, over a class period or semester, over years) and at multiple scales (activity, lesson, course, department) and these patterns should help inform effective, targeted TPD. Methods to observe classroom patterns include quantitative counts of specific behaviors and qualitative descriptions of behaviors. Quantitative approaches can measure patterns broadly and provide a baseline of information; however, these approaches don’t address why or how some phenomenon occurs (Creswell and Clark, 2017). Quantitative analyses often use data sources such as surveys, interviews, or less prescribed observations where interactions are coded to enable deep understandings, but generally of fewer subjects (Creswell and Clark, 2017).

One quantitative tool developed to describe classroom patterns are standardized external observer protocols. For instance, the Reformed Teaching Observation Protocol (RTOP; MacIsaac and Falconer, 2002), Teaching Dimensions Observations Protocol (DTOP; Hora, 2013), and Classroom Observation Protocol for Undergraduate Science (COPUS; Smith et al., 2013) all standardize classroom observation. However, RTOP gives one score and uses curriculum moves, or planned lesson structure, and so does not offer a description of behaviors (Hora, 2013). DTOP offers fine-grain data of interactions and details the types of lecture, not just a course measurement of lecture as a general behavior, but this detail requires a high level of training between multiple researchers (Hora, 2015). COPUS offer more course measurements than DTOP, but the coarseness offers ease with training and more accuracy in the less detailed measures. DTOP and COPUS both quantify student and instructor behaviors in equal amounts of time to reveal behavior patterns of each. Additionally, some tools use a quasi-qualitative and quantitative approach such as an instructor self-assessment survey, the Teaching
Practices Inventory or TPI (Wieman and Gilbert, 2014). TPI uses evidence-based weighted scores, developed from evidence presented in the literature of specific classroom practices and curriculum structure. Patterns of behavior may help identify those instructors who support more generative learning in their classrooms and those who do not. Each case provides an opportunity to understand what can be improved and how, and contrasting cases can help discern patterns that inform TPD.

In this study we examined patterns of classroom behavior over a semester across STEM fields to understand which teaching strategies were employed. We focused on undergraduate introductory STEM courses instruction because these courses provide formative experiences in science for students from diverse academic backgrounds. Specifically, we asked the following: What student and instructor behavior patterns occur in introductory STEM courses at a midsized state university? What practices do instructors implement in these courses? What instructional practices can we categorize to inform future TPD? In this study, we limited our focus of “classroom” or “learning environment” to refer specifically to the lecture component of courses, as opposed to the laboratory components. To explore patterns, we triangulated our data sources to reflect three perspectives: an external observer using COPUS (Smith et al., 2013), an instructor TPI self-assessment (Wieman and Gilbert, 2014), and an anonymous student survey. We observed classrooms three times over a semester. Instructors took the TPI once, and the student survey was administered once at the end of the semester. Our goals were to report patterns and make recommendations for next steps in TPD for those instructors experiencing pedagogical discontentment.
4.4 Research Design

4.4.1 Study School

We chose a midsized state university, ISU, to conduct our study. Located in the relatively rural Intermountain West, ISU has about 13,000 undergraduate students and 1,500 graduate students, with almost half the student body registered as part-time students. ISU has a historic commitment to student-centered teaching and innovative pedagogical practices. This commitment is exemplified by one of the few Doctorate of Arts (DA) programs in the world, established in the 1970s (Serve et al., 2013). The DA program is designed for training undergraduate faculty members in biology, political science, mathematics, and previously in English. DA graduate students achieve expertise in both research in their field of study as well as in pedagogy and practice teaching. However, despite a long-standing commitment to pedagogical practices, ISU currently offers little campuswide TPD. For instance, though a university teaching and learning center did exist, it dissolved a few years after it was established (personal correspondence with Ed Nuhfer). In a 2018 survey, 90% of instructors polled (n = 264) at the ISU identified “active learning” as an area in which they would like training. As mentioned above, in a 2018 and 2019 student attrition survey (2018 n = 161; 2019 n = 189), instruction quality was identified as a top institutional reason for students leaving ISU (42% of students identified this reason in 2019). ISU can thus be described as a university with high interest in active learning instruction, but one with little institutional support for training instructors in these techniques.
4.4.2 Data Sources

We identified instructors who were teaching introductory STEM courses in the physics, geology, biology, and chemistry departments from 2018 to 2019 and recruited volunteers willing to participate with no additional compensation. Three forms of data gave a more complete picture of the classroom than any one data source: a classroom observer/researcher, instructor self-reporting, and an anonymous student survey.

4.4.2.1 Classroom Observer.

Observation Tool. Classroom observations were conducted during the lecture component of class during three unannounced times over the semester per course at the beginning, middle, and end of the course, not including exam days. We used a standard classroom observation protocol, COPUS (Smith et al., 2013; Supplementary Figure 4.1). For COPUS procedures, observers recorded instructor and student behaviors in alternating two-minute periods of time for the entire class. Conducting observations with two observers for eight classroom observations in six courses established an interobserver reliability of 84%. Collectively, our data represent observations from 18 instructors teaching 20 different courses and 585 students at ISU.

Quantitative Behavior Analyses. Statistical analysis of COPUS behavior data occurred in three ways: as a percentage of total behavior, as a percentage of time in the classroom, and in a comparison to national data (Stains et al., 2018). Before calculating, we aggregated the data for each observed class period and then aggregated each course’s data to assess the overall patterns at ISU. First, we calculated the percentage of a given observed behavior relative to the total behaviors for students and instructors by counting
the total number of times a given behavior type was observed for both students and instructors and dividing by the total observed behavior types. We next calculated the percentage of classroom time spent on a given behavior for both instructors and students. We did this by counting the number of 2-minute observations during which a behavior was observed and dividing the intervals by the total 2-minute intervals (for either students or instructors) to get a percentage of the total class time that behavior was observed. We report the overall results as well as compare a subset of the top five TPI scoring instructors and the lowest five TPI scoring instructors. Lastly, we compared the percentage of classroom time spent on a given behavior for students or instructors at ISU to national averages reported by Stains et al. (2018). In Stains et al.’s (2018) study, 71% of the national data represented lower-level undergraduate courses. Stains et al. reported the average time spent on the top three behaviors for students and instructors and the standard deviation for those averages. Thus, we reported the same calculations for comparison.

**ICAP Framework.** When we reported the study’s COPUS behavioral analyses, we presented student and instructor behaviors organized into the ICAP framework. In order to characterize instructor and student behaviors as more generative or less generative in science classrooms, we grouped classroom behaviors as either interactive or constructed (more generative) or as active and passive (less generative; Table 4.1). For instance, student behaviors included listening, individual think time, group work, and so forth. Each of these behaviors coincided with categories from ICAP such as passive (listening), constructive (individual think time), and interactive (group work). In Table 4.1 we outlined how each COPUS behavior is categorized using the ICAP framework.
This grouping did not assess the types of questions students answered in class or that were asked on a quiz, which would be necessary to determine if either of these behaviors were more appropriately described as a constructive, generative behavior versus active (kinesthetic), less generative behavior. However, neither behavior would be interactive unless it involved student-to-student discussion or a group quiz, in which case the behavior would also be counted as “group work” by COPUS observations. While the ICAP framework was developed from a student learning and behavior perspective, instructor behaviors are classified based on the goals of such a strategy. Thus, we used these classifications as a guideline for our analyses.
Table 4.1

*Groupings of COPUS Classroom Observation Behaviors Within an ICAP Framework.*

<table>
<thead>
<tr>
<th></th>
<th>Student behaviors learning depth</th>
<th>Instructor behaviors teaching depth</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interactive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group work</td>
<td>Moving through the classroom</td>
</tr>
<tr>
<td></td>
<td>Whole class discussion</td>
<td>Follow-up</td>
</tr>
<tr>
<td></td>
<td>Clicker question</td>
<td>Asking a clicker question</td>
</tr>
<tr>
<td></td>
<td></td>
<td>One on one</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Waiting</td>
</tr>
<tr>
<td><strong>Constructive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Individual think time</td>
<td>Posing a question</td>
</tr>
<tr>
<td></td>
<td>Presenting</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Predicting</td>
<td>Listening and answering a question</td>
</tr>
<tr>
<td></td>
<td>Asking a question</td>
<td></td>
</tr>
<tr>
<td><strong>Active</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Answering a question*</td>
<td>Listening and answering a question</td>
</tr>
<tr>
<td></td>
<td>Taking a quiz/test*</td>
<td>Demo/video</td>
</tr>
<tr>
<td></td>
<td>Listening and taking notes</td>
<td>Writing</td>
</tr>
<tr>
<td></td>
<td>(note taking is not distinguished by COPUS)</td>
<td></td>
</tr>
<tr>
<td><strong>Passive</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Listening</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lecturing</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Classroom observations using COPUS (Smith et al., 2013) categorized by the ICAP framework (Chi and Wylie, 2014). The asterisk (*) indicates behaviors that could vary widely in how they are categorized where either answering verbal or written questions with short, one-word answers or multiple choice are less generative behaviors.

4.4.2.2 Instructor Practices

_Instructor Self-Reporting._ We administered an instructor self-assessment once in the semester, which balanced the external observer data. Most participating instructors (n = 17 of 18) took a Teaching Practice Inventory (TPI) survey (Wieman and Gilbert, 2014), which asked instructors to report on their use of common course components including
items related to conducting class (lecture) and those used in grading and communicating with students (Supplementary Figure 4.2) and took 10–20 minutes to complete. Instructor responses were scored with points that are weighted based on evidence supporting the use of the practice found in the science education research literature. These TPI points were called “extent of use of research-based teaching practices” (ETP) scores (Wieman and Gilbert, 2014) and the highest possible score was 67. We also included three open-ended questions that asked what an instructor’s formal teaching training had been, how many new activities that they attempted a semester, and any suggestions for supporting teaching at the institution (Supplementary Figure 4.2).

**TPI Quantitative and Qualitative Analyses.** Analyses of TPI included taking quantitative approaches to reporting the score distribution and comparing means (± SD), as well as taking a more qualitative approach of instructor responses to specific components of the self-assessment survey. We evaluated the data overall for emergent trends of similarity among all the instructors. Then, we selected the highest and lowest five scoring instructors to investigate extreme differences and similarities that may have emerged from the practices and responses within those groups. Midrange scoring instructors were also grouped (n = 8) and compared to the highest and lowest scoring instructors’ results. We used data from these groups to compare teaching practices specifically self-reported in the TPI section (Section 3; Supplementary Figure 4.2) focused on “in-class features and activities” that most closely related to the observed COPUS behaviors (Smith et al., 2014; Auerbach et al., 2018). Differences between groups were quantified using Wilcoxon sign-ranked pared tests; however, we focused our
examination on the differences between the highest and lowest scoring instructors to emphasize clear differences in COPUS behaviors and qualitative patterns.

4.4.2.3. Student Experience

*Anonymous Student Survey.* We assessed the student perspective through an anonymous student survey (n = 585) given in the last week of class. We designed the survey to be short and consume little class time; it included four questions (Supplementary Figure 4.3). The first three student questions inquired about the semester frequency of assigned group work in class (lecture), assigned group work outside class (lecture), and student questions during lecture. We considered frequent to be >6 times a semester, moderately frequent to be 3–6 times a semester, infrequent to be <3 times a semester, and none to be zero times a semester. The last question asked students to rate their overall engagement during class (lecture) as high, moderate, or low. We analyzed these data using the aggregate of all student responses per course and correlated student responses to each student survey question with the TPI scores.

4.4.3 Protection of Human Subjects

This study complies with ISU-IRB protocol and was certified as exempt due to low risk as it did not require students or instructors to do anything outside of normal classroom activity (certified exempt number IRB-FY2018-256). Students and instructors were observed within the normal classroom context. Instructors were recruited via email and participated voluntarily without the knowledge of department chairs or other
administrators. Results were reported at the college level to protect students, instructors, and departments.

4.5 Results

Overall a higher percentage of “less generative” COPUS behaviors were observed than “more generative” (Table 4.2) for both students and instructors across ISU introductory STEM classrooms. Indeed, when fitting these behaviors into the ICAP learning framework (Chi and Wylie, 2014) and grouping them according to Table 4.2, 81% of total behaviors displayed by students were less generative while 71% of total behaviors of instructors were less generative. Between students and instructors, instructors displayed a higher percentage of more generative teaching of total behaviors (26%) compared to students (16%; Figure 4.2).
Table 4.2

*Percentages of Classroom Observations of Classroom Behaviors for Students and Instructors and Percentage of Time Spent Measured by COPUS.*

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Students</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% Observed</td>
<td>Mean % of total time</td>
<td>SD</td>
<td>SE</td>
</tr>
<tr>
<td><strong>More generative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asking questions</td>
<td></td>
<td>5.7</td>
<td>8.45</td>
<td>12.33</td>
<td>3.18</td>
</tr>
<tr>
<td>Group work</td>
<td></td>
<td>4.7</td>
<td>5.00</td>
<td>7.85</td>
<td>1.43</td>
</tr>
<tr>
<td>Individual thinking</td>
<td></td>
<td>3.4</td>
<td>4.97</td>
<td>7.07</td>
<td>1.83</td>
</tr>
<tr>
<td>Whole class discussion</td>
<td></td>
<td>0.6</td>
<td>0.84</td>
<td>1.69</td>
<td>0.44</td>
</tr>
<tr>
<td>Clicker questions</td>
<td></td>
<td>0.2</td>
<td>0.28</td>
<td>0.82</td>
<td>0.21</td>
</tr>
<tr>
<td>Predicting</td>
<td></td>
<td>0.2</td>
<td>0.24</td>
<td>0.99</td>
<td>0.26</td>
</tr>
<tr>
<td>Presenting</td>
<td></td>
<td>0.2</td>
<td>0.17</td>
<td>0.71</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>Less generative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td></td>
<td>61.3</td>
<td>85.13</td>
<td>8.56</td>
<td>2.21</td>
</tr>
<tr>
<td>Answering questions</td>
<td></td>
<td>17.0</td>
<td>22.69</td>
<td>20.01</td>
<td>5.17</td>
</tr>
<tr>
<td>Waiting</td>
<td></td>
<td>2.7</td>
<td>3.75</td>
<td>4.26</td>
<td>1.10</td>
</tr>
<tr>
<td>Test/quiz</td>
<td></td>
<td>0.5</td>
<td>0.46</td>
<td>1.46</td>
<td>0.38</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>3.6</td>
<td>5.14</td>
<td>6.23</td>
<td>1.61</td>
</tr>
<tr>
<td>Behavior</td>
<td>% Observed</td>
<td>Mean % of total time</td>
<td>SD</td>
<td>SE</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>----------------------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>More generative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posing questions</td>
<td>14</td>
<td>26.44</td>
<td>18.52</td>
<td>4.78</td>
<td></td>
</tr>
<tr>
<td>Following-up</td>
<td>4.7</td>
<td>9.38</td>
<td>11.02</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td>Waiting</td>
<td>2.0</td>
<td>3.30</td>
<td>4.97</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>One on one</td>
<td>2.3</td>
<td>4.60</td>
<td>7.34</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td>Moving/guiding</td>
<td>1.6</td>
<td>3.22</td>
<td>7.51</td>
<td>4.78</td>
<td></td>
</tr>
<tr>
<td>Clicker question</td>
<td>1.2</td>
<td>2.41</td>
<td>6.51</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>Less generative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lecturing</td>
<td>43</td>
<td>82.90</td>
<td>10.27</td>
<td>2.65</td>
<td></td>
</tr>
<tr>
<td>Writing in real time</td>
<td>15.4</td>
<td>32.23</td>
<td>31.77</td>
<td>8.20</td>
<td></td>
</tr>
<tr>
<td>Answering question</td>
<td>4.7</td>
<td>10.51</td>
<td>15.13</td>
<td>3.91</td>
<td></td>
</tr>
<tr>
<td>Demo/video</td>
<td>2.3</td>
<td>4.98</td>
<td>6.94</td>
<td>1.79</td>
<td></td>
</tr>
<tr>
<td>Administrative</td>
<td>5.9</td>
<td>11.99</td>
<td>4.95</td>
<td>1.28</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>2.9</td>
<td>5.69</td>
<td>5.11</td>
<td>1.32</td>
<td></td>
</tr>
</tbody>
</table>

Notes: COPUS (Smith et al., 2013) results put into the IPAC framework of more generative to less generative (Chi and Wylie, 2014) learning or teaching. Data were collected in 20 introductory STEM courses between 2018 and 2019. The “% observed” are the percentage of a particular observation out of all observations recorded, the “mean % of total time” refers to the mean percentage time a given behavior was recorded during all 2-minute intervals (can equal over 100% because multiple behaviors are recorded per 2-minute interval, SD is standard deviation, and SE is standard error.
Figure 4.2

*Percentage of Total Aggregated Observations of Student and Instructor Behaviors.*

Notes: Aggregated COPUS (Smith et al., 2013) classroom observations of 20 introductory STEM classrooms. Each behavior was coded as more generative (green) and less generative (purple) according to the ICAP framework (Chi and Wylie, 2014). Grey on the right indicates “Other” and is not categorized.

The greatest amount of time students behaved in the less generative ways was listening (85% of the total time) and answering questions (23% of the total time). Behind these less generative behaviors was a more generative behavior, asking questions (8% of the total time). Of the total time in class, students were least often observed predicting (0.24% of the total time) or presenting (0.17% of the total time; Figure 4.3, A), which are both more generative behaviors. The majority of the time, instructors were observed lecturing (83% of the total time) and writing in real time on the board or projector (32%
of the total time); these are less generative behaviors. Conversely, the third highest amount of time was allocated to instructors asking or posing questions to the class (26% of the total time), a more generative behavior. Instructors spent little time on the more generative behaviors of moving around or guiding (3% of the total time) the students as they worked and/or posing clicker questions (2% of the total time; Figure 4.3, B).

The highest five TPI scoring instructor behaviors differed in some cases from the lowest five scoring (Figure 4.3, C, D). For students, more time was spent completing more generative tasks of group work time and individual think time, and answering instructor questions, which we classified as less generative due to the nature of most instructor questions. For instructors, asking and following up occurred more for large amounts of time in the highest scoring instructors’ classrooms, which were both generative behaviors. Writing in real time was the largest difference observed for the highest-scoring scoring instructors, which is a less generative behavior. The top behaviors in ISU students (Figure 4.4, A) and instructors (Figure 4.4, B) found in introductory STEM courses were similar for students and instructors nationally (Stains et al., 2018; Figure 4.4).
Figure 4.3

*Student and Instructor COPUS Behaviors as a Percentage of the Total Class Time.*

Notes: Percent of student (A) and instructor (B) COPUS behaviors as a percent of the total time the behavior was observed in introductory STEM courses (n = 20). Error bars are SE. (C) Student and (D) instructor behaviors observed in the classrooms of the top five (light bars) and bottom five (dark bars) instructors as scored by self-reported evidence-based teaching practices (Teaching Practices Inventory; Wieman and Gilbert, 2014) as percentage of total time the behaviors were observed.
Figure 4.4

*Most Time Spent by Students and Instructors at Idaho State University Compared to Universities Nationally.*

Notes: The top three most common behaviors of students (A) and instructors (B) in introductory STEM courses at ISU (orange) and nationally (blue) as presented by Stains et al. (2018). Error bars are SD per reported by Stains et al. (2018).

Overall instructor TPI self-assessments (Wieman and Gilbert, 2014) demonstrate that instructor scores ranged from 20 to 44 and averaged 32 ± 6.9 (± SD) ETP points (Figure 4.5), or 30% to 65%, with an average of 48%. The highest five scoring instructors scored from 39 to 44 (41 ± 2.1, mean ± SD), the lowest five scoring instructors scored from 20 to 26 (24 ± 2.5, mean ± SD), while the midranging instructors scored from 28.6 to 36 (32 ± 2.8 mean ± SD). Each scoring group differed from each other significantly when comparing the highest to lowest (p = 0.02), the highest to midrange (p = 0.004), and lowest to midrange (p = 0.004). Overall the self-reported in-class practices (TPI section 3) patterns corroborated observed classroom behavior. For example, 83% of the time instructors were observed lecturing, and on the TPI, all instructors reported they lectured >60% of the time. However, of the five highest scores, instructors reported <80% lecturing compared to the five lowest scores in which 4 out of 5 instructors
reported lecturing 80%–100%. The midrange instructors more often reported 60%–80% time spent lecturing.

The in-class practices TPI section (section 3) scores significantly differed between the highest and lowest scoring instructors. Out of 15 possible points from TPI section 3, highest scoring instructors scored 10 ± 2.6 (average ± SD), while lowest scoring instructors scored 3.2 ± 1.1 (average ± SD). Midrange instructors’ in-class activities (TPI section 3) scores were 5.7 ± 2.4 (average ± SD) points. Comparisons of the in-class activities section (TPI section 3) scores differed significantly for the highest scoring group such that highest and the lowest scoring groups had a p-value of 0.01 and the highest and midrange groups had a p-value of 0.03. However, the lowest and midrange scoring groups did not differ (p = 0.64).

Despite the quantitative differences between the highest scoring instructors’ use of evidence-based in-class practices, classrooms between groups were relatively similar. For instance, class size did not differ significantly between groups, though the highest scoring instructor class sizes tended to be smaller. On average, the highest-scoring instructors’ class size was 34 ± 11 (average ± SD) students and this did not significantly differ from the lowest-scoring instructors’ class size of 56 ± 29 (average ± SD) students (p = 0.09) or from the midrange scoring instructors’ class size of 58 ± 50 (average ± SD) students (p = 0.39); similarly the lowest did not differ from midrange class size (p = 0.61). In addition, classroom structure was similar with majority of classes from each group being held in small auditorium seating.

Qualitative differences emerged from the TPI by the highest scoring instructors compared to the lowest scoring instructors (Table 4.3). These differences included
students posing questions, discussions about the relevance of material to students, the amount of time spent of the process of science, and the amount of literature read in the field. The highest scoring instructors reported on the TPI that students posed questions at least once per class while the lowest five reported zero. Discussions about importance of the class material from the students’ perspectives occurred more than ten times in a semester as reported by the highest scored compared to three times as reported by the lowest scored. Most of the highest scored instructors reported more than 25% of class time on the process of science (all reported >10%) on the TPI, while greater than half of the lowest scored instructors reported 0%-10% time spent on the process of science. Outside of class, the highest scoring instructors reported reading literature related to the course on the TPI, and none was reported by the lowest scoring instructors.

Figure 4.5

*Distribution of ISU Instructors’ Self-Reported Practices.*

Notes: The distribution of the ISU instructors’ n = 18. self-reported practices via the teaching practices inventory (TPI; Wieman et al., 2014) scores. Scores are calculated based on the extent of the use of research-based teaching practices (ETP) on a 0–67
integer range. The grey box plot above indicates the score range (grey lines), the interquartile range (grey box), and the mean (white line) is labeled.

Table 4.3

*Differences and Similarities Synthesized from Instructors’ Self-Reported Practices.*

<table>
<thead>
<tr>
<th>Characterizing instructor self-reported teaching practices of highest and lowest five scoring instructors</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest scoring</td>
<td>Lowest scoring</td>
<td>All</td>
</tr>
<tr>
<td>Lectured &lt;80% of time</td>
<td>Most lectured 80–100%</td>
<td>“None” or “no” formal teacher training</td>
</tr>
<tr>
<td>Students posed at least one question per class</td>
<td>No student-posed questions</td>
<td></td>
</tr>
<tr>
<td>&gt;10 discussions per semester about material relevance from students’ perspective</td>
<td>3 discussions per semester about material relevance from students’ perspective</td>
<td>No student course evaluation throughout the semester</td>
</tr>
<tr>
<td>Most spent &gt;25% time on science as process (all &gt;11%)</td>
<td>&gt;1/2 spent 0%–10% on science as process</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Patterns are extracted from the Teaching Practices Inventory (TPI; Wieman and Gilbert, 2014) comparing the highest five scoring instructors and the lowest five.

Qualitatively analyzed similarities between the highest scores and lowest scores were present in the amount of formal teaching background and collection of student evaluations throughout the semester. When asked to report how much formal training in teaching they had received, 9 of 17 reported they received “none” or “no” formal teaching training. Five instructors indicated that training had occurred through teaching assistant or graduate course experience, three indicated they attended professional development, and two had high school teaching experience. Neither group reported collecting student evaluations periodically throughout the semester.

The anonymous student surveys confirmed that higher scores on TPI were positively correlated (r = 0.7) with higher frequency of group work (Figure 4.6). In
addition, anonymous student surveys suggested more frequently asked questions (Q3) were positively correlated \( (r = 0.3) \) with self-reported engagement (Q4; Figure 4.6).

Figure 4.6

*Correlations of Instructor Self-Assessment (TPI) Scores and Anonymous Student Surveys.*

Notes: Color map of instructor self-assessment score (TPI; Wieman and Gilbert, 2014) and the anonymous student survey questions correlation values are such that red indicates a positive correlation, blue indicates a negative correlation, and more saturated colors indicate stronger correlations closer to 1 or -1.

4.6 Discussion

Our study quantified undergraduate introductory STEM classroom behavior of both students and instructors using a number of different measurements and perspectives including an external observations, self-assessment, and student surveys. This triangulation of data helped us to identify broad patterns of behavior across the sciences at ISU for both instructors and students, and assess how much time is spent engaging in more generative behaviors that result in deeper learning and increased learning gains. We found that both students and instructors engaged in more of the less generative behaviors
than the more generative behaviors. Compared to national studies, the three most common behaviors at ISU were identical to those identified to be most common nationally (Smith et al., 2014; Lund et al., 2015; Stains et al., 2018). Students in STEM classrooms mostly listened to lectures, answered instructor questions, and asked questions. Instructors most commonly lectured, wrote in front of the class, and answered student questions.

### 4.6.1 Contextualizing STEM ISU Patterns of Behavior and Practice

ISU COPUS behaviors can be contextualized by comparing results to a cross-campus study at University of Maine (Smith et al., 2014) which places ISU results at mid-low range. From 1 to 4 quadrants, ISU student and instructor behaviors resembled the lower third quadrant where 1 represented the most generative learning behaviors and 4 the least as established by the range of University of Maine COPUS observation profiles (Smith et al., 2014). The majority of student behavior in quadrant 1 comprised group work, while in the lowest generative category, quadrant 4, students averaged 94% listening (Smith et al., 2014). Similarly, instructor quadrant 1 consisted of 8% lecture on average, and the highest amounts of instructor moving/guiding and “one on ones” with students (57% average), and quadrant 4 consisted of 66% lectures on average with moving/guiding and “one on ones” not represented (Smith et al., 2014). Generally, in active learning environments COPUS results have been measurably distinguished by higher amounts of instructor follow-up, moving/guiding, and “one on ones” with higher amounts of student group work (Smith et al., 2014; Lund et al., 2015; Commeford et al., 2020; Kranzfelder et al., 2020).
The quantitative analyses of “in-class activities and features” extrapolated from TPI section 3 enabled comparisons between the highest, lowest, and midrange scoring instructors, and also situated ISU instructors relative to novice/expert active learning instructors reported in the literature. Several differences between those instructors who made up the highest TPI scores from the lowest emerged in our analyses. Students in the highest scoring instructors’ classrooms spent more time completing group work, and these instructors posed more questions. We found differences in highest scoring instructors who reported a lower amount of lecture time, more student-posed questions, more discussion on science relevance to students’ lives, and more discussion about the process of science than the lowest five scoring instructors. We note that TPI scores of 100% are not necessarily the goal, but that there may be practices that are overlooked or ignored that would be helpful for developing TPD. Of the highest and lowest scoring instructors, none had formal teaching professional development, and none collected student evaluations throughout the course. Furthermore, there was a strong correlation between the amount of group work and the TPI score as well as between the amount of questions students asked and their self-reported engagement level.

Despite the higher levels of in-class evidence-based practices reported by the highest scoring instructors, ISU in-class levels compared to novice active learning instructors recruited specifically from the science education field by Auerbach et al. (2018), who scored 10 ± 2.8 (average ± SD) on TPI section 3 (Auerbach et al., 2018; compare to ISU 10 ± 2.6, average ± SD). The lowest scoring and midrange instructors fell far below the novice active learning instructors for this TPI section. Auerbach et al. (2018) considered active learning instructors to be novice if they had less than 40 TPD
hours and did not show evidence for effective student learning gains, or systematic critical reflection of teaching. Thus, the disparity in lower scoring ISU instructors’ practices potentially reflects the low TPD levels and teaching support reported as a barrier to implementing active learning. Without TPD support or engagement, highest scoring instructors are likely employing one of the other expert criteria Auerbach et al. (2018) put forth, though they may be doing so informally; this suggestion warrants further qualitative investigation. Desired outcomes reflected by the behavior pattern analyses in our results were to provide support for teaching at ISU that could specify the desired PTD from the 2018 instructor survey by making recommendations to develop TPD curriculum within existing infrastructure and to target those evidence-based practices few ISU instructors employ.

4.6.2 Least Common Behaviors: Missed Opportunities

While the most common behaviors we observed corroborated nationally reported most common behaviors (Smith et al., 2014; Lund et al., 2015; Stains et al., 2018), the least common student behaviors are potentially more important to examine. We observed that the least amount of students’ time was spent predicting and presenting, both of which reflect core scientific competencies identified for undergraduate biology education (AAAS, 2009). Predicting is integrated into the practice of science practice, and presenting is integrated into science communication. There was a disconnect in this case between the behavior observation results and the self-reporting of high-scoring TPI instructors who reported a greater amount of time devoted during class sessions for discussing the process of science, which may indicate they did not cover these topics
using more generative, student-centered techniques. Given the alignment of ISU to the national most common behaviors, it seems likely that although those data were not published by Stains et al. (2018), these least common behaviors are also similar nationally. Indeed, Smith et al. (2014) and Lund et al. (2015) reported the same low time spent for students predicting or presenting but, again, emphasized most common behaviors. Little encouragement of students predicting or communicating science in lecture sessions places the importance of science content over practice or process.

As one instructor in our study indicated, it may be argued that while these skills are often practiced during laboratory time, not engaging in them during the lecture component of courses is a missed opportunity for students to practice. Indeed, of the 20 ISU courses we observed, many (15) are structured as a 3-hour lecture and a separate 3-hour laboratory section where presumably more hands-on predicting may occur; to what degree this model of split curriculum is effective for students’ learning requires more research. However, if prediction and communication behaviors represent core science competencies, most commonly encouraged during laboratory section curriculum, and graduate student teaching assistants (TAs) often teach labs, these important competencies are mostly facilitated by the most novice teacher and scientists. In this case, TAs then require high-quality TPD (Gardner and Jones, 2011) to support their implementing these challenging teaching techniques, the assessment of which was outside this study scope. Additional research could also focus on the effectiveness of graduate student TAs compared to lecture instructors implementing active learning techniques and student learning in each setting.
Generally, we found that instructors displayed behaviors of more generative teaching than students displayed learning. This finding may imply that instructors are more comfortable enacting generative teaching behaviors than asking students in their courses to do so. These results are corroborated by Smith et al.’s (2014) data and perhaps reflect the traditional education mode that assumes instructors drive learning, as opposed to students (Jones and Brader-Araje, 2002). Thus, the instructors may have been focused on what they were doing, not what students were doing. Past work suggests that a fundamental shift from teacher-centered to student-centered teaching requires an underlying philosophy that supports this approach and is related to ways instructors view learning and knowledge theory, whether they articulate it or not. Our data were not specific enough to discern variations in lecturing or other pedagogical moves that might be used during a lecture, like the use of humor (Hora, 2015) or interaction styles observed in classrooms (Grinath and Southerland, 2019; Kranzfelder et al., 2020), which could be considered more fine grained than COPUS behaviors.

4.6.3 Implications for Student Learning Processes and Expanding STEM Participation

The student survey results suggest two things: (a) more group work in class reported by students was correlated with higher TPI scores, and (b) when students asked more questions, they felt more engaged. One of the patterns that emerged from the TPI analysis was that those with top TPI scores also all reported that students posed questions at least once per class period. This finding is important because increased student question asking over a semester has been associated with increased student content knowledge and course completion, including by underrepresented students even in active learning environments.
(Lo, 2018). That more student questions were associated with more generative class behaviors was also suggested by Smith et al. (2014) who found that students asked more questions when instructors lectured less than 66% of the time.

The implications of these findings can extend to diversity and retaining underrepresented groups in STEM fields at ISU, highlighting the importance of promoting teaching techniques that engage student dialogue, increase sense of belonging, and connect to culture and family. Specifically, ISU is located in close proximity to Fort Hall Reservation, home to the Shoshone-Bannock tribes. Of the underrepresented minority groups in higher education in the United States, Native American students are among the least represented as students and in education research literature (Guillory and Wolverton, 2008), likely due to the legacy of oppression Native American communities have faced by educational institutions in the United States (Lundberg, 2014). However, research shows that Native American students have the highest retention and graduation rates when their communities and families are involved in their education (Guillory and Wolverton, 2008; Guillory, 2009; Lundberg, 2014). Thus, institutional practices that link students’ academic experience to their home lives have been shown to support Native American students best (Lundberg, 2014), though as Lundberg (2014) points out, the connection to student learning has not been made directly (Lundberg, 2014).

Harackiewicz et al. (2016) and Canning et al. (2018) have observed achievement gap decreases between majority and minority biology students when students wrote about how classroom material related to their home lives. Harackiewicz et al. (2016) also found that underrepresented students were much more likely to be motivated by giving back to their communities than majority students and based on the literature of Native American
student retention, this motivation could be amplified for Native American students. Thus, teaching with strategies that promote student dialogue, thought to increase sense of belonging and expand students’ sense of identity in classrooms (O’Conner et al., 2015), may be particularly effective for Native American students if instructors support them connecting their cultural, family, and classroom experiences Harackiewicz et al. (2016). This suggestion warrants further investigation for this student population.

4.6.4 Addressing Barriers: Patterns Inform Teaching Professional Development

Our results can contribute to addressing barriers first by assessing teaching strategies displayed in introductory STEM courses and second by providing recommendations for TPD. More than half the instructors participating indicated they had “no” or “none” formal teaching training on the additional TPI questions that we administered; the most common source of training listed was graduate school. Workshops, trainings, generally increased institutional support, and building community around TPD were all suggested by instructors as modes of improved support. Indeed, building community and enabling collective participation around teaching has been shown to be an important component of TPD (Wilson, 2013). That graduate training was a primary source of TPD was not surprising. The importance of graduate student teaching training has been recognized (Gardner and Jones, 2011; Reeves et al., 2016). Our findings support the experience that after TA training is completed, little support other than mentors modeling teaching has typically contributed to TPD, which is typical when scientists begin their academic teaching careers (Wilson and Cole, 1991; Borg, 2002). This sets up two interesting paradoxes. First, TAs who go on to be instructors may shift
away from potentially facilitating more student predications and generative behaviors in TPD-supported laboratory classes to more traditional, less generative modes like lecturing. Second, as faculty in our study reported, most faculty who have mentored and modeled have had only a novice level of pedagogical knowledge or training (Feixas and Euler, 2012). Each of these paradoxical experiences are avenues for future research for addressing structural barriers to teaching in higher education.

4.6.5 Research in Action: Informing a New Model for Teaching Professional Development

In response to instructors’ self-reported low levels of formal training, we developed a model of TPD curriculum around active learning workshops for STEM instructors that modeled active learning (Tanner and Allen, 2006; Gardener and Jones, 2011; Brownell and Tanner, 2012) in guiding instructors to develop a lesson plan around practicing an evidence-based teaching skill (Gess-Newsomeman, 2015). One way we planned to guide instructors was through reflection, because research has identified guided reflection as a component of effective TPD (Southerland et al., 2011; Feixas and Euler, 2012; Clara et al., 2019). The research component of the TPD included a model we developed that paired ISU DA graduate students (trained in pedagogy) and instructors to implement active learning teaching practices along with supported reflective feedback practice (Bell, 2001; Gormally et al., 2014). Supported reflective feedback occurs when feedback is offered on reflections with the goal of guiding growth, not provide evaluation (Bell, 2001). We ran two pilot workshops (4.10 Appendices A-D) and developed a research plan to test the impacts on pedagogical knowledge and skill, self-efficacy,
pedagogical discontentment, and collegiality. These workshops were developed in response to the results of our findings, began to address the concerns of instructors, and provide an avenue for building on the results reported above. Since the time of this study, ISU has also implemented university-wide, non-STEM-specific short workshops as well as biology-specific teaching strategy seminars. If interventions or increased TPD at ISU should occur, our classroom data can provide a baseline for observing shifts in practices and behaviors.

Specific TPD curriculum recommendations directly informed by our data include practice instructing students to engage in generative learning behaviors, collecting student data throughout the semester, and instructor development of teaching philosophy. For instance, informal interviews we conducted in addition to our research to share individual results with interested instructors revealed instructors struggled to articulate their teaching philosophies. We did not interpret this as evidence that instructors do not have teaching philosophies, but that they lack the language and pedagogical framework to describe them. Thus, one TPD approach is through the development of a written teaching philosophy statement and guided practice of lessons that fulfill components of the developed philosophy. We continue to recommend, as Southerland et al. (2011) suggest, to focus TPD on instructors who display pedagogical discontentment. Instructors who recognize a misalignment between their teaching goals and their teaching practices were also those interested in their individual observation results and discussing teaching strategies. Furthermore, research suggests that impactful TPD is longer in duration than a workshop, lasting a year to allow for instructors to learn over time and for follow-up (Gibbs and Coffey, 2004; Stes et al., 2010; Feixas and Euler, 2012; Trigwell et al., 2016),
which would be adequate time to address the above recommendations. Given the benefits of active learning to students, it is reasonable to assert that if time and financial compensation support TAs (Gardner and Jones, 2011), instructors should also be supported and encouraged to continue TPD.

Additional improvements to professional development can be understood through the variations in implementation across a standardized but generative curriculum. Understanding how different instructors implement generative learning helps to further refine what behaviors have impacts on student learning or engagement. Lo (2018) found that differences in the amount of time for student questions and moving/guiding behaviors were associated with student outcome difference between two instructors delivering the same active learning curriculum. Even more nuanced were Grinath and Southerland’s (2019) findings that the TA differences in the questioning sequences and question types as factual or observational as they delivered a generative laboratory curriculum had consequences for eliciting student discussion that elaborated and expanded student ideas. Similarly, Kranzfelder (2020) found that in active learning curricula, instructors still relied on authoritarian interactions with students, as opposed to dialogical. These findings emphasize the point that pedagogical knowledge alone is not enough, but practice with skills is necessary as well (Gess-Newsomeman, 2015). Thus, future research into instructor implementation within a shared generative learning context can provide contributions toward developing TPD that best supports generative behaviors in STEM classrooms.
4.7 Conclusion

ISU faculty have identified a desire for active learning TPD, and the instructor self-assessment showed that indeed, across the sciences there is an opportunity to bolster the instructors’ teaching strategies. As a whole, ISU introductory science instructors seemed to be positioned as having pedagogical discontentment around teaching, given the unified request for TPD around active learning strategies and high levels of less generative classroom behaviors. The self-assessment survey corroborates with the TPD support survey in that each showed interest for increased opportunities and targeted TPD. Overall instructors displayed more generative behaviors than students did in STEM classes, suggesting instructors may need support to shift pedagogical practices in order to facilitate students being more cognitively engaged. We recommend guiding instructors in opportunities for science students to make predictions and allow for science communication either through more group work or through student presentations. Given that the majority of instructors that we surveyed didn’t collect student feedback to inform their teaching throughout the semester, did not spend much class time on the process of science, and did not report many student-posed questions to the class, TPD should also target building these practices. While ISU has started providing more opportunities for TPD, the low level of institutional support provides an opportunity to greatly increase these efforts. STEM classes are the largest on ISU’s campus, and so increasing evidence-based STEM teaching has the potential to measurably improve student retention, given the importance of instructional quality in student attrition surveys. From the perspective of diversifying STEM at ISU, the university also has a responsibility to support
underrepresented students. We think such improvements in introductory-level science classes can contribute to better experiences for all students.
4.8 Works Cited


4.9 Supplementary Materials

Supplementary Figure 4.1

*COPUS (Smith et al., 2013)* Materials.

A.

**COPUS**

classroom observation protocol for undergraduate STEM  
*Smith et al., 2013*

<table>
<thead>
<tr>
<th>Student behaviors</th>
<th>Instructor behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>Lecturing</td>
</tr>
<tr>
<td>Individual thinking</td>
<td>Real-time writing</td>
</tr>
<tr>
<td>Discussing clicker question</td>
<td>Follow-up</td>
</tr>
<tr>
<td>Working in groups</td>
<td>Posing non-clicker question</td>
</tr>
<tr>
<td>Answering an instructor question</td>
<td>Clicker question</td>
</tr>
<tr>
<td>Asks a question</td>
<td>Answering a question</td>
</tr>
<tr>
<td>Engaged in whole class discussion</td>
<td>Moving through the class; guiding</td>
</tr>
<tr>
<td>Prediction</td>
<td>students</td>
</tr>
<tr>
<td>Presentation</td>
<td>One-on-one extended conversation</td>
</tr>
<tr>
<td>Test/quiz</td>
<td>Demo/video</td>
</tr>
<tr>
<td>Waiting</td>
<td>Waiting</td>
</tr>
<tr>
<td>Other</td>
<td>Other</td>
</tr>
</tbody>
</table>

B.

1. L-Listening; Ind-Individual thinking; CG-Clicker Q-discussion; WG-Worksheet group work; OG-Other group work; A-Prd-Predicting; SP-Student present; TQ-Test/quiz; W-Waiting; O-Other

2. Lec-Lecturing; RtW-Writing; FUp-Follow-up; PQ-Pose Q; CQ-Clicker Q; AnQ-Answer Q; MG-Moving/Guiding; 101-

For each 2 minute interval, check columns to show what’s happening in each category (or draw vertical line to indicate COPUS)

Notes: A) list of behaviors for students and instructors and B) an observation form example of COPUS observation sheet where observations are taken for students and instructors in alternating 2-minute intervals for a class period. COPUS materials are available online at: http://www.cwsei.ubc.ca/resources/COPUS.htm
Supplementary Figure 4.2

Teaching Practices Inventory (Wieman and Gilbert, 2014).

Teaching Practices Inventory (TPI; A) was distributed to all instructors as well as (B) three open ended questions, which I created. Section three that was analyzed is outlined in red.

A. TPI instructors self-reported assessment. Also available online at: http://www.cwsei.ubc.ca/resources/TeachingPracticesInventory.htm
III. In-class features and activities

A. Various

Give approximate average number:

Average number of times per class: pause to ask for questions

Average number of times per class: have small group discussions or problem solving

Average number of times per class: show demonstrations, simulations, or video clips

Average number of times per class: show demonstrations, simulations, or video where students first record predictions (write down, etc.) and then afterwards explicitly compare observations with predictions

Average number of discussions per term on why material useful and/or interesting from students' perspective

Comments on above (if any): ____________________________________________

Check all that occurred in your course:

☐ Students asked to read/view material on upcoming class session
☐ Students read/view material on upcoming class session and complete assignments or quizzes on it shortly before class or at beginning of class
☐ Reflective activity at end of class, e.g. "one-minute paper" or similar (students briefly answering questions, reflecting on lecture and/or their learning, etc.)
☐ Student presentations (verbal or poster)

Fraction of typical class period you spend lecturing/talking to whole class (presenting content, deriving mathematical results, presenting a problem solution, ...)

☐ 0-20%
☐ 20-40%
☐ 40-60%
☐ 60-80%
☐ 80-100%

Considering the time spent on the major topics, approximately what fraction was spent on the process by which the theory/model/concept was developed, including the experimental methods and results that support specific theories?

☐ 0-10%
☐ 11-25%
☐ more than 25%
### B. Individual Student Responses (ISR)

If a student response method is used to collect responses from all students IN REAL TIME IN CLASS, what method is used? (check all that occurred in your course)

- raising hands
- raising colored cards
- electronic (e.g. “clickers”) with student identifier
- electronic anonymous
- written student responses that are collected and reviewed in real time
- Other (please specify)

If you selected other, please specify __________________________________________________________________________

Number of ISR questions posed followed by student-student discussion per class ____

Number of times ISR used as quiz (counts for marks and no student discussion) per class ____

### I. Assignments (check all that occurred in your course)

- Homework/problem sets assigned or suggested but did not contribute to course grade
- Homework/problem sets assigned and contributed to course grade at intervals of 2 weeks or less
- Paper or project (an assignment taking longer than two weeks and involving some degree of student control in choice of topic or design)
- Encouragement and facilitation for students to work collaboratively on their assignments
- Explicit group assignments
- Other (please specify)

If you selected other, please specify __________________________________________________________________________

### Feedback and testing: including grading policies (check all that occurred in your course)

#### A. Feedback from students to instructor during the term

- Midterm course evaluation
- Repeated online or paper feedback or via some other collection means such as clickers
- Other (please specify)

If you selected other, please specify __________________________________________________________________________

#### B. Feedback to students (check all that occurred in your course)

- Assignments with feedback from instructor, teaching assistant, or peer before grading or with opportunity to redo work to improve grade
- Students see graded assignments
- Students see assignment answer key and/or grading rubric
- Students see graded midterm exam(s)/quizzes
- Students see midterm exam(s)/quizzes answer key(s)
- Students explicitly encouraged to meet individually with you
- Other (please specify)

If you selected other, please specify __________________________________________________________________________
C. Testing and grading

Number of tests during term that reflect course expectations (e.g. midterm exams, but not final exams)

Approximate fraction of test scores from questions that required students to explain reasoning

Approximate breakdown of course grade (% in each of the following categories)

<table>
<thead>
<tr>
<th>Component</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final exam</td>
<td>%</td>
</tr>
<tr>
<td>Midterm/other exam(s)</td>
<td>%</td>
</tr>
<tr>
<td>Homework assignments</td>
<td>%</td>
</tr>
<tr>
<td>Paper(s) or project(s)</td>
<td>%</td>
</tr>
<tr>
<td>In-class activities</td>
<td>%</td>
</tr>
<tr>
<td>In-class quizzes</td>
<td>%</td>
</tr>
<tr>
<td>Online quizzes</td>
<td>%</td>
</tr>
<tr>
<td>Participation</td>
<td>%</td>
</tr>
<tr>
<td>Lab component</td>
<td>%</td>
</tr>
<tr>
<td>Other</td>
<td>%</td>
</tr>
</tbody>
</table>

If you selected other, please specify:

VI. Other (check all that occurred in your course)

- Assessment given at beginning of course to assess background knowledge
- Use of instructor-independent pre-post test (e.g. concept inventory) to measure learning
- Use of a consistent measure of learning that is repeated in multiple offerings of the course to compare learning
- Use of pre-post survey of student interest and/or perceptions about the subject
- Opportunities for students’ self-evaluation of learning
- Students provided with opportunities to have some control over their learning, such as choice of topics for course, paper, or project, choice of assessment methods, etc.
- New teaching methods or materials were tried along with measurements to determine their impact on student learning

VII. Training and guidance of Teaching Assistants (check all that occurred in your course)

- No TAs for course
- TAs must satisfy English language skills criteria
- TAs receive ½ day or more of training in teaching
- There are Instructor-TA meetings every two weeks or more frequently where student learning and difficulties, and the teaching of upcoming material are discussed.
- TAs are undergraduates
- TAs are graduate students
- Other (please specify)

If you selected other, please specify
Three additional questions we added to Wieman and Gilbert (2014):
1- How much formal training in teaching have you received?
2- How many new activities or lessons do you try a semester?
3- What changes could be made at ISU to help you teach more effectively
Supplementary Figure 4.3

Anonymous Student Survey

These 4 questions are anonymous and for research (only the researchers will see this piece of paper). The purpose of this research is to assess teaching in introductory science classes at ISU. Questions? Contact Ruth MacNeille at macnruth@isu.edu.

Please circle one answer to the following:

Do you identify as Male / Female / Non-binary gender

1) How often are you assigned group work in class (lecture not lab)?
   a) Never
   b) 1-3 times a semester
   c) 3-6 a semester
   d) More than 6 times a semester

2) How often are you assigned group work outside of class (lecture not lab)?
   a) Never
   b) 1-3 times a semester
   c) 3-6 a semester
   d) More than 6 times a semester

3) How often do you ask the instructor questions in class (in lecture not lab)?
   a) Never
   b) 1-3 times a semester
   c) 3-6 a semester
   d) More than 6 times a semester

4) Overall, on a scale from 1-3 (1 is low level and 3 is high level), how engaging do you feel in this lecture?
   a) 1 (not very engaging)
   b) 2 (moderately engaging)
   c) 3 (highly engaging)

Thank you for participating!
4.10 Appendices A-D

4.10 Appendices A-C

*Brief Research Plan.*

Research questions (4.10 Appendix A), conjecture map (4.10 Appendix B) and methodological approach (4.10 Appendix C) for teaching professional development model and workshops.
4.10 Appendix A

*Research Question.*

Is a graduate student-instructor partnership structured around a supportive reflective feedback model of professional development effective at shifting instructors’ pedagogical discontentment and both students’ and instructors’ self-efficacy with regards to implementing generative, interactive teaching strategies?
4.10 Appendix B

Conjecture Map.

Our conjecture is that a model for teaching professional development that utilizes a graduate student-instructor partnership structured around reflective feedback model will result in 1) increased pedagogical knowledge/skill, 2) decreased instructor pedagogical discontentment, 3) increased both instructor and student self-efficacy, and 4) increased collegiality.

Notes: Conjecture map outlines the conjecture being tested through 5 design elements, the processes that are structure the design elements and result in the desired outcomes. Because reflection is an integral process in each element that implements our teaching professional development model, there is an arrow going between the two.
4.10 Appendix C

Conjecture Map.

In our designed-based research model, we will iterate on our professional development model by running our workshops at least three times in a semester for three semesters. We will analyze the data and evaluate the outcomes in between each iteration so we can make informed modification and adaptations to the next workshops. Iterated workshops can have the same participants or new participants.

Notes: Modifications of our conjecture map illustrate our design-based methodology though which iterations of our teaching professional development model will be informed and adapted in response the outcomes of the previous iteration.
4.10 Appendix D

*Data Collected.*

We will collect recorded and written data from each design element at each stage (4.10 Appendix Table D). From stage 1, we will collect self-efficacy data from the D.A. student and discontentment information from faculty. In the planning workshop, the session will be recorded and the lesson plan will provide a written artifact. In the teaching implementation and observation stage, classroom observation notes will provide an artifact. Lastly, in the reflection and debrief stage we will collect a recorded debrief session between the D.A. student and faculty in addition to individually written reflections and recorded semi-structured interview.

4.10 Appendix Table D

*Research Plan.*

<table>
<thead>
<tr>
<th>Stage 1: Pairing</th>
<th>Stage 2: Plan</th>
<th>Stage 3: Teach</th>
<th>Stage 4: Reflect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faculty: discontentment semi-structured survey; DA: self-efficacy; Science Teaching Efficacy Beliefs Instrument (Riggs &amp; Enochs, 1990); Identify areas of teaching interest Tanner (2013) Paired based on identified areas of teaching interest</td>
<td>Workshop and planning session video recorded</td>
<td>Classroom observation of implementation of teaching strategy quality; video recording for instructor</td>
<td>Debrief session and semi-structured survey and interview on experience (both) /discontentment (faculty)/self-efficacy (DA student)</td>
</tr>
</tbody>
</table>

Notes: A table describing the various stages within each phase of the research project.