

HABITAT HETEROGENEITY IN NEBRASKA STREAMS AND DISTRIBUTION
PREDICTION FOR TIER-1 CYPRINIDS USING MULTI-SCALE MODELING OF
FLUVIAL AND LANDSCAPE FEATURES

by

Connor P. Hart

A THESIS

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Master of Science

Major: Natural Resource Sciences

Under the Supervision of Professor Jonathan J. Spurgeon

Lincoln, Nebraska

December 2023

HABITAT HETEROGENEITY IN NEBRASKA STREAMS AND DISTRIBUTION PREDICTION FOR TIER-1 CYPRINIDS USING MULTI-SCALE MODELING OF FLUVIAL AND LANDSCAPE FEATURES

Connor Preston Hart, M.S.

University of Nebraska, 2023

Advisor: Jonathan J. Spurgeon

Multiscale environmental processes determine in-stream habitat conditions which drive species distributions. Habitat constitutes the physical template upon which ecological processes occur and species conduct life stage activities. Habitat heterogeneity promotes biodiversity of aquatic systems. Stream classification informs freshwater conservation by providing a useful framework to account for habitat heterogeneity, often based on landscape regions of similar environmental processes. A greater understanding of landscape-based classification frameworks as means to classify stream systems may improve understanding of drivers of biodiversity. Using Nebraska as a case study, on a statewide scale, objectives were 1) to characterize habitat availability for several at-risk fish species, and 2) to assess the categorization of stream habitat heterogeneity within multiple classification systems. Species distribution models provide a tool to characterize spatial habitat patterns and quantify species-habitat relations. Species distribution models were constructed from environmental data and historical presence data for 3 at-risk cyprinid species including Flathead Chub, Plains Minnow, and Northern Redbelly Dace. Our results support current knowledge of impactful habitat features based on species traits and environmental processes shaping site-scale habitat conditions. For all 3 species, hydrological factors such as flow velocity and discharge were among the most influential habitat conditions of examined variables. Further, variables representing stream size and

position in the drainage network consistently carried high influence. We used site-level stream data collected by the Nebraska Department of Environment and Energy (NDEE) from 1995-2021 to assess stream habitat heterogeneity among Biologically Unique Landscapes (BUL) at a statewide spatial extent. Multivariate cluster analysis and statistical similarity tests were used to determine if aquatic habitat within the BULs is unique and by which features they vary. No significant differences were found between aquatic environmental conditions within and outside BULs. Evidence for unique aquatic habitat conditions among BULs was limited, shown in minimal pairwise differences among BULs in statistical similarity tests. Furthermore, the classificatory power of the BULs underperformed all other classifiers and, as such, BULs may not improve statewide stream classification.

TABLE OF CONTENTS

| | |
|---|-----|
| LIST OF TABLES | iii |
| LIST OF FIGURES | v |
| CHAPTER 1: INTRODUCTION | 1 |
| <i>Physical Template of Streams</i> | 1 |
| <i>The Importance of Habitat Heterogeneity</i> | 2 |
| <i>Anthropogenic Disturbances</i> | 4 |
| <i>Study Area</i> | 7 |
| <i>Study Objectives</i> | 8 |
| References | 9 |
| CHAPTER 2: HABITAT RELATIONS OF FLATHEAD CHUB, PLAINS MINNOW, AND NORTHERN REDBELLY DACE | 13 |
| Abstract | 13 |
| Introduction | 14 |
| <i>Study Species</i> | 18 |
| Methods | 23 |
| <i>Environmental data</i> | 23 |
| <i>Statistical Analysis</i> | 25 |
| <i>Presence:Absence Data</i> | 25 |
| <i>Model Construction and Validation</i> | 26 |
| Results | 29 |
| Discussion | 31 |
| <i>Model Interpretation</i> | 31 |
| <i>Research Implications</i> | 39 |
| Tables | 41 |
| Figures | 48 |
| References | 58 |
| CHAPTER 3: AQUATIC HABITAT REPRESENTATION WITHIN NEBRASKA'S BIOLOGICALLY UNIQUE LANDSCAPES: IMPLICATIONS FOR STREAM CLASSIFICATION | 66 |
| Abstract | 66 |
| Introduction | 67 |
| Methods | 75 |

| | |
|---|-----|
| <i>Data Collection</i> | 75 |
| <i>Data Analysis</i> | 76 |
| <i>Multivariate Cluster Analysis</i> | 77 |
| <i>Similarity Analysis</i> | 80 |
| Results | 83 |
| <i>Summary of Similarity-based Clusters</i> | 83 |
| <i>Comparing BULs to Statewide Variation</i> | 85 |
| <i>Variation among BULs</i> | 86 |
| <i>Comparing Stream Classification Systems</i> | 87 |
| Discussion | 88 |
| Tables | 94 |
| Figures | 101 |
| References | 111 |
| CHAPTER 4: MANAGEMENT RECOMMENDATIONS AND FUTURE | |
| QUESTIONS | 117 |
| Management Recommendations | 117 |
| Future Research Questions..... | 119 |

LIST OF TABLES

| | |
|---|----|
| Table 2.1. Comparison of species traits for Flathead Chub, Plains Minnow, and Northern Redbelly Dace from the U.S. Geological Survey FishTraits Database (Frimpong and Angermeier 2009). | 41 |
| Table 2.2. Descriptions, units, and sources of environmental predictor variables used to train the species distribution models..... | 42 |
| Table 2.3. Relative and cumulative influence of predictor variables from the Flathead Chub model. See Table 2.2 for variable definitions, units, and sources..... | 43 |
| Table 2.4. Relative and cumulative influence of predictor variables from the Plains Minnow model. See Table 2.2 for variable definitions, units, and sources..... | 44 |
| Table 2.5. Relative and cumulative influence of predictor variables from the Northern Redbelly Dace model. See Table 2.2 for variable definitions, units, and sources..... | 45 |
| Table 2.6. Validation summary data for the species distribution models. Sensitivity denotes the rate of true positives in model predictions, while specificity denotes the rate of false positives. Area under curve (AUC) is a metric of model fit ranging from 0-1 determined by a receiver-operating characteristic curve. True skill statistic (TSS) is a measure of model accuracy ranging from -1 to +1..... | 46 |
| Table 2.7. Proportion of total flowline with predicted presence greater than 0.7 for each species based on Figure 4, Figure 7, and Figure 10. High presence length (HPL) denotes the sum of flowline length with predicted presence greater than 0.7 in kilometers. Total length (TL) denotes the sum length of all flowlines across Nebraska in kilometers. Percentage of high presence (PHP) denotes the proportion of flowline length with predicted presence greater than 0.7 as a percentage. | 47 |
| Table 3.1. Reduced variables from Nebraska Department of Environment and Energy (NDEE) sampling sites from 1995-2021 and their definitions used in multivariate cluster analysis..... | 94 |
| Table 3.2. Nebraska Department of Environment and Energy (NDEE) sampling site count by Biologically Unique Landscapes (BUL). BULs with 0 sites were excluded from analyses. | 95 |
| Table 3.3. Dunn test results for pairwise differences between similarity-based clusters. B denotes clusters generated from the Biologically Unique Landscapes (BUL) sites, N denotes clusters generated from the sites outside BULs, and A denotes clusters generated from the statewide subset. An asterisk denotes significance achieved at $p \leq 0.10$ | 96 |
| Table 3.4. Levene's test results for homogeneity of variance among sites within Biologically Unique Landscapes (BUL) and sites outside BULs. An asterisk denotes significance achieved at $p \leq 0.10$ and $W \geq 2.72$. See Table 3.1 for variable definitions. ... | 97 |

| | |
|--|-----|
| Table 3.5. Levene's test results for homogeneity of variance among sites within BULs and sites statewide. An asterisk denotes significance achieved at $p \leq 0.10$ and $W \geq 2.71$. See Table 3.1 for variable definitions. | 98 |
| Table 3.6. Levene's test results for homogeneity of variance among all Biologically Unique Landscapes (BUL). An asterisk denotes significance achieved at $p \leq 0.10$ and $W \geq 1.38$. See Table 3.1 for variable definitions. | 99 |
| Table 3.7. Average silhouette value of sites by classification system in descending order. Stream Order only has 8 groups because no sampled sites were in stream order 9. | 100 |

LIST OF FIGURES

| | |
|--|----|
| Figure 2.1. Correlation grid showing selected environmental predictor variables, all correlation ≤ 0.7 . See Table 2.2 for variable definitions. | 48 |
| Figure 2.2. Variable importance plots from the Flathead Chub model. The y-axes of all plots are the fitted functions, denoting predicted response where values above zero are positive influence on presence and values below zero are negative influence on presence. The x-axes are variable values, with labels showing the relative influence of each variable as a percentage. The ticks on top of the x-axis denote data deciles (i.e., tenths of the total input points). See Table 2.2 for variable definitions, units, and sources. | 49 |
| Figure 2.3. The 10-fold cross validation receiver operating-characteristic curve plot of the Flathead Chub model. The y-axis is sensitivity, or the rate of true positives. The x-axis is specificity, or the rate of false positives. In the bottom right corner the area under the curve ranging from 0-1 is noted. | 50 |
| Figure 2.4. The model predicted probability of Flathead Chub occurrence mapped over the National Hydrography Dataset (NHD) Flowlines with additional metrics. Darker color denotes higher predicted probability of occurrence. | 51 |
| Figure 2.5. Variable importance plots from the Plains Minnow model. The y-axes of all plots are the fitted functions, denoting predicted response where values above zero are positive influence on presence and values below zero are negative influence on presence. The x-axes are variable values, with labels showing the relative influence of each variable as a percentage. The ticks on top of the x-axis denote data deciles (i.e., tenths of the total input points). See Table 2.2 for variable definitions, units, and sources. | 52 |
| Figure 2.6. The 10-fold cross validation receiver operating-characteristic curve plot of the Plains Minnow model. The y-axis is sensitivity, or the rate of true positives. The x-axis is specificity, or the rate of false positives. In the bottom right corner the area under the curve ranging from 0-1 is noted. | 53 |
| Figure 2.7. The model predicted probability of Plains Minnow occurrence mapped over the National Hydrography Dataset (NHD) Flowlines with additional metrics. Darker color denotes higher predicted probability of occurrence. | 54 |
| Figure 2.8. Variable importance plots from the Northern Redbelly Dace model. The y-axes of all plots are the fitted functions, denoting predicted response where values above zero are positive influence on presence and values below zero are negative influence on presence. The x-axes are variable values, with labels showing the relative influence of each variable as a percentage. The ticks on top of the x-axis denote data deciles (i.e., tenths of the total input points). See Table 2.2 for variable definitions, units, and sources. | 55 |
| Figure 2.9. The 10-fold cross validation receiver operating-characteristic curve plot of the Northern Redbelly Dace model. The y-axis is sensitivity, or the rate of true positives. The | |

x-axis is specificity, or the rate of false positives. In the bottom right corner the area under the curve ranging from 0-1 is noted.....56

Figure 2.10. The model predicted probability of Northern Redbelly Dace occurrence mapped over the National Hydrography Dataset (NHD) Flowlines with additional metrics. Darker color denotes higher predicted probability of occurrence.....57

Figure 3.1. NDEE environmental sampling sites across Nebraska represented by red points. The green polygons are the Biologically Unique Landscapes (BULs).....101

Figure 3.2. Correlation plot of selected environmental variables. Darker shading illustrates a higher Pearson's correlation coefficient. See Table 3.1 for variable definitions.102

Figure 3.3. The average silhouette value produced by clustering schemes of up to 10 groups to determine the optimal number of clusters for sites inside Biologically Unique Landscapes (BUL; A), outside BULs (B), and statewide (C). The dotted line denotes the optimal number of clusters, where the silhouette value is maximized.103

Figure 3.4. On the right side is a dendrogram depicting the overall dissimilarity of the sites inside Biologically Unique Landscapes (BUL). Height on the y-axis denotes dissimilarity, with splits in the tree showing greater dissimilarity at greater height. The dotted line at $y = 300$ shows the height at which the dendrogram was cut to produce 2 clusters. The larger black box groups sites into cluster 1, while the smaller red box groups sites into cluster 2. On the left side, a radar plot shows means generated from normalized variables. The black line illustrates cluster 1, while the red line illustrates cluster 2.104

Figure 3.5. On the right side is a dendrogram depicting the overall dissimilarity of the sites outside Biologically Unique Landscapes (BUL). Height on the y-axis denotes dissimilarity, with splits in the tree showing greater dissimilarity at greater height. The dotted line at $y = 300$ shows the height at which the dendrogram was cut to produce 2 clusters. The larger black box groups sites into cluster 1, while the smaller red box groups sites into cluster 2. On the left side, a radar plot shows means generated from normalized variables. The black line illustrates cluster 1, while the red line illustrates cluster 2.105

Figure 3.6. On the right side is a dendrogram depicting the overall dissimilarity of the sites statewide. Height on the y-axis denotes dissimilarity, with splits in the tree showing greater dissimilarity at greater height. The dotted line at $y = 500$ shows the height at which the dendrogram was cut to produce 2 clusters. The larger black box groups sites into cluster 1, while the smaller red box groups sites into cluster 2. On the left side, a radar plot shows means generated from normalized variables. The black line illustrates cluster 1, while the red line illustrates cluster 2.106

Figure 3.7. The left plot shows Dunn test results as pairwise comparisons of Biologically Unique Landscapes (BUL) as cells colored by adjusted p -value. Darker cells are those closer to significance at $p < 0.10$. The right plot illustrates Dunn test significance between BULs as a binary of significance 1 and non-significance 0. See Table 3.2 for full BUL names.107

Figure 3.8. The left plot shows Dunn test results as pairwise comparisons of Biologically Unique Landscapes (BUL) as cells colored by Z-value. Darker cells are closer to the minimum $Z = -3.5$, while redder cells are closer to the maximum $Z = 3.5$. The right plot illustrates Dunn test significance between BULs as a binary of significance 1 and non-significance 0. See Table 3.2 for full BUL names.108

Figure 3.9. Silhouette value ranging from 1 to -1 on the y-axis by site on the x-axis for all classification systems. The plots show silhouette values by similarity based clusters (A), dominant landcover (B), ecoregion (C), stream order (D), HUC12 watersheds (E), and BULs (F). Positive silhouette values indicate better fit to assigned groups, while negative silhouette values indicate better fit to other groups. Bars per site are colored by group membership.....109

Figure 3.10. Bar charts showing the proportion of National Hydrography Dataset (NHD) flowline by stream order within each BUL.....110

CHAPTER 1

INTRODUCTION

Physical Template of Streams

Stream habitat characteristics are a result of a series of complex, hierarchical interactions between broad geographic, climatic, and geologic conditions and finer-scale physical and ecological processes (Hynes 1975; Frissell et al. 1986; Ward 1998; Davies et al. 2000; Brewer et al. 2018). The underlying hierarchical structure and dynamic nature of aquatic systems are key drivers of in-stream habitat conditions at multiple spatial scales (Frissell et al. 1986; Ward 1998). Geology directly influences and is reflected by drainage and channel patterns (Frissell et al. 1986; Ward 1998). Channel patterns are constrained where geologic features such as bedrock limit channel movement leading to more straight channels or alluvial where unconsolidated sediment is eroded and forms more sinuous channels (Ward 1998). Macrohabitat, mesohabitat, and microhabitat features are determined by processes driven by drainage and channel patterns (e.g., sediment transport and deposition; Ward 1998; Benda et al. 2004). Macrohabitat refers to large-scale (i.e., 1 km – 10 km) physical features, such as geological or topographical, encompassing both meso- and microhabitats. Mesohabitats often refer to mid-scale (i.e., 10 m – 1 km) hydrological features such as riffles, runs, and pools, and can be further categorized into microhabitat at small-scales (i.e., < 10 m; Vannote et al. 1980; Able et al. 1987; Ward 1998). Overall, broad-scale geomorphological and hydrological characteristics directly influence local-scale habitat features and biodiversity patterns (Frissell et al. 1986; Ward 1998).

Hydrological and geomorphological dynamics of stream systems may shape available habitat and drive community structure and diversity (Hynes 1975; Vannote et al. 1980; Poff 1997; Frothingham et al. 2002; Benda et al. 2004). For example, the River Continuum Concept (RCC) identifies longitudinal trends in stream habitat and biotic community variation, where habitat conditions and resulting biological communities follow a gradient in stream size and position in the drainage network (Vannote et al. 1980; Minshall et al. 1985; Frothingham et al. 2002; Benda et al. 2004). Although the linear perspective of the RCC has dominated fluvial ecology in the past, modern approaches favor a perspective of streams as a discontinuum (Townsend 1989; Poole 2002; Benda et al. 2004). Discontinuum perspectives emphasize the nonuniform and patchy nature of habitat heterogeneity along stream lengths (Townsend 1989; Benda et al. 2004; Thorp et al. 2006). Freshwater biodiversity is often attributed to habitat heterogeneity maintained through stochastic processes and disturbance regimes (e.g., periodic flood and drought) that are better represented in a discontinuum perspective (Ward 1998; Benda et al. 2004; Mayor et al. 2009). Investigations of habitat heterogeneity, hierarchical scaling, and stochastic disturbance have led to arguments for viewing streams under principles of landscape ecology (Wiens 1989; Fausch et al. 2002; Benda et al. 2004). Thus, emphasizing stochasticity and habitat heterogeneity as drivers of biodiversity under a riverscape approach could enhance understanding of ecological processes in freshwater systems at large spatial scales (Ward 1998; Fausch et al. 2002).

The Importance of Habitat Heterogeneity

Heterogeneity in physical habitat sets the spatial template for multiple ecological processes (Wiens 1989; Fausch et al. 2002; McManamay et al. 2018). Habitat

heterogeneity is often described as habitat patches, where discrete areas differ in structure, composition, or function of environmental features (Fausch et al. 2002; Wohl 2016). Heterogeneity may then be assessed by number, type, size, and connectivity of habitat patches (Fausch et al. 2002; Wohl 2016). Observed habitat characteristics are determined by grain (i.e., spatiotemporal resolution of data) and extent (i.e., the overall size of the focal area) as components of the focal scale (Wiens 1989). Assessing heterogeneity in this manner is scale dependent, as the degree of heterogeneity observed is determined by the spatiotemporal grains and extents exercised (Fausch et al. 2002; Wohl 2016).

Hydrologic and geomorphologic features are generally heterogenous in streams and may influence stream community health (Wohl 2016; Stein et al. 2018). Habitat heterogeneity promotes species diversity as increased habitat types, resources, and structural complexities broaden available niches and decrease competition (Wohl 2016; Stein et al. 2018). At broad spatiotemporal extents, habitat heterogeneity increases the probability of speciation due to adaptation to diverse environmental conditions (Stein et al. 2018). Within drainage networks, habitat heterogeneity has a compounding effect where complexity upstream positively influences that downstream, especially in solutes, sediment, and organic matter (Vannote et al. 1980; Wohl 2016; McManamay et al. 2018). Habitat heterogeneity reflects environmental processes and drives ecological processes; as such, it may be used to provide insight into these processes (Wohl 2016; McManamay et al. 2018). Furthermore, understanding patterns of habitat heterogeneity helps detect changes in environmental and ecological processes (Wohl 2016; McManamay et al. 2018). Scale dependence of habitat heterogeneity may lead to a negative relationship with

species diversity, however, as the area of suitable habitat patches for a species may decrease as habitat types increase in smaller areas (Stein et al. 2018). Further, disturbances and limited connectivity can disrupt the translation of complexity of stream hydrology and geomorphology into habitat heterogeneity (Wohl 2016; Stein et al. 2018).

Anthropogenic Disturbances

Anthropogenic modification of stream habitat through fragmentation can reduce stream connectivity and disrupt natural disturbance regimes, greatly reducing a stream's capacity to sustain biodiversity (Gido et al. 2019). Fragmentation via culverts and dams can decrease biodiversity through changes to spatial configuration of drainage networks that limit connectivity among stream reaches (Jackson and Fahrig 2015; Perkin et al. 2017). Biodiversity in fragmented stream reaches is reduced given limited recolonization following local extirpation (Fausch et al. 2002; Perkin and Gido 2011; Perkin et al. 2017). Impacts on biodiversity as a result of fragmentation have been observed in pelagic-spawning cyprinids of the Great Plains, USA (Perkin and Gido 2011; Perkin et al. 2017). Fragmentation results in an increased likelihood of range restriction and local extirpation by limiting species movement and ability to conduct life-stage activities at broad spatial extents (Staude et al. 2020). The importance of movement in maintaining population connectivity is apparent in recolonization of habitat patches where fish were extirpated by drying or flooding in streams (Fausch et al. 2002; Falke and Fausch 2010). Source-sink dynamics describe how variation in habitat availability or quality, as well as species movement between variably suitable habitat patches, may affect growth or decline of spatially isolated populations of organisms (Pulliam 1988; Ziv 1998). Source populations in the highest quality habitats keep sink populations in low quality habitats

extant through immigration (Pulliam 1988; Ziv 1998). Fragmentation limits immigration, which is integral to offsetting reduced breeding success resulting from low habitat quality, endangering sink populations (Pulliam 1988; Ziv 1998; Staude et al. 2020). Furthermore, reduced range (i.e., geographic distribution) could increase inter- and intraspecies competition for source populations (Ziv 1998; Staude et al. 2020). Individuals may recolonize reaches due to their ability to track shifting mosaics of suitable habitat (Alford and Walters 2021; Rohr et al. 2021). Though some fish are well adapted to natural disturbance regimes that reflect the dynamic nature of freshwater environments, species richness often decreases following anthropogenic disturbance (Rohr et al. 2021). However, community structure may be resilient and persistent, as some species quickly return to pre-disturbance abundances within their native reaches (Lytle and Poff 2004; Alford and Walters 2021; Rohr et al. 2021). Thus, repeated extirpation and recolonization processes maintaining species distributions at different spatial and temporal scales are disrupted by habitat fragmentation (Albanese et al. 2009; Falke and Fausch 2010; Stoll et al. 2014).

Homogenization of habitat caused by anthropogenic activities has resulted in declines of freshwater biodiversity at a greater rate than most other taxa (Cardinale et al. 2002; Saunders et al. 2002; Göthe et al. 2014; Sievert et al. 2016). Stream ecosystems have arguably undergone the most dramatic homogenization of any ecosystem type (Cardinale et al. 2002). Within the United States of America, 85% of large rivers have been straightened for navigation and flood control, altering flow regimes and reducing sediment transport (Strayer and Dudgeon 2010; Vörösmarty et al. 2010; Perkin and Gido 2011; Comte et al. 2021). Homogenized streams tend to exhibit disturbed hydrology,

manifesting as increased minimum flow velocity, decreased maximum flow velocity, and decreased hydrograph variability (McManamay et al. 2012; Paukert et al. 2020; Comte et al. 2021). Homogenization lowers biodiversity of impacted stream reaches as alterations simplify hydrology, geomorphology, and natural disturbance regimes of aquatic systems (Lytle and Poff 2004; Paukert et al. 2020; Comte et al. 2021). Such shifts in conditions of aquatic ecosystems reduce the diversity, abundance, and connectivity of available habitat for some species and reduces biodiversity by extension (Radinger et al. 2017; Paukert et al. 2020; Booher and Walters 2021). For example, human development of streambanks can remove large sections of riparian vegetation and render areas uninhabitable for species dependent on pre-disturbance habitat (Comte et al. 2021).

Stream systems remain under persistent threat stemming from shifts in climate and land-use patterns (Worthington et al. 2014). Climate change is a significant threat to habitat heterogeneity and stream-fish communities (Sievert et al. 2016). Changes in maximum water temperature of only a few degrees can result in major shifts in the occurrence, abundance, survival, growth, size distribution, and reproduction of many species (Davies et al. 2000; Lyons et al. 2010). Land-use change is another major and persistent driver of biodiversity loss in freshwater systems (Carpenter et al. 2011; Comte et al. 2021). Terrestrial landscape features that influence physiochemical processes (e.g., nutrient cycling), and biotic structure within streams are altered through land-use (Hynes 1975; Frissell et al. 1986; Vörösmarty et al. 2010; Brewer et al. 2018). Human appropriation of land for agriculture and infrastructure may drive habitat fragmentation and loss, homogenization, hydrological modification, shifting suspended nutrient loads, and alteration of riparian areas (Carpenter et al. 2011; Comte et al. 2021). Stream

sinuosity and middle stream orders are correlated with biodiversity, as both contribute to habitat heterogeneity (Ward 1998). As such, it is important to recognize the high degree of influence anthropogenic alterations have on the physical features of streams as biodiversity is threatened with changing climate conditions and habitat degradation (Carpenter et al. 2011; Comte et al. 2021).

Study Area

Streams of Nebraska, USA resemble those of the larger American Great Plains (henceforth Great Plains), a level 1 ecoregion covering approximately 1.3 million km² of the central USA (Dodds et al. 2004). Streams and rivers of the Great Plains are characterized by wide variations in stream temperature, flow regime, and water quality (Dodds et al. 2004). Flow sources vary in different regions with many low-order streams being primarily groundwater fed and others being precipitation (i.e., rainfall and snowfall) fed (Dodds et al. 2004; Ojima et al. 1999; Winter 2007). The largest source of groundwater in Nebraska is the underlying Ogallala Aquifer, also known as the High Plains Aquifer (Ojima et al. 1999; Winter 2007). Overuse of the Ogallala Aquifer (e.g., irrigation and drinking water) causes a retreat of the underground water table (Sohl et al. 2012). This has contributed to a reduction of wide floodplains and wetlands historically found in the Great Plains (Dodds et al. 2004; Sohl et al. 2012). The Ogallala Aquifer reaches saturated depths of up to 390 m, with the deepest saturation near its recharge zones in the Sandhills ecoregion (Sohl et al. 2012). Precipitation events are key to maintaining flow in Great Plains streams (Ojima et al. 1999; Dodds et al. 2004). Precipitation recharges groundwater deposits through infiltration, while runoff during precipitation events also supplies water to intermittent or ephemeral streams in more arid

regions (Ojima et al. 1999; Dodds et al. 2004; Winter 2007). Snowmelt plays a critical role in supplying water to Nebraska, as well, with the headwaters of the Platte River in Colorado, USA receiving up to 1000 mm of snowfall annually (Ojima 1999; Daniels 2007). Periods with a complete loss of surface water may be common in low-order Nebraska streams (Fausch and Bramblett 1991; Dodds et al. 2004). Evidence suggests that droughts may become more frequent due to climate change, while precipitation events are also becoming more extreme and unpredictable (Ojima et al. 1999; Christian et al. 2019). The increased frequency and severity of disturbance events (i.e., drought and precipitation) may shift in-stream habitat conditions, disconnect habitat patches, and restrict species movement, further threatening aquatic ecosystems (Matthews and Marsh-Matthews 2003).

Study Objectives

Using the state of Nebraska as a case study, the overall objectives of this thesis are 1) characterize habitat availability for several at-risk fish species, and 2) to assess the categorization of stream habitat heterogeneity within multiple classification systems. Objective 1 aims to use species distribution models to inform conservation of at-risk fish species in Nebraska by identifying ecologically relevant multiscale habitat features. Objective 2 aims to assess the classification of habitat features in conservation frameworks and introduce multivariate cluster analysis as a similarity-based classification system to improve stream classification in Nebraska. Both objectives serve to increase understanding of the structure and function of stream ecosystems and translate this understanding into effective conservation frameworks.

References

- Able, K. W., D. C. Twichell, C.B. Grimes, and R.S. Jones. 1987. Sidescan sonar as a tool for detection of demersal fish habitats. *Fishery Bulletin* 85(4):725-736.
- Albanese, B., P. L. Angermeier, and J. T. Peterson. 2009. Does mobility explain variation in colonization and population recovery among stream fishes? *Freshwater Biology* 54(7):1444-1460.
- Alford, S. L., and A. W. Walters. 2021. Rapid colonization post-displacement contributes to native fish resilience. *Ecology of Freshwater Fish* 00:1-11.
- Benda, L., N. L. Poff, D. Miller, T. Dunne, G. Reeves, G. Pess, and M. Pollock. 2004. The Network Dynamics Hypothesis: How channel networks structure riverine habitats. *BioScience* 54(5):413-427.
- Booher, E. C. J., and A. W. Walters. 2021. Biotic and abiotic determinants of finescale dace distribution at the southern edge of their range. *Diversity and Distributions* 27(4):696–709.
- Brewer, S. K., T. A. Worthington, R. Mollenhauer, D. R. Stewart, R. A. McManamay, L. Guertault, and D. Moore. 2018. Synthesizing models useful for ecohydrology and ecohydraulic approaches: An emphasis on integrating models to address complex research questions. *Ecohydrology* 11(7).
- Carpenter, S. R., E. H. Stanley, and M. J. V. Zanden. 2011. State of the World's freshwater ecosystems: physical, chemical, and biological changes. *Annu. Rev. Environ. Resour.* 36:75-99.
- Comte, L., J. D. Olden, P. A. Tedesco, A. Ruhi, and X. Giam. 2021. Climate and land-use changes interact to drive long-term reorganization of riverine fish communities globally. *Proceedings of the National Academy of Sciences* 118:1-9.
- Davies, N. M., R. H. Norris, and M. C. Thoms. 2000. Prediction and assessment of local stream habitat features using large-scale catchment characteristics: Prediction and assessment of local stream habitat. *Freshwater Biology* 45(3):343–369.
- Dodds, W.K., K. Gido, M.R. Whiles, K.M. Fritz, and W.J. Matthews. 2004. Life on the edge: the ecology of Great Plains prairie streams. *Bioscience* 54(3):205-216.
- Falke, J. A., and K. D. Fausch. 2010. From metapopulations to metacommunities: Linking theory with empirical observations of the spatial population dynamics of stream fishes. *American Fisheries Society Symposium* 73:207–233.
- Fausch, K. D., C. E. Torgersen, C. V. Baxter, and H. W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* 52(6):483.

- Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management* 10(2):199–214.
- Frothingham, K. M., B. L. Rhoads, and E. E. Herricks. 2002. A multiscale conceptual framework for integrated ecogeomorphological research to support stream naturalization in the agricultural Midwest. *Environmental Management* 29(1):16–33.
- Gido, K. B., D. L. Propst, J. E. Whitney, S. C. Hedden, T. F. Turner, and T. J. Pilger. 2019. Pockets of resistance: Response of arid-land fish communities to climate, hydrology, and wildfire. *Freshwater Biology* 64(4):761–777.
- Göthe, E., N. Friberg, M. Kahlert, J. Temnerud, L. Sandin. 2014. Headwater biodiversity among different levels of stream habitat hierarchy. *Biodiversity Conservation* 23:63–80.
- Hynes, H.B. 1975. The stream and its valley. *Verh. Internat. Verein. Limnol.* 19: 1–15.
- Jackson, H. B., and L. Fahrig. 2015. Are ecologists conducting research at the optimal scale?: Is research conducted at optimal scales? *Global Ecology and Biogeography* 24(1):52–63.
- Lyons, J., J. S. Stewart, and M. Mitro. 2010. Predicted effects of climate warming on the distribution of 50 stream fishes in Wisconsin, U.S.A. *Journal of Fish Biology* 77(8):1867–1898.
- Lytle, D. A., and N. L. Poff. 2004. Adaptation to natural flow regimes. *Trends in Ecology & Evolution* 19(2):94–100.
- Matthews, W.J. and E. Marsh-Matthews. 2003. Effects of drought on fish across axes of space, time, and ecological complexity. *Freshwater Biology* 48:1232–1253.
- Mayor, S. J., D. C. Schneider, J. A. Schaefer, and S. P. Mahoney. 2009. Habitat selection at multiple scales. *Écoscience* 16(2):238–247.
- Mažeika, S., P. Sullivan, and M.C. Watzin. 2004. Understanding stream geomorphic state in relation to ecological integrity: Evidence using habitat assessments and macroinvertebrates. *Ecological Management* 34(5):669–683.
- McManamay, R.A., D.J. Orth, and C.A. Dolloff. 2012. Revisiting homogenization of dammed rivers in the Southeastern US. *Journal of Hydrology* 424–425:217–237.
- McManamay, R.A., M.J. Troia, C.R. DeRolph, A.O. Sheldon, A.R. Barnett, S.C. Kao, and M.G. Anderson. 2018. A stream classification system to explore the physical habitat diversity and anthropogenic impacts in riverscapes of the eastern United States. *PLoS ONE* 13(6):e0198439.

- Minshall, G.W., R.C Petersen, K.W. Cummins, T.L. Bott, J.R. Sedell, C.E. Cushing, and R.L. Vannote. 1983. Interbiome comparison of stream ecosystem dynamics. *Ecological Monographs* 53(1):1-25.
- Paukert, C. P., E. R. Kleekamp, and R. W. Tingley. 2020. Identifying candidate reference reaches to assess the physical and biological integrity of Wadeable streams in different ecoregions and among stream sizes. *Ecological Indicators* 111:105966.
- Perkin, J. S., and K. B. Gido. 2011. Stream fragmentation thresholds for a reproductive guild of Great Plains fishes. *Fisheries* 36(8):371–383.
- Perkin, J. S., K. B. Gido, J. A. Falke, K. D. Fausch, H. Crockett, E. R. Johnson, and J. Sanderson. 2017. Groundwater declines are linked to changes in Great Plains stream fish assemblages. *Proceedings of the National Academy of Sciences* 114(28):7373–7378.
- Poff, N.L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society* 16(2):301-409.
- Poole, G.C. 2002. Fluvial landscape ecology: addressing uniqueness within the river discontinuum. *Freshwater Biology* 47:641-660.
- Pulliam, H. R. 1988. Sources, Sinks, and Population Regulation. *The American Naturalist* 132(5):652–661.
- Radinger, J., F. Essl, F. Hölker, P. Horký, O. Slavík, and C. Wolter. 2017. The future distribution of river fish: The complex interplay of climate and land use changes, species dispersal and movement barriers. *Global Change Biology* 23(11):4970–4986.
- Rohr, J.M., S.J. Meiners, T.D. Thomas, and R.E. Colombo. 2021. Recovery of riverine fish assemblages after anthropogenic disturbances. *Ecosphere* 12(4):e03459.
- Saunders, D. L., J. J. Meeuwig, and A. C. J. Vincent. 2002. Freshwater protected areas: strategies for conservation. *Conservation Biology* 16(1):30–41.
- Schneider, R., K. Stoner, G. Steinauer, M. Panella, and M. Humpert (Eds.). 2011. *The Nebraska Natural Legacy Project: State Wildlife Action Plan*. 2nd ed. The Nebraska Game and Parks Commission, Lincoln, NE.
- Sievert, N. A., C. P. Paukert, Y.-P. Tsang, and D. Infante. 2016. Development and assessment of indices to determine stream fish vulnerability to climate change and habitat alteration. *Ecological Indicators* 67:403–416.
- Sohl T. L., B. M. Sleeter, K. L. Sayler, M. A. Bouchard, R. R. Reke, S.L. Bennett, R. R. Sleeter, R. L. Kanengieter, Z. Zhu. 2012. Spatially explicit land-use and land-cover scenarios for the Great Plains of the United States. *Agriculture, Ecosystems, and Environment* 153:1-15.

- Staude, I. R., L. M. Navarro, and H. M. Pereira. 2020. Range size predicts the risk of local extinction from habitat loss. *Global Ecology and Biogeography* 29(1):16–25.
- Stein, A. K. Gerstner, and H. Kreft. 2018. Environmental heterogeneity as a universal driver of species richness across taxa, biomes, and spatial scales. *Ecology Letters* 17:866-880.
- Stoll, S., J. Kail, A. W. Lorenz, A. Sundermann, and P. Haase. 2014. The importance of the regional species pool, ecological species traits and local habitat conditions for the colonization of restored river reaches by fish. *PLoS ONE* 9(1):e84741.
- Strayer, D. L., and D. Dudgeon. 2010. Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society* 29(1):344–358.
- Thorp, J.H., M.C. Thoms, and M.D. DeLong. 2006. The riverine ecosystem synthesis: biocomplexity in river networks across space and time. *River Research and Applications* 22:123-147.
- Townsend, C.R. 1989. The patch dynamics concept of stream community ecology. *Journal of North American Benthological Society* 8(1):36-50.
- Vannote, R. L., G.W. Minshall, K.W. Cummins, J. R. Sedell, and C.E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37:130-137.
- Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global threats to human water security and river biodiversity. *Nature* 467(7315):555–561.
- Ward, J. V. 1998. Riverine landscapes: biodiversity patterns, disturbance regimes, and aquatic conservation. *Biological Conservation* 83(3):269–278.
- Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* 3(4):385.
- Winter, T.C. 2007. The role of ground water in generating streamflow in headwater areas and in maintaining base flow. *Journal of the American Water Resources Association* 43(1):15-25.
- Wohl, E. 2016. Spatial heterogeneity as a component of river geomorphic complexity. *Progress in Physical Geography* 40(4):598-615.
- Worthington, T. A., S. K. Brewer, T. B. Grabowski, and J. Mueller. 2014. Backcasting the decline of a vulnerable Great Plains reproductive ecotype: identifying threats and conservation priorities. *Global Change Biology* 20(1):89–102.

CHAPTER 2

HABITAT RELATIONS OF FLATHEAD CHUB, PLAINS MINNOW, AND
NORTHERN REDBELLY DACE**Abstract**

Multiscale environmental processes determine in-stream habitat conditions selected by fish at local spatial scales. Habitat includes the physical template upon which ecological processes occur and species conduct life-stage activities. Spatiotemporal availability of habitat is a fundamental driver of species distribution whereby presence of a species is determined by the structure of habitat patches. Reduction in habitat availability drives species declines. Understanding habitat features driving species presence in a stream informs conservation and management decision making. Species distribution models provide a tool to characterize spatial habitat patterns and quantify species-habitat relations. Boosted regression tree models of species distributions were constructed from environmental data and historical presence data for three at-risk cyprinid species including Flathead Chub (*Platygobio gracilis*), Plains Minnow (*Hybognathus placitus*), and Northern Redbelly Dace (*Chrosomus eos*). For all three species, hydrological factor (i.e., flow velocity and discharge) and stream size and position in the drainage network (i.e., stream order and upstream drainage area) were among the most influential habitat conditions of examined variables. Further, our results may indicate ecological uniqueness among selected species. For example, network connectivity showed high relative influence for Flathead Chub presence but not for that of Plains Minnow or Northern Redbelly Dace. Understanding multiscale environmental influences on species presence may target research of fine-scale species-habitat relations

and direct conservation and management efforts for at-risk fish species. Our results support current knowledge of impactful habitat features based on species traits and environmental processes shaping site-scale habitat conditions.

Introduction

Understanding the abiotic environmental context in which habitat conditions exist is of importance to quantifying species-habitat relations (Poizat and Pont 1996; Aarts et al. 2013). In ecology, habitat may be defined as the range of environmental conditions that produce occupancy, including those necessary for an organism to conduct life-history stages (Krausman and Morrison 2016). Spatiotemporal habitat heterogeneity in fluvial systems is largely driven by hydrological and geological processes (Frissell et al. 1986; Poizat and Pont 1996; Benda et al. 2004). For example, the geology of a catchment determines stream sinuosity (Frissell et al 1986; Ward 1998). Where a catchment with more consolidated sediments will exhibit straighter stream channels, a catchment with unconsolidated sediments will contain more sinuous streams as sediment is more easily transported by flowing water (Ward 1998). Stream sinuosity may drive riffle and pooling characteristics at smaller scales, while riffle and pooling characteristics may determine the deposition of debris serving as shelter at even smaller scales (Vannote et al. 1980; Able et al. 1987; Ward 1998). As such, the hierarchical structure and dynamic nature of aquatic systems are key drivers of local (i.e., site-scale) in-stream habitat conditions (Frissell et al. 1986; Ward 1998; Benda et al. 2004). Representing the patterns of stream habitat heterogeneity at site-scales has proved to be a fundamental challenge to aquatic conservation (Fausch et al. 2002; Falke and Fausch 2010). A site is defined as a sampling unit of stream reaches with a minimum extent of 150 meters and a maximum extent of

300 meters, often including a grain of $1/10^{\text{th}}$ the sampled extent (USEPA 2019; NDEE 2020). Due to the highly dynamic and heterogenous nature of stream systems, reporting of ever-changing environmental conditions at a site-scale (e.g., at a temporal grain of one day over an extent of one year for multiple sites) can become resource and time intensive (Fausch et al. 2002; Brewer et al. 2018). This leads to potential scale mismatches in the application of environmental data to conservation and management decisions (Poizat and Pont 1996; Helmsetter et al. 2021). Nonetheless, such efforts are integral to representing the full hierarchy of environmental processes driving local conditions from which species select habitat.

Species presence in a given stream reach is determined via a hierarchical habitat filtering and selection processes whereby large-scale processes such as geomorphology or elevation provide an initial filter with respect to physiological constraints of species (Poff 1997; Mayor et al. 2009; Keller 2023). For example, a coolwater fish species will not exist in warmwater streams because large-scale climatic filters create ecosystems non-conducive to the species physiology (Mayor et al. 2009; Keller 2023). Thus, the environmental processes acting as filters to species presence do so in driving site-scale habitat conditions (Poff 1997; Poizat and Pont 1996; Benda et al. 2004; Aarts et al. 2013). Individuals select site-scale habitat conditions to conduct life-stage activities (e.g., feeding, reproduction, refuge seeking) within the larger-scale environmental context (Mayor et al. 2009; Keller 2023). Major drivers of habitat selection by individuals include thermal variability, hydrologic regime, spatial configuration of streams, water quality, and associated biota (Durance et al. 2006; Keller 2023). Stream fish species also require multiple habitat types throughout their life stages (spawning, rearing, and refuge)

as larval and juvenile individuals may require different habitat conditions compared to adults of the same species (Poff 1997; Falke and Fausch 2010). As such, species-habitat relations may be based on hierarchical processes shaping available habitats at different times (e.g., life stages) as well as the complimentary structure and spatial configurations of those habitats.

The interplay between habitat availability and species dispersal ability governs species distribution (Aarts et al. 2013). Habitat for a species may exist at spatial scales such that individuals must move among diverse habitat patches to conduct life stages (Albanese et al. 2009; Mayor et al. 2009; Falke and Fausch 2010; Aarts et al. 2013). The metapopulation concept describes this paradigm wherein habitat patches support populations within a matrix of unsuitable habitat (Falke and Fausch 2010). Movement of individuals among patches is necessary to maintain population size and health as well as support recolonization of patches by individuals following disturbance-driven extirpation (Albanese et al. 2009; Falke and Fausch 2010). Further, unsuitable habitat may need to be traversed as individuals move between patches of suitable habitat, especially in fluvial systems where habitat availability may fluctuate seasonally with hydrological regimes (Albanese et al. 2009; Falke and Fausch 2010). As such, habitat availability for a species is also dependent on temporal contexts, wherein species presence necessitates individual ability to track and move among shifting habitat patches in time (Falke and Fausch 2010; Aarts et al. 2013). Thus, species presence in a given stream reach may imply habitat suitability for a given life stage or dispersal pathways with connection to suitable habitat in spatiotemporal contexts (Albanese et al. 2009; Mayor et al. 2009; Aarts et al. 2013). Where species presence denotes occurrence at a site-scale, species distribution represents

the entire geographical range of a species across multiple sites (Fausch et al. 2002; Sowa et al. 2008). As species distribution is also dictated by habitat availability and species dispersal at large scales, species distribution may then provide insight into habitat distribution and availability (Sowa et al. 2008; Falke and Fausch 2010; Aarts et al. 2013).

Species distribution models (SDMs) are powerful tools with diverse applications in ecology (Elith et al. 2008; Helmsetter et al. 2021). SDMs have been used to explore climate impacts on species, establish conservation areas, guide restoration, aid in invasive species management, and as components in risk analyses (Guisan and Zimmermann 2000; Kery et al. 2010; Jarnevich et al. 2015). In most instances, SDMs may be viewed as exercises in modeling spatial patterns in habitat conditions and their influence on species presence or abundance (Sowa et al. 2008). Approaching SDMs in such a manner identifies environmental processes influencing site-scale habitat conditions and emphasizes representing these processes in environmental data (Fausch et al. 2002; Sowa et al. 2008). Representing multiscale environmental processes increases certainty in ecological interpretation of species-habitat relations by addressing the different scales at which species select habitat (Wiens 1989; Poizat and Pont 1996; Mayor et al. 2009; Helmsetter et al. 2021; Keller 2023). SDMs use detection (i.e., presence or absence observations) and environmental data to quantify species-habitat relations and predict the occurrence of species at multiple spatiotemporal scales (Davies et al. 2000; Guisan and Zimmermann 2000; Fausch et al. 2002; Helmsetter et al. 2021). These predictions can then be used to map environmental suitability and target conservation and management of target species (Guisan and Zimmermann 2000; Fausch et al. 2002; Helmsetter et al. 2021). SDMs are also useful for the ability to extrapolate environmental scenarios

spatiotemporally, wherein models may be used to predict habitat suitability at greater extents (Guisan and Zimmermann 2000; Kery et al. 2010; Helmsetter et al. 2021). Thus, SDMs identify habitat variables influencing species distributions and use mapping of influential habitat features to facilitate focused conservation of in-stream habitat and resulting fish populations.

As in-stream habitat conditions continue to change due to natural and anthropogenic disturbances, assessing the distribution patterns of at-risk fish species becomes integral to their conservation (Lytle and Poff 2004; Worthington et al. 2014; Paukert et al. 2020; Martin 2022). Habitat loss is the primary driver of species imperilments and reductions in distributions as habitat patches become reduced in abundance and connectivity (Lytle and Poff 2004; Jelks et al. 2008; Mayor et al. 2009). Shifts in habitat availability and resulting fish assemblage structures have been attributed to altered hydrology, habitat homogenization, and nonnative species introduction (Rahel and Thel 2004; Jelks et al. 2008; Perkin and Guido 2012; Senecal et al. 2015; Perkin et al. 2017). The high degree of shifting habitat in North America has resulted in approximately 39% of all native fishes having received some form of conservation status (Jelks et al. 2008; Senecal et al. 2015). In the USA, at-risk species are categorized by the degrees of their imperilments and risk of extinctions as endangered, threatened, or vulnerable (Jelks et al. 2008). Reduction in the distribution of a species, often viewed over 50-year periods, is key to at-risk listing (Jelks et al. 2008; Senecal et al. 2015).

Study Species

The Nebraska Natural Legacy Project (NNLP) developed a two-tiered approach to listing at-risk species (Schneider et al. 2011). Tier-1 fish species are defined by the NNLP

as species found in Nebraska that are globally or nationally at-risk of extinction, whereas tier-2 fish are those that are at-risk in Nebraska but have healthy populations elsewhere in their global or national distributions (Schneider et al. 2011; Schneider et al. 2018). Tier-1 listing is attributed to state or federal listing as threatened or endangered, declining or disjunct populations, and endemism to Nebraska (Schneider et al. 2011; Schneider et al. 2018). From the tier-1 list, three cyprinids including Flathead Chub (*Platygobio gracilis*), Plains Minnow (*Hybognathus placitus*), and Northern Redbelly Dace (*Chrosomus eos*), were chosen as case studies for these SDMs because of differing habitat preferences across their life-stages. These species select for different habitat conditions that result in unique distribution patterns that may reflect a range of habitat types and indicate differing conservation needs (Table 1; Midway and Peoples 2019). Though habitat associations for these species are theoretically known, these associations have yet to be or have not in a contemporary context been quantified through methods such as SDMs (Rahel and Thel 2004; Rees et al. 2005; Stringer and Clancy 2020). Further, gaps in knowledge of these species' life-histories exist that may be filled with targeted research informed by habitat relations (Rahel and Thel 2004; Frimpong and Angermeier 2009; Stringer and Clancy 2020). Understanding the relations between these species and their habitats is imperative to prevent future listings and help population recovery (Schneider et al. 2011; Schneider et al. 2018).

Flathead Chub occupy turbid rivers and larger tributaries from Canada to Texas (Hesse 1994; Rahel and Thel 2004; Frimpong and Angermeier 2009). Flathead Chub are most commonly associated with small substrate like sand and gravel (Hesse 1994; Rahel and Thel 2004). Though Flathead Chub are found in all habitat types including main

channels, side channels, and backwaters, adults show preference for faster flowing main channels (Rahel and Thel 2004). Flathead Chub are thought to move upstream for spawning from mid-July to mid-September (Hesse 1994; Frimpong and Angermeier 2009; Walters et al. 2014). Females age two and older lay an average of 4,974 eggs in a spawning season, though there is some debate on this value based on the maturity of the eggs (Rahel and Thel 2004). Though found across Nebraska in the past, Flathead Chub are now at-risk due to reduction in spawning habitat availability including connected stretches of river to support drifting eggs (Worthington et al. 2014; Perkin et al. 2017). The estimated minimum threshold of stream fragment length for Flathead Chub is 183 km (Perkin and Gido 2011; Walters et al. 2014). Anthropogenic modification in habitat fragmentation by culverts, dams, and roads has reduced the availability of long reaches of uninterrupted fluvial ecosystems, which has greatly restricted Flathead Chub range (Worthington et al. 2014; Perkin and Gido 2011; Walters et al. 2014; Perkin et al. 2017).

Plains Minnow is found throughout the Great Plains from southern Saskatchewan, Canada to Texas, USA (Osterhaus and Martin 2019). Plains Minnow is associated with perennial streams with dense vegetation, sand substrate, and show preference for water temperatures around 17°C (Bryan et al. 1984; Rees et al. 2005; Osterhaus and Martin 2019). Within their resident streams, Plains Minnow select for slower flow (0.05 m/s) areas such as pools and backwaters (Rees et al. 2005; Osterhaus and Martin 2019). Plains Minnow spawn at the downstream edges of pools between late April and early August (Rees et al. 2005; Osterhaus and Martin 2019). Females release 417-4,134 (mean = 817) semi-buoyant eggs per spawning season that are communally scattered (Taylor and Miller 1990; Rees et al. 2005; Frimpong and Angermeier 2009). Spawning events appear

to be intermittent and revolve around periods of high flow (Taylor and Miller 1990; Rees et al. 2005; Frimpong and Angermeier 2009). Females mature rapidly, reaching peak sexual maturity by June of their first year (Lehtinen and Layzer 1988). As pelagic spawners, Plains Minnow have experienced declines in geographic range reflective of decline in available habitat (Hesse 1994; Osterhaus and Martin 2019). Habitat fragmentation also impacts Plains Minnow populations, with the estimated species' minimum threshold of fragment length being 115 km (Perkin and Gido 2011). Further, tributary desiccation and mainstream alterations have extirpated and isolated populations (Rees et al. 2005). In Kansas, declines in the distribution and abundance of Plains Minnow have been attributed to anthropogenic stressors and the species has been listed as endangered (Osterhaus and Martin 2019).

Northern Redbelly Dace are found in southern Canada and the northern USA, with southern populations in Colorado, South Dakota, and Nebraska (Stringer and Clancy 2020). The Nebraska population of Northern Redbelly Dace is restricted to headwater streams in the Sandhills Ecoregion and disjunct from core distributions (Schnieder et al. 2011). Northern Redbelly Dace select for slow-flowing, clear, cool-water streams with fine substrate and an abundance of filamentous algae in which they lay eggs (Stringer and Clancy 2020). Unfortunately, there are many gaps in knowledge of Northern Redbelly Dace life-history, though it is known that sexual maturity is typically reached by first summer and third summer at the latest (Frimpong and Angermeier 2009; Stringer and Clancy 2020). Phytophilic spawning (i.e., egg depositing in vegetation) occurs from May to August, in which females lay batches of up to 30 non-adhesive eggs in algal mats that hatch in 8-10 days after fertilization by multiple males (Frimpong and Angermeier 2009;

Stringer and Clancy 2020). Habitat fragmentation, climate change, and increasing water temperatures caused by anthropogenic activities may negatively influence Northern Redbelly Dace distribution (Whittier et al. 1997; Schneider et al. 2011). Studies in Montana have shown little co-occurrence of Northern Redbelly Dace with larger piscivores, such as Northern Pike (*Esox lucius*), indicating that the species may be particularly vulnerable to predation by invasive species (Stringer and Clancy 2020).

The primary goal of this study was to assess distributional patterns among the ecologically unique Flathead Chub, Plains Minnow, and Northern Redbelly Dace and quantify their species-habitat relations. Study objectives included 1) determine which environmental features most influence the presence of these species in river and stream reaches across Nebraska, and 2) determine where these features exist in Nebraska. Based on differential species traits, I predict that 1) Flathead Chub presence will be most related with high flow velocity and discharge, greater connectivity, and larger upstream drainage area apparent in higher order rivers and streams; 2) Plains Minnow presence will be most related with moderate flow velocity and discharge and greater connectivity found in higher order rivers and tributaries; and 3) Northern Redbelly Dace presence will be highest with features conducive to headwaters, such as lower flow velocity and discharge, higher groundwater input, and lower upstream drainage area consistent with headwater streams (Table 1). In a broader sense, I predict that variables representing hydrological processes and anthropogenic disturbances will most influence the presence of all three species. Both environmental processes and anthropogenic disturbances largely drive site-scale habitat conditions in the modern environmental context, which will be reflected in their influence on the distribution of all three species. A greater understanding of species-

habitat relations of these at-risk species will inform their conservation in establishing ecologically relevant habitat conditions and direct conservation efforts in the spatial locations of these conditions.

Methods

Environmental data

Multi-scale environmental data for Nebraska streams were acquired from the National Hydrography Dataset (NHD) Plus Version 2 (McKay et al. 2012; U.S. Geological Survey 2022) and the StreamCat Dataset (Hill et al. 2015). The NHD Plus Version 2 is a publicly available geospatial framework consisting of modeled flowlines generated from the original NHD, National Elevation Dataset (NED), and Watershed Boundary Dataset (WBD; McKay et al. 2012; U.S. Geological Survey 2022). The StreamCat dataset supplements the NHD Plus Version 2 with over 600 metrics compiled or derived from environmental data (Carlisle et al. 2009; Falcone et al. 2010; Wang et al. 2011; Hill et al. 2015). Soil data were acquired from the Natural Resource Conservation Service State Soil Geographic (STATSGO2) database (Schwarz and Alexander 1995). Groundwater depth for wells across Nebraska was acquired from the University of Nebraska-Lincoln School of Natural Resource's Conservation and Survey Division (CSD) interactive data map (CSD 2023). Groundwater depth for statewide wells was averaged across years (variable temporal range by well) to generate a mean annual depth to groundwater (MADTW) predictor variable.

Three additional metrics were added to represent anthropogenic influences in the form of land-use and fragmentation. Landcover data from 2019 were acquired from the National Land Cover Database (NLCD; Dewitz et al. 2021). Two land-use metrics

including dominant (i.e., greatest area) landcover class and Shannon's diversity index were generated within 12-digit hydrological unit code (HUC12) subwatersheds. Pixel counts of landcover classes within subwatersheds were generated using the summarize raster tool in ArcGIS Pro (ESRI Inc. 2022). The class of highest proportional cover (i.e., most pixels) was noted as a categorical variable of matrix (i.e., dominant) landcover class. Shannon's Diversity Index was calculated for landcover classes within subwatersheds using vegan package in program R (R Core Team 2021; Okansen et al. 2022). Shannon's diversity index (H) represents the diversity of landcover as $H = -\sum_{i=1}^s p_i \ln(p_i)$, where s is the number of landcover classes and p_i is the proportion of pixels belonging to i landcover class (Spellerberg and Fedor 2003). Further, Shannon's diversity index better represents rare habitat types than the alternative Simpson's diversity index (Spellerberg and Fedor 2003). A discontinuity index (DCI) was derived from the NHD flowlines and a dam layer including all recorded dams ≥ 1.8 m tall via the Nebraska Department of Natural Resources (McKay et al. 2012; USDA 2015). Flowlines and dams were organized by subwatershed using the spatial join tool in ArcGIS Pro (ESRI Inc. 2022). A DCI was then calculated for each HUC12 subwatershed by:

$$DCI = \frac{n}{L_t},$$

where n is the number of dams and L_t is the total length (km) of stream flowlines (an adaptation of methods by Prato et al. 2011). Overall, these environmental data were chosen to represent processes driving site-scale habitat conditions from which species select habitat, such as hydrology, drainage network connectivity, position in the drainage network, and anthropogenic alteration.

Statistical Analysis

Correlation analysis was conducted on the environmental data and calculated metrics to identify collinearity among predictor variables and inform variable reduction (Figure 1). For pairs of variables with Pearson correlation of $r \geq 0.70$, the variable with lower correlations to other variables in the dataset was retained. Furthermore, environmental variables that may not have been highly correlated but provided similar information were retained based on prevalence or descriptiveness. For example, despite a Pearson correlation of only $r = 0.56$ between incision height and incision width-to-depth ratio, only incision width-to-depth ratio was retained as it is more descriptive (Tang et al. 2014; Borboudakis and Tsamardinos 2019). The selected predictor variables represent environmental conditions that may impact species distributions at relatively large spatial scales, such as hydrologic regimes (flow velocity and discharge) and water source (baseflow, runoff; Table 2.2; Durance et al. 2006; Keller 2023). A rule of thumb for variable selection in species distribution modeling is to maintain a sample size 10 times the number of predictors (Breiner et al. 2015). In this case, with 15 selected predictors a sample size of at least 150 training data points was maintained.

Presence:Absence Data

Historical species presence data for Flathead Chub, Plains Minnow, and Northern Redbelly Dace were contributed by the Nebraska Game and Parks Commission (NGPC), the Nebraska Natural Heritage Program (NNHP), the Nebraska Department of Environment and Energy (NDEE), and acquired from the online Biodiversity Information Serving Our Nation (BISON) database (BISON 2022). Only presence points postdating 1969 were maintained as most large-scale river and stream alterations (i.e., dam,

reservoir, and levee construction and channelization) occurred before this time, especially in the Missouri, Platte, and Republican rivers and their larger tributaries (Schneiders 1996; Murphy et al. 2004). Alteration continues today, but the scale of alteration has been reduced and implemented water-management strategies have grown more sophisticated (Schneiders 1996). Where necessary, presence points were remapped to overlap with the NHD flowlines using the Near tool in ArcGIS Pro (ESRI Inc. 2022). As the provided data was presence-only, the spsurvey package was used to generate spatially balanced (i.e., evenly dispersed in space) pseudoabsences along the NHD flowlines in Program R (R Core Team 2021; Dumelle et al. 2023). Spatial balance is desirable for SDM practices in minimizing biases and yielding more precise predictor-response relationships (Cerasoli et al. 2017; Dumelle et al. 2023). The SDMs constructed with spatially balanced pseudoabsences explain more deviance in predictor-response relationships than randomly generated pseudoabsences using presence only datasets (Barbet-Massin et al. 2012; Cerasoli et al. 2017). Pseudoabsences were generated by a 1:10 presence to pseudoabsence ratio as this ratio yields adequate predictive accuracy for boosted regression tree models (Barbet-Massin et al. 2012).

Model Construction and Validation

Boosted Regression Tree (BRT) models of species distributions were constructed using the dismo package in Program R (R Core Team 2021; Hijmans et al. 2022). Using both statistical and machine learning approaches, BRTs establish patterns between predictors and a response variable (Elith et al. 2008). Preparation and selection of input variables are simplified under a BRT framework because predictor variables can be of any type (i.e., numeric, categorical, binary, etc.; De'ath 2007; Elith et al. 2008). Input

data consisting of a species presence or pseudoabsence response variable and 16 spatially related predictor variables were used to train the model for each species. BRTs recursively construct decision trees on random subsets of the input data (De'ath 2007; Elith et al. 2008). Decision trees partition the predictor space using split points, or variable values at which the response changes, to classify ranges of predictor values with homogenous responses (Elith et al. 2008). Over the boosting process, different combinations of variables and split points are tested to produce the decision tree with the lowest error (De'ath 2007; Elith et al. 2008). The first decision tree is built to minimize error, then subsequent decision trees are iteratively built from the error of the previous tree (i.e., second tree uses residuals from the first, third tree uses residuals from the second, etc.) in a manner emphasizing poorly modeled observations (De'ath 2007; Elith et al. 2008). Each subsequent tree compensates for and corrects error in the previous trees by focusing on poorly characterized interactions between ranges of predictor variable values and the response (De'ath 2007; Elith et al. 2008). The final tree is considered a linear combination of all previous trees in which the overall error is maximally reduced and the relationship between predictor and response variables is best characterized (De'ath 2007; Elith et al. 2008). Tree complexity, or the number of split values in each tree, was set at five for the Flathead Chub and Plains Minnow models and at three for the Northern Redbelly Dace model. The selected tree complexity optimizes model performance for the sample size of input data, 5159 points for Flathead Chub, 3234 points for Plains Minnow, and 814 points for Northern Redbelly Dace including both presences and pseudoabsences (Elith et al. 2008; Elith and Leathwick 2017). Furthermore, a learning rate of 0.01 was used as standard with boosted regression tree

models constructed from datasets of comparable type and size to our own (Elith et al. 2008; Elith and Leathwick 2017).

Models were internally validated using 10-fold cross validation wherein input data were broken into ten equally sized data subsets (Hastie et al. 2001; Allouche et al. 2006; Elith et al. 2008; Hao et al. 2020). Nine of these subsets were training datasets used in model constructions and the tenth subset was a testing dataset used to test the sensitivity and specificity of the models (Hastie et al. 2001; Allouche et al. 2006; Hao et al. 2020). Sensitivity quantifies omission error, or the ability of the models to accurately predict presences, whereas specificity quantifies commission error, or the ability of the models to accurately predict absences (Hastie et al. 2001; Allouche et al. 2006; Cerasoli et al. 2017). Both sensitivity and specificity measure accuracy on a 0 to 1 scale (Hastie et al. 2001; Allouche et al. 2006; Cerasoli et al. 2017). A receiver-operating characteristic (ROC) curve plotting sensitivity against false positives (i.e., $1 - \text{specificity}$) then assesses model fit, wherein the area under curve (AUC) ranging from 0 to 1 denotes better model fit at higher values (Hastie et al. 2001; Allouche et al. 2006; Cerasoli et al. 2017; Artaev 2023). However, AUC is threshold-independent and for models with presence:absence predictions accuracies should be complemented with threshold-dependent metrics (Allouche et al. 2006; Artaev 2023). A true skill statistic (TSS) is a threshold-dependent measure of model accuracy using $\text{TSS} = \text{Sensitivity} + \text{Specificity} - 1$ (Allouche et al. 2006; Artaev 2023). TSS ranges from -1 to +1, where more positive values reflect more accuracy and values 0 or below indicate model performance is no better than random chance (Allouche et al. 2006; Artaev 2023). In general, models with TSS above 0.7 are considered good quality (Artaev 2023).

Results

The Flathead Chub model was trained on 5159 data points comprised of 469 presences and 4690 absences. The Flathead Chub model had flow velocity, discharge, upstream drainage area, and organic substrate as the most influential predictor variables at 72.65% cumulative influence (Table 3). Flow velocity (30.50%) and discharge (29.94%) have the greatest relative influences on Flathead Chub presence, while upstream drainage area (6.87%) and percent organic substrate (5.34%) carry strong influence as well (Table 3). Discharge had a spike in effect on presence at approximately 200 m³/s, then gradually decreases in effect until increasing above 1000 m³/s (Figure 2). Flow velocity shows spikes in positive effect around 0.50 m/s and 0.75 m/s, and generally higher effect above 1.00 m/s (Figure 2). Total upstream drainage area shows a positive effect spike around 200,000 km² and plateaus after a small drop in effect (Figure 2). Percent organic substrate has peak effect at approximately 1.50%, dropping to a plateau in effect just after (Figure 2). The Flathead Chub model performed well under 10-fold cross validation, exhibiting high accuracy with sensitivity = 0.86 and specificity = 0.89 (Table 6). The model also has strong fit and accuracy with AUC = 0.94 and TSS = 0.75 (Figure 3; Table 6).

The Plains Minnow model was trained on 3234 data points comprised of 294 presences and 2940 absences. The Plains Minnow model has discharge (30.11%), upstream drainage area (20.86%), stream order (10.49%), flow velocity (9.09%), landcover diversity (6.72%), and slope (6.36%) as the most influential predictor variables at 83.63% cumulative influence (Table 4). Discharge has a positive effect on presence across its range of values, with the consistently strongest effect above 1000 m³/s (Figure

5). Total upstream drainage area starts with strongly negative effect at low values, increasing to positive effect past 200,000 km² (Figure 5). Effect on presence increases with stream order, with orders 7 and above showing positive effect and stream order 8 having the strongest effect (Figure 5). Flow velocity reaches positive effect around 0.50 m/s, increases up to 0.65 m/s, and then decreases until a spike at 1.00 m/s (Figure 5). Landcover diversity has positive effect at lower values and slowly decreases to negative effect across its range (Figure 5). Slope exhibits stronger effect at low values (Figure 5). The Plains Minnow model also performed well under 10-fold cross validation, reaching high accuracy with sensitivity = 0.84 and specificity = 0.88 (Table 6). The model also has strong fit and accuracy with AUC = 0.94 and TSS = 0.72 (Figure 6; Table 6).

The Northern Redbelly Dace model was trained on 814 data points comprised of 74 presences and 740 absences. The Northern Redbelly Dace model had flow velocity (19.99%), dominant landcover class (14.71%), stream order (12.03%), upstream drainage area (10.03%), discharge (8.76%), mean annual precipitation (7.81%), and slope (6.10%) as the most influential predictor variables at 79.43% cumulative influence (Table 5). Flow velocity shows positive effect on presence above 0.25 m/s with peak effect at approximately 0.40 m/s (Figure 8). The effect on presence is positive for low-intensity developed and grassland dominated subwatersheds and negative for cropland dominated subwatersheds (Figure 8). Stream order 7 has the strongest effect on presence, with orders 1 and 2 having positive effect as well (Figure 8). Total upstream drainage area reaches positive effect and plateaus around 68,000 km² (Figure 8). Discharge has positive effect on presence above approximately 20 m³/s (Figure 8). Mean annual precipitation reaches positive effect at -7.5 mm/year and remains positive afterwards (negative values

denote more evapotranspiration than precipitation in a given year; Figure 8). Slope has positive effect at lower values but decreases to remain neutral past 0.005 m/m (rise over run in meters; Figure 8). Despite this, the Northern Redbelly Dace model has the best performance under 10-fold cross validation of the three models, reaching high accuracy with sensitivity = 0.91 and specificity = 0.94 (Table 6). The model also has strong fit and accuracy with AUC = 0.98 and TSS = 0.85 (Figure 9; Table 6). While the total sample size is greater than 10 times the number of predictor variables, the Northern Redbelly Dace model may exhibit overfitting when considering the presences sample size of 74 (Breiner et al. 2015).

Discussion

Model Interpretation

Influential features from the Flathead Chub SDM support theory on environmental processes driving local habitat conditions and habitat use by Flathead Chub. Adult Flathead Chub are known to occupy main channels in large rivers, which is supported by the high influence of flow velocity and discharge at high values (Hesse 1994; Rahel and Thel 2004; Frimpong and Angermeier 2009). The positive effect of greater upstream drainage area further supports the selection of high order rivers, as the most positively impactful value at approximately 200,000 km² is relatively large within the variables range for Flathead Chub presences. Though Flathead Chub select for sand substrate, substrate diversity may also be important given the positive effect of organic material from approximately 1.37% to 1.53% (Hesse 1994; Rahel and Thel 2004; Frimpong and Angermeier 2009). The necessity of uninterrupted stretches of fluvial habitat for spawning is supported by the high influence of DCI at lower values (Hesse

1994; Rahel and Thel 2004). This relation reinforces the imperilment of Flathead Chub as a result of anthropogenic disturbance in the form of habitat fragmentation (Schneider et al. 2018; Perkin and Gido 2017). The positive effect of depth to groundwater at lower values may indicate that groundwater supply is an important driver of local conditions from which Flathead Chub select habitat. The role of precipitation in water supply is also captured in the similar influences and positive effects of precipitation and runoff at high values, though these variables were less influential than direct representatives of hydrology such as flow velocity and discharge (Dodds et al. 2004; Winter 2007). The relative influence of groundwater and precipitation related variables supports findings of their importance in supplying water to Nebraska rivers and streams, which goes on to shape habitat availability at small scales (Dodds et al. 2004; Winter 2007). Shannon's diversity index of subwatershed-scale landcover shows a positive effect on Flathead Chub presence at relatively high values. This may indicate that in-stream habitat heterogeneity positively effects Flathead Chub presence because diversity in the landscape drives that of the riverscape (Frissell et al. 1986; Fausch et al. 2002). Stream orders 4, 6, and 7 have positive effect on Flathead Chub presence, indicating that individuals may prefer habitat in tributaries rather than high order mainstems (Hesse 1994; Rahel and Thel 2004). The low influence and neutral effect of flood frequency across its range may also support findings of resilience of pelagic spawners to disturbance regimes (Worthington et al. 2014; Perkin and Gido 2017). The influences of slope, mean summer stream temperature, and subwatershed-scale dominant landcover did not fluctuate as values changed. Thus, the range of these environmental features in Nebraska streams may not impact habitat selection by Flathead Chub. Overall, the high

relative influences of hydrological variables (i.e., flow velocity and discharge) and position in the drainage network (i.e., upstream drainage area) support the selection of high order rivers and tributaries by Flathead Chub.

Influential features from the Plains Minnow SDM generally support processes driving site-scale conditions and theory on habitat use by Plains Minnow. Plains Minnow are thought to occupy pools in large rivers and tributaries, which is somewhat supported by the positive effects of discharge, upstream drainage area, stream order, and flow velocity on presence at high values (Hesse 1994; Rees et al. 2005; Osterhaus and Martin 2019). However, the positive relation with high values of discharge and flow velocity contradicts selection of slower flowing pools, which may result from data being too coarse grain to characterize pool mesohabitat within segments. Slope shows positive effect at lower values, likely representing the flat topography of landscapes that Nebraska rivers and streams traverse (Dodds et al. 2004). The role of groundwater in water supply is also captured in the positive effect of depth to groundwater at low values, though the low relative influence of depth to groundwater shows this may be an indirect driver of Plains Minnow presence (Dodds et al. 2004; Winter 2007). The connectivity of fluvial habitat may not impact Plains Minnow as much as Flathead Chub due to the low relative influence of DCI and gradual increase of effect with DCI value (Rees et al. 2005; Osterhaus and Martin 2019). This relation may indicate Plains Minnow are not as dependent on movement for spawning and may contradict their imperilment as a result of stream habitat fragmentation (Schneider et al. 2018; Osterhaus and Martin 2019). The macrophyte based diet of Plains Minnow may also be represented in the positive effect of percent organic material in the substrate at values above approximately 1.50% (Frimpong

and Angermeier 2009; Osterhaus and Martin 2019). The low relative influence and neutral effect of flood frequency across its range may also support findings of resilience of pelagic spawners to disturbance regimes (Worthington et al. 2014; Perkin and Gido 2017). Plains Minnow do not appear to respond to landscape features in the same way as Flathead Chub, as Plains Minnow presence shows little fluctuation as values of Shannon's diversity of landcover and dominant landcover class change. Mean annual precipitation and runoff have little influence, thus may not have a meaningful role in shaping available habitat for Plains Minnow selection. Plains Minnow show resilience to temperature shifts in the low influence and neutral effect of mean summer stream temperature in the majority of its range and positive effect at high values, which may contradict previous studies findings of cooler water selection (Osterhaus and Martin 2019). Overall, variable influences align with what is thought of as suitable habitat for Plains Minnow in uninterrupted lengths of higher order rivers and streams, which would have naturally high flow velocity and discharge.

Influential features from the Northern Redbelly Dace SDM somewhat support previous studies on processes driving site-scale conditions habitat use by Northern Redbelly Dace. Northern Redbelly Dace selection of headwater habitat is supported by the high influence of flow velocity and positive effect at low values around 0.4 m/s (Frimpong and Angermeier 2009; Stringer and Clancy 2020). However, the positive effect of stream order 7 and upstream drainage area at high values and negative influence of discharge at low values directly contradicts Northern Redbelly Dace selection of headwater streams (Frimpong and Angermeier 2009; Stringer and Clancy 2020). Land use may impact Northern Redbelly Dace presence as indicated by the negative relation of

cultivated cropland dominated subwatersheds (class 82) and positive relation with grassland (class 71; Dewitz et al. 2021). Slope shows positive effect at lower values, likely representing the flat topography of landscapes through which Nebraskan rivers and streams run (Dodds et al. 2004). The positive effect of depth to groundwater at low values highlights the importance of groundwater in supplying water to the headwater streams for which Northern Redbelly Dace select (Dodds et al. 2004; Winter 2007; Frimpong and Angermeier 2009; Stringer and Clancy 2020). Further, depth to groundwater was more influential on Northern Redbelly Dace presence than Flathead Chub and Plains Minnow presence, reinforcing the differential selection of headwaters and high order rivers respectively. The role of precipitation in water supply is captured in the relative influence and positive effect of precipitation at high values (Dodds et al. 2004; Winter 2007). However, runoff shows little influence and effect as values change, indicating groundwater may be a more critical water supplier to streams that Northern Redbelly Dace occupy. Organic material in the substrate shows little influence and effect as values change, which may reflect the more generalist approach Northern Redbelly Dace have to substrate selection (Frimpong and Angermeier 2009). Northern Redbelly Dace selection of headwater streams may be reflected in the positive effect of lower mean summer stream temperature (Stringer and Clancy 2020). Flood frequency shows little relative influence that did not fluctuate in effect as values changed, thus it may not effect habitat selection by Northern Redbelly Dace. Connectivity may not influence Northern Redbelly Dace presence as indicated by the low influence of DCI, though some positive effect is exhibited at lower values. Though land use shows high influence on Northern Redbelly Dace presence, landcover diversity shows little relation across its values.

Dominant landcover class, baseflow index, mean annual precipitation, and slope align with what is thought of as suitable habitat for Northern Redbelly Dace in the intermittent, groundwater- and precipitation-fed headwaters in the Sandhills Ecoregion. However, the Northern Redbelly Dace model is not as conducive with the literature as the previous models, wherein flow velocity, stream order, and total upstream drainage area imply some suitable habitat in high order streams where suitability in low order headwaters would be expected.

Although SDMs are powerful tools it is important to note their limitations, namely that inferring habitat availability has the caveats of representing incomplete data, maintaining biases in the training data, and requiring validations of their predictions (Kery et al. 2010; Newbold 2010; Jarnevich et al. 2015). The accuracies of species identification and presence location in historical data are often uncertain (Newbold 2010). Historical datasets also often carry spatial, temporal, environmental, and taxonomic biases in their sampling methods (Newbold 2012). Use of historical datasets in SDMs is still commonplace, given recognition that model predictions are inherently theoretical and open to verification or refutation (Labay et al. 2011; Jarnevich et al. 2015; Senecal et al. 2015). Given the speed of environmental alteration and degradation in the modern day, there is an urgency to understand species-habitat relations and translate this understanding into effective management (Newbold 2012; Labay et al. 2011; Schneider et al. 2011). As such, historical datasets representing already accrued knowledge, such as the data used in this study, are still useful for informing species distributions and conservation (Newbold 2010).

Local-scale environmental data were not collected in conjunction with species presence data used in this study, which is common in historical datasets (Newbold 2010; Labay et al. 2011). As such, there is a disconnect between species presences and associated environmental conditions in spatiotemporal contexts. This introduces a large degree of uncertainty to model interpretations, as it is highly likely that the environmental data informing the species-habitat relations do not match the local-scale habitat selected by individuals at a given time. More robust modelling practices may backcast historical environmental conditions to compensate for missing environmental data (Labay et al. 2011; Worthington et al. 2014). Some approaches view SDMs as snapshots in time and space. Thus, presence data must constitute both species presence and habitat conditions at a given site (Jarnevich et al. 2015). Nonetheless, species-habitat relations can still be established with historical presences related to larger-scale variables, such as those on the subwatershed or catchment scales. Larger-scale variables represent the environmental processes shaping local scale habitat conditions and can still inform how these processes impact species distributions (Poff 1997; Fausch et al. 2002). Thus, though a spatial and temporal disconnect exists between the environmental and species presence data used, the SDMs presented in this study may offer insight into species-habitat relations.

The SDMs presented in this research were internally validated, which cannot overcome biases in their assessment of model fit through validity metrics such as AUC and TSS (Hao et al. 2020). External validation, however, uses out-of-sample data to test the accuracy of SDM predictions (Hao et al. 2020). Some validation approaches withhold subsets of datasets for testing of model fit, as done in this study, but data subsets from the same overall sample maintain the same biases (Hao et al. 2020). External validation of

SDM predictions using field validation test the adaptability of the model to outside biases, thus offers a better performing assessment under broader model applicability and should be used when possible (Jarnevich et al. 2015; Hao et al. 2020). For example, field validation may not observe a species in a predicted suitable stream reach. Barring detection probability, this may be indicative that the model needs more data to better characterize species habitat relations. In such cases, newly sampled data may be contributed to the current model to improve performance. A benefit of boosted regression tree models is that new data can easily be added to the current model iterations to improve model certainty and interpretations of species-habitat relations (Hijmans et al. 2022). However, lack of species detection at high predicted probability sites may also indicate that a site with habitat suitability for a target species is not accessible to individuals in space and time. In such a situation, a common conservation practice of translocation of individuals may help bolster species abundance at larger scales (George et al. 2009; Tarkan et al. 2015).

Rare species pose a challenge to species distribution modeling techniques because of the low number of historical presences available in data resulting from either lower detection probabilities or true lower population numbers (Breiner et al. 2015; Martin et al. 2022). All three study species could be considered rare given the relatively low proportion of suitable fluvial habitat in the SDM predictions (Table 7). Especially in combination with high numbers of predictors, few presences can lead to model overfitting where the fit of the model predictions is misleadingly strong (Breiner et al. 2015). In particular, the Northern Redbelly Dace model appears to exhibit overfitting due to the rarity of presence points in the historical datasets used. For example, 5 out of 74 of

the Northern Redbelly Dace presence points occurred in stream order 7, yet the SDM for Northern Redbelly Dace showed some strong relation of presence to characteristics of higher order rivers. Certainty in interpreted species-habitat relations from SDMs can be improved with the introduction of more spatially and temporally robust environmental and species presence data (Breiner et al. 2015; Martin et al. 2022). The addition of new data reduces overfitting by increasing species presence points, while also increasing certainty by building relationships from more spatially and temporally appropriate environmental data (Breiner et al. 2015; Martin et al. 2022). Overall, the models presented in this study should not be perceived as absolute definitions of species-habitat relations, rather tools for their examination to be iteratively improved through direction of future sampling.

Research Implications

SDMs map habitat suitability for target species, spatially directing conservation and management actions to reaches where a species presence is likely. For a rare and at-risk species, where habitat availability may be unknown or decreasing in the greater environmental context, maps of habitat suitability may expedite the conservation process. This may be especially useful given time is of concern to build understanding of these species and take effective conservation actions (Newbold 2010). For example, maps of habitat suitability for Flathead Chub, Plains Minnow, and Northern Redbelly Dace produced in this study may be used for targeted future research of these species and sampling of their habitats (Figure 4; Figure 7; Figure 10). Where gaps in knowledge exist for the life-histories of these at-risk minnow species, habitat suitability maps can be used to inform targeted studies to fill these gaps. SDMs are easily constructed for any target

species, given appropriate data for model training, to serve similar functions (Elith et al. 2008). Further, various model types with different applications are available including boosted regression tree, random forest, general additive, and general linear models (Guisan and Zimmerman 2000). For example, SDMs can be used to predict and prevent encroachment of invasive species through mapping habitat suitability along an invasion front (Marras et al. 2015).

Characterizing species habitat-relations provides insight into patterns of environmental features for consideration in management of habitat for these at-risk species. Hydrological variables such as flow velocity and discharge were highly influential for all three species. This may reflect the role of large-scale hydrological processes in shaping site-scale habitat conditions (Fausch et al. 2002; Dodds et al. 2004). As such, consideration of hydrological processes may benefit conservation of at-risk fish species. For example, positive relations between flow and presence of a target species could serve as the goal for successful stream restoration efforts or development of flow standards (Radinger and Wolter 2015). Further, species distributions may reflect patterns in habitat heterogeneity (Sowa et al. 2008). Using the Flathead Chub model to examine this concept, stream reaches with high predicted presence likely align with influential values of selected variables (Sowa et al. 2008). Where predicted presence of Flathead Chub is high, one may observe discharge above 1000 m³/s, flow velocity of approximately 0.5 m/s to 0.8 m/s, and 1.50% organic matter in the substrate (Sowa et al. 2008). Field validation of model predictions would allow further insight into the accuracy of translation of species distribution to habitat heterogeneity patterns.

Tables

Table 2.1. Comparison of species traits for Flathead Chub, Plains Minnow, and Northern Redbelly Dace from the U.S. Geological Survey FishTraits Database (Frimpong and Angermeier 2009).

| Species | Diet | Longevity (Years) | Spawning Season | Spawning Strategy | Flow | Substrate |
|---------------------------|--|------------------------------|----------------------------|--|-------------------|---|
| Flathead Chub | macrophytes, invertebrates, smaller fish | 6 | July - September | Nonguarding, Lithopelagophillic (sand) | Moderate- Fast | clay, silt, sand |
| Plains Minnow | algae, phytoplankton, macrophytes, detritus | 2 | April - July | Nonguarding, Lithopelagophillic | Slow- Moderate | clay, silt, sand |
| Northern Redbelly Dace | algae, phytoplankton, invertebrates | 3 | May - August | Nonguarding, Phytophillic | Slow | muck, clay, silt, sand, gravel, cobble, boulder, detritus |

Table 2.2. Descriptions, units, and sources of environmental predictor variables used to train the species distribution models.

| Variable | Description | Units | Source |
|-----------------|---|-------------------|---------------|
| StreamOrde | Stream order | Strahler's | NHD+ v2 |
| SLOPE | Slope of flowline | m/m | NHD+ v2 |
| TotDASqKM | Total upstream drainage area | km ² | NHD+ v2 |
| V0001E | Reach-scale flow velocity estimate with gage adjustments | m/s | NHD+ v2 |
| Q0001E | Reach-scale flow discharge estimate with gage adjustments | m ³ /s | NHD+ v2 |
| Precip_Min | Mean annual precipitation minus evapotranspiration within catchment | mm | StreamCat |
| MSST_2014 | Mean summer stream temperature | °C | StreamCat |
| RunoffWs | Mean annual runoff within watershed | mm | StreamCat |
| Mx_class | Dominant landcover class within HUC12, class 11 is open water, class 22 is developed (low intensity), class 71 is grassland, and class 82 is cropland | | Calculated |
| SDHI | Shannon's diversity index for landcover classes within HUC12 | | Calculated |
| DCI | Discontinuity Index within HUC12 | | Calculated |
| OM | Organic matter in substrate | % | STATSGO |
| AFLDFREQ | Annual flood frequency (1 = frequent (>50% chance), 2 = occasional (5-50% chance), 3 = rare (<5% chance)) | | STATSGO |
| MADTW | Annual mean depth to groundwater | m | Calculated |

Table 2.3. Relative and cumulative influence of predictor variables from the Flathead Chub model. See Table 2.2 for variable definitions, units, and sources.

| Variable | Relative Influence (%) | Cumulative Influence (%) |
|-----------------|-------------------------------|---------------------------------|
| Q0001E | 30.50 | 30.50 |
| V0001E | 29.94 | 60.44 |
| TotDASqKM | 6.87 | 67.31 |
| OM | 5.34 | 72.65 |
| DCI | 4.10 | 76.75 |
| SLOPE | 3.93 | 80.68 |
| MADTW | 3.91 | 84.59 |
| SHDI | 3.64 | 88.23 |
| StreamOrde | 3.18 | 91.41 |
| MSST_2014 | 2.48 | 93.89 |
| Precip_Min | 2.20 | 96.09 |
| RunoffWs | 2.07 | 98.16 |
| AFLDFREQ | 1.06 | 99.22 |
| Mx_class | 0.77 | 100.00 |

Table 2.4. Relative and cumulative influence of predictor variables from the Plains Minnow model. See Table 2.2 for variable definitions, units, and sources.

| Variable | Relative Influence (%) | Cumulative Influence (%) |
|-----------------|-------------------------------|---------------------------------|
| Q0001E | 30.11 | 30.11 |
| TotDASqKM | 20.86 | 50.97 |
| StreamOrde | 10.49 | 61.46 |
| V0001E | 9.09 | 70.55 |
| SHDI | 6.72 | 77.27 |
| SLOPE | 6.36 | 83.63 |
| MADTW | 3.71 | 87.34 |
| DCI | 3.50 | 90.84 |
| OM | 2.48 | 93.32 |
| AFLDFREQ | 2.08 | 95.40 |
| Mx_class | 1.90 | 97.30 |
| RunoffWs | 1.75 | 99.05 |
| Precip_Min | 0.57 | 99.62 |
| MSST_2014 | 0.38 | 100.00 |

Table 2.5. Relative and cumulative influence of predictor variables from the Northern Redbelly Dace model. See Table 2.2 for variable definitions, units, and sources.

| Variable | Relative Influence (%) | Cumulative Influence (%) |
|-----------------|-------------------------------|---------------------------------|
| V0001E | 19.99 | 19.99 |
| Mx_class | 14.71 | 34.70 |
| StreamOrde | 12.03 | 46.73 |
| TotDASqKM | 10.03 | 56.76 |
| Q0001E | 8.76 | 65.52 |
| Precip_Min | 7.81 | 73.33 |
| SLOPE | 6.10 | 79.43 |
| MADTW | 4.63 | 84.06 |
| OM | 4.34 | 88.40 |
| MSST_2014 | 3.78 | 92.18 |
| RunoffWs | 3.43 | 95.61 |
| AFLDFREQ | 2.03 | 97.64 |
| DCI | 1.23 | 98.87 |
| SHDI | 1.15 | 100.00 |

Table 2.6. Validation summary data for the species distribution models. Sensitivity denotes the rate of true positives in model predictions, while specificity denotes the rate of false positives. Area under curve (AUC) is a metric of model fit ranging from 0-1 determined by a receiver-operating characteristic curve. True skill statistic (TSS) is a measure of model accuracy ranging from -1 to +1.

| Model Species | Sensitivity | Specificity | AUC | TSS |
|------------------------|--------------------|--------------------|------------|------------|
| Flathead Chub | 0.85 | 0.87 | 0.93 | 0.73 |
| Plains Minnow | 0.85 | 0.88 | 0.93 | 0.73 |
| Northern Redbelly Dace | 0.83 | 0.94 | 0.97 | 0.77 |

Table 2.7. Proportion of total flowline with predicted presence greater than 0.7 for each species based on Figure 4, Figure 7, and Figure 10. High presence length (HPL) denotes the sum of flowline length with predicted presence greater than 0.7 in kilometers. Total length (TL) denotes the sum length of all flowlines across Nebraska in kilometers. Percentage of high presence (PHP) denotes the proportion of flowline length with predicted presence greater than 0.7 as a percentage.

| Species | HPL (km) | TL (km) | PHP (%) |
|------------------------|-----------------|----------------|----------------|
| Flathead Chub | 3199.01 | 165615.90 | 1.93 |
| Plains Minnow | 3993.13 | 165615.90 | 2.41 |
| Northern Redbelly Dace | 6050.51 | 165615.90 | 3.65 |

Figures

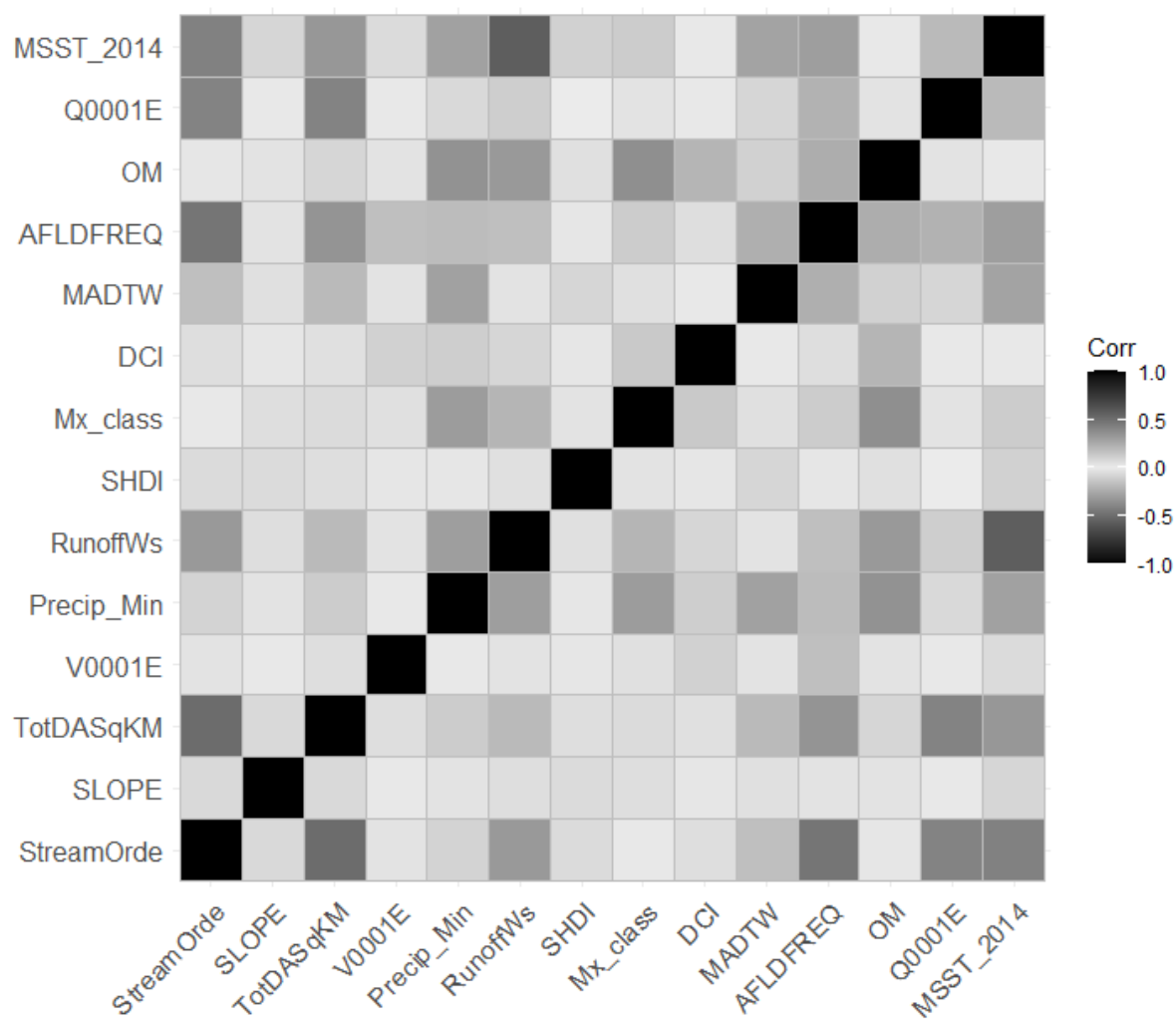


Figure 2.1. Correlation grid showing selected environmental predictor variables, all correlation ≤ 0.7 . See Table 2.2 for variable definitions.

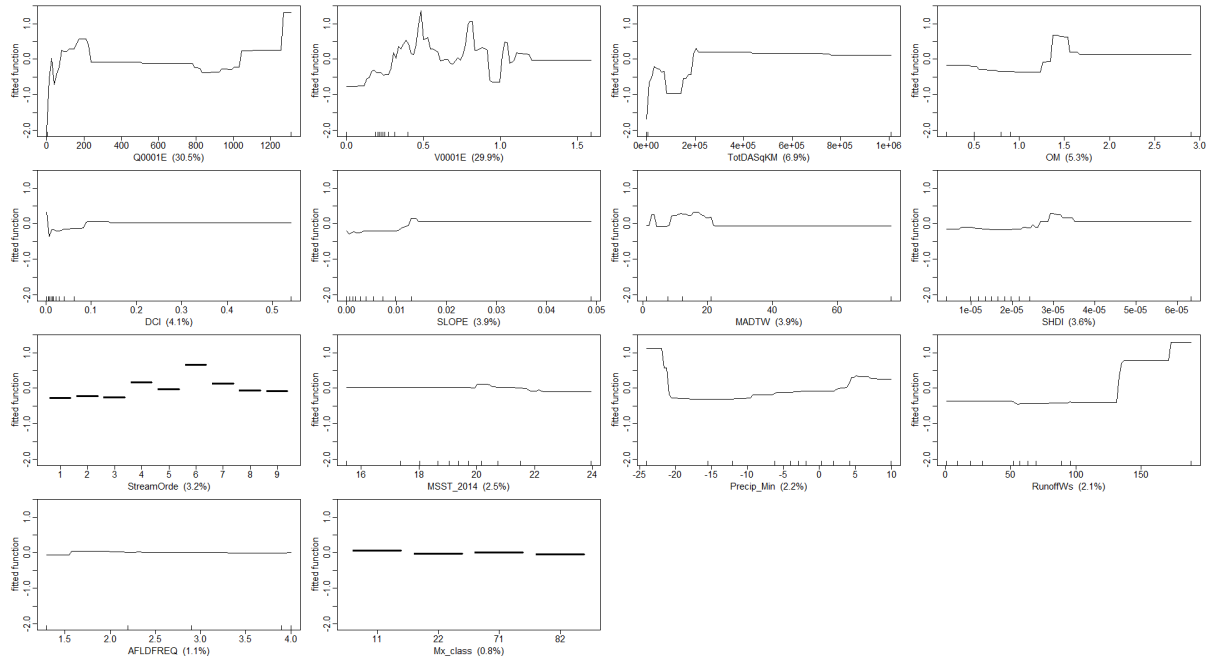


Figure 2.2. Partial Dependence plots from the Flathead Chub model ordered by relative influence. The y-axes of all plots are the fitted functions, denoting predicted response where values above zero are positive influence on presence and values below zero are negative influence on presence. The x-axes are variable values, with labels showing the relative influence of each variable as a percentage. The ticks on top of the x-axis denote data deciles (i.e., tenths of the total input points). See Table 2.2 for variable definitions, units, and sources.

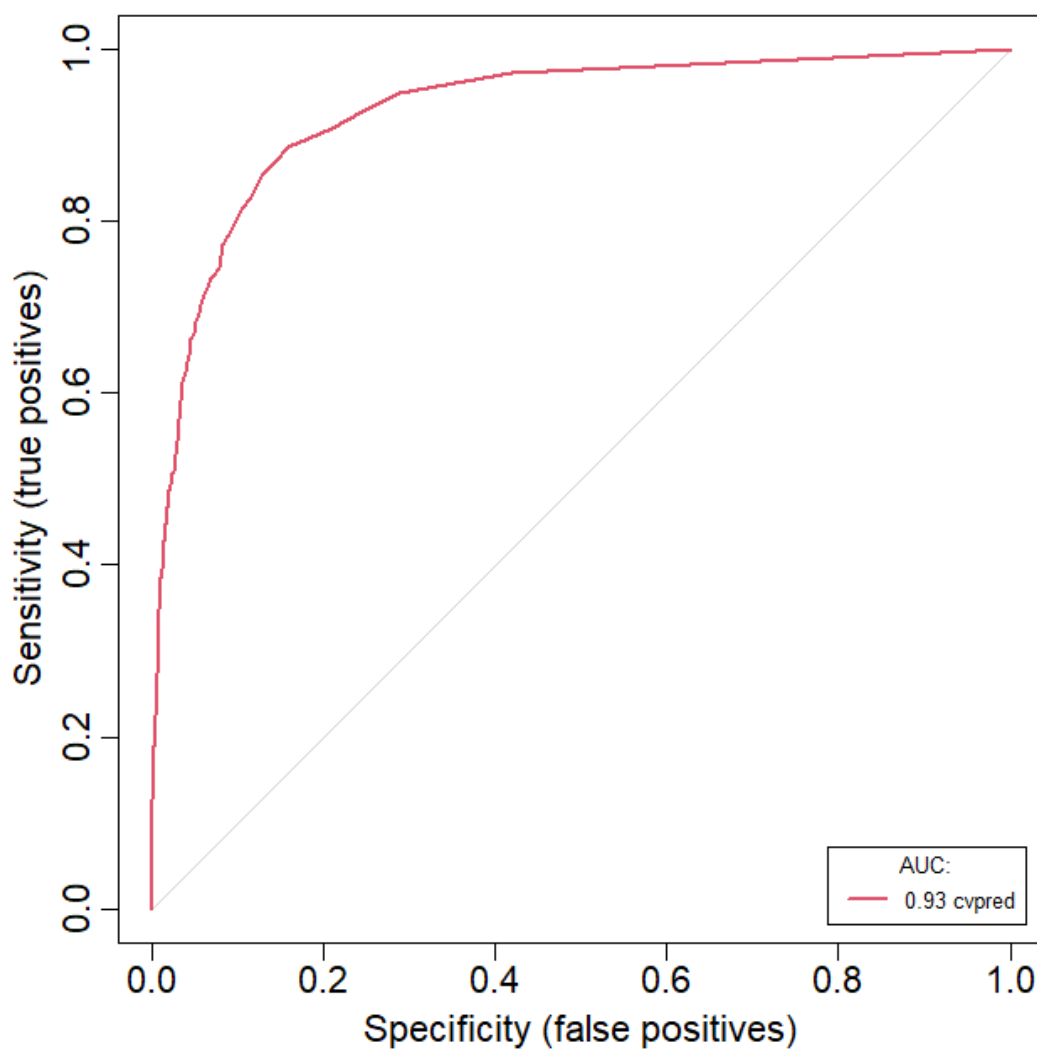


Figure 2.3. The 10-fold cross validation receiver operating-characteristic curve plot of the Flathead Chub model. The y-axis is sensitivity, or the rate of true positives. The x-axis is specificity, or the rate of false positives. In the bottom right corner the area under the curve ranging from 0-1 is noted.

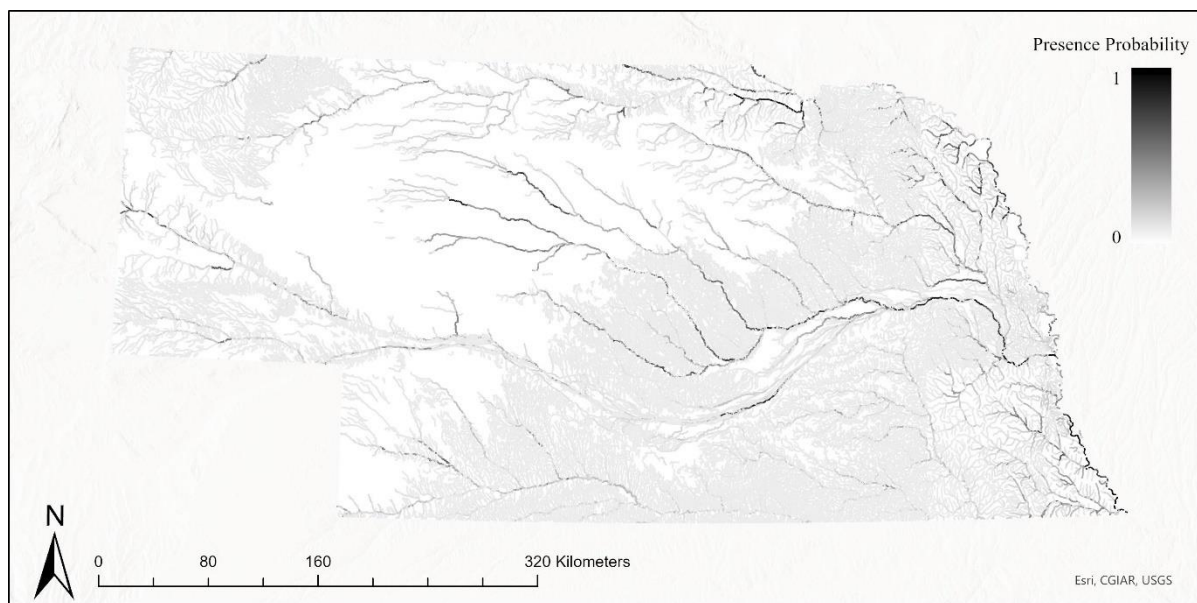


Figure 2.4. The model predicted probability of Flathead Chub occurrence mapped over the NHD Flowlines with additional metrics. Darker color denotes higher predicted probability of occurrence.

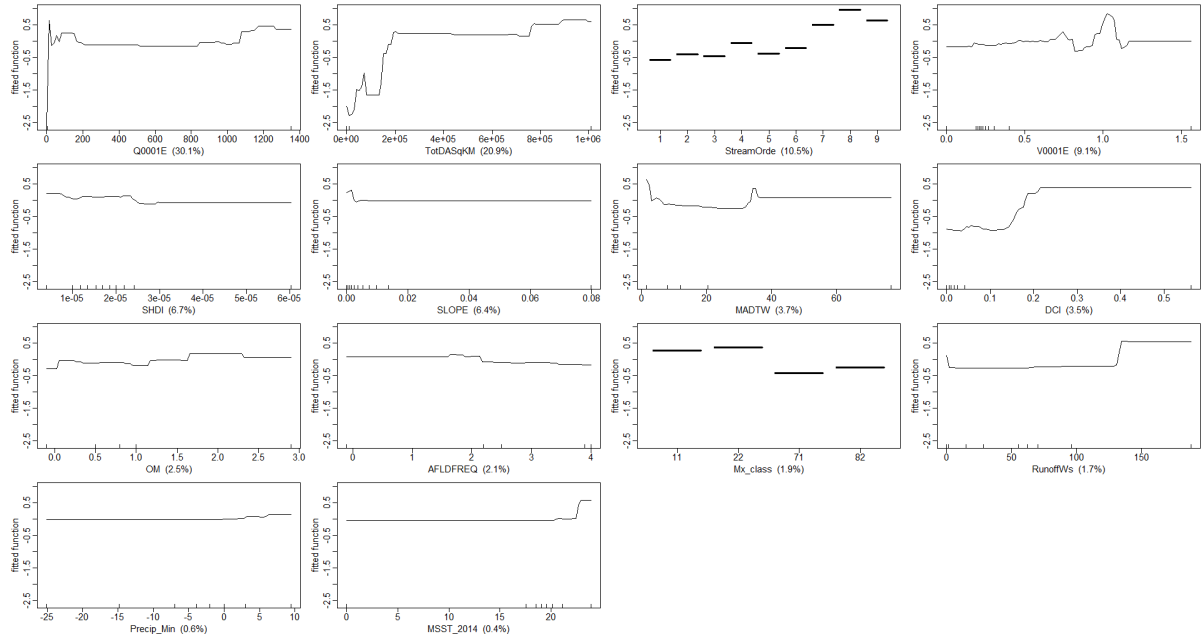


Figure 2.5. Partial dependence plots from the Plains Minnow model ordered by relative influence. The y-axes of all plots are the fitted functions, denoting predicted response where values above zero are positive influence on presence and values below zero are negative influence on presence. The x-axes are variable values, with labels showing the relative influence of each variable as a percentage. The ticks on top of the x-axis denote data deciles (i.e., tenths of the total input points). See Table 2.2 for variable definitions, units, and sources.

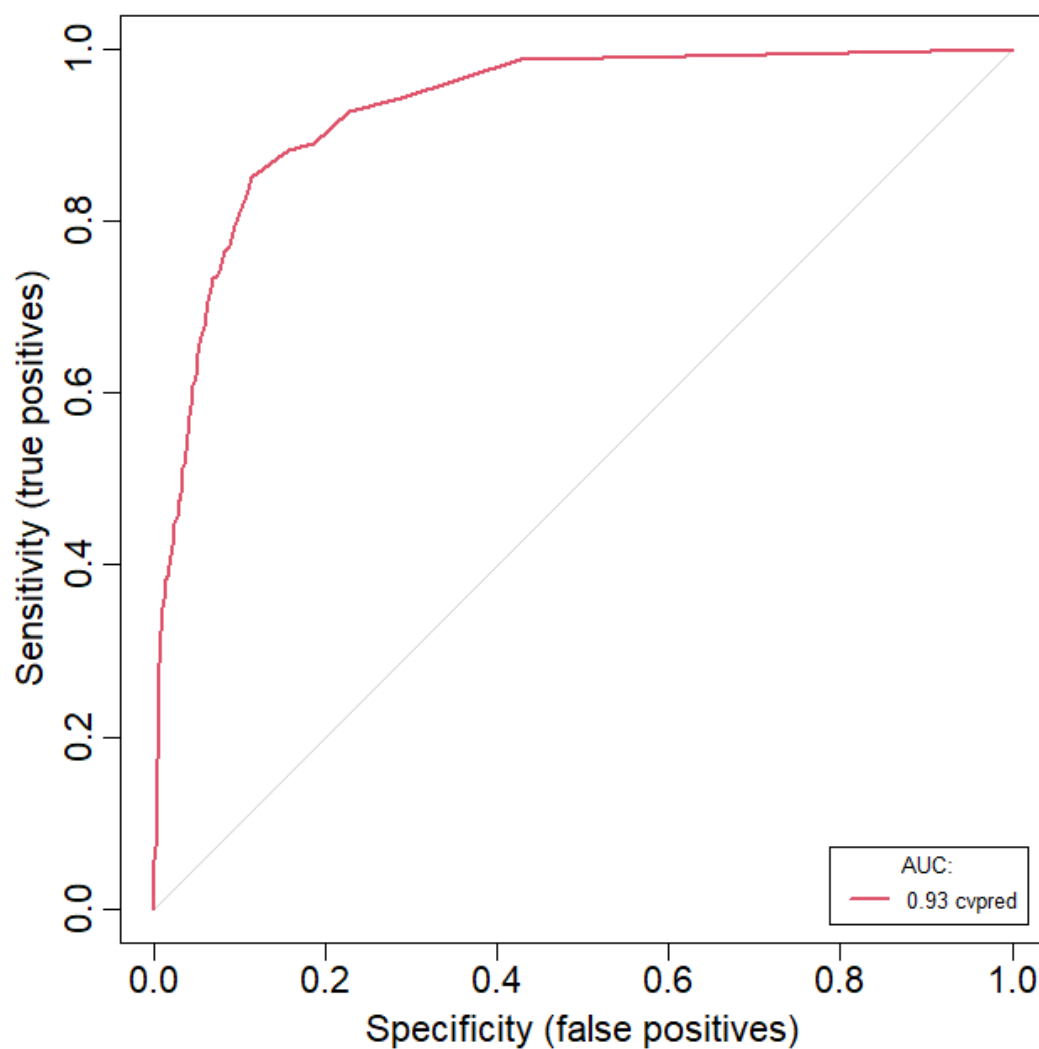


Figure 2.6. The 10-fold cross validation receiver operating-characteristic curve plot of the Plains Minnow model. The y-axis is sensitivity, or the rate of true positives. The x-axis is specificity, or the rate of false positives. In the bottom right corner the area under the curve ranging from 0-1 is noted.

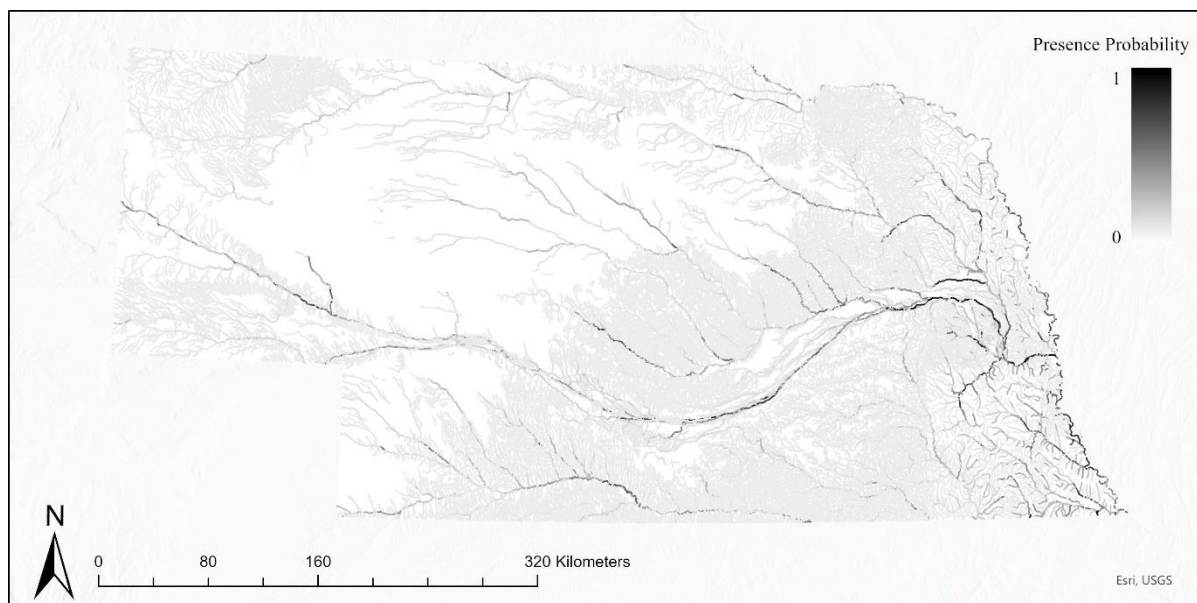


Figure 2.7. The model predicted probability of Plains Minnow occurrence mapped over the NHD Flowlines with additional metrics. Darker color denotes higher predicted probability of occurrence.

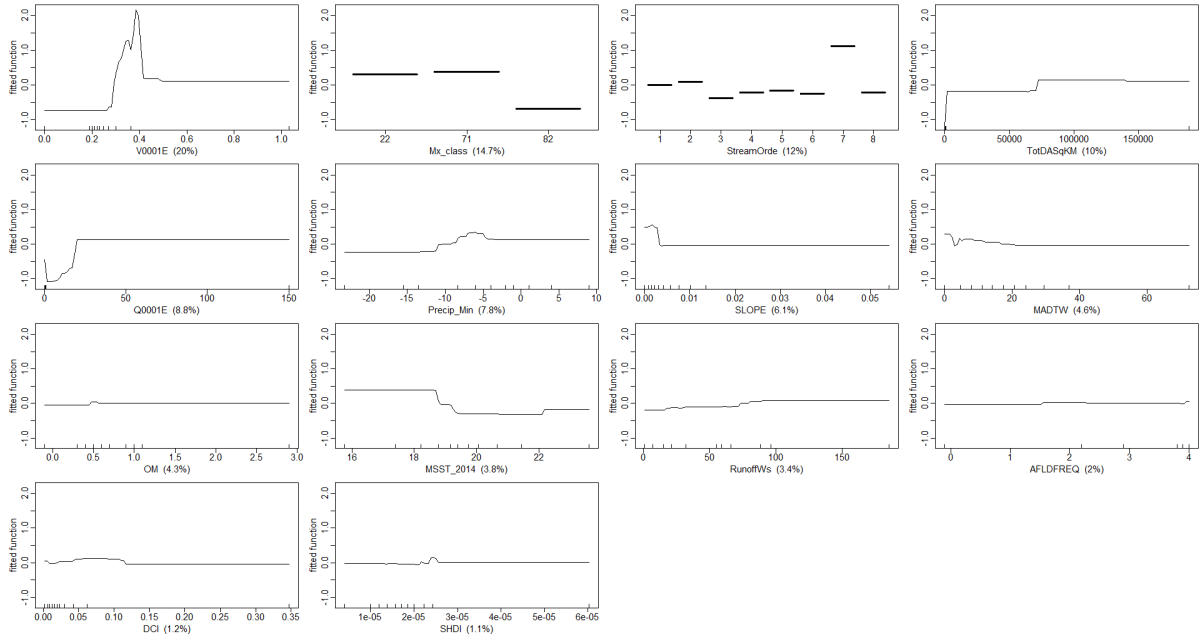


Figure 2.8. Partial dependence plots from the Northern Redbelly Dace model ordered by relative influence. The y-axes of all plots are the fitted functions, denoting predicted response where values above zero are positive influence on presence and values below zero are negative influence on presence. The x-axes are variable values, with labels showing the relative influence of each variable as a percentage. The ticks on top of the x-axis denote data deciles (i.e., tenths of the total input points). See Table 2.2 for variable definitions, units, and sources.

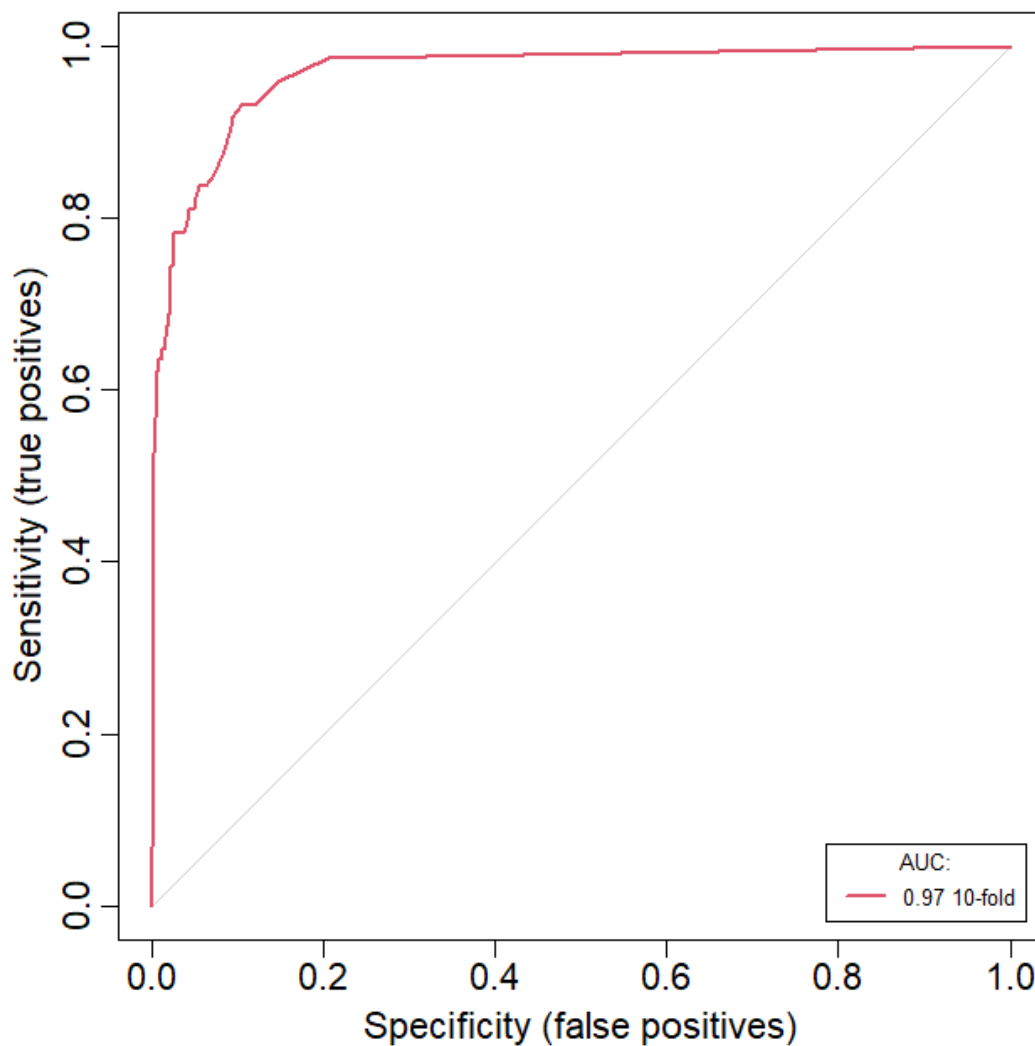


Figure 2.9. The 10-fold cross validation receiver operating-characteristic curve plot of the Northern Redbelly Dace model. The y-axis is sensitivity, or the rate of true positives. The x-axis is specificity, or the rate of false positives. In the bottom right corner the area under the curve ranging from 0-1 is noted.

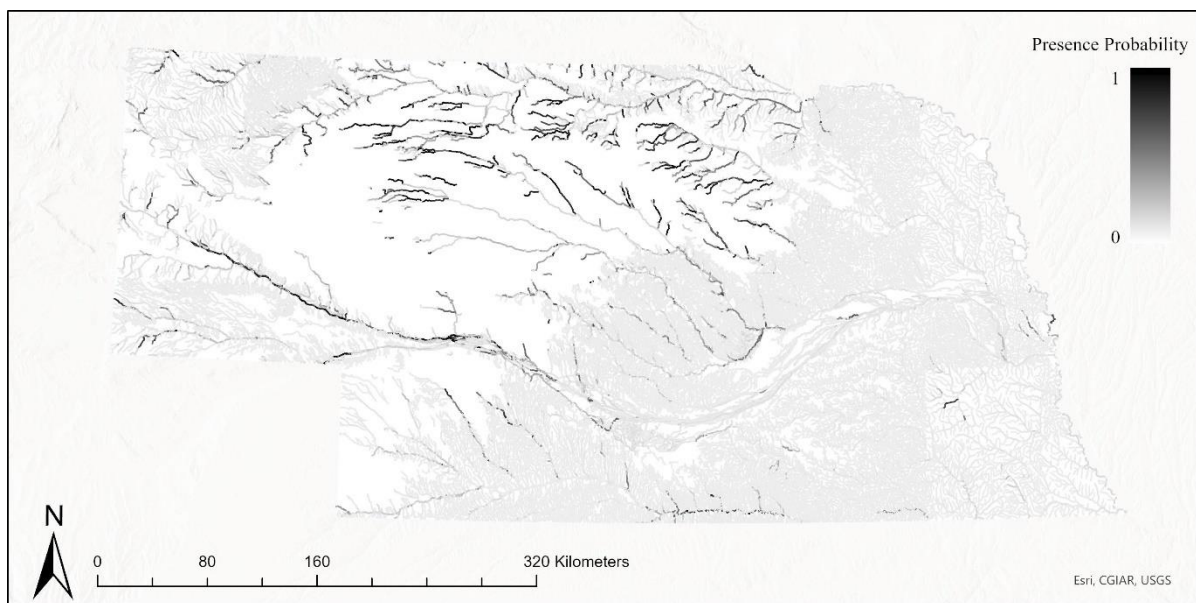


Figure 2.10. The model predicted probability of Northern Redbelly Dace occurrence mapped over the NHD Flowlines with additional metrics. Darker color denotes higher predicted probability of occurrence.

References

- Aarts, G., J. Fieberg, S. Brasseur, and J. Matthiopoulos. 2013. Quantifying the effect of habitat availability on species distributions. *Journal of Animal Ecology* 82:1135-1145.
- Albanese, B., P. L. Angermeier, and J. T. Peterson. 2009. Does mobility explain variation in colonization and population recovery among stream fishes? *Freshwater Biology* 54(7):1444-1460.
- Alford, S. L., and A. W. Walters. 2021. Rapid colonization post-displacement contributes to native fish resilience. *Ecology of Freshwater Fish* 00:1-11.
- Allouche, O., A. Tsoar, and R. Kadmon. 2006. Assessing the accuracy of species distribution models: prevalence, kappa, and the true skill statistic (TSS). *Journal of Applied Ecology* 43:1223-1232.
- Artaev, O. 2023. Prediction of current and future suitable habitats for three invasive freshwater fish species in Europe. *Water* 15:2091-2106.
- Barbet-Massin, M., F. Jiguet, C.H. Albert, and W. Thuiller. 2012. Selecting pseudo-absences for species distribution models: Where and how many? *Methods in Ecology and Evolution* 3:327-338.
- BISON, 2022, Biodiversity Information Serving Our Nation: accessed 2022, at <https://bison.usgs.gov>.
- Borboudakis, G. and I. Tsamardinos. 2019. Forward-backward selection with early dropping. *Journal of Machine Learning Research* 20:1-39.
- Breiner, F.T., A. Guisan, A. Bergamini, and M.P. Nobis. 2015. Overcoming limitations of modelling rare species by using ensembles of small models. *Methods in Ecology and Evolution* 6:1210-1218.
- Brewer, S. K., T. A. Worthington, R. Mollenhauer, D. R. Stewart, R. A. McManamay, L. Guertault, and D. Moore. 2018. Synthesizing models useful for ecohydrology and ecohydraulic approaches: An emphasis on integrating models to address complex research questions. *Ecohydrology* 11(7).
- Bryan, J.D., L.G. Hill, and W.H. Neill. 1984. Interdependence of acute temperature preference and respiration in the Plains Minnow. *Transactions of the American Fisheries Society* 113(5):557-562.
- Cardinale, B.J., M.A. Palmer, C.M. Swan, S. Brooks, and N.L. Poff. 2002. The influence of substrate heterogeneity on biofilm metabolism in a stream ecosystem. *Ecology* 83(2):412-422.

- Carlisle, D. J. Falcone, and M. Meador. 2009. Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. *Environmental Monitoring and Assessment* 151:143-160.
- Carpenter, S. R., E. H. Stanley, and M. J. V. Zanden. 2011. State of the World's freshwater ecosystems: physical, chemical, and biological changes. *Annu. Rev. Environ. Resour.* 36:75-99.
- Cerasoli, F., M. Iannella, P. D'Alessandro, and M. Biondi. 2017. Comparing pseudo-absences generation techniques in boosted regression trees models for conservation purposes: A case study on amphibians in a protected area. *PLoS ONE* 12(11):e0187589.
- Comte, L., J. D. Olden, P. A. Tedesco, A. Ruhi, and X. Giam. 2021. Climate and land-use changes interact to drive long-term reorganization of riverine fish communities globally. *Proceedings of the National Academy of Sciences* 118:1-9.
- Conservation Survey Division, University of Nebraska Lincoln. 2023. WLProd, Map Server. https://csdgis.unl.edu/csdgis/rest/services/Production_Services/WLProd/MapServer/0. Using ESRI ArcGIS Pro. Version 11.
- Davies, N. M., R. H. Norris, and M. C. Thoms. 2000. Prediction and assessment of local stream habitat features using large-scale catchment characteristics: Prediction and assessment of local stream habitat. *Freshwater Biology* 45(3):343-369.
- De'ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88(1): 243-251.
- Dewitz, J., and U.S. Geological Survey. 2021. National Land Cover Database (NLCD) 2019 Products. U.S. Geological Survey data release.
- Dodds, W.K., K. Gido, M.R. Whiles, K.M. Fritz, and W.J. Matthews. 2004. Life on the edge: the ecology of Great Plains prairie streams. *Bioscience* 54(3):205-216.
- Dumelle, M., T. Kincaid, A.R. Olsen, and M. Weber. 2023. spsurvey: Spatial Sampling Design and Analysis in R. *Journal of Statistical Software* 105(3):1-29.
- Durance, I., C. Lepichon, and S. J. Ormerod. 2006. Recognizing the importance of scale in the ecology and management of riverine fish. *River Research and Applications* 22(10):1143-1152.
- Elith, J., J.R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77: 802-813.
- Elith, J., and J.R. Leathwick. 2017. Boosted regression trees for ecological modeling. R Documentation. <https://cran.r-project.org/web/packages/dismo/vignettes/brt.pdf>.

- Esri Inc. 2022. *ArcGIS Pro* (Version 3.0). Esri Inc. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>.
- Falcone, J.A., D.M. Carlisle, and L.C. Weber. 2010. Quantifying human disturbance in watersheds: variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. *Ecological Indicators* 10:264-273.
- Falke, J. A., and K. D. Fausch. 2010. From metapopulations to metacommunities: Linking theory with empirical observations of the spatial population dynamics of stream fishes. *Americans Fisheries Society Symposium* 73:207–233.
- Fausch, K. D., C. E. Torgersen, C. V. Baxter, and H. W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* 52(6):483.
- Fischer, J. R., and C. P. Paukert. 2008. Habitat relationships with fish assemblages in minimally disturbed Great Plains regions. *Ecology of Freshwater Fish* 17(4):597–609.
- Frimpong, E. A., and P. L. Angermeier. 2009. Fish traits: A database of ecological and life-history traits of freshwater fishes of the United States. *Fisheries* 34(10):487–495.
- Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management* 10(2):199–214.
- George, A.L., B.R. Kuhajda, J.D. Williams, M.A. Cantrell, P.L. Rakes, and J.R. Shute. 2009. Guidelines for propagation and translocation for freshwater fish conservation. *Fisheries* 34(11):529-545.
- Guisan, A. and N.E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147-186.
- Hao, T., J. Elith, J.J. Lahoz-Monfort, and G. Guillera-Aroita. 2020. Testing whether ensemble modelling is advantageous for maximizing performance of species distribution models. *Ecography* 43:549-558.
- Hastie, T., R. Tibshirani, and J.H. Friedman. 2001. *The elements of statistical learning: data mining, inference, and prediction*. Springer-Verlag, New York.
- Helmsetter, N.A., C.J. Conway, B.S. Stevens, and A.R. Goldberg. 2021. Balancing the transferability and complexity of species distribution models for rare species conservation. *Diversity and Distributions* 27:95-108.
- Hesse, L. 1994. The status of Nebraska fishes in the Missouri River, 5 selected chubs and minnows (Cyprinidae): Sicklefing Chub (*Macrhybopsis meeki*), Sturgeon Chub (*M. geuda*), Silver Chub (*M. Storeriana*), Speckled Chub (*M. aestivaus*), Flathead

- Chub (*Platygobio gracilis*), Plains Minnow (*Hybognathus placitus*), and Western Silvery Minnow (*H. argyritis*). Transactions of the Nebraska Academy of Sciences and Affiliated Societies 21:99–108.
- Hijmans, J.R., S. Phillips, J. Leathwick, J. Elith. 2022. dismo: Species Distribution Modeling. R package version 1.3-9.
- Hill, R.A., M.H. Weber, S.G. Leibowitz, A.R. Olsen, and D.J. Thornbrugh. 2015. The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. Journal of the American Water Resources Association 52:120-128.
- Jackson, H. B., and L. Fahrig. 2015. Are ecologists conducting research at the optimal scale?: Is research conducted at optimal scales? Global Ecology and Biogeography 24(1):52-63.
- Jarnevich, C.S, T.J. Stohlgren, S. Kumar, J.T. Morisette, and T.R. Holcombe. 2015. Caveats for correlative species distribution modeling. Ecological Informatics 29:6-15.
- Jelks, H.L., S.J. Walsh, N.M. Burkhead, S. Contreras-Balderas, E. Diaz-Pardo, D.A. Hendrickson, J. Lyons, N.E. Mandrak, F. McCormick, J.S. Nelson, S.P. Platania, B.A. Porter, C.B. Renaud, J.J. Schmitter-Soto, E.B. Taylor, and M.L. Warren Jr. 2008. Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries 33:372-407.
- Krausman, P.R. and M.L. Morrison. 2016. Another plea for standard terminology. Journal of Wildlife Management 80(7):1143-1144.
- Keller, D.H. 2023. Scale-dependence of fish habitat associations in a North American river system. Freshwater Biology 68:12-24.
- Kery, M., B. Gardner, and C. Monnerat. 2010. Predicting species distributions from checklist data using site-occupancy models. Journal of Biogeography 37:1851-1862.
- Labay, B., A.E. Cohen, B. Sissel, D.A. Hendrickson, F.D. Martin, and S. Sarkar. 2011. Assessing historical fish community composition using surveys, historical collection data, and species distribution models. PLoS One 6(9):e25145.
- Lehtinen, S. F., and J. B. Layzer. 1988. Reproductive cycle of the Plains Minnow, *Hybognathus placitus* (Cyprinidae), in the Cimarron River, Oklahoma. The Southwestern Naturalist 33(1):27.
- Lyons, J., J. S. Stewart, and M. Mitro. 2010. Predicted effects of climate warming on the distribution of 50 stream fishes in Wisconsin, U.S.A. Journal of Fish Biology 77(8):1867-1898.

- Lytle, D. A., and N. L. Poff. 2004. Adaptation to natural flow regimes. *Trends in Ecology & Evolution* 19(2):94-100.
- Marras, S., A. Cucco, F. Antognarelli, E. Azzurro, M. Milazzo, M. Bariche, M. Butenschön, S. Kay, M. Di Bitettio, G. Quattrocchi, M. Sinerchia, and P. Domenici. 2015. Predicting the future thermal habitat suitability of competing native and invasive fish species: from metabolic scope to oceanographic modelling. *Conservation Physiology* 3(1): doi:10.1093/conphys/cou059.
- Mayor, S. J., D. C. Schneider, J. A. Schaefer, and S. P. Mahoney. 2009. Habitat selection at multiple scales. *Écoscience* 16(2):238-247.
- McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea. 2012. NHDPlus Version 2: User Guide. US Geological Survey.
- Martin, M.E., M.S. Delheimer, K.M. Moriarty, D.A. Early, K.A. Hamm, J.N. Pauli, T.L. McDonald, and P.N. Manley. 2022. Conservation of rare and cryptic species: challenges of uncertainty and opportunities for progress. *Conservation Science and Practice* 4:e12809.
- Midway, S. R., and B. K. Peoples. 2019. Effects of life-history traits on stream fish abundances across spatial scales. *Ecology of Freshwater Fish* 28(4):639–649.
- Murphy, P.J., T.J. Randle, L.M. Fotherby, and J.A. Daraio. 2004. The Platte River channel: history and restoration. U.S. Department of the Interior.
- Newbold, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. *Progress in Physical Geography* 34(1):3-22.
- Oksanen, J., G.L. Simpson, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, P. Solymos, M. Henry, H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Caceres, S. Durand, H.B. Antoniazzi Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M.O. Hill, L. Lahti, D. McGlinn, M. Ouellette, E.R. Cunha, T. Smith, A. Stier, C.J.F. Ter Braak, and J. Weedon. (2022). *Vegan: Community Ecology Package*. R package version 2.6-4.
- Osterhaus, D.M. and E.C. Martin. 2019. Trends in distribution of Plains Minnow (*Hybognathus placitus*) in Kansas from 1964 to 2017. *The American Midland Naturalist* 182(2):203-215.
- Paukert, C. P., E. R. Kleekamp, and R. W. Tingley. 2020. Identifying candidate reference reaches to assess the physical and biological integrity of wadeable streams in different ecoregions and among stream sizes. *Ecological Indicators* 111:e105966.
- Perkin, J. S., and K. B. Gido. 2011. Stream fragmentation thresholds for a reproductive guild of Great Plains fishes. *Fisheries* 36(8):371–383.

- Perkin, J. S., K. B. Gido, J. A. Falke, K. D. Fausch, H. Crockett, E. R. Johnson, and J. Sanderson. 2017. Groundwater declines are linked to changes in Great Plains stream fish assemblages. *Proceedings of the National Academy of Sciences* 114(28):7373–7378.
- Poff, N.L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society* 16(2):301-409.
- Poizat, G., and D. Pont. 1996. Multi-scale approach to species–habitat relationships: juvenile fish in a large river section. *Freshwater Biology* 36(3):611-622.
- Prato, E.P., C. Comoglio, and O. Calles. 2011. A simple management tool for planning the restoration of river longitudinal connectivity at watershed level: priority indices for fish passes.
- Pulliam, H. R. 1988. Sources, Sinks, and Population Regulation. *The American Naturalist* 132(5):652–661.
- Radinger, J. and C. Wolter. 2015. Disentangling the effects of habitat suitability, dispersal, and fragmentation on the distribution of river fishes. *Ecological Applications* 25(4):914-927.
- Radinger, J., F. Essl, F. Hölker, P. Horký, O. Slavík, and C. Wolter. 2017. The future distribution of river fish: The complex interplay of climate and land use changes, species dispersal and movement barriers. *Global Change Biology* 23(11):4970–4986.
- Rahel, F.J. and L.A. Thel. 2004. Flathead Chub (*Platygobio gracilis*): a technical conservation assessment. USDA Forest Service, Rocky Mountain Region.
- R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rees, D.E., R.J. Carr, and W.J. Miller. 2005. Plains Minnow (*Hybognathus placitus*): a technical conservation assessment. USDA Forest Service, Rocky Mountain Region.
- Rohr, J.M., S.J. Meiners, T.D. Thomas, and R.E. Colombo. 2021. Recovery of riverine fish assemblages after anthropogenic disturbances. *Ecosphere* 12(4):e03459.
- Schneider, R., K. Stoner, G. Steinauer, M. Panella, and M. Humpert. 2011. The Nebraska Natural Legacy Project: State Wildlife Action Plan. 2nd ed. The Nebraska Game and Parks Commission, Lincoln, NE.
- Schneider, R., M. Fritz, J. Jorgenson, S. Schainost, R. Simpson, G. Steinauer, and C. Rothe-Groleau. 2018. Revision of the tier 1 and 2 lists of species of greatest conservation need: a supplement to the Nebraska Natural Legacy Project State Wildlife Action Plan. The Nebraska Game and Parks Commission, Lincoln, NE.

- Schneiders, B. 1996. The myth of environmental management: the Corps, the Missouri River, and the channelization project. *Agricultural History* 70(2):337-350.
- Schwarz, G.E. and R.B., Alexander. 1995. Soils data for the conterminous United States derived from the NRCS State Soil Geographic (STATSGO) database. U.S. Geological Survey data release.
- Senecal, A.C., A.W. Walters, and W.A. Hubert. 2015. Historical data reveal fish assemblage shifts in an unregulated prairie river. *Ecosphere* 6(12):287. <http://dx.doi.org/10.1890/ES14-00361.1>.
- Sievert, N. A., C. P. Paukert, Y.-P. Tsang, and D. Infante. 2016. Development and assessment of indices to determine stream fish vulnerability to climate change and habitat alteration. *Ecological Indicators* 67:403–416.
- Spellerberg I.F. and P.J. Fedor. 2003. A tribute to Claude Shannon (1916-2001) and a plea for more rigorous use of species richness, species diversity, and the ‘Shannon-Wiener’ Index. *Global Ecology and Biogeography* 12:177-179.
- Staude, I. R., L. M. Navarro, and H. M. Pereira. 2020. Range size predicts the risk of local extinction from habitat loss. *Global Ecology and Biogeography* 29(1):16–25.
- Stoll, S., J. Kail, A. W. Lorenz, A. Sundermann, and P. Haase. 2014. The importance of the regional species pool, ecological species traits and local habitat conditions for the colonization of restored river reaches by fish. *PLoS ONE* 9(1):e84741.
- Stringer, A., and N. Clancy. 2020. Status of Northern Redbelly Dace (*Chrosomus eos*) in Montana. U.S. Forest Serv:4.
- Tang, J., S. Alelyani, and H. Liu. 2014. Feature selection for classification: A review. *Data Classification: Algorithms and Applications* 37:62.
- Tarkan, A.S., S.M. Marr, and F.G. Ekmekçi. 2015. Non-native and translocated fish species in Turkey. *Fishes in Mediterranean Environments* 3:1-28.
- Taylor, C. M., and R. J. Miller. 1990. Reproductive ecology and population structure of the Plains Minnow, *Hybognathus placitus* (Pisces: Cyprinidae), in Central Oklahoma. *American Midland Naturalist* 123(1):32.
- United States Department of Agriculture (USDA). 2019. Technical release 210-60: Earth dams and reservoirs.
- U.S. Geological Survey. 2022. National Hydrography Dataset v2.
- Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. R. Liermann, and P. M. Davies. 2010. Global threats to human water security and river biodiversity. *Nature* 467(7315):555–561.

- Walters, D. M., R. E. Zuellig, H. J. Crockett, J. F. Bruce, P. M. Lukacs, and R. M. Fitzpatrick. 2014. Barriers impede upstream spawning migration of Flathead Chub. *Transactions of the American Fisheries Society* 143(1):17–25.
- Wang, L., D. Infante, P. Esselman, A. Cooper, D. Wu, W. Taylor, D. Beard, G. Whelan, and A. Ostroff. 2011. A hierarchical spatial framework and database for the National River Fish Habitat Condition Assessment. *Fisheries* 36:436-499.
- Ward, J. V. 1998. Riverine landscapes: biodiversity patterns, disturbance regimes, and aquatic conservation. *Biological Conservation* 83(3):269–278.
- Whittier, T. R., D. B. Halliwell, and S. G. Paulsen. 1997. Cyprinid distributions in Northeast U.S.A. lakes: evidence of regional-scale minnow biodiversity losses. *Canadian Journal of Fisheries and Aquatic Sciences* 54:1593-1607.
- Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* 3(4):385.
- Winter, T.C. 2007. The role of ground water in generating streamflow in headwater areas and in maintaining base flow. *Journal of the American Water Resources Association* 43(1):15-25.
- Worthington, T. A., S. K. Brewer, T. B. Grabowski, and J. Mueller. 2014. Backcasting the decline of a vulnerable Great Plains reproductive ecotype: identifying threats and conservation priorities. *Global Change Biology* 20(1):89–102.
- Ziv, Y. 1998. The effect of habitat heterogeneity on species diversity patterns: a community-level approach using an object-oriented landscape simulation model (SHALOM). *Ecological Modelling* 111:135-170.

CHAPTER 3

AQUATIC HABITAT REPRESENTATION WITHIN NEBRASKA'S BIOLOGICALLY UNIQUE LANDSCAPES: IMPLICATIONS FOR STREAM CLASSIFICATION

Abstract

Stream classification informs freshwater conservation by providing a useful framework to account for natural habitat heterogeneity, establish ecologically relevant standards for comparison, and generalize ecological processes within larger spatial scales. Stream classification efforts are often based on landscape features and attempt to encompass the breadth of physical and chemical variability of streams across physiographic regions. A greater understanding of landscape-based classification approaches as means to classify site-scale stream habitat heterogeneity may improve understanding of drivers of biodiversity. We used site-level stream data collected by the Nebraska Department of Environment and Energy (NDEE) from 1995 to 2021 to assess stream habitat heterogeneity among Biologically Unique Landscapes (BUL) at a statewide spatial scale. Multivariate cluster analysis and statistical similarity tests were used to determine if aquatic habitats within the BULs are different from those outside and statewide and by which habitat characteristics. Silhouette values were also calculated using similarity-based clusters, level 3 ecoregions, hydrologically unique code (HUC) 12 watersheds, Strahler's stream order, matrix (i.e., dominant) landcover class, and BULs to provide a direct comparison of the classification power of the BUL construct to

commonly used classifiers. No significant differences were suggested between aquatic environmental conditions within BULs, outside BULs, and statewide. This implies that the BULs adequately capture habitat heterogeneity of selected metrics. Evidence for unique aquatic habitat conditions among BULs was limited, shown in minimal pairwise differences among BULs in statistical similarity tests, suggesting that the BULs may have minimal benefit in grouping their contained streams. Furthermore, the classificatory power of the BULs underperformed all other classifiers and, as such, BULs may not improve statewide stream classification. The similarity-based approach outperformed assessed landscape-based approaches, indicating that similarity-based stream classification may be considered in the development of standard stream classification systems in the future.

Introduction

Stream classifications provide frameworks to account for natural habitat heterogeneity, establish ecological standards for comparison, and generalize fine-scale processes across greater spatial scales (Hawkins and Vinson 2000; Mažeika et al. 2004; McManamay et al. 2018; McManamay and DeRolph 2019). Applications of stream classification systems may also serve conservation planning, direct environmental management programs, inform monitoring through reference sites, and reduce variables in model building (Wang et al. 2006; McManamay and DeRolph 2019). Streams contain variable physical templates driving biological processes (Vannote et al. 1980; Mažeika et al. 2004; McManamay et al. 2018). Broadly, stream classification systems group streams of similar physical characteristics and separate streams with different physical characteristics (Hawkins and Vinson 2000; McManamay et al. 2018). Stream

classifications are fundamental to understanding the diversity of stream habitat across large areas and the influence habitat heterogeneity has on biological community structure (Hawkins et al. 2000; Wang et al. 2006; McManamay and DeRolph 2019). An ideal classification system will account for environmental heterogeneity whereby shifts in observed and expected conditions will reveal potential system changes (Hawkins et al. 2000). As such, strong stream classifications may be derived from reference sites within a region of interest (Hawkins et al. 2000). A site is defined as a sampling unit of stream reaches with a minimum length of 150 meters and a maximum length of 300 meters (USEPA 2019; NDEE 2020). Reference sites often exhibit environmental conditions reminiscent of the most natural conditions possible given human alteration of the surrounding landscape (Hawkins et al. 2000). Classification systems focused on site-scale conditions in streams may properly represent habitat heterogeneity and inform conservation decision making (Nel et al. 2011; Melles et al. 2012). Studies have shown that classifications based on site-level environmental conditions outperform those based on landscapes in regard to within group homogeneity and between group heterogeneity (Van Sickle and Hughes 2000; Caruso 2014). Representing site-scale heterogeneity in both sampling and classification is necessary to recognize degradation and define freshwater conservation areas (Hawkins et al. 2000; McManamay et al. 2018; McManamay and DeRolph 2019).

Multiple classification approaches exist including those based on landscapes, stream order, and statistical similarity. Large-scale landscape-based classification systems are geographical in nature and group streams by climatic and geological features expected to drive similar biological community structures (Sokal 1974; Melles et al.

2012). Classification based on stream order groups streams by size and position in the drainage network assumed to be consistent with similar hydrological and geomorphological features (Vannote et al. 1980; Van Sickle and Hughes 2000; Melles et al. 2012; Caruso 2014). Similarity-based classifications are not often applied in fluvial ecology but offer an approach of grouping streams in a non-geographic context (McNeil et al. 2005). Classification systems generally function at single scales, depending on the management purposes leading towards their development. However, hierarchical (i.e., multiscale) classifications are increasing in popularity (Melles et al. 2012). Hierarchical classifications assume large-scale processes set constraints on small-scale processes, representing both multiple scales in nested groupings that may reflect the true hierarchical nature of aquatic systems (Melles et al. 2012).

Landscape-based classifications have been applied to group freshwater systems by boundaries present across the landscape, such as ecoregions or watersheds (Hawkins et al. 2000; Heatherly et al. 2014). Landscape-based classifications assume streams can be grouped by landscape features because stream habitat conditions are determined directly and indirectly by the surrounding terrestrial environment (Hynes 1975; Baxter and Hauer 2000; Hawkins et al. 2000; Saunders et al. 2002). Ecoregions have been used as a fundamental biogeographic classification unit for freshwater conservation wherein aquatic features are assumed to correspond with terrestrial features such as landforms, vegetation, soils, and land use (Hughes et al. 1994; Hawkins et al. 2000; Heatherly et al. 2014; Hughes et al. 2015). Hughes et al. (1994) provided evidence for the strength of ecoregion-based classification for aquatic environmental features like water quality and faunal assemblages (Hughes et al. 1994). Though ecoregion-based classification is useful

for its generality and inclusivity, finer-scale differentiation is often necessary to better represent the site-level heterogeneity of freshwater systems (Hawkins et al. 2000; Nel et al. 2011; Caruso 2014; Hermoso et al. 2015).

Inclusion of hydrological units at the watershed and sub-watershed scales within landscape-based classification may represent spatially explicit landscape influence on streams (Hawkins et al. 2000; Nel et al. 2011; Heatherly et al. 2014). Watershed-based classifications assume that stream physical attributes and ecological responses are influenced by properties of their catchments including geologic and hydrologic (Hynes 1975; Caruso 2014; Heatherly et al. 2014). In statistical assessments, watershed-based classification is weaker than ecoregion-based classification at larger spatial scales and stronger at smaller spatial scales. Causes of scale-dependent performance of watershed-based classification may include the lower degree of variation within smaller catchments compared to larger catchments (Hawkins et al. 2000). However, ecoregions and watersheds as landscape classifiers poorly explain differences among streams and their communities in practice (Hawkins et al. 2000; Hawkins and Vinson 2000; Wang et al. 2006; Heatherly et al. 2014). The statistical weakness of landscape-based classifications has been attributed to their coarseness as smaller-scale heterogeneity impacting biota is not accounted (Hawkins et al. 2000; Hawkins and Vinson 2000; Van Sickle and Hughes 2000). This may also explain the disparity in watershed classifiers of varying size where smaller watersheds may better represent site-level features compared to larger watersheds. Though the generality of large catchments and ecoregions are initially useful for large-scale classification, they are unlikely to provide fine-scale specificity for any

single ecosystem feature including stream physical conditions (Sokal 1974; Hawkins et al. 2000; Van Sickle and Hughes 2000).

Stream order serves as a classification framework and is positively related to biological community diversity and structure (Van Sickle and Hughes 2000; Caruso 2014). Stream order is often used to group streams of similar hydrology and geomorphology (Melles et al. 2012; Caruso 2014). Stream order classifies streams starting with headwaters serving as stream order 1 (Vannote et al. 1980; Frissell et al. 1986; Ward 1998). The confluence of two streams of the same order (i.e., two first-order streams) results in the formation of a larger-ordered stream downstream (i.e., second-order stream; Vannote et al. 1980; Ward 1998). Broadly, low-order headwaters are proposed to have narrower channels, higher inputs from allochthonous sources, lower temperatures, less discharge, macroinvertebrates adapted to feeding on coarse particulate organic matter (CPOM), and low fish community diversity (Vannote et al. 1980). Mid-order streams (4-6) are proposed to be larger, have more autochthonous energy input, a higher proportion of fine particulate organic matter (FPOM), warmer temperatures, more discharge, more macroinvertebrates adapted to feeding on FPOM, and the highest fish diversity (Vannote et al. 1980). High-order streams (7-9) are proposed to have the largest channel, more input from autochthonous sources, higher temperatures, greatest discharge, macroinvertebrates adapted to collecting FPOM, and reduced fish diversity (Vannote et al. 1980). While this pattern may not exist for all freshwater systems, classifications based on stream order generally assume that streams can be grouped by size and environmental conditions and follow the above longitudinal gradient (Vannote et al. 1980; McCormick et al. 2000). Stream classifications based on stream order have similar

or better statistical strength than that of large watersheds (McCormick et al. 2000; Van Sickle and Hughes 2000). However, McCormick et al. (2000) noted that stream classification based on stream order still underperformed classifiers based on taxonomic groups. As such, there may still be disparity between fine-scale environmental conditions captured by classifiers based on stream size and those experienced by stream biota (McCormick et al. 2000).

Similarity-based classification frameworks differ from landscape-based classification frameworks in that they are not reliant on physiographic constraints (Van Sickle and Hughes 2000; Heatherly et al. 2014). Similarity-based classification can be particularly useful as a framework for management decisions, where sites of similar conditions or quality are grouped for similar management (Van Sickle 1997; Van Sickle and Hughes 2000). Similarity-based approaches to classification may better represent site-level environmental conditions than solely landscape-based classification in classifying sites by similarity of observed conditions, rather than by assumed similarity by spatial proximity (Hughes et al. 1994; Van Sickle and Hughes 2000; Caruso 2014; Heatherly et al. 2014). Further, statistical assessment of classification strength at ecoregion, watershed, and stream order scales using site-level data is commonplace (Van Sickle 1997; Hawkins and Vinson 2000; McCormick et al. 2000; Van Sickle and Hughes 2000). An example of similarity-based classification is hierarchical cluster analysis, wherein statistical dissimilarity approaches are used to group sites into classes where similarity is maximized within groups and minimized among groups (Van Sickle and Hughes 2000; McNeil et al. 2005; Long et al. 2010). In landscape analyses, cluster analysis is considered spatially neutral because sites are clustered by quantified

characteristics including physical features (Van Sickle 1997; Van Sickle and Hughes 2000). Thus, integrating statistical methods such as cluster analysis may make for more robust and applicable classifications that are also based on spatial association.

The high degree of linear connectivity of drainage networks limits the effectiveness of landscape- and stream order-based conservation (Nel et al. 2011; Melles et al. 2012; Hermoso et al. 2015). Freshwater habitat is often incidentally included in protected terrestrially derived reserves, which may fail to properly address aquatic habitat heterogeneity (Saunders et al. 2002; Linke et al. 2011; Melles et al. 2012; Hermoso et al. 2015). Traditional conservation planning focused on classifying highly biodiverse terrestrial ecosystems and setting them aside as protected reserves (Sokal 1974; Van Sickle 1997; Fausch et al. 2002; Nel et al. 2011; Melles et al. 2012). Reserves containing protected ecosystems rarely overlap well with hydrological units to capture the entirety of catchments or sub-catchments, so that some segments of aquatic systems go unprotected or are downstream of unprotected waters (Nel et al. 2011; Melles et al. 2012; Hermoso et al. 2015). Thus, protected aquatic ecosystems are susceptible to changes from upstream, downstream, and surrounding landscapes, which may not fully protect entire aquatic ecosystems and promote biodiversity (Nel et al. 2011; Hermoso et al. 2015).

Stream classifications based on ecoregions and watersheds have been the predominant approaches used globally (Stoddard et al. 2008; Melles et al. 2012; Heatherly et al. 2014). This may be attributable to the claims of Hynes (1975), wherein stream ecosystems and their characteristics are determined by the valley (i.e., the landscape) through which they flow. This approach shifted emphasis from site-level stream characteristics to large-scale ecological processes (Nel et al. 2011; Melles et al.

2012). While some later approaches emphasized the importance of habitat heterogeneity, stochasticity, disturbance regimes, and hierarchical scaling of streams, many stream classification efforts continued to prioritize generalizations based on the influence of large-scale landscape features (Frissell et al. 1986; Townsend and Hildrew 1994; Melles et al. 2012). While ecoregions and watersheds are useful for classifying streams with regards vertebrate assemblages, landscape-based approaches may not be as effective as similarity-based approaches (Van Sickle 1997; Van Sickle and Hughes 2000).

Stream classification efforts within Nebraska, USA have been based on landscape features and fish communities (Heatherly et al. 2014). However, a more holistic approach to stream classification including larger biological communities may advance statewide aquatic conservation efforts. In 2005, the Nebraska Game and Parks Commission implemented such an approach with the Nebraska Natural Legacy Project to identify Biologically Unique Landscapes (henceforth BUL) that encompass a majority of Nebraska's biodiversity (Schneider et al. 2011). The Nebraska Natural Legacy Project addresses aquatic conservation by inclusion of freshwater species as components of BUL delineation. Each BUL is delineated using ecoregions and known occurrences of at-risk species and biological communities with the end goal of biodiversity preservation (Figure 2; Schneider et al. 2011). While the Nebraska Natural Legacy Project provides a strong framework for joint landscape- and species-based conservation, a greater understanding of the effectiveness of the BUL framework as a means to group heterogeneity in Nebraska's stream systems may aid the development of a standard statewide classification system. Further, the State Wildlife Action Plan states the development of a statewide classification system for freshwater habitats as a priority conservation

challenge (Schneider et al. 2011). An assessment of the BULs ability to classify site-level variation in aquatic habitat will determine their effectiveness as a form of stream classification. The objectives of this study were to 1) compare in-stream habitat conditions within the BULs to those outside and statewide, 2) assess habitat heterogeneity among BULs, and 3) compare classification strength of traditional classification systems to a similarity-based approach. I predict that a similarity-based approach to stream classification will better represent site-level habitat heterogeneity of stream environments across the state.

Methods

Data Collection

Site-level stream data collected by the Nebraska Department of Environment and Energy (NDEE) from 1995 to 2021 were used to assess stream habitat heterogeneity among BULs (NDEE 2019). There were 564 useable sites collected as part of both the Stream Biological Monitoring Program (SBMP) and National Rivers and Streams Assessment (NRSA; Figure 1; Paulsen et al. 2008; NDEE 2019). SBMP sampling follows a roving survey design, wherein 33-40 randomly selected sites are sampled in 1-3 river basins annually (NDEE 2019). During a six-year cycle, all 13 major river drainage basins in Nebraska are sampled (NDEE 2019). For instance, 61 sites were sampled in the summers of 2018 and 2019 and covered each major drainage basin in Nebraska (NDEE 2019). Physical and chemical characteristics were measured across 11 sampling transects subdividing 200-meter sites using United States Environmental Protection Agency (US EPA) standardized methods for either wadeable or navigable streams (NDEE 2020). For

the purposes of this study, NDEE sites were assigned to the BULs in which they occur in ArcGIS Pro (ESRI Inc. 2022).

Data Analysis

Variable reduction improves statistical analyses by mitigating irrelevant or redundant data (Tang et al. 2014; Borboudakis and Tsamardinos 2019). The original 216 environmental variables present in the NDEE data were reduced via correlation analysis and the remaining variables were used to represent the physical conditions of sampled rivers and streams in a multivariate cluster analysis of Nebraska's BULs (Table 1; Figure 2; Tang et al. 2014). For pairs of variables with Pearson correlation of $r \geq 0.7$, the variable with lower correlation to other variables in the dataset was selected.

Furthermore, environmental variables that may not have been highly correlated but provide similar information were selected based on prevalence or descriptiveness. For example, despite a Pearson correlation of only $r = 0.56$ between incision height and incision width/depth ratio, only incision width/depth ratio was retained as it is more descriptive (Tang et al. 2014; Borboudakis and Tsamardinos 2019). Furthermore, an exception was made for the sinuosity index and habitat quality variables because their high correlation is due to biases in data where any site missing one value is missing the other. Selected environmental variables aim to represent thermal variability, hydrologic regimes, geomorphology, and water quality as these variable types may be key drivers of aquatic ecosystem structure at multiple spatial scales (Frissell et al. 1986; Durance et al. 2006). Prior to analyses, sites missing large amounts of data or with unexpectedly extreme values of variables were dropped. For example, sites on the mainstem Platte River are expected outliers in some metrics, having naturally larger width-to-depth ratios

and discharges, while a low-order tributary with a comparable value was dropped as recorded metrics may be erroneous.

Three additional metrics were calculated to supplement the features collected by NDEE and introduce multiscale variables for analyses. Groundwater depth for wells across Nebraska was acquired from the University of Nebraska-Lincoln School of Natural Resource's Conservation and Survey Division (CSD) interactive data map (CSD 2023). As measurements ranged from the early 20th century to present day, groundwater depth was averaged across years to generate a mean annual depth to groundwater (MADTW) variable. Two land-use metrics including dominant landcover class and Shannon's diversity index were generated within 12-digit hydrological units (HUC12) using landcover data retrieved from the National Land Cover Database (Dewitz et al. 2021). Pixel counts of landcover classes within HUC12s were generated using the summarize raster tool in ArcGIS Pro (ESRI Inc. 2022). The class of highest proportional cover (i.e., most pixels) was noted as a categorical variable of dominant, or matrix, landcover class (Meshesha et al. 2016). For example, in a HUC12 with majority agricultural land use any stream would have cropland assigned as dominant landcover. Shannon's diversity index was calculated for landcover classes within HUC12s using vegan package in Program R (R Core Team 2021; Okansen et al. 2022).

Multivariate Cluster Analysis

Numerical data were standardized to zero mean and unit variance using:

$$z_{if} = \frac{x_{if} - m_f}{s_f},$$

where z_{if} is the standardized value for observation i of variable f , x_{if} is the observed value, m_f is the mean of variable f , and s_f is the standard deviation of variable f (Long et al. 2010). The environmental data were then broken into three subsets to determine if stream habitat within the BULs carries similar trends in variation to those observed outside the BULs and statewide. The subsets included 230 sites within BULs, 334 sites outside BULs, and all 564 sampled sites. Using variation as a proxy for heterogeneity, the BUL construct would encompass adequate heterogeneity if variation between the three subsets is similar, allowing for stream habitat within the BULs to represent stream habitat statewide. When comparing sites in BULs to sites outside BULs, similar variation indicates that heterogeneity outside of BULs is not different from that within.

Furthermore, when comparing sites in BULs to sites statewide, similar variation indicates statewide trends in heterogeneity are captured in the BUL construct. Another necessity for cluster analysis is the calculation of a measure of separation, in this case Euclidean distance as all selected environmental data are on a ratio or interval scale (Kaufman and Rosseeuw 1987; Long et al. 2010). Euclidean distance, given by:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2},$$

measures the distance between two points in multiple dimensions (i.e., variables; Kaufman and Rosseeuw 1987). A Euclidean distance matrix was generated between all objects (i.e., sampling sites) within each subset using the *factoextra* package in Program R (Kassambara and Mundt 2020; R Core Team 2021).

Data were hierarchically clustered using Ward's minimum variance method in the *factoextra* package in Program R (Ward 1963; Kaufman and Rosseeuw 1987; Kassambara and Mundt 2020; R Core Team 2021). Ward's minimum variance method is

commonly used in cluster analysis because it is computationally efficient for larger datasets, functions particularly well in multivariate space, and produces more homogenous clusters than other methods (Ward 1963; Kaufman and Rosseeuw 1987; Murtagh 2014). Ward's method begins by placing each datum object in its own cluster; clusters are then merged stepwise in a manner minimizing the within-cluster variance of the resulting cluster (Ward 1963; Kaufman and Rosseeuw 1987; Murtagh 2014). An output cluster dendrogram visualizing the hierarchical distance between objects is then cut to the appropriate height to produce the selected number of clusters, in this case two clusters for each subset. For the statewide sites subset the dendrogram was cut to a height of 500, whereas the inside and outside BULs subsets were cut to heights of 300 (Figure 4, Figure 5, Figure 6). Descriptive statistics were then generated for the resulting clusters by their contained sampling sites. Cluster means and standard deviations were calculated to observe variation in physical conditions of streams across Nebraska, as well as cluster-defining differences.

An optimal number of clusters for each subset was determined from the distance matrix using a silhouette method in the *cluster* and *factoextra* packages in Program R (Kassambara and Mundt 2020; R Core Team 2021; Maechler et al. 2022). A silhouette value illustrates how similar an object is to its own cluster when compared to other clusters on a -1 to +1 scale, with high values denoting more similarity to the objects resident cluster (Kaufman and Rosseeuw 1987; Rosseeuw 1987). A silhouette value was calculated for each object i by:

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}},$$

where $a(i)$ is a measure of within cluster dissimilarity, $b(i)$ is a measure of between cluster similarity, and $\max\{a(i), b(i)\}$ is the observed maximum dissimilarity for object i either within or between clusters (Rosseeuw 1987). An average silhouette value was calculated for theoretical clustering in up to 10 groups to determine an optimal clustering scheme (i.e., that with an average silhouette value closest to 1). 2-group clustering was selected for each subset, as further generated clusters of exceedingly low membership and greatly reduced the average silhouette width (Figure 3). Silhouette values were calculated for every site using similarity-based clusters, level 3 ecoregions, HUC12 watersheds, stream order, matrix (i.e., dominant) landcover, and BULs as classification systems. This allowed for direct comparison of the classification power of the BUL construct to commonly used classifiers to address objective 3 (Wang et al. 2017).

Similarity Analysis

Similarity analyses were conducted among site groupings (i.e., inside BULs, outside BULs, and statewide) to determine if stream habitat conditions within the BULs accurately represent those found across Nebraska. A Kruskal-Wallis test is a non-parametric test for similarity between multiple groups. Kruskal-Wallis tests work well when data does not fit a normal distribution. Kruskal-Wallis tests begin by ranking all data, ignoring group membership. The test statistic is given by:

$$H = (N - 1) \frac{\sum_{i=1}^g n_i (\bar{r}_i - \bar{r})^2}{\sum_{i=1}^g \sum_{j=1}^{n_i} (r_{ij} - \bar{r})^2},$$

where N is the total number of observations, g is the number of groups, n_i is the number of observations in group i , r_{ij} is the rank of observation j in group i , \bar{r}_i is the average rank of observations in group i , and \bar{r} is the average of all r_{ij} (Kruskal and Wallis 1952; Dinno

2015). A null hypothesis of no difference between groups is then rejected or accepted based on the test statistic H . Kruskal-Wallis tests were run between subsets and clusters of different subsets, where rejection of the null hypothesis is warranted at $H \geq 4.605$; $df = 2$. To address objective 1, subsets were tested for similarity by their means and standard deviations to determine if habitat characteristics were different within BULs, outside BULs, and statewide. To address objective 2, a Kruskal-Wallis test was used to determine if there were differences among mean environmental features of BULs.

Kruskal-Wallis tests only denote whether there is difference between any groups in a dataset and further tests must be applied to determine pairwise group differences. Dunn's test was run among the 31 analyzed BULs to determine if aquatic habitat among them is unique. Like the Kruskal-Wallis test, Dunn's test is a non-parametric pairwise multiple comparisons procedure based on rank sums (Dunn 1964; Dinno 2015). The test statistic of Dunn's test is calculated with:

$$Z_i = \frac{W_i - W_j}{\sqrt{\left(\frac{N(N+1)}{12}\right) - (\sum T_s^3 - T_s)/(12(N-1)) / ((1-n_i) + (1-n_j))}},$$

where W_i and W_j are the average sum of ranks for groups i and j respectively, N is the total number of observations, T_s is the number of ties (data points in both groups), and n_i and n_j are the numbers of objects in groups i and j respectively (Dunn 1964; Dinno 2015). Positive Z -values indicate the first group i has higher mean values than the second group j , while negative Z -values indicate the opposite. Further, the $|Z|$ -value relative to all comparisons indicates the overall strength of the difference. To account for the greater chance of type-1 errors in multiple comparisons, p -values for pairwise comparisons were adjusted using the Benjamini-Hochberg method (Dinno 2015). For both the Kruskal-

Wallis and Dunn's tests, a null hypothesis of no statistical differences between groups was tested at $p \leq 0.10$.

Levene's tests for homogeneity of variance were conducted post-clustering to determine which environmental variables vary differently within BULs, outside BULs, statewide, and among BULs. Levene's test follows the Kruskal-Wallis and Dunn's tests to further examine differences by determining features of heterogeneous variation. Levene's test was chosen over the alternative Bartlett test because our numerical data does not fit a normal distribution (Levene 1960; Brown and Forsythe 1974). In this case, Levene's test was conducted using the median of each variable because doing so retains statistical power while performing better than the mean for non-normal data (Brown and Forsythe 1974; Gastwirth et al. 2009). Levene's test assesses the homogeneity of variance between groups of data, in this case subsets, clusters, and BULs in separate tests. Testing a null hypothesis of equal variance against an alternative hypothesis of unequal variance, significant difference between clustering schemes is achieved at $p \leq 0.10$. The test statistic of Levene's test is defined as:

$$W = \frac{(N-k)}{(k-1)} \cdot \frac{\sum_{i=1}^k N_i (Z_i - Z_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{N_i} N_i (Z_{ij} - Z_i)^2},$$

where k is the number of groups, N_i is the number of observations in group i , N is the total number of observations, Z_{ij} is the absolute deviation of observation j in group i , Z_i is the mean absolute deviation for group i , and $Z_{..}$ is the mean of all absolute deviations (Z_{ij} ; Levene 1960; Brown and Forsythe 1974; Gastwirth et al. 2009). The end result produces a value for both p and W denoting the homogeneity of variance by variable among tested groups (Levene 1960; Brown and Forsythe 1974; Gastwirth et al. 2009). In Levene's tests

between clusters of subsets, significant findings denote which variables are useful in grouping sites. In Levene's tests between subsets, these values provide insight into environmental heterogeneity differences of streams within BULs, outside of BULs, and statewide.

Results

Summary of Similarity-based Clusters

The sites inside BULs clustered into two groups. Cluster 1 of the BUL sites subset has variable means close to the overall mean (Figure 4). Cluster 2 of the BUL sites has discharge, woody riparian coverage, conductivity, and wetted width to thalweg depth ratio greater than the overall means; while bank angle, turbidity, incision width to depth ratio, riparian zone width, and depth to groundwater are below the overall means (Figure 4). A Levene's test between clusters of the BUL sites subset quantifies the differences wherein discharge, wetted width to thalweg depth ratio, conductivity, substrate diameter, and depth to groundwater have heterogenous variance between clusters at $p \leq 0.10$ and $W \geq 2.72$. All other environmental variables have homogenous variance between clusters. Conductivity, discharge, and wetted width to thalweg depth ratio differentiate streams within the BULs as these variables have the highest degree of heterogenous variance.

The sites outside BULs clustered best into two groups. Cluster 1 was characterized by environmental variable means generally close to the overall means (Figure 5). Cluster 2 is differentiated by woody riparian coverage, no riparian coverage, pH, turbidity, incision width to depth ratio, and depth to groundwater above the overall means and bank angle, percent fish cover, and HUC12 landcover diversity below the

overall means (Figure 5). A Levene's test between clusters of the outside BULs subset quantifies these differences where percent fish cover, percent organic substrate, depth to groundwater, no riparian coverage, dissolved oxygen, turbidity, incision width to depth ratio, woody riparian coverage, slope, discharge, riparian zone width, substrate diameter, temperature, pH, and sinuosity have heterogenous variance between clusters at $p \leq 0.10$ and $W \geq 2.72$. All other environmental variables have homogenous variance between clusters. Especially of note, percent fish refugia, percent organic substrate, depth to groundwater, no riparian coverage, dissolved oxygen, turbidity, incision width to depth ratio, woody riparian coverage, slope, and discharge differentiate streams outside of BULs as these variables had the highest degree of heterogenous variance.

The statewide sites clustered into two groups. Cluster 1 of the statewide subset has variable values close to the overall means (Figure 6). In contrast, Cluster 2 has discharge, woody riparian coverage, no riparian coverage, conductivity, turbidity, and wetted width to thalweg depth ratio higher than the overall means, whereas annual mean flood frequency and depth to groundwater are lower than the overall means (Figure 6). In particular, mean discharge and wetted width to thalweg depth ratio are much higher in cluster 2 than cluster 1. A Levene's test between clusters of the statewide subset quantifies differences where discharge, wetted width to thalweg depth ratio, conductivity, no riparian coverage, turbidity, HUC12 landcover diversity, riparian zone width, and depth to groundwater have heterogenous variance between clusters at $p \leq 0.10$ and $W \geq 2.71$. All other environmental variables have homogenous variance between clusters. Especially of note, discharge, conductivity, and wetted width to thalweg depth ratio differentiate streams statewide as these variables have the highest degree of heterogenous

variance. As such, sites statewide appear to be classifying by discharge, conductivity, and wetted width to thalweg depth ratio. Thus, statewide sites appear to vary by the same metrics as sites within BULs, suggesting that statewide stream habitat heterogeneity trends are present within the BUL construct.

Comparing BULs to Statewide Variation

Kruskal-Wallis tests were used to further quantify difference among subsets, addressing the first objective. The Kruskal-Wallis test between the BUL and outside BUL subsets did not achieve significance at $p = 0.81$, meaning there is no overall significant difference between environmental features within and outside the BULs. As such, a post-hoc Dunn test among clusters of both subsets was not performed. A Levene's test was used to determine if there are differences in the variation of specific environmental features that were eclipsed in the overall comparison of the Kruskal-Wallis test. In the Levene's test comparing the BUL sites subset and the sites outside BULs subset, significant difference ($p \leq 0.10$, $W \geq 2.71$) in variance was achieved for conductivity ($p = 0.05$), dissolved oxygen ($p < 0.01$), discharge ($p < 0.01$), groundwater depth ($p = 0.01$), percent organic matter in substrate ($p = 0.07$), incision width to depth ratio ($p = 0.03$), mean substrate diameter ($p < 0.01$), mean bank angle ($p = 0.05$), and wetted width to thalweg depth ratio ($p < 0.01$; Table 4). Thus, the variances of these metrics are heterogenous between sites within BULs and sites outside BULs, indicating that there is heterogeneity not captured by the BUL construct for those aforementioned environmental variables. Overall, the Kruskal-Wallis test and Levene's test suggest that stream habitat characteristics within BULs are not different from those outside the BULs.

The comparison between statewide sites and BUL sites is meant to determine if BULs adequately contain statewide heterogeneity. The Kruskal-Wallis test between the BUL and statewide subsets did not achieve significance at $p = 0.62$, meaning there is no overall significant difference between environmental features inside the BULs and statewide. This was followed with a Levene's test to determine if there are differences in the variation of specific environmental features. The Levene's test between the BUL sites subset and the statewide sites subset showed significant differences ($p \leq 0.10$, $W \geq 2.71$) in the variance of dissolved oxygen ($p = 0.04$), discharge ($p = 0.02$), groundwater depth ($p = 0.09$), mean substrate diameter ($p = 0.04$), and wetted width to thalweg depth ratio ($p = 0.05$; Table 5). Thus, the variances of these metrics are heterogeneous between sites within BULs and sites statewide, indicating that there is statewide heterogeneity not captured by the BUL construct. A significant difference was not achieved for all other environmental variables, thus variance is homogeneous between BUL sites and statewide sites and heterogeneity is encompassed.

Variation among BULs

The Kruskal-Wallis test suggested differences among the aquatic habitat of all BULs ($p = 0.01$). In the post-hoc Dunn test for pairwise differences between BULs, the only significant comparisons at $p \leq 0.10$ are the Lower Loup River (LLR) and Verdigris-Brazile (VBZ; $Z = -3.39$, $p = 0.08$) and LLR and the Middle Niobrara (MNI; $Z = -3.23$, $p = 0.09$; Figure 7, Figure 8). Negative Z -values indicate that LLR has generally lower means than both VBZ and MNI, while the magnitude of both as $|Z|$ -values are among the highest of all comparisons (Figure 8). While no other comparisons reached significance, many approached $p = 0.10$ such as the Central Platte River (CPR) and Elkhorn River

Headwaters (ERH; $Z = 2.72$, $p = 0.11$; Figure 7, Figure 8). However, a positive Z -value indicates that CPR has higher variable means than ERH, while the magnitude of its $|Z|$ -value put it relatively high in comparison strength. Comparisons with a high p -value and nearing $Z = 0$ may also indicate BULs with similar conditions. For example, ERH and the Kimball Grasslands (KGL $Z = 0.003$, $p = 0.50$; Figure 7, Figure 8) may have largely similar environmental conditions. The Levene's test comparing all BULs suggested difference ($p \leq 0.10$, $W \geq 1.38$) in variance for conductivity ($p < 0.01$), dissolved oxygen ($p < 0.01$), discharge ($p < 0.01$), groundwater depth ($p = 0.05$), percent organic matter in substrate ($p < 0.01$), incision width to depth ratio ($p < 0.01$), mean substrate diameter ($p < 0.01$), percent fish cover ($p < 0.01$), and wetted width to thalweg depth ratio ($p < 0.01$; Table 6).

Comparing Stream Classification Systems

Addressing objective 3, similarity-based clusters were the strongest classification system with the average silhouette value closest to 1 of $\bar{s}(i) = 0.33$ (Table 7, Figure 9). The only other classification system with a positive average silhouette value is matrix landcover with $\bar{s}(i) = 0.013586$ (Table 7, Figure 9). Ecoregion is the strongest classification system, of those with negative average silhouette values at $\bar{s}(i) = -0.01$ (Table 7, Figure 9). Strahler's stream order underperformed ecoregions at $\bar{s}(i) = -0.03$ (Table 7, Figure 9). The second worst performing classification system was HUC12 watersheds at $\bar{s}(i) = -0.16$ (Table 7, Figure 9). Finally, BULs underperformed all other classifiers of aquatic habitat conditions at $\bar{s}(i) = -0.22$ (Table 7, Figure 9).

Discussion

Heterogenous variance was noted for dissolved oxygen, discharge, mean substrate diameter, and wetted width to thalweg depth ratio among the inside BUL, outside BUL, and statewide subsets. As such, these features may be useful in grouping sites statewide or indicating differences in impactful environmental processes. For example, mean substrate diameter and wetted width to thalweg depth ratio can be seen as representing geomorphological processes, while discharge may represent hydrological processes. The use of these features would account for the impact of these processes on site-level environmental variation, approaching stream classification in a manner addressing fine-scale heterogeneity. Such an approach may be reminiscent of Rosgen's stream classification system in classifying streams by breaks in the gradient of these highly variable features (Rosgen 1994; Melles et al. 2012). Taking a discharge-based approach as an example, a split point could be determined of either high or low discharge and sites may then be grouped by their discharge (Rosgen 1994). Rosgen's stream classification approach has been criticized for neglecting to consider the surrounding catchment and stochasticity of feature variation across space and time (Juracek and Fitzpatrick 2003; Gordon et al. 2004). However, it could also be thought that certain features represent processes operating at higher scales, such as catchments, and address such criticisms (Van Sickle and Hughes 2000; Hawkins et al. 2000; Paulsen et al. 2008).

Defining references for stream quality statewide may benefit from assessment of environmental features exhibiting limited variation. Extreme values for less variable features could be indicative of good or poor stream health (Stoddard et al. 2008; Melles et al. 2012; Heatherly et al. 2014). For example, in the Levene's tests sinuosity index

typically exhibited low heterogeneity in each subset shown in high p values that indicate greater homogeneity. Thus, comparing the sinuosity of sites or incorporating sinuosity in habitat quality metrics may serve to establish reference sites (i.e., those with a relatively high sinuosity index) and imperiled sites in need of recuperative management (i.e., those with a relatively low sinuosity index; Stoddard et al. 2008; Heatherly et al. 2014). This framework could be applied to any low heterogeneity variable, given it is recorded at enough sites to characterize statewide variation. However, the scale and intensity of data collection limits the applicability of such a comparative framework. For instance, turbidity was missing measurements at 65 of 564 sites. Robust sampling and classification with respect to underlying processes can properly account for challenges of representing freshwater ecosystems in conservation planning. Stream classification systems for temperature regimes and geomorphological features are uncommon, even more so are classification systems combining multiple habitat features (McManamay et al. 2018). This is likely due to the rarity of data on physical habitat features or systematic standardized collection of them (McManamay et al. 2018). Sampling protocols occurring frequently over time may account for temporal heterogeneity in stream conditions (Mažeika et al. 2004). Spatially balanced and high-resolution standardized sampling provides the most representative data inputs for analyzing riverscape heterogeneity at large spatial scales (Mažeika et al. 2004; McManamay et al. 2018). Such sampling efforts are essential to account for fine scale heterogeneity of streams within classification systems (Mažeika et al. 2004; McManamay et al. 2018). The classificatory strength of sampled environmental features could be expanded upon in exploratory modeling, where

features could be tested as predictors of a measure of habitat quality (Nijboer and Verndonschot 2004; Jorgenson et al. 2009).

Site-scale aquatic habitat conditions within BULs vary similarly with those outside BULs and statewide, suggesting that the BUL construct contains adequate statewide heterogeneity in the assessed metrics. This bodes well for the role of BULs in conservation and management decisions, where their contained landscapes and biological communities can serve as a basis for decision making without concern for missing unique aquatic habitat statewide (Van Sickle and Hughes 2000). Ultimately, conditions within BULs are not unique from conditions outside BULs and carry the same variation trends as statewide from an aquatic habitat perspective. However, the similarity of conditions within and outside BULs calls into question the uniqueness of aquatic habitat conditions among BULs. While similarity-based groups generated through cluster analysis are assumed to be unique from one another due to the clustering process, this may not necessarily be the case. In Dunn test comparisons of clusters, the two clusters of any subset approached but did not achieve significant difference. This may indicate a statewide lack of heterogeneity in assessed metrics, likely as a result of environmental homogenization from land use conversion (Cardinale et al. 2002).

Site-scale aquatic habitat was not unique among BULs, indicating that aquatic habitat conditions may be similar among BULs. In an ideal classification system, between-group similarity is minimized so that desirable management and conservation actions are applied to the correct areas (Hawkins et al. 2000; Melles et al. 2012). It could be thought that the unique biological communities from which BULs are delineated reflect unique terrestrial environments (Schneider et al. 2011). As river and stream

conditions are known to be influenced by the terrestrial environments surrounding them, it logically follows that aquatic environment conditions within BULs are unique as well (Hynes 1975; Frissell 1986; Fausch et al. 2002). However, few unique aquatic environments existed among BULs (Figure 7, Figure 8). Thus, as it pertains to site-level environmental features in Nebraska, there is a disconnect or disruption within the representation of environmental processes in their classification. However, the BULs contain unique proportions of stream orders, thus further study relating stream order to in-stream habitat conditions may reveal unique habitat structure among the BULs (Figure 10). The high degree of anthropogenic land use alterations within the BULs may homogenize aquatic habitats that would have been unique under natural conditions (Hawkins et al. 2000; Schneider et al. 2011). This hypothesis is supported by the relatively high average silhouette value of the matrix landcover classification, though a specific study would better reveal such dynamics (Figure 9). Furthermore, as some BULs overlap or are in close spatial proximity to one another, the BULs may install boundaries for aquatic systems where there are none or poorly overlap with those that do exist (Figure 1). This is a consistent challenge to landscape-based classification systems that may be overcome in similarity-based or hierarchical classification systems (Nel et al. 2011; Melles et al. 2012; Hermoso et al. 2015).

The current BUL classification system underperforms all other schemes (i.e., ecoregion, HUC12 watershed, stream order, and dominant landcover) when comparing average silhouette values. As such, BULs may not improve stream classification in the face of the better performing traditional methods. However, all traditional classification systems underperformed given the generally negative average silhouette values (Table 6).

This may be expected as many traditional classification systems are landscape-based and the applicability of such classification systems has been called into question before (Hawkins 2000; Hawkins and Vinson 2000; Van Sickle and Hughes 2000). For classification systems with more groups, such as the BULs and HUC12 watersheds, data resolution becomes a concern with some groups having few or no sampled sites (Table 2; Van Sickle 1997; Hawkins and Vinson 2000). Low membership leads to poor characterization of conditions within these groups, while potentially inflating the weakness of their classification (Van Sickle 1997; Hawkins and Vinson 2000). Otherwise, the underperformance of landscape-based classification systems may indicate that while heterogeneity of aquatic habitat conditions is spatially contained, it is not well grouped by the BUL construct and other landscape-based classification systems.

Fine-scale watersheds, such as HUC12, have been found to outperform both ecoregions and stream order as a stream classification system in many past assessments of physiographic classifiers (Hawkins et al. 2000; Van Sickle and Hughes 2000). However, HUC12 watersheds had a lower average silhouette depth than both ecoregions and stream order in this study (Table 7). This may also be a result of limited data resolution, wherein the average silhouette value of the HUC12 classification system is limited by a silhouette value of $s(i) = 0$ for any HUC12 with only 1 sampled site (Van Sickle 1997; Hawkins and Vinson 2000). Nonetheless, as the majority of sites have negative silhouette values in their watershed, site-level environmental conditions are poorly classified by HUC12 watersheds (Figure 9). Thus, the most heterogeneous environmental features, dissolved oxygen, discharge, mean substrate diameter, and wetted width to thalweg depth ratio, may vary on larger scales than HUC12 watersheds.

This study provides an assessment of the classificatory power of traditional and novel stream classification systems. This sets groundwork for the establishment of a standard for stream classification statewide while opening discussion for how stream classification is thought about. Statistically generated clusters outperform all other classification systems with the highest average silhouette value. As such, cluster analysis of reach-scale data offers a similarity-based alternative to the landscape- and biological community-based classifications currently used in Nebraska. Ecoregions and HUC12s are geographically constrained, so similarity occurring within and between their groups is spatially linked (Hughes et al. 1994; Hawkins et al. 2000; Hawkins and Vinson 2000). However, statistical similarity groups sites regardless of spatial similarity (Van Sickle 1997; Van Sickle and Hughes 2000). While spatially similar sites may still be similar in a statistical framework, this may not always be the case. Consider a pair of streams in the same HUC12 watershed, one highly altered and one unaltered. Under a landscape-based classification they would be grouped together, while a statistical similarity-based classification would separate them based on reach-scale differences. Such a scenario is likely in the agriculturally dominated landscapes of Nebraska, thus integrating similarity-based classification with traditional classification systems may bolster stream classification statewide. While applying traditional classification methods like ecoregions, HUC12 watersheds, and stream order is still encouraged, their application should consider supplementation with statistical similarity to better address site-level heterogeneity (Hawkins et al. 2000; Van Sickle and Hughes 2000; Nel et al. 2011). Furthermore, multiple traditional and novel classification systems may be combined to generate a more robust hierarchical classification (Nel et al. 2011; Melles et al. 2012).

Tables

Table 3.1. Reduced variables from NDEE sampling sites from 1995-2021 and their definitions. These variables were used in multivariate cluster analysis of sites across Nebraska.

| Variable | Definition | Unit |
|-----------------|---|---------------------|
| Dischrg | Discharge | m ³ /sec |
| Temp | Water temperature | °C |
| XCMGW | Riparian woody vegetation canopy+mid+ground mean cover | % |
| xgb | Mean riparian barren ground | % |
| xbka | Mean bank angle | |
| Cndctvt | Conductivity | µs/cm |
| pH | pH | |
| DO | Dissolved oxygen | ppm |
| Turb | Turbidity | NTU |
| Habscore5 | Habitat quality index, derived from multiple hydrological and geomorphological measurements | |
| inciswd | Incision width/depth ratio | |
| xslope | Mean stream slope | m/m |
| xwd_rat | Mean wetted width/thalweg depth ratio | m |
| X_HALL | Mean riparian zone width, all vegetation types | m |
| pfc_all | % any type fish coverage present | % |
| lsb_dmm | log10 transformed mean substrate diameter | mm |
| sinu | Sinuosity index | |
| SHDI | Shannon's diversity index for landcover classes in HUC12 | |
| OM | Organic matter in substrate | % |
| AFLDFREQ | Annual flood frequency (1 = frequent (>50% chance), 2 = occasional (5-50% chance), 3 = rare (<5% chance)) | |
| MADTW | Annual mean depth to groundwater | m |

Table 3.2. NDEE sampling site count by BUL. BULs with 0 sites were excluded from analyses. Abbreviations for the BUL names used in later figures appear in parentheses after the full names.

| BUL | N |
|--------------------------------|----------|
| Central Loess Hills (CLS) | 6 |
| Central Platte River (CPR) | 5 |
| Cherry County Wetlands (CCW) | 12 |
| Elkhorn Confluence (EHC) | 1 |
| Elkhorn River Headwaters (ERH) | 6 |
| Indian Cave Bluffs (ICB) | 1 |
| Keya Paha (KPA) | 4 |
| Kimball Grasslands (KGL) | 2 |
| Loess Canyons (LOC) | 1 |
| Lower Loup Rivers (LLR) | 5 |
| Lower Niobrara River (LNR) | 3 |
| Lower Platte River (LPR) | 6 |
| Middle Niobrara (MNI) | 9 |
| Missouri River (MOR) | 2 |
| North Platte River (NPR) | 11 |
| Oglala Grasslands (OGL) | 5 |
| Panhandle Prairies (PHP) | 1 |
| Pine Ridge (PIR) | 12 |
| Platte Confluence (PLC) | 7 |
| Ponca Bluffs (PCB) | 4 |
| Rainwater Basin (RWB) | 26 |
| Saline Wetlands (SWL) | 17 |
| Sandhills Alkaline Lakes (SAL) | 1 |
| Sandsage Prairie (SGP) | 7 |
| Sandstone Prairies (STP) | 12 |
| Snake River (SNR) | 1 |
| Southeast Prairies (SEP) | 17 |
| Thurston-Dakota Bluffs (TDB) | 1 |
| Upper Loup Rivers (ULR) | 22 |
| Upper Niobrara River (UNR) | 8 |
| Verdigris-Bazile (VBZ) | 15 |
| Willow Creek Prairies | 0 |
| Rulo Bluffs | 0 |
| Dismal River Headwaters | 0 |
| Wildcat Hills | 0 |

Table 3.3. Dunn test results for pairwise differences between similarity-based clusters. B denotes clusters generated from the BUL sites, N denotes clusters generated from the sites outside BULs, and A denotes clusters generated from the statewide subset. An asterisk denotes significance achieved at $p \leq 0.10$.

| Comparison | <i>p</i> |
|-------------------|-----------------|
| A1 - A2 | 0.11 |
| A1 - B1 | 0.24 |
| A2 - B1 | 0.07* |
| A1 - B2 | 0.21 |
| A2 - B2 | 0.46 |
| B1 - B2 | 0.16 |
| A1 - N1 | 0.43 |
| A2 - N1 | 0.12 |
| B1 - N1 | 0.23 |
| B2 - N1 | 0.23 |
| A1 - N2 | 0.11 |
| A2 - N2 | 0.30 |
| B1 - N2 | 0.06* |
| B2 - N2 | 0.39 |
| N1 - N2 | 0.15 |

Table 3.4. Levene's test results for homogeneity of variance among sites within BULs and sites outside BULs. An asterisk denotes significance achieved at $p \leq 0.10$ and $W \geq 2.72$. See Table 3.1 for variable definitions.

| Variable | <i>W</i> | <i>p</i> |
|-----------------|-----------------|-----------------|
| AFLDFREQ | 0.36 | 0.55 |
| Cndctvt | 3.89 | 0.05* |
| DO | 9.75 | <0.01* |
| Dischrg | 16.34 | <0.01* |
| Habscore5 | 1.18 | 0.28 |
| MADTW | 6.11 | 0.01* |
| OM | 3.39 | 0.07* |
| SHDI | 0.05 | 0.82 |
| Temp | 2.11 | 0.15 |
| Turb | 1.44 | 0.23 |
| XCMGW | 0.39 | 0.53 |
| X_HALL | 1.64 | 0.20 |
| Xslope | 1.07 | 0.30 |
| inciswd | 4.54 | 0.03* |
| lsb_dmm | 9.93 | <0.01* |
| pH | 0.28 | 0.59 |
| pfc_all | 2.62 | 0.11 |
| sinu | 0.09 | 0.76 |
| xbka | 3.97 | 0.05* |
| xgb | 0.01 | 0.91 |
| xwd_rat | 12.52 | <0.01* |

Table 3.5. Levene's test results for homogeneity of variance among sites within BULs and sites statewide. An asterisk denotes significance achieved at $p \leq 0.10$ and $W \geq 2.71$. See Table 3.1 for variable definitions.

| Variable | <i>W</i> | <i>p</i> |
|-----------------|-----------------|-----------------|
| AFLDFREQ | 0.84 | 0.36 |
| Cndctvt | 0.91 | 0.34 |
| DO | 4.46 | 0.04* |
| Dischrg | 5.32 | 0.02* |
| Habscore5 | 0.48 | 0.49 |
| MADTW | 2.89 | 0.09* |
| OM | 1.55 | 0.21 |
| SHDI | <0.01 | 0.95 |
| Temp | 0.84 | 0.36 |
| Turb | 0.70 | 0.40 |
| XCMGW | 0.17 | 0.68 |
| X_HALL | 0.72 | 0.40 |
| Xslope | 0.44 | 0.51 |
| inciswd | 2.46 | 0.12 |
| lsb_dmm | 4.46 | 0.03* |
| pH | 0.12 | 0.73 |
| pfc_all | 1.08 | 0.30 |
| sinu | 0.04 | 0.84 |
| xbka | 1.59 | 0.21 |
| xgb | <0.01 | 0.94 |
| xwd_rat | 3.92 | 0.05* |

Table 3.6. Levene's test results for homogeneity of variance among all BULs. An asterisk denotes significance achieved at $p \leq 0.10$ and $W \geq 1.38$. See Table 3.1 for variable definitions.

| Variable | <i>W</i> | <i>p</i> |
|-----------------|-----------------|-----------------|
| AFLDFREQ | 1.04 | 0.42 |
| Cndctvt | 2.10 | <0.10* |
| DO | 2.74 | <0.10* |
| Dischrg | 4.88 | <0.10* |
| Habscore5 | 0.77 | 0.80 |
| MADTW | 1.53 | 0.05* |
| OM | 5.70 | <0.10* |
| SHDI | 1.28 | 0.16 |
| Temp | 0.96 | 0.54 |
| Turb | 1.04 | 0.41 |
| XCMGW | 1.05 | 0.40 |
| X_HALL | 0.90 | 0.62 |
| Xslope | 1.20 | 0.23 |
| inciswd | 3.66 | <0.10* |
| lsb_dmm | 2.56 | <0.10* |
| pH | 1.30 | 0.15 |
| pfc_all | 2.54 | <0.10* |
| sinu | 0.91 | 0.61 |
| xbka | 1.16 | 0.27 |
| xgb | 1.26 | 0.18 |
| xwd_rat | 3.06 | <0.10* |

Table 3.7. Average silhouette value of sites by classification system in descending order. Stream Order only has 8 groups because no sampled sites were in stream order 9.

| Classifier | Number of Groups | Average Silhouette Value |
|-------------------|-------------------------|---------------------------------|
| Cluster | 2 | 0.33 |
| Matrix Landcover | 3 | 0.01 |
| Ecoregion | 4 | -0.01 |
| Stream Order | 8 | -0.03 |
| HUC12 | 429 | -0.16 |
| BUL | 32 | -0.22 |

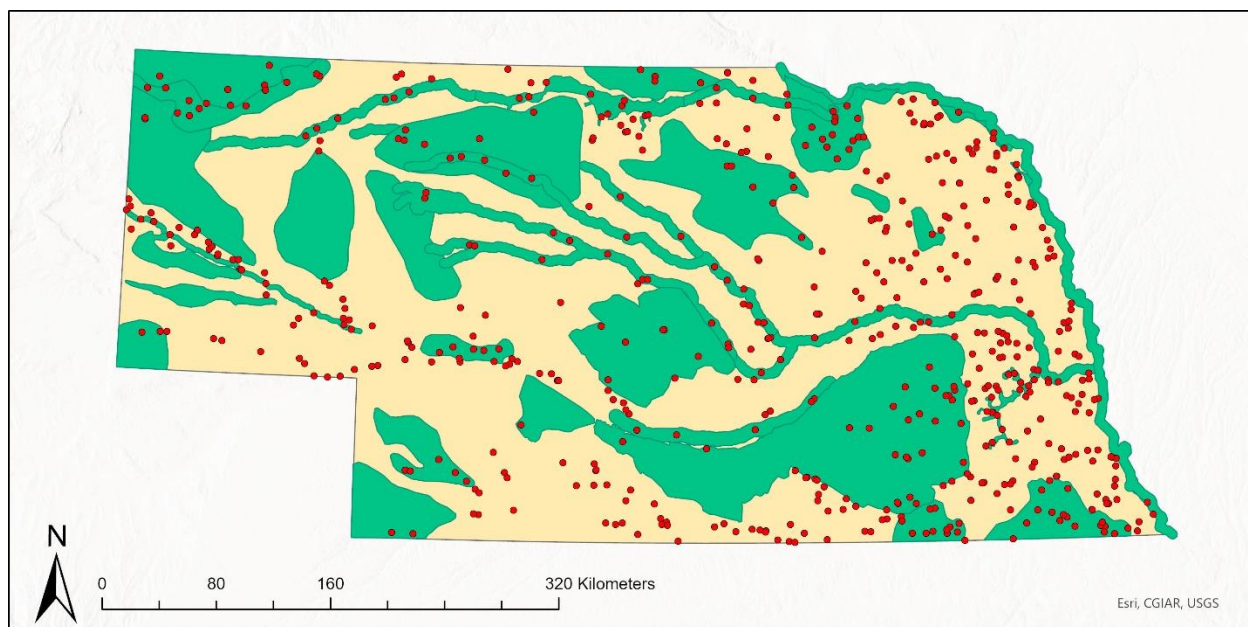
Figures

Figure 3.1. NDEE environmental sampling sites across Nebraska are represented by red points. The green polygons are the Biologically Unique Landscapes (BULs).

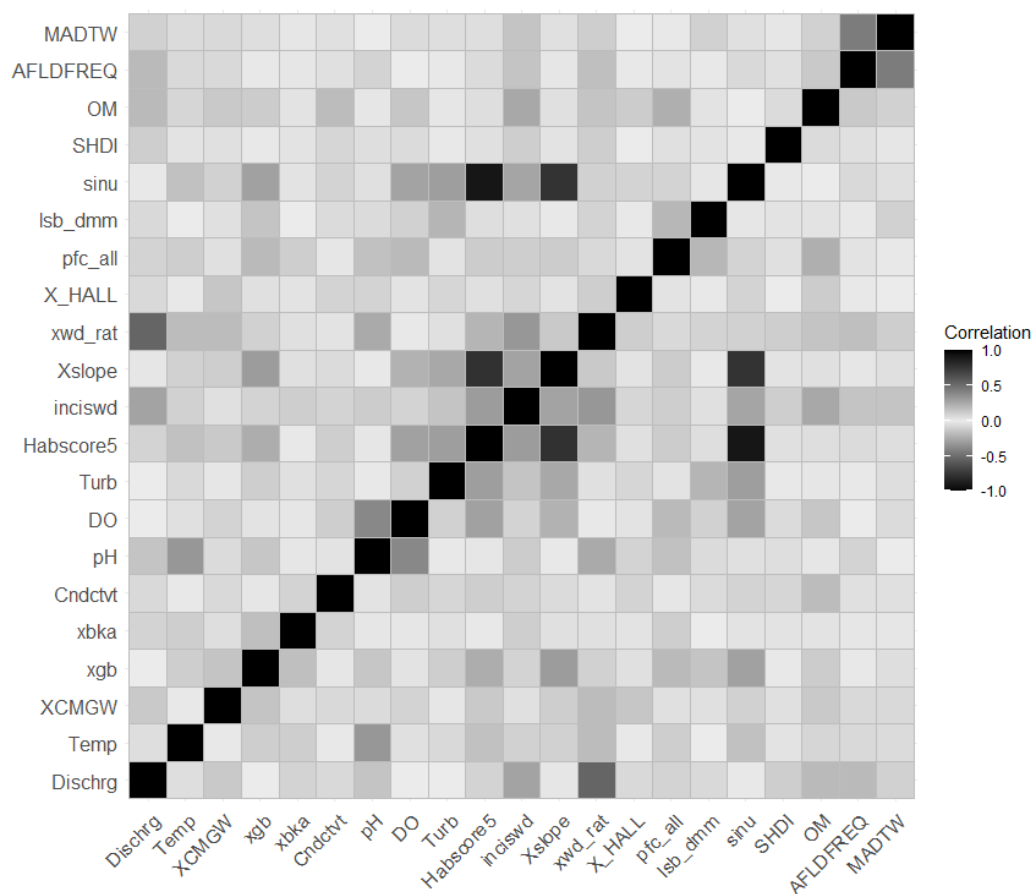


Figure 3.2. Correlation plot of selected environmental variables. Darker shading illustrates a higher Pearson's correlation coefficient. See Table 3.1 for variable definitions.

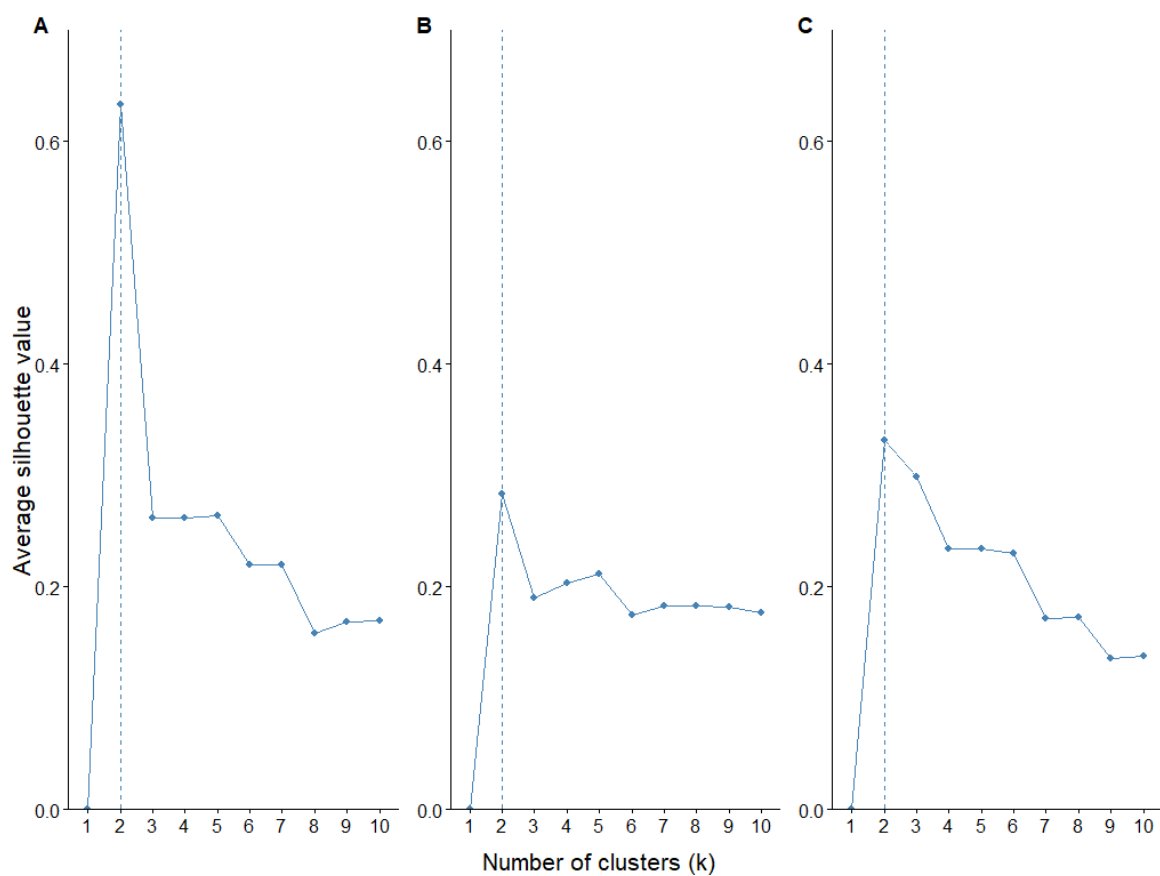


Figure 3.3. The average silhouette value produced by clustering schemes of up to 10 groups to determine the optimal number of clusters for sites inside BULs (A), outside BULs (B), and statewide (C). The dotted line denotes the optimal number of clusters, where the silhouette value is maximized.

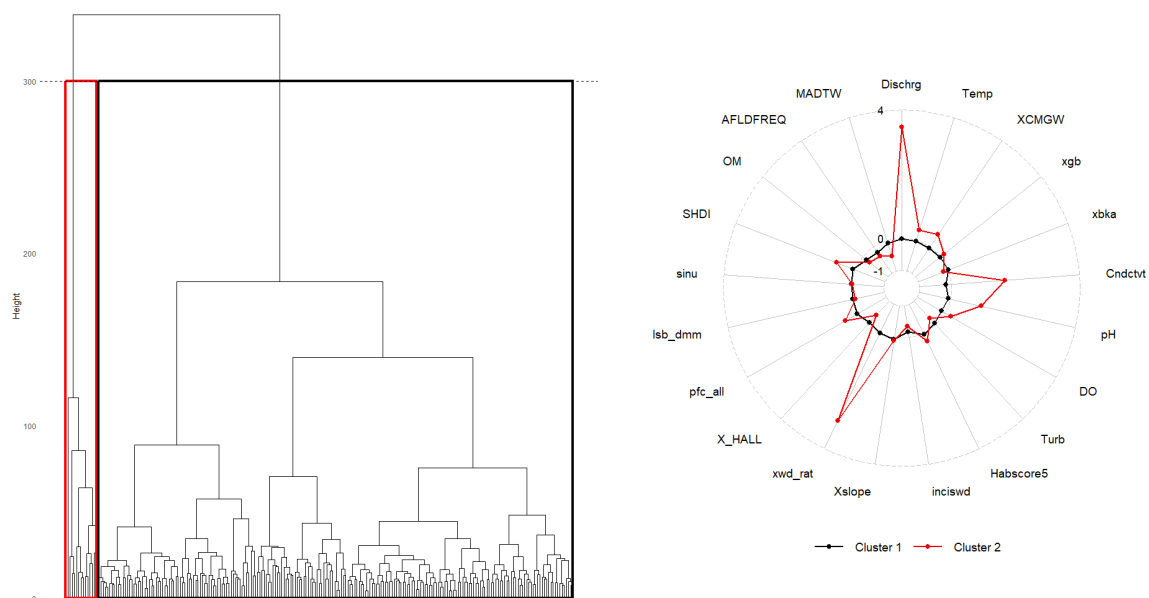


Figure 3.4. On the right side is a dendrogram depicting the overall dissimilarity of the sites inside BULs. Height on the y-axis denotes dissimilarity, with splits in the tree showing greater dissimilarity at greater height. The dotted line at $y = 300$ shows the height at which the dendrogram was cut to produce 2 clusters. The larger black box groups sites into cluster 1, while the smaller red box groups sites into cluster 2. On the left side, a radar plot shows means generated from normalized variables. The black line illustrates cluster 1, while the red line illustrates cluster 2.

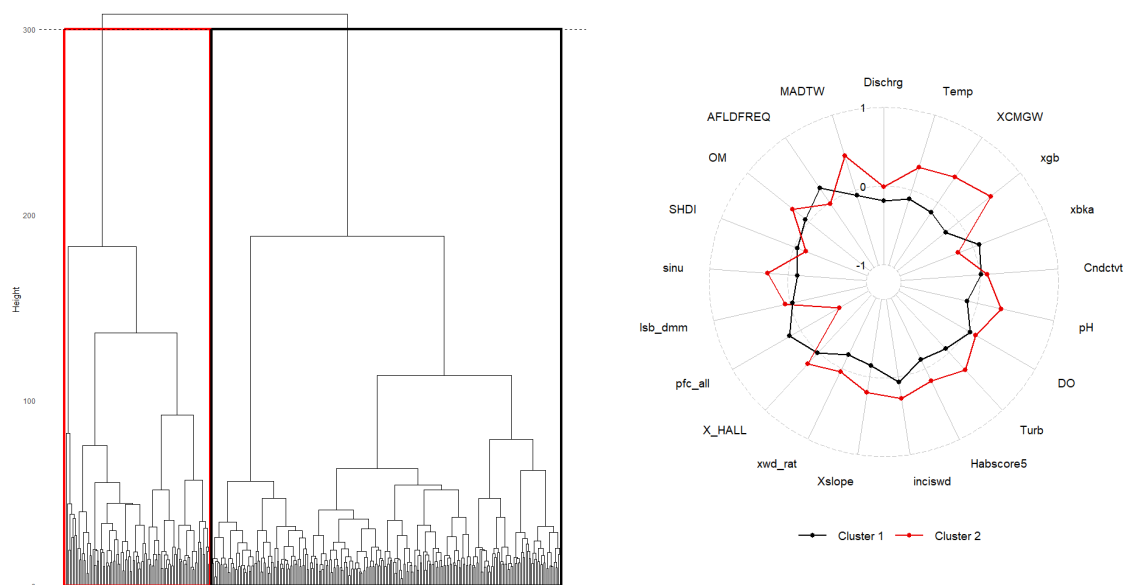


Figure 3.5. On the right side is a dendrogram depicting the overall dissimilarity of the sites outside BULs. Height on the y-axis denotes dissimilarity, with splits in the tree showing greater dissimilarity at greater height. The dotted line at $y = 300$ shows the height at which the dendrogram was cut to produce 2 clusters. The larger black box groups sites into cluster 1, while the smaller red box groups sites into cluster 2. On the left side, a radar plot shows means generated from normalized variables. The black line illustrates cluster 1, while the red line illustrates cluster 2.

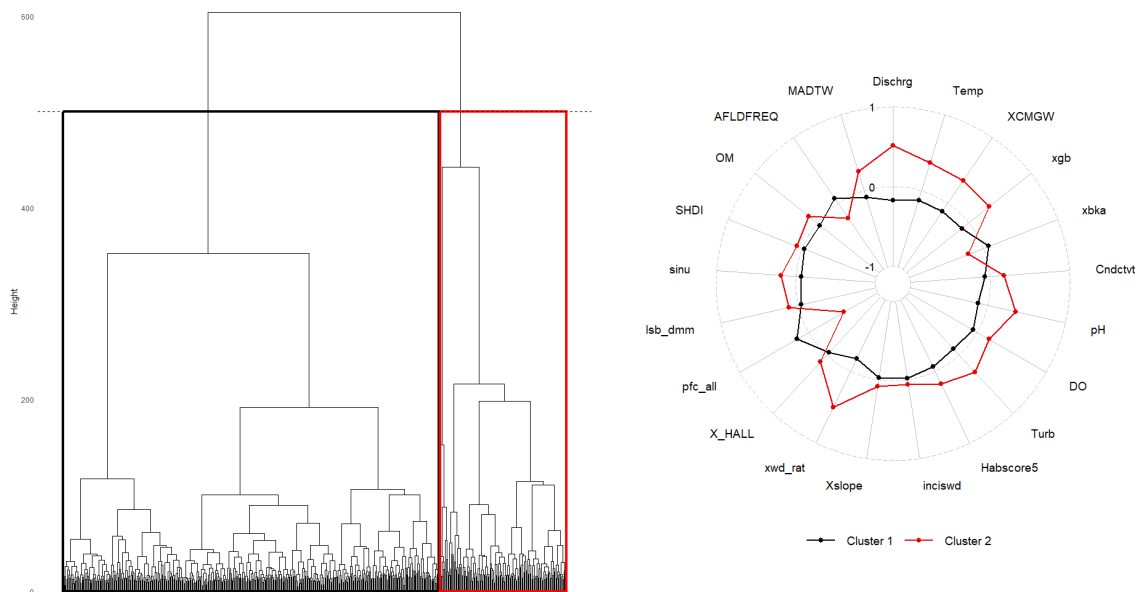


Figure 3.6. On the right side is a dendrogram depicting the overall dissimilarity of the sites statewide. Height on the y-axis denotes dissimilarity, with splits in the tree showing greater dissimilarity at greater height. The dotted line at $y = 500$ shows the height at which the dendrogram was cut to produce 2 clusters. The larger black box groups sites into cluster 1, while the smaller red box groups sites into cluster 2. On the left side, a radar plot shows means generated from normalized variables. The black line illustrates cluster 1, while the red line illustrates cluster 2.

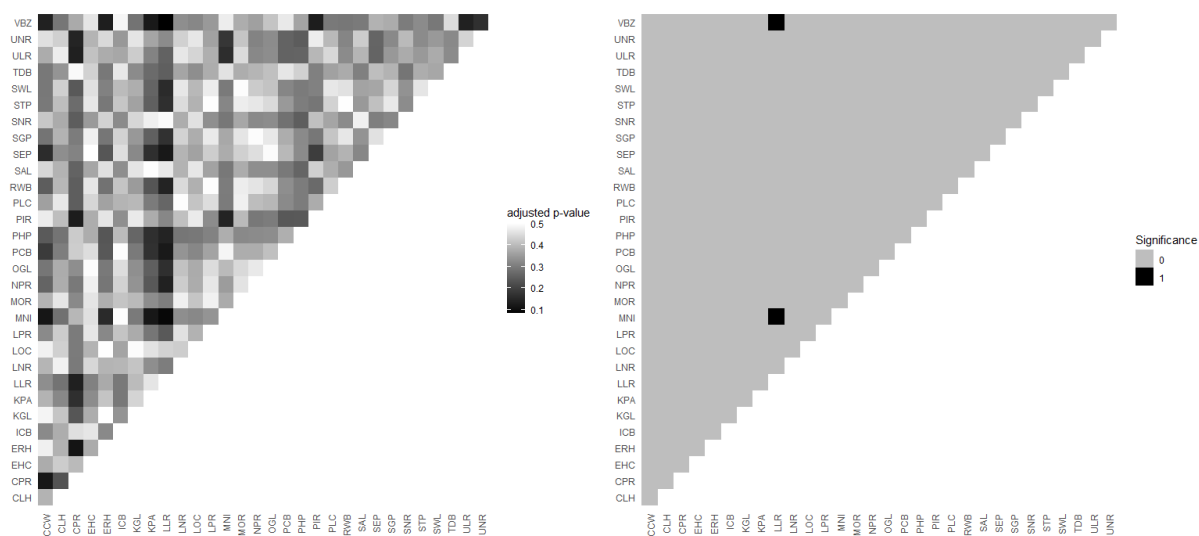


Figure 3.7. The left plot shows Dunn test results as pairwise comparisons of BULs as cells colored by adjusted p -value. Darker cells are those closer to significance at $p < 0.10$. The right plot illustrates Dunn test significance between BULs as a binary of significance 1 and non-significance 0. See Table 3.2 for full BUL names.

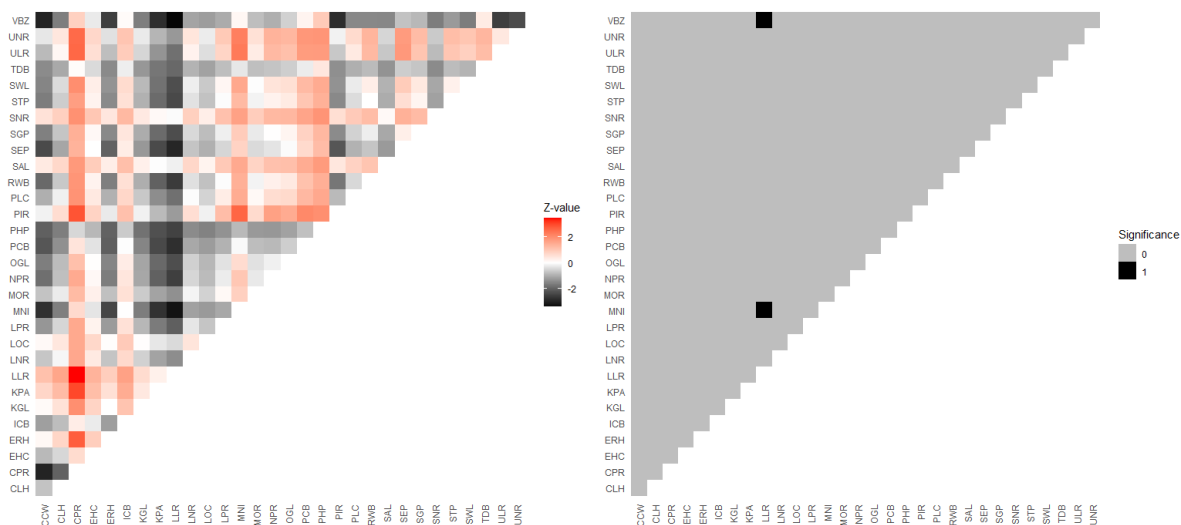


Figure 3.8. The left plot shows Dunn test results as pairwise comparisons of BULs as cells colored by Z-value. Darker cells are closer to the minimum $Z = -3.5$, while redder cells are closer to the maximum $Z = 3.5$. The right plot illustrates Dunn test significance between BULs as a binary of significance 1 and non-significance 0. See Table 3.2 for full BUL names.

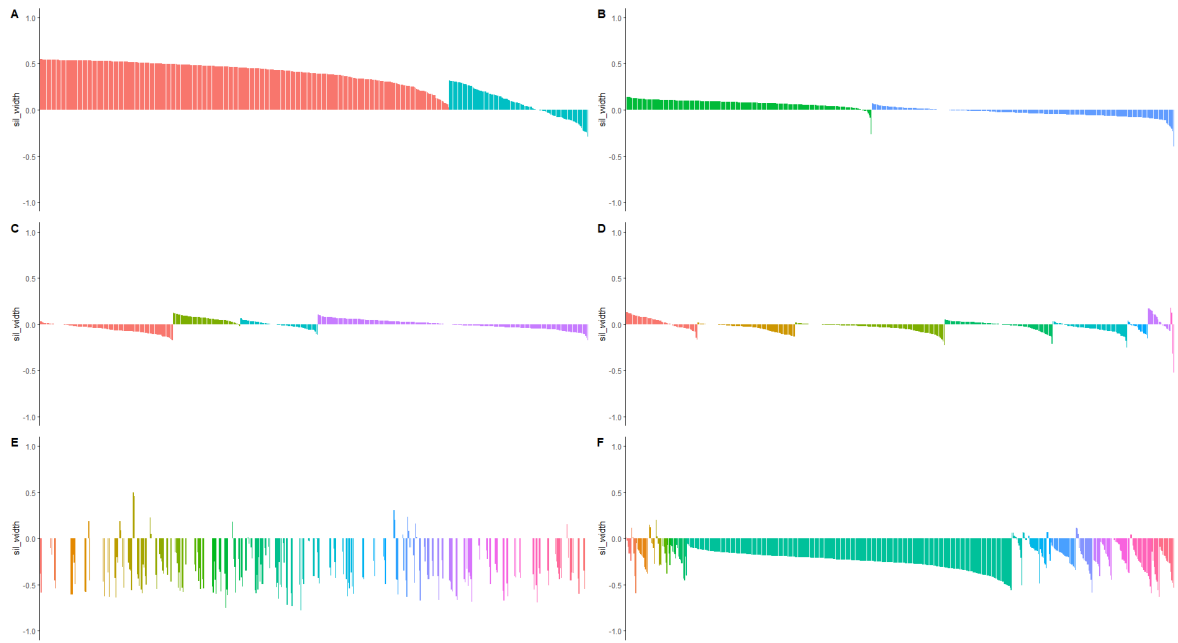


Figure 3.9. Silhouette value ranging from 1 to -1 on the y-axis by site on the x-axis for all classification systems. The plots show silhouette values by similarity-based clusters (A), dominant landcover (B), ecoregion (C), stream order (D), HUC12 watersheds (E), and BULs (F). Positive silhouette values indicate a better fit to assigned groups, while negative silhouette values indicate a better fit to other groups. Bars per site are colored by group membership.

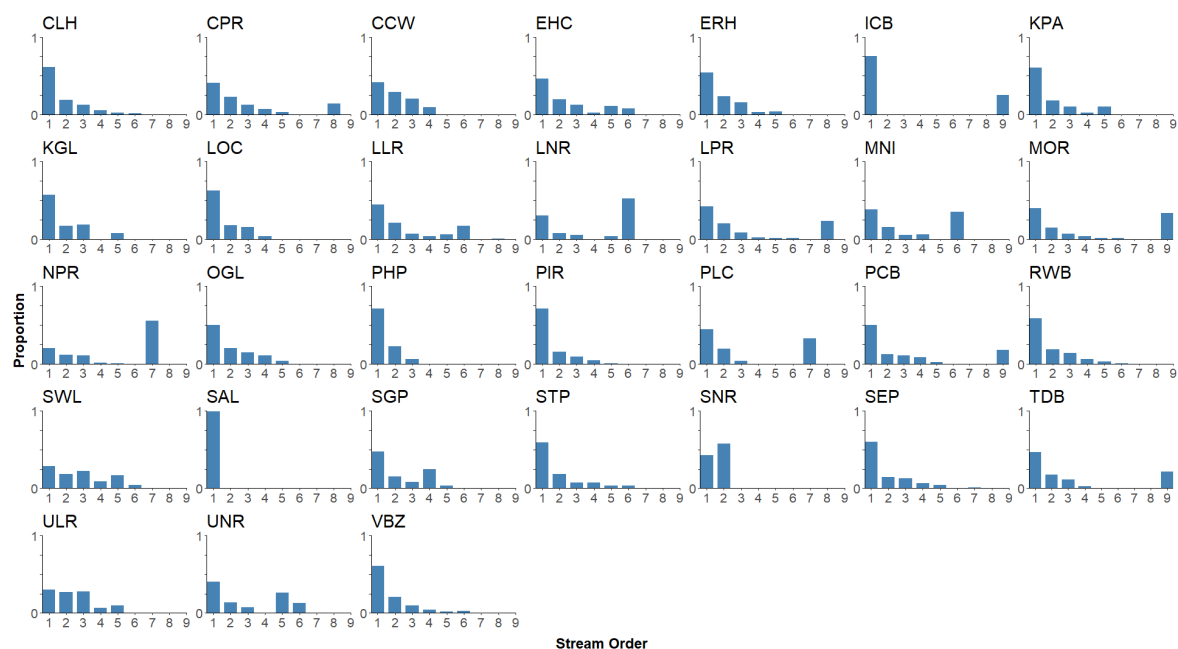


Figure 3.10. Bar charts showing the proportion of NHD flowline by stream order within each BUL, labeled by BUL abbreviation. See Table 3.2 for full names of BULs.

References

- Baxter, C. V., and F. R. Hauer. 2000. Geomorphology, hyporheic exchange, and selection of spawning habitat by bull trout. *Canadian Journal of Fisheries and Aquatic Sciences* 57:1470–1481.
- Borboudakis, G. and I. Tsamardinos. 2019. Forward-backward selection with early dropping. *Journal of Machine Learning Research* 20:1–39.
- Brown, M.B. and A.B. Forsythe. 1974. Robust tests for the equality of variance. *Journal of the American Statistical Association* 69:364–367.
- Cardinale, B.J., M.A. Palmer, C.M. Swan, S. Brooks, and N.L. Poff. 2002. The influence of substrate heterogeneity on biofilm metabolism in a stream ecosystem. *Ecology* 83(2):412–422.
- Caruso, B.S. 2014. GIS-based stream classification in a mountain watershed for jurisdictional evaluation. *Journal of the American Water Resources Association* 50(5):1304–1324.
- Christian, J.I., J.B. Basara, J.A. Otkin, E.D. Hunt, R.A. Wakefield, P.X. Flanagan, and X. Xiao. 2019. A methodology for flash drought identification: application of flash drought frequency across the United States. *Journal of Hydrometeorology* 20:833–846.
- Clarke, K.R. 1993. Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*. 18(1): 117–143.
- Conservation Survey Division, University of Nebraska-Lincoln. 2023. WLProd, Map Server.
https://csdgis.unl.edu/csdgis/rest/services/Production_Services/WLProd/MapServer/0. Using ESRI ArcGIS Pro. Version 3.0.
- Cross, F.B., R.L. Mayden, and J.D. Stewart. 1986. Fishes in the Western Mississippi Basin (Missouri, Arkansas, and Red Rivers). Pages 363–400 in C.H. Hocutt and E.O. Wiley, editors. *The Zoogeography of North American Freshwater Fishes*. John Wiley & Sons, New York, New York.
- Daniels, J.M. 2007. Flood hydrology of the North Platte River headwaters in relation to precipitation variability. *Journal of Hydrology* 344:70–81.
- Dewitz, J., and U.S. Geological Survey. 2021. National Land Cover Database (NLCD) 2019 Products. U.S. Geological Survey data release.
- Dinno, A. 2015. Nonparametric pairwise multiple comparisons in independent groups using Dunn’s test. *The Stata Journal* 15(1):292–300.
- Dodds, W.K., K. Gido, M.R. Whiles, K.M. Fritz, and W.J. Matthews. 2004. Life on the edge: The ecology of Great Plains prairie streams. *BioScience* 54(3): 205–216.

- Dodds, W.K. and R.M. Oakes. 2006. Controls on nutrients across a prairie stream watershed: land use and riparian cover effects. *Environmental Management* 37(5):634–646.
- Dunn, O.J. 1964. Multiple comparisons using rank sums. *Technometrics* 6(3): 241–252.
- Durance, I., C. Lepichon, and S. J. Ormerod. 2006. Recognizing the importance of scale in the ecology and management of riverine fish. *River Research and Applications* 22(10):1143–1152.
- Esri Inc. 2022. *ArcGIS Pro* (Version 3.0). Esri Inc. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>.
- Falke, J.A., K.D. Fausch, R. Magelky, A. Aldred, D.S. Durnford, L.K. Riley, and R. Oad. 2011. The role of groundwater pumping and drought in shaping ecological futures for stream fishes in a dryland river basin of the western Great Plains, USA. *Ecohydrology* 4:682–697.
- Fausch, K.D. and R.G. Bramblett. 1991. Disturbance and fish communities in intermittent tributaries of a Western Great Plains river. *American Society of Ichthyologists and Herpetologists* 1991:659–674.
- Fausch, K. D., C. E. Torgersen, C. V. Baxter, and H. W. Li. 2002. Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes. *BioScience* 52(6):483–498.
- Frissell, C. A., W. J. Liss, C. E. Warren, and M. D. Hurley. 1986. A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management* 10(2):199–214.
- Gastwirth, J.L., Y.R. Gel, and W. Mao. 2009. The impact of Levene’s Test of equality of variances on statistical theory and practice. *Statistical Science* 24(3):343–360.
- Gordon, N.D., T.A. McMahon, B.L. Finlayson, C.J. Gippel, and R.J. Nathan. 2004. Stream hydrology: an introduction for ecologists, pp. 233–286. John Wiley & Sons, Ltd., Chichester, West Sussex.
- Hawkins, C.P., R.H. Norris, J. Gerritsen, R.M. Hughes, S.K. Jackson, R.K. Johnson, and R. Jan Stevenson. 2000. Evaluation of the use of landscape classifications for the prediction of freshwater biota: Synthesis and recommendations. *Journal of the North American Benthological Society* 19:501–517.
- Hawkins, C.P. and M.R. Vinson. 2000. Weak correspondence between landscape classifications and stream invertebrate assemblages: implications for bioassessment. *Journal of the North American Benthological Society* 19(3):501–517.

- Heatherly, Thomas II, K. Bazata, D. Schumacher, and E. Taylor. 2014. A comparison of fish-based classification schemes for streams and rivers in Nebraska. *Journal of Environmental Quality* 43:1004–1012.
- Hermoso, V., A. F. Filipe, P. Segurado, and P. Beja. 2015. Filling gaps in a large reserve network to address freshwater conservation needs. *Journal of Environmental Management* 161:358–365.
- Holliday, V.T. 1990. Soils and landscape evolution of eolian plains: The Southern High Plains of Texas and New Mexico. *Geomorphology* 3:489–515.
- Hughes, R.M., S.A. Heiskary, W.J. Matthews, and C.O. Yoder. 1994. Use of ecoregions in biological monitoring. Lewis Press, Chelsea, Michigan.
- Hughes, R.M., A.T. Herlihy, and J.C. Sifneos. 2015. Predicting aquatic vertebrate assemblages from environmental variables at three multiscale geographic extents of the western USA. *Ecological Indicators* 57:546–556.
- Hynes, H.B. 1975. The stream and its valley. *Verh. Internat. Verein. Limnol.* 19: 1–15.
- Jorgenson, J.C., J.M. Honea, M.M. McClure, T.D. Cooney, K. Engie, and D.M. Holzer. 2009. Linking landscape-level change to habitat quality: an evaluation of restoration actions on the freshwater habitat of spring-run Chinook salmon. *Freshwater Biology* 54: 1560-1575.
- Juracek, K.E. and F.A. Fitzpatrick. 2003. Limitations and implications of stream classification. *Journal of the American Water Resources Association* 39:659-670.
- Kassambara, A. and F. Mundt. 2020. Factoextra: Extract and visualize the results of multivariate data analyses. R package version 1.0.7.
- Kaufman, L. and P.J. Rosseeuw. 1987. Finding groups in data. Wiley, New York.
- Kruskal, W.H. and W.A. Wallis. 1952. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association* 47:583–621.
- Lamba, J., A.M. Thompson, K.G. Karthikeyan, and F.A. Fitzpatrick. 2015. Sources of fine sediment stored in agricultural lowland stream, Midwest, USA. *Geomorphology* 236(1):44–53.
- Linke, S., E. Turak, and J. Nel. 2011. Freshwater conservation planning: the case for systematic approaches: The case for systematic conservation planning. *Freshwater Biology* 56(1):6–20.
- Legendre, P. and L. Legendre. 1998. Numerical ecology, 2nd English edition. Elsevier Science, Amsterdam, Netherlands.
- Levene, H. 1960. Contributions to probability and statistics: Essays in honor of Harold Hotelling. Stanford University Press, Palo Alto. 278–279.

- Long, J., T. Nelson, and M. Wulder. 2010. Regionalization of landscape pattern indices using multivariate cluster analysis. *Environmental Management* 46:134–142.
- Maechler M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik (2022). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.4.
- Mažeika, S., P. Sullivan, and M.C. Watzin. 2004. Understanding stream geomorphic state in relation to ecological integrity: Evidence using habitat assessments and macroinvertebrates. *Ecological Management* 34(5):669–683.
- McCormick, F.H., D.V. Peck, and D.P. Larsen. 2000. Comparison of geographic classification schemes for Mid-Atlantic stream fish assemblages. *Journal of the North American Benthological Society* 19(3):385–404.
- McManamay, R.A., M.J. Troia, C.R. DeRolph, A.O. Sheldon, A.R. Barnett, S.C. Kao, and M.G. Anderson. 2018. A stream classification system to explore the physical habitat diversity and anthropogenic impacts in riverscapes of the eastern United States. *PLoS ONE* 13(6):e0198439.
- McManamay, R.A. and C.R. DeRolph. 2019. A stream classification system for the conterminous United States. *Scientific Data* 6, 190017. <https://doi.org/10.1038/sdata.2019.17>.
- McNeil, V.H., M.E. Cox, and M. Preda. 2005. Assessment of chemical water types and their spatial variation using multi-stage cluster analysis, Queensland, Australia. *Journal of Hydrology* 310:181–200.
- Melles, S.J., N.E. Jones, and B. Schmidt. 2012. Review of theoretical developments in stream ecology and their influence on stream classification and conservation planning. *Freshwater Biology* 57:415–434.
- Meshesha, T.W., S.K. Tripathi, and D. Khare. 2016. Analyses of land use and land cover change dynamics using GIS and remote sensing during 1984 and 2015 in the Beresa Watershed Northern Central Highland of Ethiopia. *Modeling Earth Systems and Environment* 2:1–12.
- Murtagh, F. 2014. Ward’s hierarchical agglomerative clustering method: Which algorithms implement ward’s criterion? *Journal of Classification* 31:274–295.
- Nebraska Department of Environment and Energy – Water Quality Division. 2020. Nebraska Water Monitoring Programs Report 2018 & 2019.
- Nel, J. L., E. Turak, S. Linke, and C. Brown. 2011. Integration of environmental flow assessment and freshwater conservation planning: a new era in catchment management. *Marine and Freshwater Research* 62(3):290–299.
- Nijboer, R.C., and P.F.M. Verndonschot. 2004. Variable selection for modelling effects of eutrophication on stream and river ecosystems. *Ecological Modelling* 177: 17–39.

- Ojima, D., L. Garcia, E. Elgaali, K. Miller, T.G.F. Kittel, and J. Lackett. 1999. Potential climate change impacts on water resources in the Great Plains. *Journal of the American Water Resources Association* 35(6):1443–1454.
- Oksanen, J., G.L. Simpson, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, P. Solymos, M. Henry, H. Stevens, E. Szoecs, H. Wagner, M. Barbour, M. Bedward, B. Bolker, D. Borcard, G. Carvalho, M. Chirico, M. De Caceres, S. Durand, H.B. Antoniazi Evangelista, R. FitzJohn, M. Friendly, B. Furneaux, G. Hannigan, M.O. Hill, L. Lahti, D. McGlinn, M. Ouellette, E.R. Cunha, T. Smith, A. Stier, C.J.F. Ter Braak, and J. Weedon. (2022). *Vegan: Community Ecology Package*. R package version 2.6-4.
- Paulsen, S.G. A. Mayo, D.V. Peck, J.L. Stoddard, E. Tarquinio, S.M. Holdsworth, J. Van Sickle, L.L. Yuan, C.P. Hawkins, A.T. Herlihy, P.R. Kaufmann, M.T. Barbour, D.P. Larsen, and A.R. Olsen. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. *Journal of the American Benthological Society* 27(4):812-821.
- R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rosgen, D.L. 1994. A classification of natural rivers. *Catena* 22:169-199
- Rousseeuw, P.J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65.
- Saunders, D. L., J. J. Meeuwig, and A. C. J. Vincent. 2002. Freshwater protected areas: strategies for conservation. *Conservation Biology* 16(1):30–41.
- Schneider, R., K. Stoner, G. Steinauer, M. Panella, and M. Humpert (Eds.). 2011. *The Nebraska Natural Legacy Project: State Wildlife Action Plan*. 2nd ed. The Nebraska Game and Parks Commission, Lincoln, NE.
- Sohl T.L., B.M. Sleeter, K.L. Sayler, M.A. Bouchard, R.R. Reker, S.L. Bennett, R.R. Sleeter, R.L. Kanengieter, and Z. Zhu. 2012. Spatially explicit land-use and land-cover scenarios of the Great Plains of the United States. *Agriculture, Ecosystems, and Environment* 153:1–15.
- Sokal, R.R. 1974. Classification: Purposes, principles, progress, prospects. *American Association for the Advancement of Science* 185:1115–1123.
- Stoddard, J.L., A.T. Herlihy, D.V. Peck, R.M. Hughes, T.R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the American Benthological Society* 27(4):878-891.
- Strayer, D. L. and D. Dudgeon. 2010. Freshwater biodiversity conservation: recent progress and future challenges. *Journal of the North American Benthological Society* 29(1):344–358.

- Tang, J., S. Alelyani, and H. Liu. 2014. Feature selection for classification: A review. *Data Classification: Algorithms and Applications* 37:62–86.
- Townsend, C.R. and A.G. Hildrew. 1994. Species traits in relation to a habitat templet for river systems. *Freshwater Biology* 31:265-275.
- Trimble, D.E. 1980. The geologic story of the Great Plains. *Geologic Survey Bulletin* 1493. Washington DC: Government Printing Office.
- USEPA. 2019. National rivers and streams assessment 2018/19: field operations manual wadeable. http://www.epa.gov/sites/default/files/2019-05/documents/nrsa_1819_fom_wadeable_version_1.2_0.pdf
- Vannote, R. L., G.W. Minshall, K.W. Cummins, J. R. Sedell, and C.E. Cushing. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* 37:130–137.
- Van Sickle, J.V. 1997. Using mean similarity dendrograms to evaluate classifications. *Journal of Agricultural, Biological, and Environmental Statistics* 2(4):370–388.
- Van Sickle, J.V. and R.M. Hughes. 2000. Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. *Journal of the North American Benthological Society* 19(3):370–384.
- Wang, L., P.W. Seelbach, and R.M. Hughes. 2006. Introduction to landscape influences on stream habitats and biological assemblages. *American Fisheries Society Symposium* 48:1–23.
- Wang, F., H. Franco-Penya, J.D. Kelleher, J. Pugh, and R. Ross. 2017. An analysis of the application of simplified silhouette to the evaluation of k-means clustering validity. *Machine Learning and Data Mining in Pattern Recognition* 1:295-301.
- Ward, J.H. Jr. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58(308):236–244.
- Winter, T.C. 2007. The role of groundwater in generating streamflow in headwaters and in maintaining baseflow. *Journal of the American Water Resources Association* 43:15–25.
- Young, F.J. and R.D. Hammer. 2000. Defining geographic soil bodies by landscape position, soil taxonomy, and cluster analysis. *Soil Science Society of America* 64(3):989–998.

CHAPTER 4

MANAGEMENT RECOMMENDATIONS AND FUTURE RESEARCH QUESTIONS

Management Recommendations

I established species-habitat relations for Flathead Chub, Plains Minnow, and Northern Redbelly Dace as representatives of at-risk minnow species (Chapter 2). Variables representing hydrological conditions, stream size and position in the drainage network, and groundwater connectivity had relatively high influence on presence for each species. As such, these environmental processes may be the most influential on at-risk species distributions and should take priority in aquatic sampling and conservation efforts. For example, flow velocity of approximately 0.4 m/s was the most positively influential on Northern Redbelly Dace presence. If habitat restoration is conducted on a stream site with resident Northern Redbelly Dace, achieving a flow velocity of 0.4 m/s may serve as a goal for the project. However, models presented in chapter 2 were constructed on limited historical datasets and would benefit from validation on field-sampled data. Overall, I recommend an expansion of modeling approaches used in this study to include more species and more spatially and temporally robust environmental data. Ecologically relevant sampling of environmental data may consider a focus of hydrological characteristics of streams as these were the most influential characteristics for all three study species. Further, expansion into community modeling may provide more insight into the structure and function of aquatic environments, including linkages with habitat, and further inform conservation decisions.

Species-habitat relations described in chapter 2 were used to map habitat suitability for Flathead Chub, Plains Minnow, and Northern Redbelly Dace across Nebraska. Stream reaches with high predicted presence for any species should be validated with field sampling. Such an approach would serve to inform future sampling in what environmental metrics to collect and where, as well as to adaptively refine the models with this sampling to improve the characterization of species-habitat relations. Further, if target species are discovered in reaches of high predicted presence, these reaches may then serve for targeted study of these species life histories. For instance, there are gaps in knowledge of Flathead Chub, Plains Minnow, and Northern Redbelly Dace life history post-spawning. As such, targeted study of populations during and after spawning may further inform preferential habitat use by different life stages and a more robust understanding of their ecology. Targeted study will also allow for collection of fine-scale habitat conditions at the time of species presence, further informing ecologically relevant habitat conditions for at-risk species.

In chapter 3, I discovered the BUL construct captured statewide trends in stream habitat heterogeneity. Furthermore, stream habitat within the BULs was similar to stream habitat outside the BULs. Thus, the BULs would likely have a positive role in conserving the habitat heterogeneity of aquatic systems in Nebraska. However, stream habitat was not unique among BULs by selected metrics. The lack of unique in-stream conditions among BULs limits their applicability as a form of stream classification. As such, while the BULs encompass statewide stream habitat heterogeneity, I would not recommend their use as stream classification units. Furthermore, it may be advisable to update the current BUL construct to reflect unique in-stream habitats and the aquatic biological

communities they drive. This may be done in a hierarchical fashion through the revision of BUL boundaries on similarity-based clusters and stream order, both of which may capture fine-scale habitat heterogeneity.

In chapter 3, it was determined that similarity-based stream classification outperformed traditional geographical classifications (i.e., ecoregions, subwatersheds, stream order, dominant landcover) in categorizing site-scale habitat conditions. Similarity-based methods may be considered in the establishment of a standard statewide stream classification system in Nebraska. However, hierarchical classification systems may better represent the hierarchy of environmental processes driving site-scale habitat conditions. I recommend pursuing the development of a hierarchical stream classification system incorporating both similarity- and landscape-based groupings, such as similarity-based clusters, ecoregions, and stream order (as the two strongest performing landscape-based classifiers). Furthermore, expansion to include assessment of classification systems based on biological community structure may serve to inform stream classifications in Nebraska. Developing a strong stream classification system using aforementioned approaches would serve to adaptively refine the representation of stream habitat heterogeneity in environmental models and conservation frameworks.

Future Research Questions

- How does habitat use differ between individuals of different life stages for Flathead Chub, Plains Minnow, and Northern Redbelly Dace?
- How do anthropogenic alterations shift habitat structure of stream ecosystems in temporal contexts? Which alterations most impact community structure and diversity?

- Which environmental processes most influence distribution of large-bodied at-risk fish species? Which environmental processes most influence local in-stream biological community structure?
- How does the classification strength of hierarchical stream classification approaches compare to that of similarity- and landscape-based approaches? Which classification hierarchies have the strongest performance (e.g., ecoregions + subwatersheds or ecoregions + stream order)?
- How does the classification strength of biological community-based classification systems compare to that of similarity- and landscape-based approaches?
- How does drainage network connectivity influence landscape-based classification strength? Do landscape-based classification approaches partition drainage networks in a manner not representative of connected systems?
- How can Nebraska's Biologically Unique Landscapes represent aquatic biological communities? Are site-scale species assemblages or in-stream habitat conditions more representative of unique stream environments?