

**Ecological Drivers of Fish Metacommunity Structure
in Boreal Shield Lakes**

A thesis presented to
The Faculty of Graduate Studies
of
Lakehead University
by
Wyatt Beach

In partial fulfillment of requirements
for the degree of
Master of Science in Biology

January 15, 2025

Abstract

Fish community composition in freshwater lakes is shaped by a range of biotic and abiotic factors, including environmental conditions, species interactions, and spatial connectivity between waterbodies. While aquatic community ecology studies historically treated lakes as isolated systems, recent research has increasingly embraced a metacommunity framework, integrating spatial connectivity with environmental and biological predictors of community composition. Despite this shift, few studies have thoroughly examined the relative roles of spatial connectivity, environmental factors, and species interactions in shaping lake fish communities. To address this gap, I conducted a study across 81 lakes distributed within two quaternary watersheds at the IISD Experimental Lakes Area in northwest Ontario. Using Joint Species Distribution Modeling (JSDM) alongside spatial eigenvector mapping techniques—Asymmetric Eigenvector Mapping (AEM) and Moran’s Eigenvector Mapping (MEM)—drivers of fish community composition were investigated. Results indicate that spatial variables—specifically lake connectivity, stream flow direction, and the maximum gradient along connecting streams—are primary drivers of fish metacommunity composition. In presence-absence models, these spatial factors explained more variation than environmental variables and species co-occurrence patterns (potentially reflecting species interactions). Conversely, relative abundance models (conditional on presence) performed poorly across all ecological models evaluated. These findings provide valuable insights into the role of spatial connectivity relative to other factors in shaping fish community structure on a presence-absence basis, emphasizing the importance of applying a metacommunity approach in community analyses.

Land Acknowledgment

I respectfully acknowledge that Lakehead University is located on the traditional lands of Fort William First Nation, signatory to the Robinson-Superior Treaty of 1850. My fieldwork for this research was conducted in Northwestern Ontario, at the IISD Experimental Lakes Area, within the traditional territories of the Anishinaabe, Ojibwe, Cree, and Métis peoples, under Treaty 3. I extend my gratitude for the opportunity to carry out my research on these lands.

Acknowledgement

First and foremost, I would like to thank my thesis advisor, Dr. Michael Rennie, for your guidance, support, and countless brainstorming sessions throughout my degree. Your expertise and mentorship have been instrumental in shaping this work.

I would also like to thank my committee members, Dr. Adam Algar and Dr. Stephanie Melles, for all of your help, especially with ecological modeling and Bayesian statistics.

Thank you to all of my lab mates at the CEE Lab, especially my field assistants, Kelvi Toskovich, Leif Buxton, and Connor Kubinec for all of the long days you put in at the ELA surveying lakes, assessing streams, and bushwacking to areas very few have ventured to go.

Thank you to everyone at the IISD Experimental Lakes Area, both past and present. My research heavily relied on the data collected over decades at the ELA, and this project would not have been possible without the foundation laid by those before me. Special thanks to Lauren Hayhurst, Lee Hrenchuk, Michelle Matson, and Marco Haust for your help during the field season.

I would also like to acknowledge the Ontario Ministry of Natural Resources for providing additional data on several lakes featured in this study.

Finally, I want to thank Claire Matthews, for your unwavering love and support throughout my degree. You've listened to me ramble about my research more than anyone else, and I am deeply grateful for your encouragement and support.

Table of Contents

Abstract.....	ii
Land Acknowledgment.....	iii
Acknowledgement	iv
List of Tables	vii
List of Figures.....	viii
Introduction.....	1
1.1 Background.....	1
1.2 Environmental influences on fish communities.....	2
1.3 Biological interactions in fish communities	5
1.4 Spatial influences on fish communities	6
1.5 Metacommunity modeling approaches	9
1.6 Joint Species Distribution Modeling.....	11
1.7 IISD Experimental Lakes Area	12
1.8 Objectives	13
Methods.....	14
2.1 Study Area	14
2.2 Lake Selection.....	16
2.3 Fish Abundance Surveys, Water Chemistry Sampling, and Estimation.....	16
2.4 Landcover Data.....	20
2.5 Lake Connectivity Surveys.....	21
2.6 Statistical Approach.....	23
2.7 Spatial-only JSJM Modeling	24
Results.....	29
3.1 Fish Abundance	29
3.2 Environmental Variables	29
3.3 Spatial-only JSJM Model	31
3.3.1 Presence-Absence	31
3.3.2 Relative Abundance.....	35
3.4 Full Dataset JSJM Models.....	37
3.4.1 Presence-Absence Models	37
3.4.2 Relative Abundance Models.....	39
3.5 Model ESC Presence-Absence Analyses.....	41
3.5.1 Overall Exploratory Power	41
3.5.2 Spatial Eigenvectors.....	44
3.5.3 Species Responses to Environmental and Spatial Variables	48
3.5.4 Environmental Drivers of Fish Presence	49
3.5.5 Spatial Drivers of Fish Presence	56
3.5.6 Species Associations (i.e., co-occurrences)	62

Discussion	63
References	78
Appendix A	88
Appendix B	104
Appendix C	105
Appendix D	107
Appendix E	109
Appendix F	113
Appendix References	115

List of Tables

Table 3.1. Fish species included in JSMD modeling, with a minimum of 10 occurrences (i.e., present in at least 10 lakes).	29
Table 3.2. Explanatory (Mean Tjur R^2 , AUC) and predictive power (10-fold Cross Validation Mean Tjur R^2 , 10-fold Cross Validation AUC, WAIC) of Spatial JSMD models for presence-absence of fish species. Spatial JSMD models are assessed to identify appropriate spatial descriptors for Full Dataset JSMD models. Models with poor cross-validation values suggest poor model fit.	33
Table 3.3. Explanatory (Mean SR^2) and predictive power (10-fold Cross Validation Mean SR^2 , and WAIC) of Spatial-only JSMD Models for relative abundance (conditional on presence) of fish species. Models with poor cross-validation values suggest poor model fit.	36
Table 3.4. Summary of Full Dataset JSMD models (with optimized spatial predictors from SM1) for presence-absence of fish species. Model performance is ranked based on explanatory (Mean Tjur R^2 , AUC) and predictive (10-fold Cross Validation Mean Tjur R^2 , 10-fold Cross Validation AUC, WAIC) power of each model. Models with poor cross-validation values suggest poor model fit.	38
Table 3.5. Summary of Full Dataset JSMD models (with optimized spatial predictors) for relative abundance (conditional on presence) of fish species. Model performance is ranked based on explanatory (Mean Tjur R^2 , AUC) and predictive (10-fold Cross Validation Mean Tjur R^2 , 10-fold Cross Validation AUC, WAIC) power of each model.	40
Table 3.6. Main explanatory variables and associated explained variance (%) values for each fish species included in Model ESC for presence-absence.	41

List of Figures

Figure 2.1. Distribution of the 81 waterbodies and associated watercourses included in this study, shaded in dark blue. Quaternary watershed boundaries are defined by red lines. The Eagle Watershed is located in the northeast region and the Dryberry Watershed is located in the southwest region.

..... 15

Figure 2.2. Lake connectivity diagram used for producing Moran's Eigenvector Mapping (MEM) and Asymmetric Eigenvector Mapping (AEM). Dark blue boxes represent lakes included in Joint Species Distribution Models (JSDMs). Light blue boxes represent lakes not included in modeling but necessary to include for eigenvector mapping. Site 0 and associated connections represent a theoretical site upstream of all study lakes, necessary for AEM exclusively. The hypothetical site 0 introduces directionality into the diagram and establishes asymmetric spatial variables. To maximize the variance explained by spatial associations and minimize the likelihood of selecting eigenvectors describing only random spatial patterns, four spatial eigenvectors associated with the highest eigenvalues were selected for each model

..... 25

Figure 3.1. Correlation between \log_{10} values of maximum depth (m) and lake surface area (ha) amongst 81 study lakes ($r = 0.82$).

..... 30

Figure 3.2. Correlation between \log_{10} values of total nitrogen ($\mu\text{g/L}$) and total phosphorus ($\mu\text{g/L}$) amongst 81 study lakes ($r = 0.55$).

..... 31

Figure 3.3. Variance partitioning (%) relative to presence-absence Model ESC's Tjur's R^2 for each fish species. The height of bars indicates the Tjur's R^2 value associated with each species. Green represents cumulative environmental variables, blue represents cumulative spatial variables (i.e., selected spatial eigenvectors), and orange represents species co-occurrence latent variables.

..... 42

Figure 3.4. Variance partitioning (%) relative to presence-absence Model ESC's Tjur's R^2 for each fish species. The height of bars indicates the Tjur's R^2 value associated with each species. Green shades represent environmental variables, blue shades (1-4) represent spatial eigenvector variables, and orange represents species co-occurrence latent variables.

..... 44

Figure 3.5. Visual representation of Spatial AEM eigenvectors 1-4 (a-d) used in Model ESC. AEM eigenvectors represent lake connectivity, flow direction of connecting watercourses, and the maximum gradient along a connecting watercourse. Each dot represents a lake. Green dots are positive, red dots are negative. The size of the dot

increases as the absolute eigenvector values increase. The width of the light blue polylines represent the maximum stream gradient found along with connecting watercourse.

..... 46

Figure 3.6 Correlation between lake order and spatial eigenvector 2 amongst 81 study lakes ($r_s = -0.73, p < 0.05$).

..... 48

Figure 3.7. Posterior distributions (means and 95% credible intervals) of regression coefficients for environmental and spatial eigenvector variables, by species. The coefficients are derived from the presence-absence Model ESC. Green indicates a variable has a strong positive association with a species. Red indicates a variable has a strong negative association with a species. Species with neutral environmental associations are shown in black.

..... 49

Figure 3.8. Response of lake trout presence probability to (a) \log_{10} value of maximum lake depth (m) and (b) the \log_{10} value of dissolved organic carbon (mg/L). Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

..... 51

Figure 3.9. Actual and predicted distribution of lake trout across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where lake trout were predicted to be present but were not observed. Purple represents lakes where lake trout were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

..... 52

Figure 3.10. Response of burbot presence probability to (a) \log_{10} value of maximum lake depth (m) and (b) \log_{10} value of dissolved organic carbon (mg/L). Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

..... 53

Figure 3.11. Actual and predicted distribution of burbot across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where burbot were predicted to be present but were not observed. Purple represents lakes where burbot were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

..... 54

Figure 3.12. Response of slimy sculpin presence probability to \log_{10} value of maximum lake depth (m). Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

..... 55

Figure 3.13. Actual and predicted distribution of slimy sculpin across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where slimy sculpin were predicted to be present but were not observed. Purple represents lakes where slimy sculpin were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

..... 56

Figure 3.14. Response of northern pike presence probability to spatial eigenvector 2. Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

..... 57

Figure 3.15. Actual and predicted distribution of northern pike across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where northern pike were predicted to be present but were not observed. Purple represents lakes where northern pike were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

..... 58

Figure 3.16. Response of white sucker presence probability to spatial eigenvector 2. Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

..... 59

Figure 3.17. Actual and predicted distribution of white sucker across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where white sucker were predicted to be present but were not observed. Purple represents lakes where white sucker were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

..... 60

Figure 3.18. Response of finescale dace presence probability to spatial eigenvector 2. Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

..... 61

Figure 3.19. Actual and predicted distribution of finescale dace across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where finescale dace were predicted to be present but were not observed. Purple represents lakes where finescale dace were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

..... 62

Figure 3.20. Species-to-species associations based on species co-occurrence latent variables produced in presence-absence Model ESC. Blue refers to positive and red to negative associations amongst species. Size of circles represent posterior probability.

..... 63

Introduction

1.1 Background

Aquatic ecosystems worldwide are experiencing rapid and unprecedented rates of biodiversity loss (Desforges et al., 2022; Wudu et al., 2023). This decline is primarily driven by anthropogenic activities, including climate change, the introduction of non-native species, and pressures from commercial fishing (Cooke et al., 2023; Desforges et al., 2022; Dextrase & Mandrak, 2006). Collectively, these factors pose a major threat to the structure, function, and stability of aquatic systems and their associated ecological communities. As environmental pressures intensify, understanding the biotic and abiotic factors that shape aquatic communities is essential for making informed management decisions, predicting the influence of future ecological changes, and identifying key considerations for restoration efforts. Despite the need to take a more comprehensive and holistic approach to studying aquatic communities, research that considers or includes metapopulation dynamics and connectivity between ecosystems remains relatively limited (Leibold et al., 2004).

Early studies on drivers of aquatic community composition primarily focused on the influence of local environmental suitability. These studies investigated environmental conditions or biological interactions between species, often treating lakes as isolated systems. However, recent research has increasingly adopted a metacommunity framework, which emphasizes the importance of considering habitat connectivity along with environmental and biological factors when studying ecological communities (Guimarães et al., 2014; Leibold et al., 2004; Olden et al., 2001).

Metacommunity theory is defined as a set of local communities connected by the dispersal of interacting species (Leibold et al., 2004). This framework emphasizes the need to evaluate effects of both local and regional processes on community composition (Thompson et al., 2020; Wilson, 1992). In the context of fish composition in lakes, local processes refer to species interactions and lake-specific environmental conditions (Brown et al., 2011), whereas regional processes correspond to the means of species dispersal and movement as the degree of connectivity between waterbodies (Brown et al., 2011; Guimarães et al., 2014). By integrating both local and regional processes in biological analyses, a metacommunity approach aims to provide a deeper understanding of how various factors influence community structure.

The metacommunity framework emphasizes the importance of both environmental suitability and habitat connectivity. While environmental suitability sustains species within their habitats, effective dispersal and colonization heavily rely on connectivity between habitats. An environment may provide optimal conditions, yet without sufficient connectivity, species may face isolation, limiting genetic exchange and increasing vulnerability to localized disturbances. Conversely, in less favourable environments, species may face negative growth rates and potential extirpation unless immigration, facilitated by connectivity to neighboring communities, offsets these hostile conditions. (Chase et al., 2020; Moritz et al., 2013).

1.2 Environmental influences on fish communities

Extensive research has been conducted over the past several decades investigating associations between local environmental conditions and the composition of fish communities (Chu et al., 2004; Johnson et al., 1977; Mehner et al., 2005; Öhman et al.,

2006; Plumb & Blanchfield, 2009; Rahel, 1984; Robinson & Tonn, 1989). Lake morphology, including surface area and lake depth, plays a major role in shaping the structure and ecological dynamics of aquatic environments, influencing the distribution and behaviour of various fish species. The depth of a lake affects thermal stratification, dissolved oxygen levels and light penetration, all of which are critical factors influencing the presence and abundance of fish populations. Certain fish species, such as lake trout (*Salvelinus namaycush*) and cisco (*Coregonus artedii*), are particularly sensitive to water temperature and oxygen concentrations associated with specific depth zones (Fang et al., 2009; Johnson et al., 1977). Lake trout specifically are strongly correlated with lakes supporting higher dissolved oxygen levels, along with greater water clarity, greater depths, and lower levels of total dissolved solids (Johnson et al., 1977; Sellers et al., 1998). Deep, cold-water lakes provide an ideal habitat for these cold-water species, offering refuge during warmer months and suitable conditions for foraging. In contrast, shallow lakes provide preferred habitat for cool and warmwater fish species that thrive in more sunlit environments with abundant vegetation and cover (Mushet et al., 2023).

In addition to lake morphology, water chemistry properties such as acidity, productivity, and light attenuation are also known to influence fish abundance and distribution (Benoît et al., 2016; Hossain et al., 2019; Jackson et al., 2001; Johnson et al., 1977; Mills et al., 1987). Acidic conditions in lakes can result from natural processes, such as organic matter decomposition, as well as human activities, including the historic impacts of acid rain and runoff from mining operations (Beamish, 1976). Even slight deviations of 1-2 units from neutral pH can affect the survival, reproduction, and distribution of aquatic organisms, including both fish and invertebrates (Baker &

Christensen, 1991; Findlay et al., 1999; Matuszek & Beggs, 1988; Mills et al., 1987).

During an 8-year acidification study at the Experimental Lakes Area, major shifts in the abundance of several fish species, including fathead minnow (*Pimephales promelas*), northern pearl dace (*Margariscus nachtriebi*), slimy sculpin (*Cottus cognatus*), white sucker (*Catostomus commersonii*), and lake trout, were observed to occur within 1 to 1.5 pH units of the natural state of the lake, leading to substantial changes in community structure (Mills et al., 1987).

Lake productivity, often represented by nitrogen and phosphorus levels, also greatly influence the ecological dynamics of lakes. Total phosphorus (TP) serves as a key indicator of eutrophication, influencing the growth of aquatic plants, algae, and other primary producers within freshwater ecosystems (Schindler et al., 1971, 1978). Elevated levels of TP can stimulate algal blooms, leading to decreased water clarity, oxygen depletion, and alterations in aquatic habitat structure (Hossain et al., 2019; Orihelet al., 2017). Excessive nutrient inputs, often originating from agricultural runoff, sewage discharge, and atmospheric deposition, can exacerbate nutrient loading in lakes, promoting the proliferation of algae and reducing habitat suitability for many fish species (Orihel et al., 2017). Conversely, lakes with more balanced nutrient levels and moderate productivity typically support more diverse aquatic communities, contributing to overall ecosystem resilience and stability (Correll, 1998).

Light attenuation, often associated with dissolved organic carbon (DOC), has also been shown to influence lake ecosystems and the fish communities they support (Schindler et al., 1997; Stasko et al., 2012). DOC is typically discharged into oligotrophic lakes from their surrounding catchment area (Schindler et al., 1997). Due to its effect on

light attenuation, elevated levels of DOC are often associated with reductions in the depth of euphotic zones, shallower and more stable thermoclines, and increased protection of aquatic organisms from harmful UV wavelengths (Schindler et al., 1997). Increased DOC levels are associated with higher epilimnetic temperatures, lower hypolimnetic temperatures, reduced oxygen in the hypolimnion, darker environments, decreased macrophyte cover, reduced benthic biomass production, and heightened protection against pollutants (Stasko et al., 2012). The influence of DOC on aquatic ecosystems affects the abundance and distribution of many fish species; previous research observed a decrease in the relative abundance of walleye (*Sander vitreus*) with increasing DOC levels, along with a decline in early growth rates for both walleye and lake trout (Benoit, 2014).

1.3 Biological interactions in fish communities

Along with environmental factors, biological interactions including competition, predation, and mutualism are known to influence metacommunity structure (COSEWIC, 2009; Englund et al., 2009; Hulsman et al., 2016; Scheibel et al., 2016). Predation can have a large structuring effect on community composition; for example, northern pike (*Esox lucius*) exhibit negative associations with several species and can eliminate most other fish species when introduced to naïve environments (Byström et al., 2007; Nicholson et al., 2015). Similarly, competition between species can also impact fish community structure, as has been observed among larval coregonids (lake whitefish and lake herring; Davis & Todd, 1998). Although past studies highlight the potential influence species interactions can have on fish community composition, efforts to

effectively incorporate these interactions into community modeling have been relatively limited (Wagner et al., 2020).

1.4 Spatial influences on fish communities

Spatial connectivity, within the metacommunity framework, is essential for facilitating regional dispersal between different waterbodies and hydrological systems (Brown et al., 2011; Urban, 2004). Lakes with higher degrees of connectivity increase the likelihood of fish dispersal between waterbodies, reducing the risk of local extinction while increasing the likelihood of homogenization amongst fish communities (Olden et al., 2001). Conversely, lakes that are more isolated face an increased risk of extinction and lower rates of recolonization (Fullerton et al., 2010; Olden et al., 2001). In addition to the presence of a connection, the degree of connectivity between lakes is primarily influenced by factors including the length and number of adjoining watercourses, the flow direction of these watercourses, and the presence of physical barriers along their paths (Olden et al., 2001).

Although hydrological connectivity has historically been overlooked when researching fish community in lakes, recent research has begun to emphasize its importance (King et al., 2021; Laske et al., 2016). Research in the Midwest United States demonstrated that surface water connectivity positively influenced species richness in both lakes and streams, explaining 23% of the observed species composition across the region (King et al., 2021). Similarly, studies of Arctic lakes have identified connectivity as a primary driver of species richness, further emphasizing its importance in shaping fish communities (Laske et al., 2016).

Despite evidence highlighting the importance of spatial connectivity in ecological studies, one of the major challenges in evaluating this factor is appropriately integrating spatial patterns into ecological models. Depending on the model's objectives, spatial patterns can be viewed either as key contributors to ecological structures or as potential sources of bias when investigating specific processes (Dray et al., 2006). In aquatic ecosystems, incorporating spatial patterns is particularly challenging because species dispersal is often restricted to specific pathways, such as watercourses between connected lakes, rendering overland proximity metrics inappropriate. A promising solution to this issue is the application of spatial eigenvector mapping techniques, such as Moran's eigenvector mapping (MEM) and asymmetric eigenvector mapping (AEM). These approaches can better represent specific connections among lakes by incorporating watercourse distance or other metrics that may significantly influence fish dispersal (Blanchet et al., 2008b; Dray et al., 2006).

Moran's eigenvector mapping (MEM) was initially developed by Borcard and Legendre (2002) in the form of Principal Coordinates of Neighbour Matrices (PCNM) and was later extended more broadly as MEM eigenfunctions by Dray et al. (2006). MEM is an eigenvector approach used to depict spatial patterns within a landscape. MEM eigenfunctions are derived through a principal coordinate analysis applied to a matrix among sampling locations (Dray et al., 2006). This matrix may be binary, thus exploring connectivity between sites, or weighted, where distance (or any other form of weight, such as watercourse distance, elevational gradients, etc.) can be applied. The resulting eigenvectors corresponding to the matrix can subsequently function as spatial predictor variables in multivariate regression models (Dray et al., 2006). MEM eigenvectors with

high eigenvalues (i.e., the first eigenvectors produced) typically capture broad-scale spatial patterns and explain the most variation in the associated spatial matrix.

Eigenvectors with lower eigenvalues, on the other hand, tend to capture fine-scale spatial patterns and explain much less variation in the associated spatial matrix (Dray et al., 2006; Heino et al., 2013).

Blanchet et al. (2008) introduced Asymmetric Eigenvector Mapping (AEM) as an alternative to MEM, differing primarily in its ability to account for directionality and the influence of asymmetric processes on the distribution of species (Blanchet et al., 2008b). By doing so, ecological processes related to aquatic systems can more accurately be investigated, allowing for the incorporation of river networks and the flow direction of waterways into spatial models. Similar to MEM, AEM is also able to include variables as weights in the spatial matrix.

The construction of AEM eigenfunctions begins with a directional connection diagram (e.g. Figure 2.2) displaying all relevant sites (i.e., lakes) and edges (i.e., hydrological connections), reflecting the degree of connectivity amongst sites of interest. To enforce directionality into the diagram, a theoretical site 0 is added “upstream” of the study area. This site 0, within an aquatic standpoint, connects to all headwater lakes in a study area (Blanchet et al., 2008b). The patterns found in the connection diagram are then transformed into a sites-by-edges asymmetric matrix. Within an AEM matrix, non-zero values are present where a site is connected upstream to the corresponding edge. Conversely, a zero is present when the site is not connected upstream to the edge. For AEM matrices, weights will generally reflect the difficulty of upstream movement from low elevation to higher elevation sites (Blanchet et al., 2008b).

Following the construction of a sites-by-edges matrix, a singular value decomposition (SVD) is conducted, producing spatial eigenvectors, summarizing the variability found within the connectivity matrix. Like any other SVD and similar to MEM eigenfunctions, the first eigenvectors produced (those with the highest eigenvalues) explain the most variation in the matrix and summarize more broad-scale spatial patterns while those produced last (those with lower eigenvalues) explain the least variation and reflect fine-scale spatial patterns (Blanchet et al., 2008b; Dray et al., 2006).

1.5 Metacommunity modeling approaches

Given the complexity of fish communities in freshwater lakes and the many potential factors at play, a multivariate modeling approach is often required. Although previous efforts have been made to model the influence of environmental attributes, biological interactions, and hydrological connectivity together, these efforts have typically left out certain aspects of hydrological connectivity or species interactions. Sharma et al. (2011) provided insight into spatial and environmental factors influencing fish communities across all lakes in Ontario, but lacked detailed hydrological data, instead modelling connections among tertiary watersheds rather than among lakes (due primarily to computational limitations at the time) and excluding flow direction and potential fish movement barriers from their analysis (Sharma et al., 2011). Similar modeling efforts by Olden et al. (2001) found strong relationships between fish community composition and lake isolation and morphology in a collection of south-central Ontario lakes, yet did not account for the influence of piscivorous species or natural fish movement barriers such as waterfalls, chutes, and cascades between adjoining lakes.

Despite the recognition that fish movement barriers likely have an influence of fish community composition in lakes in literature, few studies have effectively quantified their impact in conjunction with other factors when considering spatial connectivity. In studies that have included fish barriers in their analyses, often other factors were not analyzed. Mozzaquattro et al. (2020), for example, highlighted the importance of physical barriers, spatial distance, flow direction, and habitat features in shaping fish metacommunities in southern Brazil streams but were limited in their consideration of environmental and biological variables. To date, I have found no studies that comprehensively analyze fish barriers, flow direction and watercourse distance, along with environmental and biological factors together when considering fish community composition in freshwater lakes.

Along with the challenges of modeling hydrological connectivity and the limitations of previous studies, past efforts to incorporate the influence of species interactions on community assemblage have also been limited (Leibold et al., 2021; Pichler & Hartig, 2021; Wagner et al., 2020). Two of the most common methods used in community modeling, ordination and species distribution modeling (SDMs), do not account for species interactions (Pichler & Hartig, 2021). To overcome these limitations, a relatively new modeling technique called Joint Species Distribution Modeling (JSDMs) has become increasingly popular, which is a multivariate regression model approach that considers species-specific environmental preferences, spatial autocorrelation, and covariances among species simultaneously to explain metacommunity structure (Leibold et al., 2021; Tikhonov et al., 2020). Unlike Species Distribution Models (SDMs), JSDMs integrate species-level models into a single model, fitted simultaneously to all community

data, allowing for species co-occurrences to be included in the final model (Ovaskainen & Abrego, 2020).

1.6 Joint Species Distribution Modeling

Although numerous JSDM approaches have been established, Hierarchical Modeling of Species Communities (HMSC) has been recognized for its strong predictive performance (Norberg et al., 2019). HMSC is a Bayesian-fitted JSDM and a multivariate hierarchical generalized linear mixed model (GLMM; Ovaskainen & Abrego, 2020). The HMSC framework is versatile in utilizing ecological data to model processes that structure community composition. It has the ability to incorporate environmental variables, phylogenetic relationships, potential biological interactions (represented as species co-occurrences), and spatial data into its analysis as well as accept community composition data as both presence-absence and species abundance. The fitting of HMSC models to data are achieved through Bayesian inference, employing Markov chain Monte Carlo (MCMC) estimation techniques (Tikhonov et al., 2020).

Within the HMSC framework, latent variables are used to capture species co-occurrence patterns that remain unexplained by the model's specified fixed effects, such as environmental variables. These residual co-occurrence patterns can suggest potential biotic interactions or other unmeasured processes, which are then summarized in a residual species association matrix. Consequently, when two species are observed together more or less frequently than expected based on their environmental niches, they are respectively deemed as positively or negatively associated (Ovaskainen & Abrego, 2020). Although these species co-occurrences may be the result of biological interactions, it is important to note that these patterns could also be the result of other undefined

factors not included in the JSMD analysis. For this reason, species co-occurrences are unable to be properly modeled alone or without the inclusion of established fixed environmental and spatial variables (Ovaskainen & Abrego, 2020).

1.7 IISD Experimental Lakes Area

The rapid decline in biodiversity within aquatic ecosystems highlights a critical need to understand the driving factors that shape metacommunity structure. Given the complexities associated with investigating the influence of both environmental and biological factors on community composition while also accounting for spatial connectivity, it is clear that a new holistic approach can and should be taken for freshwater lake ecosystems. The International Institute for Sustainable Development (IISD) Experimental Lakes Area (ELA) provides an excellent opportunity to conduct such a study due to its rich collection of data among hydrologically connected waterbodies, providing a solid foundation to investigate drivers of fish metacommunity structure in a region representative of the boreal shield ecozone.

Following the establishment of the ELA in 1968, a total of 109 lakes were surveyed initially, providing information on the identity and relative abundance of fish populations, water chemistry metrics, and lake morphology data (maximum lake depth and surface area) for all surveyed lakes (Beamish et al., 1976; Cleugh & Hauser, 1971). Since this initial survey, several additional lakes in the region have also been sampled using similar methodology, thus providing a rich dataset for exploring watershed level fish metacommunity structure.

In addition to developing a rich dataset of the fish communities within the ELA, Beamish et al. (1976) also noted several anecdotal differences in fish species composition

between the two main quaternary watersheds included in these early surveys; the Eagle watershed to the northwest and Dryberry watershed to the southeast (Beamish et al., 1976). Lake whitefish (*Coregonus clupeaformis*), for example, were reportedly rare in lakes surveyed within the Dryberry watershed but abundant in lakes within the Eagle watershed. Conversely, lake herring (*Coregonus artedii*) were reported to be relatively common in the Dryberry watershed but less prevalent in the Eagle watershed. Johnny darters (*Etheostoma nigrum*) were also present in several surveyed lakes in the Dryberry watershed yet completely absent from lakes sampled in the Eagle watershed at the time.

1.8 Objectives

The main objectives of this thesis were:

- A. To determine the influence of lake connectivity and associated spatial variables (i.e., flow direction of connecting watercourses, stream gradient, stream sinuosity, and fish barriers such as waterfalls, chutes, and cascades) on fish community composition in freshwater lakes.
- B. To evaluate the relative importance of lake connectivity compared with environmental conditions and potential biological interactions (i.e., species co-occurrences) on fish community composition; and
- C. To identify key variables influencing the distribution of individual fish species in the region of the study.

Methods

2.1 Study Area

To determine the influence of lake connectivity, environmental variables, and species co-occurrence on fish metacommunity composition, 81 lakes spanning across 2 neighbouring quaternary watersheds were studied (Figure 2.1). The majority of the selected lakes are found near the ELA, located in northwest Ontario, approximately 52 km southeast of Kenora, ON. The ELA is surrounded by hundreds of freshwater lakes, all of which are found within two quaternary watersheds; the Eagle Watershed, located in the northeast region of study, and the Dryberry Watershed, located in the southwest region of study (Beamish et al., 1976). Lakes within the Eagle Watershed drain north into Eagle Lake, eventually joining the English River system (Schindler et al., 1996). Lakes within the Dryberry Watershed drain south, first into Dryberry Lake and then into the Lake of the Woods (Schindler et al., 1996). Both ultimately drain into the Winnipeg River, which then flows into Lake Winnipeg (Brunskill & Schindler, 1971; Schindler et al., 1996).

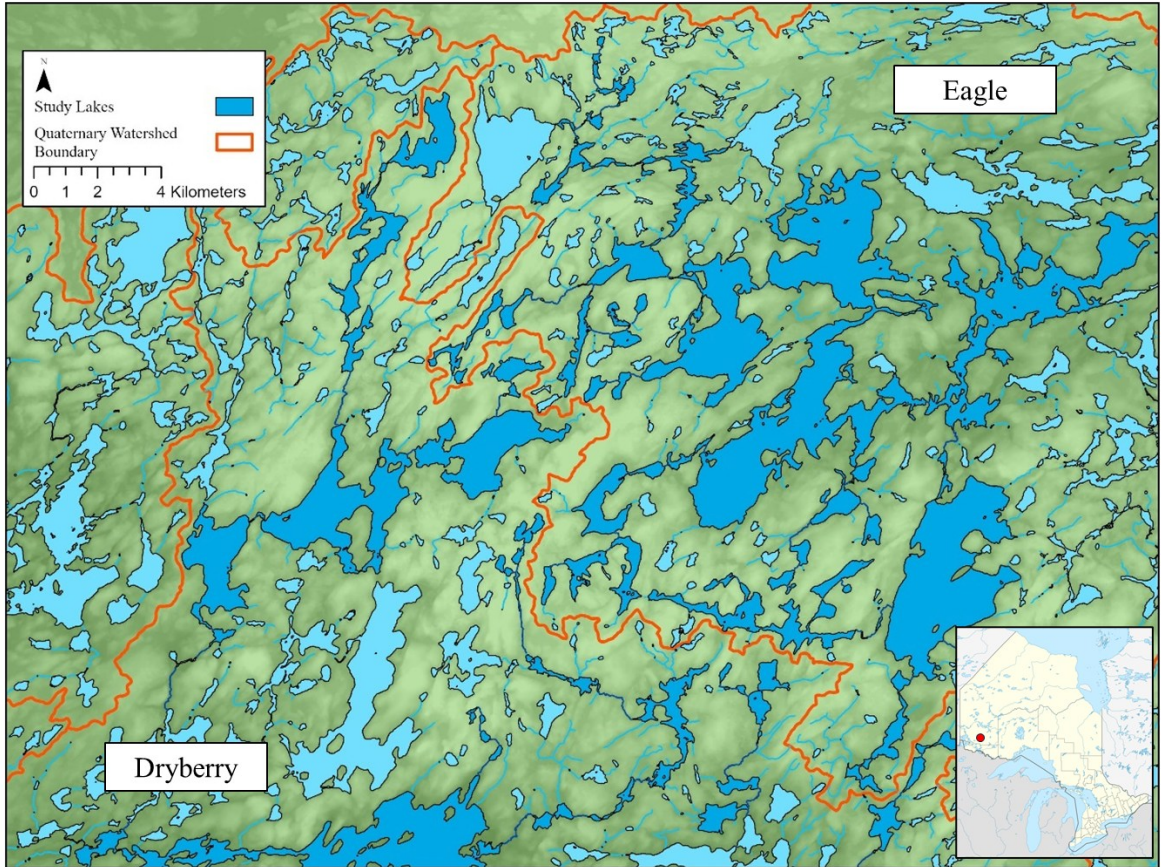


Figure 2.1. Distribution of the 81 waterbodies and associated watercourses included in this study, shaded in dark blue. Quaternary watershed boundaries are defined by red lines. The Eagle Watershed is located in the northeast region and the Dryberry Watershed is located in the southwest region.

Approximately 14,000 years ago, the region now encompassing the ELA was covered in glacial ice (Brunskill & Schindler, 1971). As the ice began to melt, a large lake formed, Lake Agassiz, with the eastern margin of the lake situated over where the ELA is now (Brunskill & Schindler, 1971). Over time, water began to recede, forming the lake basins and watersheds present today. As a result, lakes between the Eagle and Dryberry watersheds have been separated for thousands of years, with lakes within each watershed varying in degrees of connectivity to one another.

Due to the ELA's rich collection of lake data and previous observations of fish metacommunity diversity, the area provides an excellent opportunity to investigate fish metacommunity composition and the potential drivers of fish metacommunity structure.

2.2 Lake Selection

Lakes chosen for this study were primarily selected based on their level of connectivity to other surrounding waterbodies. Since my research objectives primarily focused on understanding the influence of lake connectivity and stream-related variables (such as flow direction and the presence of fish movement barriers) in shaping fish community structure, I specifically targeted lakes that are part of longer lake chains, defined as a series of connecting lakes. Lake chains where historical fish surveys and water sampling had previously been conducted on a majority of lakes were prioritized. Additional surveys were conducted to include previously unsurveyed lakes within identified chains, ensuring the inclusion of as many lakes as possible within selected lake chains (Appendix A).

2.3 Fish Abundance Surveys, Water Chemistry Sampling, and Estimation

Data from both historical and recent surveys were utilized to provide information for each of the 81 study lakes. Fish community composition, water chemistry, and lake morphology data for each lake were collected between 1971 to 2023 (Appendix A). Historical data from 1971 to 2021 were available for 60 lakes, gathered by various organizations including IISD-ELA, Ontario Ministry of Natural Resources (OMNR), and the Department of Fisheries and Oceans (DFO).

Historic fish surveys used various sampling methods, including gillnets, minnow traps, trap netting and seining (Appendix A). In cases where lakes underwent multiple

fish surveys, all available survey data were used to present a comprehensive depiction of the lake's species composition. This approach was implemented to account for variations in sampling methods and efforts across lakes that were sampled multiple times over the years, thereby reducing bias in species presence and abundance. Prior to 2014, all collections were made under the authority of Fisheries and Oceans Canada. Following 2014, collections were made under the approval of the OMNR by way of scientific collection permits and with the approval of the Canada Council for Animal Care by way of the Lakehead University Animal Care Committee (AUP 1465945, 1469342).

In 2022 (June-August), with the help of several Lakehead University research assistants, I conducted broad-scale monitoring (BSM) fish surveys for the remaining 10 unsurveyed study lakes. Surveys were based on the *IISD-ELA Broad-Scale Monitoring Protocol*, which is a modified version of OMNR's *Manual of Instructions for Broad-Scale Fish Community* (IISD, 2018; Sandstrom et al., 2013). Surveys typically involved setting 2-3 large mesh gills nets (8 mesh sizes per gang x 3.1 m panels) and 2-3 small mesh gill nets (5 mesh sizes per gang x 2.5 m panels) per sampling day across pre-defined depth strata in the lake (see below), with set durations lasting 16-22 hours (IISD, 2018).

Prior to setting nets, the maximum depth of each lake was measured using a handheld depth sounder. Temperature and oxygen profiles were recorded at the deepest part of the lake using a YSI Pro ODO unit to identify potential anoxic depths. Secchi depth measurements were also taken at this location, along with two 500 ml water samples collected for subsequent laboratory analysis to determine water quality metrics including pH, dissolved organic carbon (DOC), and total phosphorus (TP).

During BSM surveys, depth strata of 1-3m, 3-6m, 6-12m, 12-20m, 20-35m, and 35-50m were sampled, with sampling efforts among depth strata conducted proportionally to the area of each stratum in the lake, excluding depths where anoxic (<2 ppm of dissolved oxygen) conditions were observed (IISD, 2018). In addition to gillnets, minnow traps were placed along the shoreline of the lake, with at least three traps set per sampling day, as minnows often prefer downed trees and other underwater plants/structures for habitat (IISD, 2018).

Fish community composition for each of the 81 study lakes was represented as either presence-absence or relative abundance, where relative abundance was categorized broadly as absent, rare, moderately abundant, or abundant based on survey catches. These relative abundance measures, initially introduced by Beamish et al. (1976), have been a common practice for rapidly summarizing fish population data for ELA lakes since 1969 using limited data and has recently been demonstrated to accurately describe variation in catch per unit effort (CPUE) for several fish species (Littlefair et al., 2024). Relative abundance data were unavailable for 3 of our study lakes (lakes L254, L454, L589), resulting in a dataset of 78 lakes for relative abundance-specific analyses and 81 lakes for presence-absence analyses (Appendix A).

Despite variations in survey years, fish community composition was assumed to have remained fairly consistent over time due to the relatively non-invasive survey methods deployed when assessing lakes, limited angling pressure in the region, and the limited amount of development in the area. This assumption is supported by previous long-term monitoring studies (Milling, 2020; Rennie et al., 2019; Slongo, 2022), which reported minimal changes in fish species abundance over extended periods in several

ELA lakes not subjected to experimental manipulation. Due to the broad metrics used for quantifying fish species' abundance at each lake, a major change in a lake's composition or management would likely have to occur to affect a species' estimated relative abundance or presence-absence.

Historic water sampling of the study lakes often varied in the frequency of sampling. For lakes that had been sampled multiple times, water chemistry metrics were averaged, while single-sampled lakes retained original values. Periods of data collected during whole-lake experiments that altered typical lake characteristics were excluded from the analysis. Water samples used for water chemistry metrics were also restricted to those sampled between a depth of 0-3m and those occurring during the open water season (i.e. between May and October).

In addition to variation in water sampling frequency, the types of water chemistry metrics evaluated amongst historically sampled lakes also varied. In some cases, missing metrics were estimated from auxiliary data using published relationships (Beamish et al., 1976; Chow-Fraser, 1991; Molot & Dillon, 1997). In the absence of pH data, lake surface area was used to estimate pH based on the observed regression found during the original ELA fish surveys (Beamish et al., 1976):

(A) $\text{Log pH} = 0.77 + 0.04 [\text{log Lake Surface Area}]$ ($r = 0.57, P < 0.001$); where lake surface area is in units of 10^4m^2 .

For total phosphorus (TP), conductivity and lake depth were used to calculate an estimate value (Chow-Fraser, 1991):

(B) $\log TP = 0.473 \log MEI + 0.483$ ($r^2 = 0.58$); where TP is total phosphorus ($\mu\text{g/L}$) and MEI is defined below.

(C) $MEI = \text{Conductivity} \div \text{Mean Depth}$; where MEI refers to morphoedaphic index (a metric used to predict fish production in lakes), conductivity is in $\mu\text{S/cm}$, and mean depth is in meters (m), defined in (D) below.

(D) $\log \text{Mean Depth} = 0.940 [\pm 0.058] \log \text{Maximum Depth} - 0.349$ ($r^2 = 0.82$, $n = 60$, $P = 0.0001$); where mean depth and maximum depth are in meters (m).

To estimate dissolved organic carbon (DOC), we used water colour to provide an estimated value (Molot & Dillon, 1997):

(E) $DOC = 1.735 + 0.070 [\text{colour}]$ ($R^2 = 0.93$); where DOC refers to dissolved organic carbon (mg/L) and colour refers to water colour (Hu).

2.4 Landcover Data

All landcover-related analyses were performed using ArcGIS Pro 3.1. The surface area of study lakes was determined using the Ontario Integrated Hydrology Data (OIHD) v1.0 (OMNRF, 2012), in conjunction with high-resolution imagery and digital elevation models (DEMs) from a 2017 aerial survey of the area, provided by the ELA. Any discrepancies in watercourse polylines and lake polygons from OIHD were corrected using both high-resolution imagery and field assessment observations (Appendix B).

Lake connectivity metrics, including average stream gradient, maximum stream gradient, stream sinuosity, and watercourse length, were calculated using tools in ArcGIS Pro (Appendix C). In cases where the maximum depth of a study lake was not available from historical data, I utilized a GIS approach established by Heathcote et al., (2015).

This method estimates maximum depth using the lake's surface area, elevation, and surrounding topography (Heathcote et al., 2015):

(F) $D = 2 * \sqrt{(A \div \pi)}$; where D is the equivalent diameter (m) and A is the lake area (km²).

(G) *25% buffer* = 25% of D (m)

(H) $\log_{10} Z_{\max} = 0.35 + \log_{10} \text{elevation change}_{25} * 0.79$ ($R^2 = 0.52$, $F_{[1,431]} = 473$, $P < 0.001$); Z_{\max} is the estimated maximum depth of the lakes (m), $\text{elevation change}_{25}$ is the difference between mean elevation within the *25% buffer* and the elevation of the lake surface (m).

Instances where maximum lake depth was estimated using this method are indicated in Appendix A.

2.5 Lake Connectivity Surveys

Connectivity between study lakes was examined via direct observation from May to August 2022. Preliminary research, which included reviewing hydrological maps, aerial imagery, digital elevation models (DEMs), and consulting with researchers at the ELA, guided this fieldwork.

During spring and summer months, field assessments were conducted to confirm and evaluate previously identified connections between lakes. Drones (DJI mini-2) were deployed along remote and longer streams to verify the connection path between lakes and identify potential fish barriers, such as waterfalls, chutes, or cascades. While temporary barriers, including beaver dams and logjams, exist on the landscape and may impact fish movement in the short-term (i.e. tens of years), our investigation focused on more permanent barriers on the landscape like waterfalls, chutes and cascades (i.e.,

hundreds to thousands of years; see below for definitions) and therefore having a more permanent impact on fish metacommunity structure (Davis et al., 2018).

The assessment of potential fish barriers followed the Alberta Conservation Association's (ACA) *Guide to Waterfall Fish Barrier Assessment* (Blackburn et al., n.d.). Given the absence of standardized assessment methods for fish barriers at both the provincial and federal levels in Ontario, I opted to utilize ACA's methodology due to its comprehensive approach for evaluating natural fish barriers. While originally designed for trout species, it was considered to be a conservative method for assessing upstream fish passage in our study area, given the limited ability of most fish species in the region to migrate upstream in watercourses (Gardunio, 2014; Spens & Ball, 2008).

Fish barriers eligible for assessment included waterfalls, chutes, and cascades. Waterfalls were defined by shifts in water velocity, resulting in the detachment of water from the streambed typically exceeding 23 degrees and with a minimum height of 0.75 meters. Chutes were characterized by steep gradients, usually less than 23 degrees, with water largely in contact with the streambed. Cascades were recognized as stream sections featuring a sequence of waterfalls and/or chutes of varying intensity and turbulence. Subsequent evaluations determined how the upstream movement of fish might be impeded, defined as a barrier mode. Four main barrier modes were evaluated: (1) leaping barriers, (2) swimming velocity barriers, (3) swimming depth barriers, and (4) swimming turbulence barriers (Blackburn et al., n.d.).

Following this assessment, barriers were assigned scores ranging from 0 to 3, with a score of 3 indicating highly unlikely fish passage upstream. After evaluating all watercourses connecting adjoining lakes, each received a score, and in cases of multiple

barriers on a single between-lake connection, the highest fish barrier score was chosen as the spatial weight during eigenvector mapping.

2.6 Statistical Approach

To evaluate the impact of lake connectivity and identify key spatial variables influencing fish metacommunity composition, I employed a Joint Species Distribution Model (JSDM) approach, following the Hierarchical Modeling of Species Communities (HMSC) framework outlined by Ovaskainen & Algebra (2020). To determine which spatial variables related to lake connectivity most strongly explained fish metacommunity composition amongst the study lakes in the absence of other factors (environmental, species associations), spatial-only JSDMs were developed. These models consisted exclusively of spatial eigenvectors representing different types of spatial variables related to lake connectivity (Blanchet et al., 2008b; Dray et al., 2006). Models were executed and ranked based on their explanatory and predictive power. Spatial variables identified by spatial-only JSDMs as having the greatest influence on fish community structure were retained for later modeling efforts (see below).

To determine the role of spatial variables (identified above) relative to environmental factors and species co-occurrence on fish metacommunity composition, a second JSDM approach was conducted. Full dataset JSDMs consisting of (A) environmental, spatial, and species co-occurrence latent variables, (B) exclusively environmental variables, (C) exclusively spatial variables, and (D) environmental and spatial variables were formulated and ranked based on their explanatory and predictive performance to determine the most effective combination of variables yielding the best performing models. Spatial variables selected from our best performing spatial-only

JSDM in the previous analysis were incorporated as the spatial variables in this assessment. After determining which set of variables produced the best performing model, overall and species-specific variance partitioning was conducted, identifying the relative role of each variable in explaining community and species-specific patterns of occurrence. Due to the nature of species co-occurrence latent variables, co-occurrence is unable to be modeled alone or without the inclusion of established fixed environmental and spatial variables (see Section 1.6; Ovaskainen & Abrego, 2020).

2.7 Spatial-only JSDM Modeling

Spatial-only JSDM models were implemented using R-packages *aem*, *adespatial*, and *Hmsc* (Blanchet et al., 2008b; Dray et al., 2006; Ovaskainen et al., 2017). Moran's eigenvector mapping (MEM) and asymmetric eigenvector mapping (AEM) were used to represent several different spatial metrics I wanted to investigate (Blanchet et al., 2008b; Dray et al., 2006). MEM eigenvectors were used to represent overland distance between lakes and hydrological connectivity between lakes (not considering directional flow) whereas AEM eigenvectors were used to represent hydrological connectivity in conjunction with the flow direction of watercourses.

Spatial weights were integrated into hydrological connectivity-based MEM and AEM eigenvectors to assess factors such as watercourse length, stream gradient, stream sinuosity, and physical fish barriers (waterfalls, chutes, and cascades quantified by fish barrier assessment scores; Appendix C). These weights were converted using a concave-down function ($f = 1 - d_{ij} / \max(d_{ij})$), with d_{ij} denoting a specified spatial weight between n sampling locations, and $\max(d_{ij})$ representing the largest spatial weight within the dataset (Blanchet et al., 2008b; Dray et al., 2006). Following the transformation of the lake

connectivity diagram (Figure 2.2) into a site-by-edges matrix (AEM) or site-by-site matrix (MEM), converted spatial weights were multiplied to each matrix, then converted into spatial eigenvectors through singular-value decomposition (AEM) or principal coordinate analysis (MEM). Spatial weights solely affecting fish movement upstream, such as stream gradient and fish barriers, were exclusively applied to AEM eigenvectors (Appendix D). Stream sinuosity, while potentially reducing a stream’s water velocity, may also promote habitat complexity and therefore facilitate bidirectional fish movement between lakes (Kaufmann & Robison, 1998). For this reason, sinuosity was considered an appropriate spatial weight for inclusion in both AEM and MEM eigenvectors.

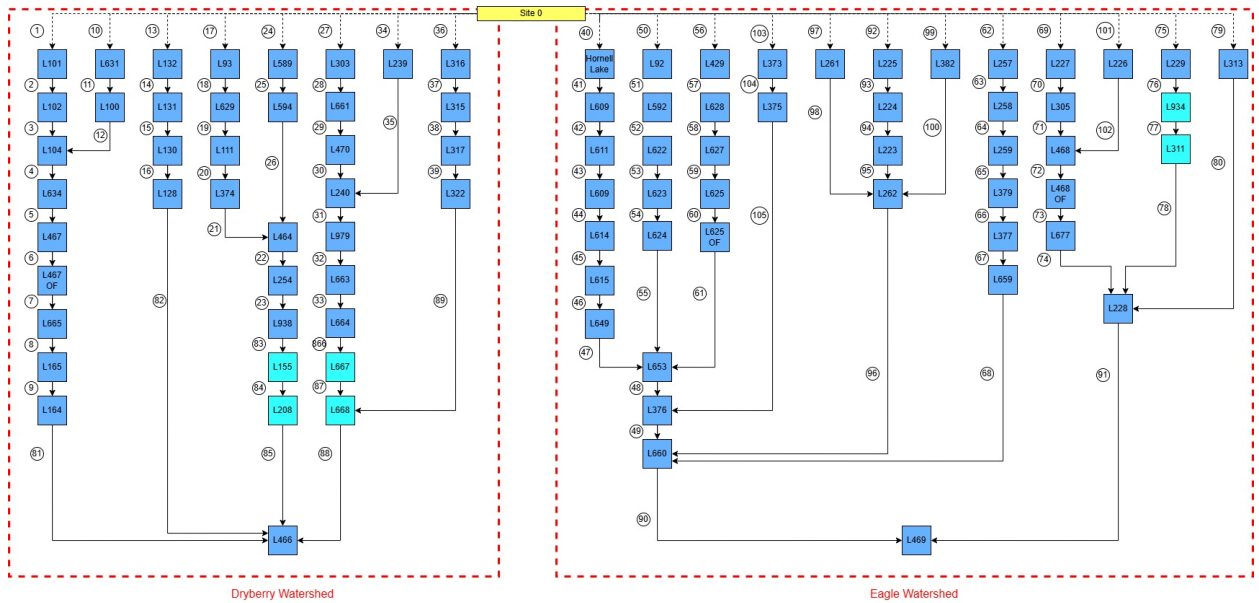


Figure 2.2. Lake connectivity diagram used for producing Moran’s Eigenvector Mapping (MEM) and Asymmetric Eigenvector Mapping (AEM). Dark blue boxes represent lakes included in Joint Species Distribution Models (JSDMs). Light blue boxes represent lakes not included in modeling but necessary to include for eigenvector mapping. Site 0 and associated connections represent a theoretical site upstream of all study lakes, necessary for AEM exclusively. The hypothetical site 0 introduces directionality into the diagram and establishes asymmetric spatial variables. To maximize the variance explained by spatial associations and minimize the likelihood of selecting eigenvectors describing only random spatial patterns, four spatial eigenvectors associated with the highest eigenvalues were selected for each model (Dray et al., 2006).

A hurdle model approach was applied to analyze presence-absence and relative abundance of fish communities within the HMSC framework, given the prevalence of zeros due to species absences. These models consist of two components: one addressing presence-absence and the other focusing on abundance conditional on presence. For the presence-absence model, species abundance data were transformed by converting non-zero values to one, with absences retained as zeros, and applying a binomial model with a probit link function to each species (probit model is easier to fit to data than a logistic model in HMSC; Ovaskainen & Abrego, 2020). Subsequently, for the abundance model, a separate dataset was created treating zeros as missing values, modeling abundances conditional of presence using a Poisson model. By running separate presence-absence and abundance (conditional on presence) analyses, we were able identify distinct drivers of a species presence and abundance (Odriozola et al., 2021). Relative abundance data, originally categorized as either rare, moderately abundant, or abundant, were converted into numerical values of 1, 2, and 3, respectively.

During model fitting, the posterior distribution was sampled using four Markov Chain Monte Carlo (MCMC) chains, each comprising 15,000 iterations, with the initial 5,000 iterations discarded as burn-in. To achieve 1,000 posterior samples per chain and a total of 4,000 posterior samples, the iterations were thinned by 200. Convergence of the MCMC chains was assessed by computing the potential scale reduction factor (PSRF) for parameters related to species responses to spatial variables, ensured that PSRF values did not exceed 1.2 (Brooks & Gelman, 1998; Gelman & Donald, 1992).

I evaluated the performance of spatial-only JSDM models based on indicators of both explanatory (mean Tjur R^2 , mean SR^2 , and mean Area under curve) and predictive

power (WAIC and mean 10-fold cross-validation: Tjur R^2 , SR^2 , Area under curve). Tjur R^2 (a pseudo- R^2) and Area under the curve (AUC) are measures of discrimination, asking how well the occurrence probabilities discriminate sampling units as either occupied or empty. The units of AUC and Tjur R^2 are different; a model that behaves ‘as well as by chance’ will yield an AUC of 0.5 and a Tjur R^2 of 0, whereas a model that perfectly discriminates empty and occupied sampling units will have an AUC of 1 and a Tjur R^2 of 1. SR^2 is considered a pseudo- R^2 . While R^2 in a linear model is calculated as the squared Pearson correlation between predicted and actual values, SR^2 is computed as the squared Spearman correlation between observed and predicted values. 10-fold cross validation refers to a technique used to evaluate the performance of a model by dividing the data into ten equal parts or folds. The model is trained on nine folds and tested on the remaining one, and this process is repeated ten times, with each fold used exactly once as the test data. The results from each iteration are averaged to provide an overall performance metric. The Widely Applicable Information Criterion (WAIC) is a measure used to evaluate the performance of statistical models, particularly in Bayesian analysis. WAIC estimates the out-of-sample prediction error by considering both the goodness of fit and the complexity of the model. A lower WAIC value indicates a better model, as it suggests a good balance between fitting the data well and maintaining simplicity to avoid overfitting.

For presence-absence data analysis, models with the highest explanatory and predictive mean Tjur R^2 and mean Area under Curve (AUC) values and lowest WAIC score were deemed as our best performing model. Conversely, for the relative abundance dataset, models yielding the highest explanatory and predictive mean SR^2 values and

lowest WAIC scores were considered best performing. Predictive power was calculated to identify overparameterization for any specific model.

2.7 Full Dataset JSMD Modeling

Full dataset JSMD models were conducted using the same framework and model specifications as the spatial-only JSMD models. Four environmental variables were selected for modeling: lake size (maximum lake depth), productivity (total phosphorus), acidity (pH), and light attenuation (dissolved organic carbon), based on their documented influence on fish presence and abundance (Beamish et al., 1976; Matuszek & Beggs, 1988; Mills et al., 1987; Rennie et al., 2019; Stasko et al., 2012).

Utilizing the most explanatory and predictive spatial variables identified from the spatial-only JSMD modeling, spatial eigenvectors for presence-absence modeling and relative abundance (conditional on presence) modeling were selected independently (Dray et al., 2006). Additionally, latent variables were introduced to account for species co-occurrences not explained by fixed variables. These latent variables were generated by integrating a community-level random effect at the sampling level (i.e., lake).

Similar to spatial-only JSMD modeling, a hurdle model approach was applied, following MCMC sampling as described previously. Explanatory and predictive power were also used to identify our best performing model as described above, which was then analyzed for further investigation of the region's metacommunity structure.

Results

3.1 Fish Abundance

A total of 44 fish species were found among the 81 sampled lakes included in the analysis. Species richness varied greatly among study lakes, ranging from 1 to 27 species (mean = 6, SD = 6). To avoid modeling issues due to low sample size, only species present in at least 10 lakes were selected for analysis; consequently, 16 fish species were included in JDSM modelling analyses (Table 3.1). Among these selected species, white sucker, yellow perch, and northern pearl dace were most frequently observed, occurring in 70, 58, and 53 lakes, respectively.

Table 3.1. Fish species included in JDSM modeling, with a minimum of 10 occurrences (i.e., present in at least 10 lakes).

Common Name	Latin Name	Number of occurrences
Blacknose Shiner	<i>Notropis heterolepis</i>	26
Burbot	<i>Lota lota</i>	15
Fathead Minnow	<i>Pimephales promelas</i>	27
Finescale Dace	<i>Chrosomus neogaeus</i>	25
Johnny Darter	<i>Etheostoma nigrum</i>	13
Lake Herring (Cisco)	<i>Coregonus artedi</i>	24
Lake Trout	<i>Salvelinus namaycush</i>	36
Lake Whitefish	<i>Coregonus clupeaformis</i>	20
Longnose Dace	<i>Rhinichthys cataractae</i>	19
Northern Pearl Dace	<i>Margariscus nachtriebi</i>	53
Northern Pike	<i>Esox lucius</i>	19
Northern Redbelly Dace	<i>Chrosomus eos</i>	18
Slimy Sculpin	<i>Cottus cognatus</i>	29
Spottail Shiner	<i>Notropis hudsonius</i>	15
White Sucker	<i>Catostomus commersonii</i>	70
Yellow Perch	<i>Perca flavescens</i>	53

3.2 Environmental Variables

Lake size varied greatly, with surface areas ranging from 0.4 ha to 9,546.0 ha (mean = 259.8 ha, SD = 1122.6 ha) and maximum depths ranging from 0.5 m to 167.0 m

(mean = 23.3 m, SD = 30.9 m). Maximum lake depth was chosen as the sole metric representing lake size due to its strong correlation ($r = 0.82$) with surface area and its known influence on various fish species (Fang et al., 2009; Johnson et al., 1977). To address the skewed distribution of maximum lake depth, the \log_{10} values of maximum lake depth were used in the models.

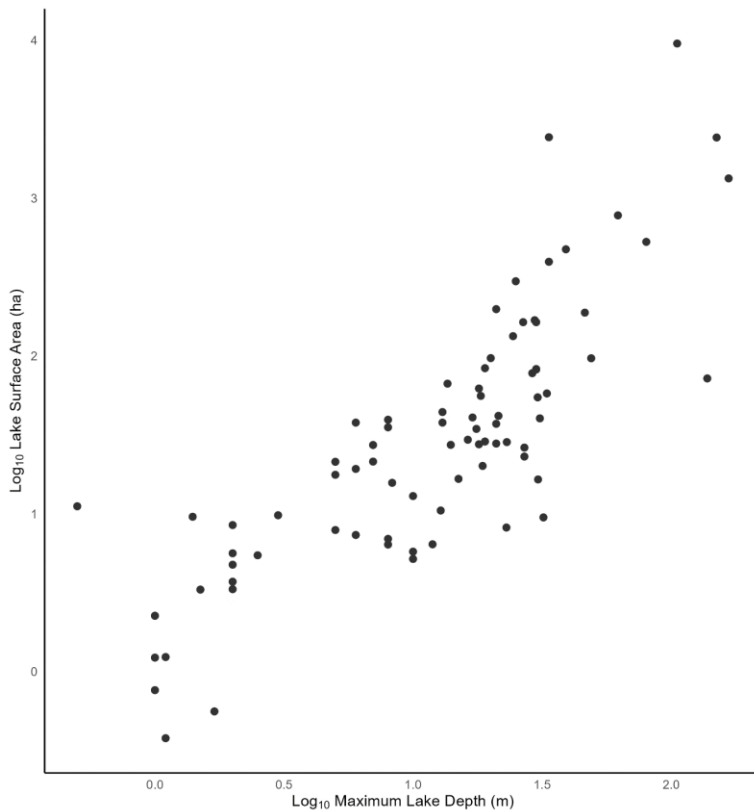


Figure 3.1. Correlation between \log_{10} values of maximum depth (m) and lake surface area (ha) amongst 81 study lakes ($r = 0.82$).

Acidity, represented by mean pH levels, varied from 5.15 to 7.88 across study lakes (mean = 6.77, SD = 0.46). Lake productivity, represented by mean TP levels, ranged from 3.0 $\mu\text{g/L}$ to 41.5 $\mu\text{g/L}$ (mean = 10.1 $\mu\text{g/L}$, SD = 6.7 $\mu\text{g/L}$). Due to the relatively strong correlation between TP and TN levels, TP was chosen as the sole variable to represent lake productivity in the analyses ($r = 0.55$; Figure 3.2). Given the

skewed distribution of TP, \log_{10} values of TP were used. Mean DOC levels, representing light attenuation, ranged from 3.30 mg/L to 15.01 mg/L (mean = 7.67 mg/L, SD = 2.64 mg/L). To address the skewed distribution of DOC, \log_{10} values of DOC were used. All lake-specific values of environmental factors can be found in Appendix E.

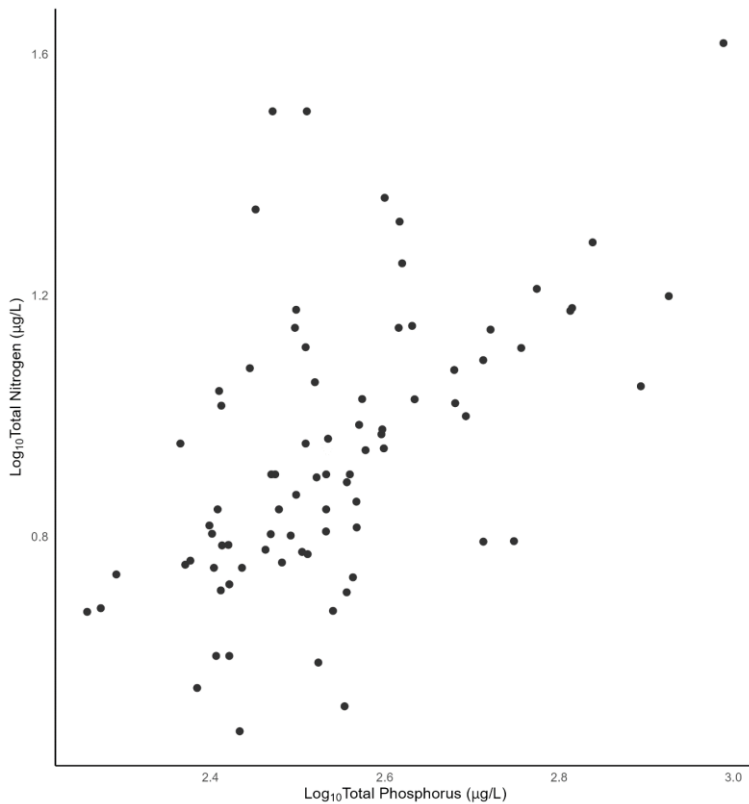


Figure 3.2. Correlation between \log_{10} values of total nitrogen ($\mu\text{g/L}$) and total phosphorus ($\mu\text{g/L}$) amongst 81 study lakes ($r = 0.55$).

3.3 Spatial-only JSDM Model

3.3.1 Presence-Absence

A total of 11 spatial-only JSDM models (SM1–SM11, where SM stands for spatial-only model; Table 3.2) were run to assess fish presence-absence. Spatial-only JSDM models, designed to identify key spatial variables influencing the presence or absence of fish species, demonstrated varying levels of predictive and explanatory power,

with model SM1 (an AEM model using maximum stream gradient as a weighting variable; Table 3.2) exhibiting the highest performance in both metrics. As an AEM model, SM1 consisted of spatial eigenvectors representing lake connectivity, stream flow direction, and included the maximum stream gradient along a watercourse as a spatial weight. The first four spatial eigenvectors selected for SM1 explained 85.33% of the spatially weighted connectivity matrix variance, effectively capturing most of the spatial variance related to lake connectivity, flow direction, and maximum stream gradients. Models using AEM eigenvectors (SM1-SM7) consistently outperformed all MEM (SM8-SM11) models, highlighting the importance of including stream flow direction when investigating fish metacommunities. Additionally, Spatial-only JSDM models utilizing either AEM and MEM eigenvectors that included spatially weights by stream sinuosity demonstrated the poorest performance compared to other models, indicating stream sinuosity as a poor explanatory variable for estimating fish metacommunity structure.

Table 3.2. Explanatory (Mean Tjur R², AUC) and predictive power (10-fold Cross Validation Mean Tjur R², 10-fold Cross Validation AUC, WAIC) of Spatial JSDM models for presence-absence of fish species. Spatial JSDM models are assessed to identify appropriate spatial descriptors for Full Dataset JSDM models. Models with poor cross-validation values suggest poor model fit.

Spatial Model	Eigenvector Type	Spatial Weight	Mean Tjur R ² ¹	10-fold Cross Validation Mean Tjur R ² ²	AUC ¹	10-fold Cross Validation AUC ²	WAIC ³
SM1	AEM	Maximum stream gradient along watercourse	0.1737	0.1226	0.7235	0.6416	8.3550
SM2	AEM	Hydrological connectivity only	0.1722	0.1127	0.7224	0.6053	8.3644
SM3	AEM	Average stream gradient along watercourse	0.1719	0.1206	0.7229	0.6382	8.3652
SM4	AEM	Hydrological watercourse length	0.1683	0.1089	0.7186	0.6032	8.3981
SM5	AEM	Maximum fish barrier assessment score along watercourse	0.1678	0.1168	0.7194	0.6332	8.4091
SM6	AEM	Presence of fish barrier	0.1678	0.1169	0.72034	0.6332	8.4161

SM7	AEM	Stream sinuosity along watercourse	0.1358	0.0797	0.70445	0.5901	8.7475
SM8	MEM	Hydrological watercourse length	0.0743	0.0279	0.65218	0.4877	9.2910
SM9	MEM	Overland distance between lakes	0.0723	0.0144	0.6812	0.5126	9.2258
SM10	MEM	Hydrological connectivity only	0.0641	0.0161	0.65192	0.4594	9.4358
SM11	MEM	Stream sinuosity along watercourse	0.0420	-0.0248	0.6114	0.3825	9.4027

¹ Tjur R² and Area under the curve (AUC) are measures of discrimination, asking how well the occurrence probabilities discriminate sampling units as either occupied or empty. The units of AUC and Tjur R² are different; a model that behaves ‘as well as by chance’ will yield an AUC of 0.5 and a Tjur R² of 0, whereas a model that perfectly discriminates empty and occupied sampling units will have an AUC of 1 and a Tjur R² of 1.

² 10-fold cross validation refers to a technique used to evaluate the performance of a model by dividing the data into ten equal parts or folds. The model is trained on nine folds and tested on the remaining one, and this process is repeated ten times, with each fold used exactly once as the test data. The results from each iteration are averaged to provide an overall performance metric.

³ The Widely Applicable Information Criterion (WAIC) is a measure used to evaluate the performance of statistical models, particularly in Bayesian analysis. WAIC estimates the out-of-sample prediction error by considering both the goodness of fit and the complexity of the model. A lower WAIC value indicates a better model, as it suggests a good balance between fitting the data well and maintaining simplicity to avoid overfitting.

1 **3.3.2 Relative Abundance**

2 Spatial-only JSDM models were also evaluated for assessing the relative
3 abundance (conditional on presence) of fish species. The best performing model in terms
4 of explanatory power was SM12, which incorporated MEM eigenvectors describing
5 overland distance between study lakes (regardless of connectivity). Apart from model
6 SM12, models using AEM eigenvectors (SM13-SM19) consistently outperformed other
7 MEM (SM20-SM22) models, similar to presence-absence spatial models. Despite these
8 apparent patterns in explanatory power, all spatial-only JSDM relative abundance models
9 exhibited very poor predictive power, suggesting poor model fit. The negative values of
10 10-fold cross validation mean SR^2 for all relative abundance models indicate that these
11 models' ability to predict fish metacommunity composition is worse than a null model,
12 which predicts the overall mean of the response variable (community composition)
13 without accounting for predictor variables (spatial eigenvectors). Although a SR^2 is not
14 actually able to be negative, HMSC reverts the sign if the correlation coefficient is
15 negative to indicate poor model fit (Ovaskainen & Abrego, 2020).

Table 3.3. Explanatory (Mean SR^2) and predictive power (10-fold Cross Validation Mean SR^2 , and WAIC) of Spatial-only JSDM Models for relative abundance (conditional on presence) of fish species. Models with poor cross-validation values suggest poor model fit.

Model	Eigenvector Type	Spatial Weight	Mean SR^2 ¹	10-fold Cross Validation Mean SR^2 ²	WAIC ³
SM12	MEM	Overland distance between lakes	0.2112	-0.0313	41.2293
SM13	AEM	Average stream gradient along watercourse	0.1976	-0.0129	41.2872
SM14	AEM	Stream sinuosity along watercourse	0.1937	-0.0126	41.2801
SM15	AEM	Hydrological connectivity only	0.1931	-0.0146	41.2802
SM16	AEM	Maximum stream gradient along watercourse	0.1864	-0.0183	41.2879
SM17	AEM	Hydrological watercourse length	0.1842	-0.0132	41.2794
SM18	AEM	Maximum fish barrier assessment score along watercourse	0.1793	-0.0498	41.2974
SM19	AEM	Presence of fish barrier	0.1761	-0.0509	41.2975
SM20	MEM	Hydrological connectivity only	0.1667	-0.0448	41.2884
SM21	MEM	Hydrological watercourse length	0.1414	-0.0644	41.2873
SM22	MEM	Stream sinuosity along watercourse	0.1000	-0.1029	41.2092

¹ SR^2 is considered a pseudo- R^2 . While R^2 in a linear model is calculated as the squared Pearson correlation between predicted and actual values, SR^2 is computed as the squared Spearman correlation between observed and predicted values.

² 10-fold cross validation refers to a technique used to evaluate the performance of a model by dividing the data into ten equal parts or folds. The model is trained on nine folds and tested on the remaining one, and this process is repeated ten times, with each fold used exactly once as the test data. The results from each iteration are averaged to provide an overall performance metric.

³ The Widely Applicable Information Criterion (WAIC) is a measure used to evaluate the performance of statistical models, particularly in Bayesian analysis. WAIC estimates the out-of-sample prediction error by considering both the goodness of fit and the complexity of the model. A lower WAIC value indicates a better model, as it suggests a good balance between fitting the data well and maintaining simplicity to avoid overfitting.

3.4 Full Dataset JSDM Models

3.4.1 Presence-Absence Models

Among the Full Dataset JSDM models evaluating the presence and absence of fish species, Model ESC (Environment + Space + Co-occurrence), which incorporated environmental, spatial (SM1), and species co-occurrence latent variables, had the highest explanatory power and one of the highest predictive powers of all models considered (Table 3.4; Appendix F). The notable explanatory and predictive performance of Model ESC suggests that considering environmental, spatial, and species co-occurrence latent variables collectively are necessary for investigating fish metacommunity composition (as presence-absence), providing justification for its use in subsequent analyses.

Table 3.4. Summary of Full Dataset JSJM models (with optimized spatial predictors from SM1) for presence-absence of fish species. Model performance is ranked based on explanatory (Mean Tjur R², AUC) and predictive (10-fold Cross Validation Mean Tjur R², 10-fold Cross Validation AUC, WAIC) power of each model. Models with poor cross-validation values suggest poor model fit.

Model	Eigenvector Type	Mean Tjur R ² ¹	10-fold Cross Validation Mean Tjur R ² ²	AUC ¹	10-fold Cross Validation AUC ²	WAIC ³
Environment + Spatial + Co-occurrence (ESC)	AEM Max Gradient	0.4012	0.1949	0.9044	0.7165	6.9068
Environment + Spatial (ES)	AEM Max Gradient	0.2965	0.2091	0.8358	0.7196	7.5980
Spatial Only (S)	AEM Max Gradient	0.1737	0.1143	0.7233	0.6187	8.3574
Environment Only (E)	AEM Max Gradient	0.1660	0.1091	0.7466	0.6372	8.2480

¹ Tjur R² and Area under the curve (AUC) are measures of discrimination, asking how well the occurrence probabilities discriminate sampling units as either occupied or empty. The units of AUC and Tjur R² are different; a model that behaves 'as well as by chance' will yield an AUC of 0.5 and a Tjur R² of 0, whereas a model that perfectly discriminates empty and occupied sampling units will have an AUC of 1 and a Tjur R² of 1.

² 10-fold cross validation refers to a technique used to evaluate the performance of a model by dividing the data into ten equal parts or folds. The model is trained on nine folds and tested on the remaining one, and this process is repeated ten times, with each fold used exactly once as the test data. The results from each iteration are averaged to provide an overall performance metric.

³ The Widely Applicable Information Criterion (WAIC) is a measure used to evaluate the performance of statistical models, particularly in Bayesian analysis. WAIC estimates the out-of-sample prediction error by considering both the goodness of fit and the complexity of the model. A lower WAIC value indicates a better model, as it suggests a good balance between fitting the data well and maintaining simplicity to avoid overfitting.

3.4.2 Relative Abundance Models

Among the Full Dataset JSMD models evaluating the relative abundance of fish species, Model C had the highest explanatory power (Table 3.5). However, similar to the relative abundance spatial-only models, all tested models for relative abundance (including Model C) demonstrated poor predictive power (10-fold cross validation SR^2 and WAIC), suggesting poor model fit generally (Table 3.5). The large difference between explanatory and predictive performance suggests that relative abundance, conditional on presence, is poorly explained by the tested variables and should therefore not be analyzed further. The negative values of 10-fold cross validation mean SR^2 (see Section 3.3.2) for all relative abundance models suggests that the ability of these models to predict fish metacommunity composition is worse than a null model (i.e. one that predicts the overall mean of the response variable (community composition) without considering the predictor variables). This suggests poor model fit, overparameterization within the models, and/or selection of inappropriate variables. The especially high mean SR^2 value observed with Model C is due to the model solely using latent variables as explanatory variables, meaning no environmental or spatial variable were included. These latent variables were produced to explain the most variation in the dataset and do not necessarily reflect any actual variables at play influencing relative abundance.

Table 3.5. Summary of Full Dataset JSDM models (with optimized spatial predictors) for relative abundance (conditional on presence) of fish species. Model performance is ranked based on explanatory (Mean Tjur R^2 , AUC) and predictive (10-fold Cross Validation Mean Tjur R^2 , 10-fold Cross Validation AUC, WAIC) power of each model.

Model	Eigenvector Type	Mean SR^2 ¹	10-fold Cross Validation Mean SR^2 ²	WAIC ³
Environment + Spatial + Co-occurrence (ESC)	MEM connectivity	0.3664	-0.0062	41.5807
Environment + Spatial (ES)	MEM connectivity	0.2868	-0.0181	41.5182
Spatial Only (S)	MEM connectivity	0.1910	-0.0746	41.2835
Environment Only (E)	MEM connectivity	0.1871	-0.0974	41.2248

¹ SR^2 is considered a pseudo- R^2 . While R^2 in a linear model is calculated as the squared Pearson correlation between predicted and actual values, SR^2 is computed as the squared Spearman correlation between observed and predicted values.

² 10-fold cross validation refers to a technique used to evaluate the performance of a model by dividing the data into ten equal parts or folds. The model is trained on nine folds and tested on the remaining one, and this process is repeated ten times, with each fold used exactly once as the test data. The results from each iteration are averaged to provide an overall performance metric.

³ The Widely Applicable Information Criterion (WAIC) is a measure used to evaluate the performance of statistical models, particularly in Bayesian analysis. WAIC estimates the out-of-sample prediction error by considering both the goodness of fit and the complexity of the model. A lower WAIC value indicates a better model, as it suggests a good balance between fitting the data well and maintaining simplicity to avoid overfitting.

3.5 Model ESC Presence-Absence Analyses

3.5.1 Overall Exploratory Power

Due to the high explanatory and predictive power of Model ESC in assessing the presence and absence of fish species, further analyses with this model were deemed appropriate. Of the explained variance by Model ESC, 52.6% was explained by spatial variables 36.0% by environmental variables, and 11.4% by species co-occurrence latent variables (Figure 3.3). Of the species modeled, lake trout, northern redbelly dace, and northern pike were best explained overall, whereas blacknose shiner was the least explained, with respective Tjur R^2 values of 0.62, 0.61, 0.58, and 0.12 respectively (Figure 3.3).

Explained variance of each factor varied considerably between species (Figure 3.4). Overall, spatial eigenvector 2 and the maximum lake depth (i.e., lake size) explained the most variance across all species at 32.7% and 20.4% each. Total phosphorus (i.e., lake productivity) and DOC (i.e., light attenuation) explained the least variance across all species at 4.7% each. Lake trout, lake herring, slimy sculpin, burbot, and lake whitefish, many of which are considered cold-water species, were best explained by environmental variables, with maximum lake depth (i.e. lake size) being the most important environmental explanatory variable (Table 3.6). Northern pike, yellow perch, white sucker, johnny darter, finescale dace, spottail shiner, blacknose shiner, and longnose dace were best explained by spatial variables, specifically spatial eigenvector 2, except longnose dace, which was best explained by 3 (eigenvectors defined below). Northern redbelly dace and fathead minnow were best explained by species co-occurrence latent variables.

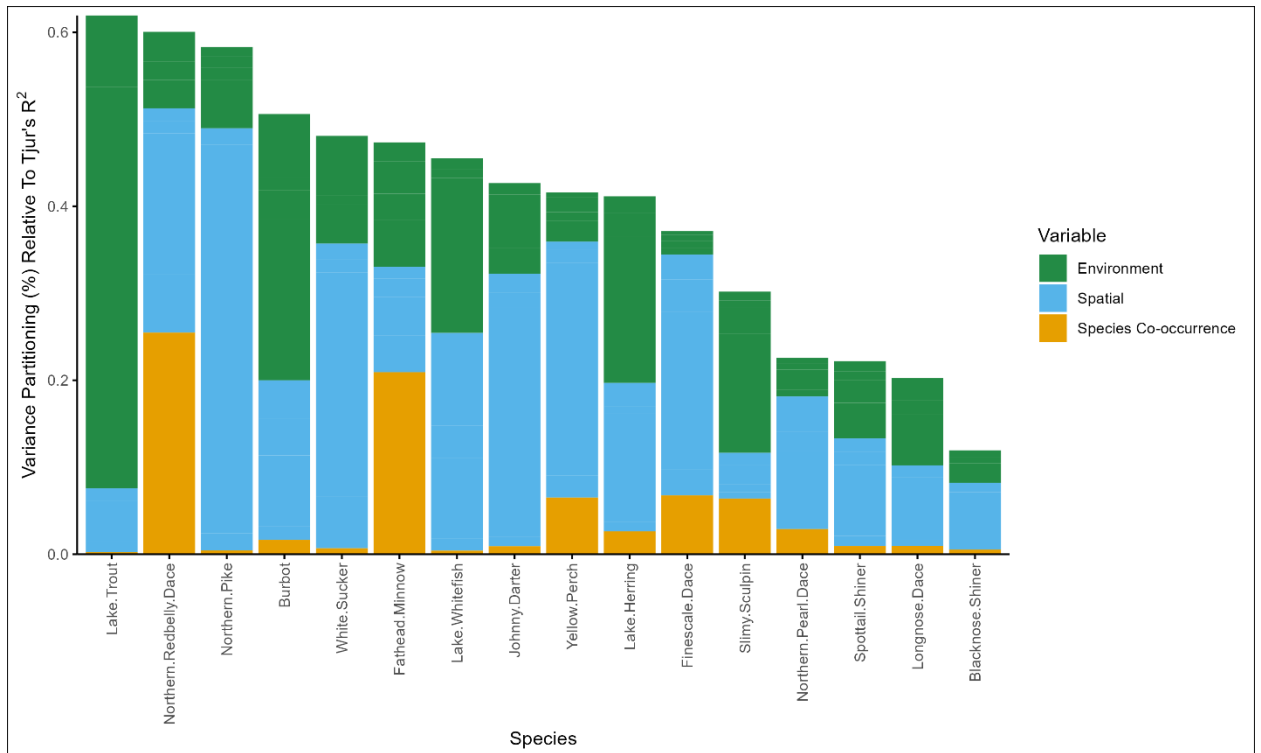


Figure 3.3. Variance partitioning (%) relative to presence-absence Model ESC's Tjur's R^2 for each fish species. The height of bars indicates the Tjur's R^2 value associated with each species. Green represents cumulative environmental variables, blue represents cumulative spatial variables (i.e., selected spatial eigenvectors), and orange represents species co-occurrence latent variables.

Table 3.6. Main explanatory variables and associated explained variance (%) values for each fish species included in Model ESC for presence-absence.

Species	Highest Explanatory Variable Type		Highest Explanatory Variable	
Blacknose Shiner	Spatial	64.2%	Spatial Eigenvector 2	34.4%
Burbot	Environment	60.7%	Maximum Lake Depth	37.0%
Fathead Minnow	Species Co- occurrence	44.41%	Species Co-occurrence	44.4%
Finescale Dace	Spatial	75.30%	Spatial Eigenvector 2	49.7%
Johnny Darter	Spatial	73.4%	Spatial Eigenvector 2	61.8%
Lake Herring	Environment	50.9%	Maximum Lake Depth	40.0%
Lake Trout	Environment	87.8%	Maximum Lake Depth	65.0%
Lake Whitefish	Environment	42.5%	Maximum Lake Depth	29.3%
Longnose Dace	Spatial	47.3%	Spatial Eigenvector 3	29.0%
Northern Pearl Dace	Spatial	68.3%	Spatial Eigenvector 2	33.9%
Northern Pike	Spatial	83.4%	Spatial Eigenvector 2	73.1%
Northern Redbelly Dace	Species Co- occurrence	44.1%	Species Co-occurrence	44.1%
Slimy Sculpin	Environment	62.5%	Maximum Lake Depth	46.0%
Spottail Shiner	Spatial	55.2%	Spatial Eigenvector 2	36.3%
White Sucker	Spatial	73.4%	Spatial Eigenvector 2	54.4%
Yellow Perch	Spatial	69.7%	Spatial Eigenvector 2	58.1%

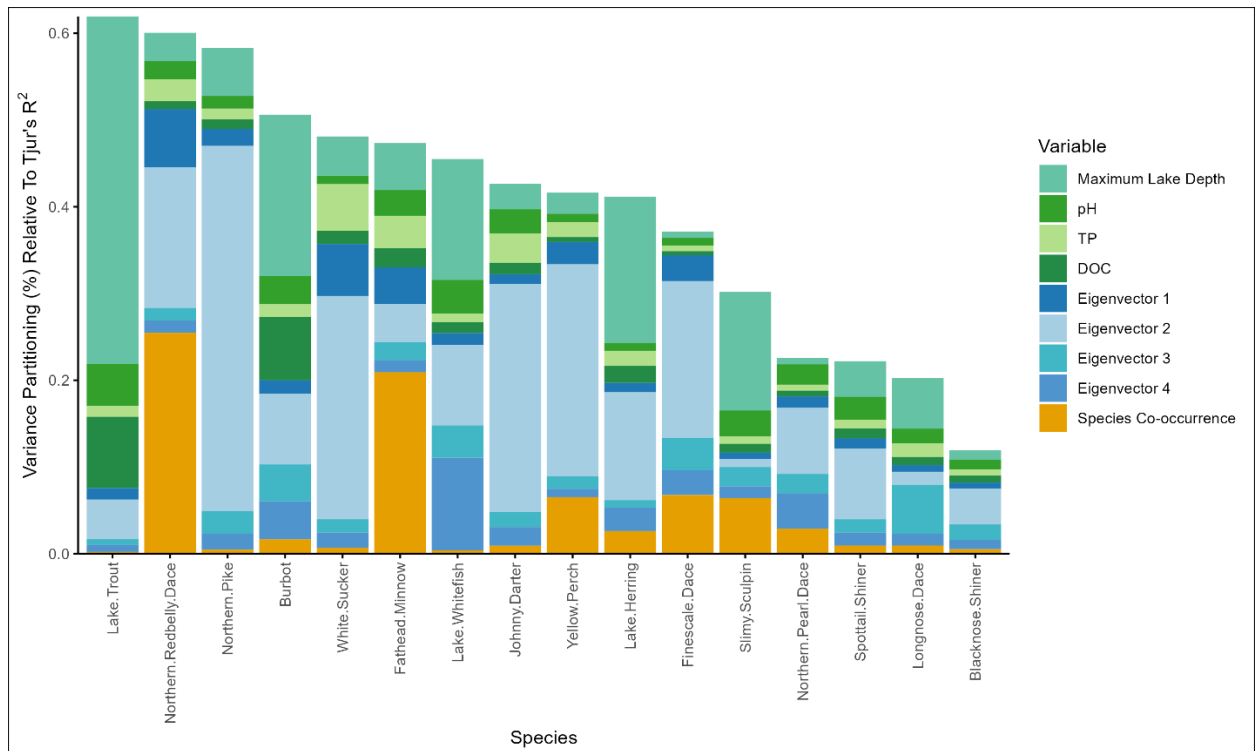


Figure 3.4. Variance partitioning (%) relative to presence-absence Model ESC's Tjur's R^2 for each fish species. The height of bars indicates the Tjur's R^2 value associated with each species. Green shades represent environmental variables, blue shades (1-4) represent spatial eigenvector variables, and orange represents species co-occurrence latent variables.

3.5.2 Spatial Eigenvectors

Although spatial eigenvectors 1-4 in Model ESC collectively represent lake connectivity, stream flow direction, and maximum stream gradient, each eigenvector individually highlights key specific spatial patterns. To interpret visually the influence of each of these eigenvectors in describing spatial variation, values of each spatial eigenvector were plotted in relation to the spatial occurrence of the lakes which they describe. Spatial eigenvector 1 appears to highlight quaternary watershed boundaries found in the study area, with positive values representing Eagle watershed and negative values representing Dryberry watershed (Figure 3.5). Additionally, sites associated with higher absolute eigenvalues appear to be further downstream in lake chains, whereas

lower absolute eigenvalues seem to be located further upstream (Figure 3.5). Spatial eigenvector 2 appears to more strongly emphasize differences between lakes furthest downstream to headwater lakes and mid-level lakes, as highlighted by the strong negative correlation between lake order and eigenvector 2's values ($r_s = -0.73$; Figure 3.6). Spatial eigenvectors 3 and 4 appear to represent distinct chains of lakes present in each study area (Figure 3.5). Identifying which spatial eigenvectors most represent the influence of maximum stream gradient along connecting watercourses appears to be somewhat challenging. Of the four spatial eigenvectors included in Model ESC, eigenvector 2 had the strongest explanatory power, suggesting a strong association between lake order and community composition.

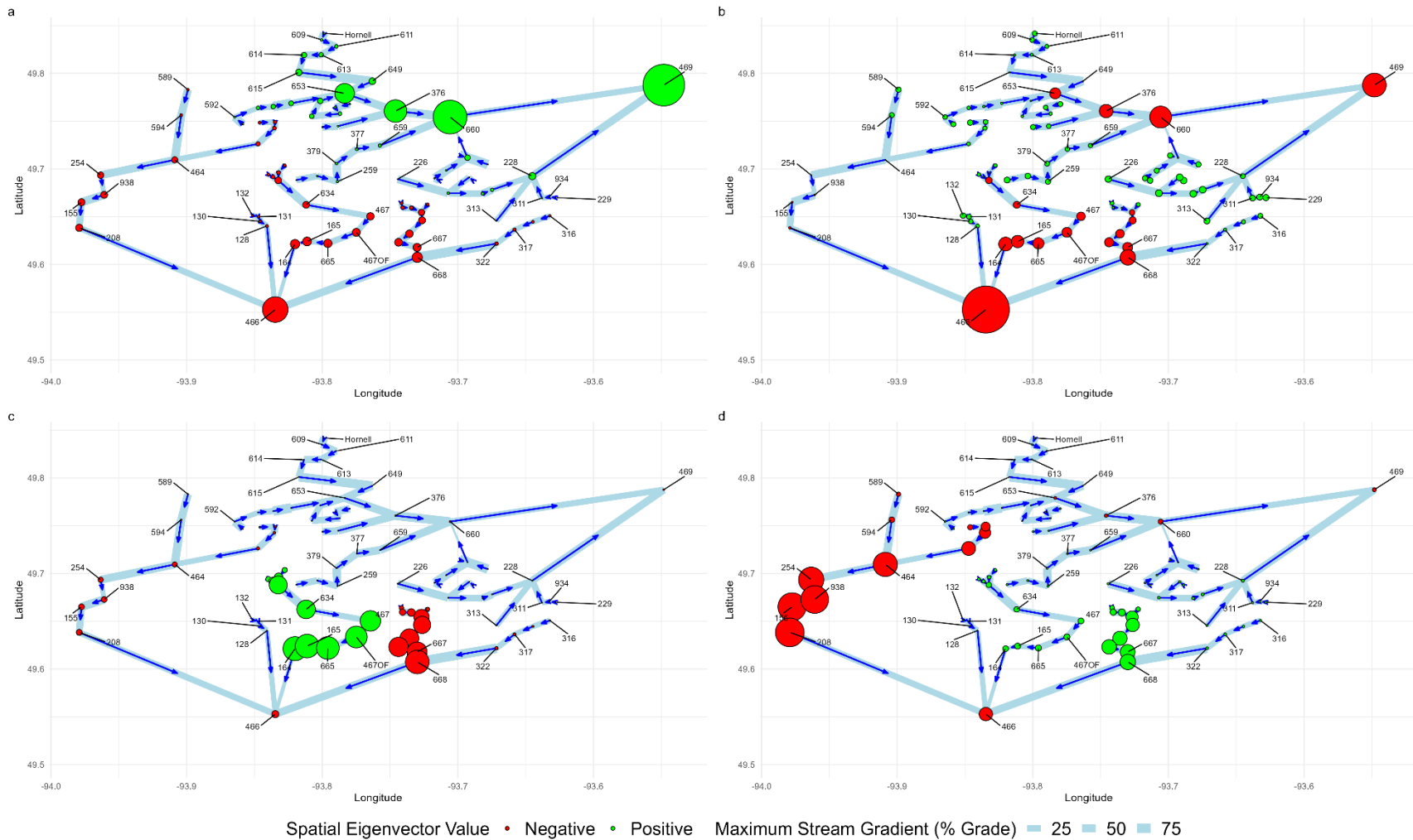


Figure 3.5. Visual representation of Spatial AEM eigenvectors 1-4 (a-d) used in Model ESC. AEM eigenvectors represent lake connectivity, flow direction of connecting watercourses, and the maximum gradient along a connecting watercourse. Each dot

represents a lake. Green dots are positive, red dots are negative. The size of the dot increases as the absolute eigenvector values increase. The width of the light blue polylines represent the maximum stream gradient found along with connecting watercourse.

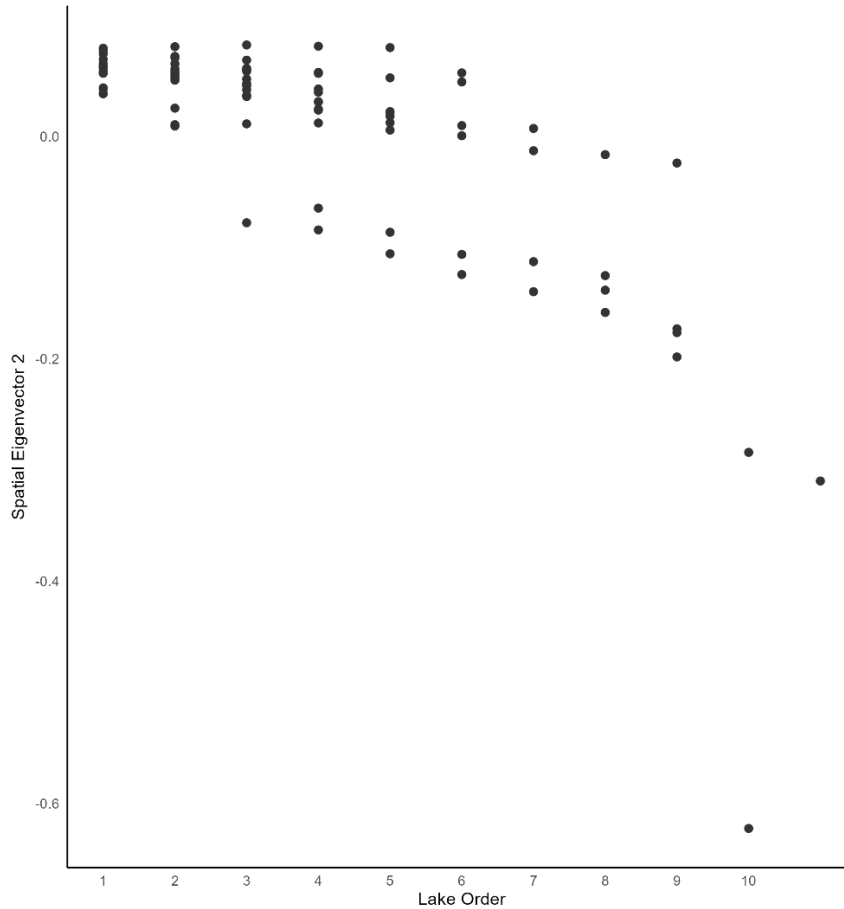


Figure 3.6 Spearman correlation between lake order and spatial eigenvector 2 amongst 81 study lakes ($r_s = -0.73$, $p < 0.05$).

3.5.3 Species Responses to Environmental and Spatial Variables

Nearly every environmental and spatial variable in Model ESC, excluding pH, had a strong positive or negative association with at least one species (>95% posterior probability; Figure 3.7). Maximum lake depth and spatial eigenvector 2 specifically had the most positive or negative associations with 9 and 11 out of 16 species, respectively.

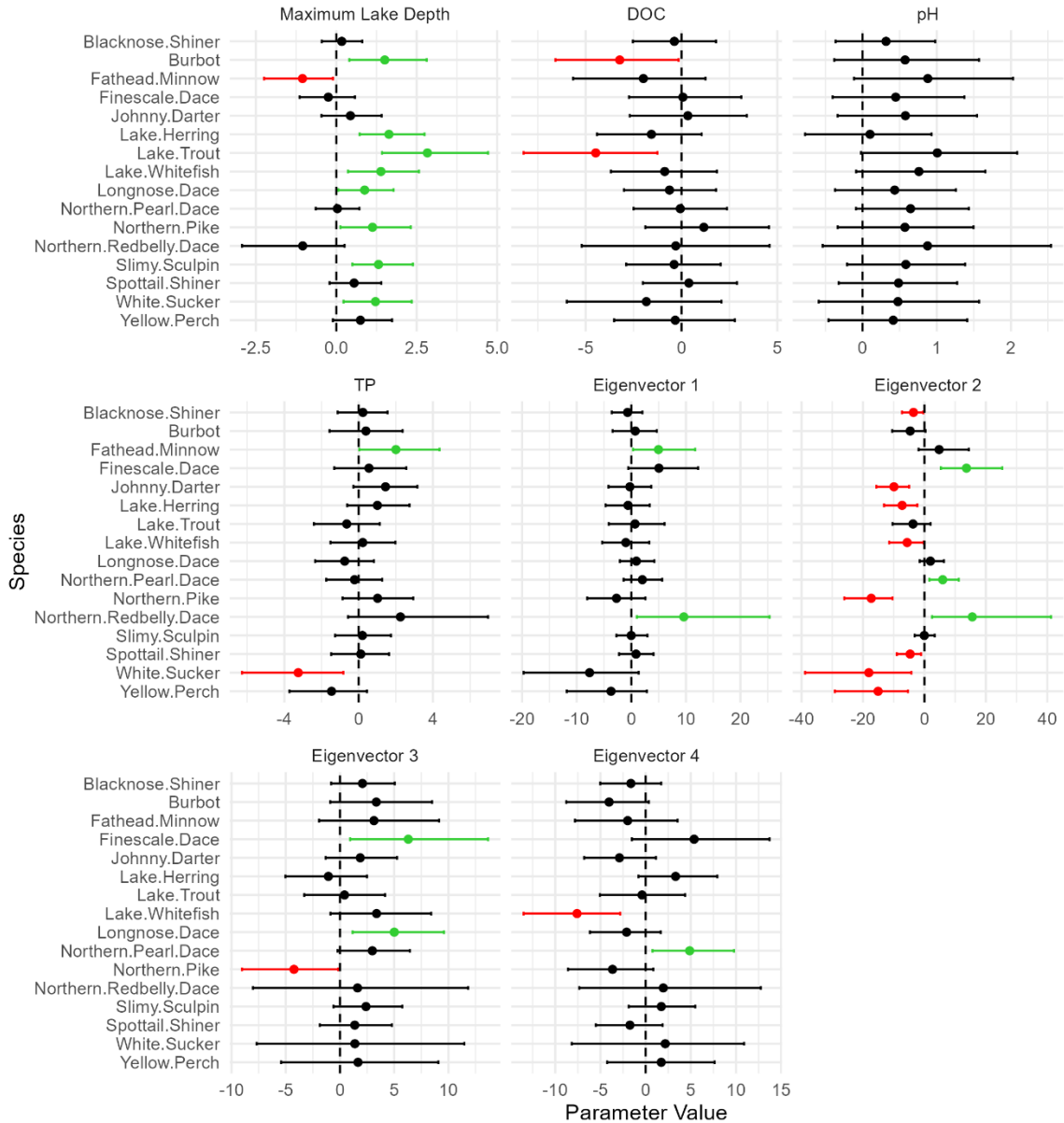


Figure 3.7. Posterior distributions (means and 95% credible intervals) of regression coefficients for environmental and spatial eigenvector variables, by species. The coefficients are derived from the presence-absence Model ESC. Green indicates a variable has a strong positive association with a species. Red indicates a variable has a strong negative association with a species. Species with neutral environmental or spatial associations are shown in black.

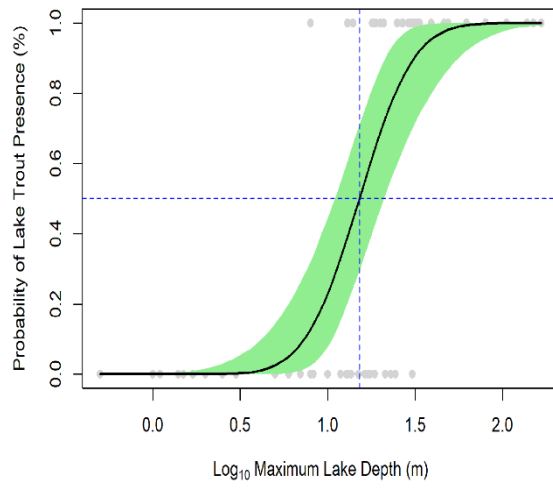
3.5.4 Environmental Drivers of Fish Presence

Environmental variables accounted for 36.0% of the total variance explained across all species (Figure 3.4). Of this explained variance, maximum lake depth —

representing lake size in Model ESC — contributed 20.4%, with lake trout, burbot, and slimy sculpin showing particularly strong associations with this variable. Dissolved organic carbon (DOC), representing light attenuation, explained 4.7%, with lake trout and burbot being especially influenced by this factor. pH and total phosphorus contributed less to the overall variance, explaining 6.2% and 4.7% of the total explained variance, respectively, with few species showing particularly strong positive or negative associations with either variable.

Lake trout ($T_{\text{jur}} R^2 = 0.619$, 10-fold Cross Validation $T_{\text{jur}} R^2 = 0.521$) was best explained by maximum lake depth and DOC (Figures 3.8). Based on presence-absence Model ESC, lake trout has strong positive associations with maximum lake depth and a negative association with DOC. The \log_{10} maximum lake depth value where presence probability reaches 0.5 was 1.182 (95% CI: 1.055 to 1.308), equal to an actual maximum lake depth of 15.2 m (95% CI: 11.4 m to 20.3 m; Figure 3.8). The \log_{10} DOC value where presence probability reaches 0.5 was 2.744 (95% see CI: 2.694 to 2.785) equal an actual DOC value of 6.66 mg/L (95% CI: 5.94 mg/L to 7.31 mg/L; Figure 3.8). The actual and predicted distribution of lake trout based on all model predictors is displayed in Figure 3.9. If we use 0.5 probability as the cutoff for likely species presence, Model ESC correctly predicted the presence and absence of lake trout for 71 of the 81 study lakes.

a



b

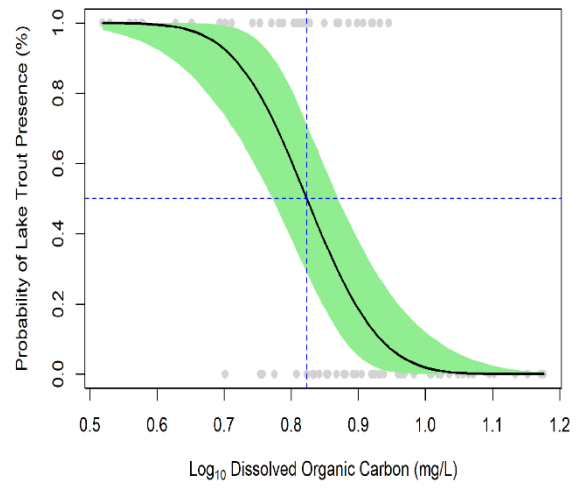


Figure 3.8. Response of lake trout presence probability to (a) log₁₀ value of maximum lake depth (m) and (b) the log₁₀ value of dissolved organic carbon (mg/L). Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

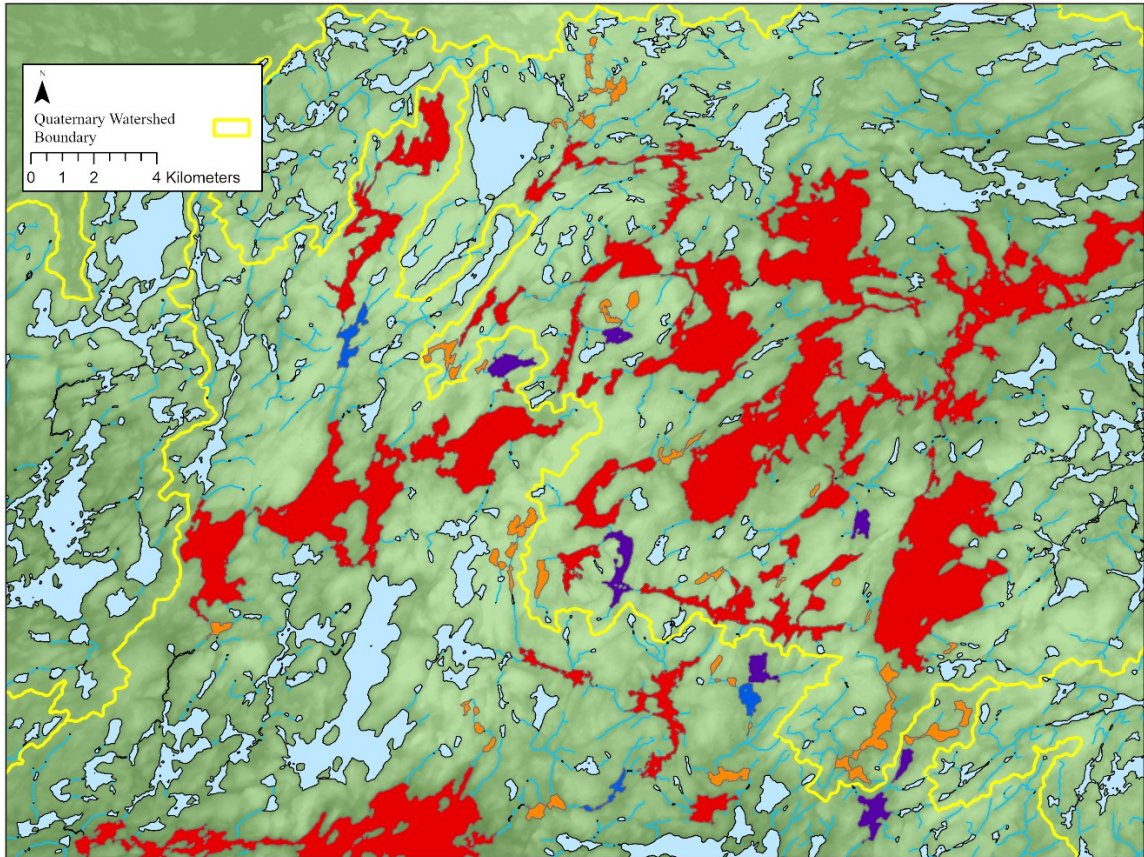


Figure 3.9. Actual and predicted distribution of lake trout across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where lake trout were predicted to be present but were not observed. Purple represents lakes where lake trout were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

For burbot (Tjur $R^2 = 0.506$, 10-fold Cross Validation Tjur $R^2 = 0.376$), maximum lake depth and DOC were also the key explanatory variables (Figure 3.10). The presence-absence Model ESC suggests that burbot strongly favours deeper lakes and lower DOC levels. The \log_{10} maximum lake depth where the likelihood of presence reaches 0.5 is 1.718 (95% CI: 1.402 to 2.222), translating to an actual depth of 52.2 m (95% CI: 25.3 m to 167.0 m; Figure 3.10). Given the correlation between maximum depth and surface area (see Figure 3.1), the corresponding lake surface area at this probability level is 174.4 ha

(95% CI: 73.0 ha to 703.5 ha). For DOC, the \log_{10} value for 0.5 presence probability is 2.579 (95% CI: 2.439 to 2.678), equivalent to an actual concentration of 6.65 mg/L (95% CI: 4.40 mg/L to 7.82 mg/L; Figure 3.10). The actual and predicted distributions of burbot, incorporating all model predictors, are shown in Figure 3.11. With a 0.5 probability cutoff, Model ESC accurately predicted burbot presence and absence in 74 of the 81 study lakes.

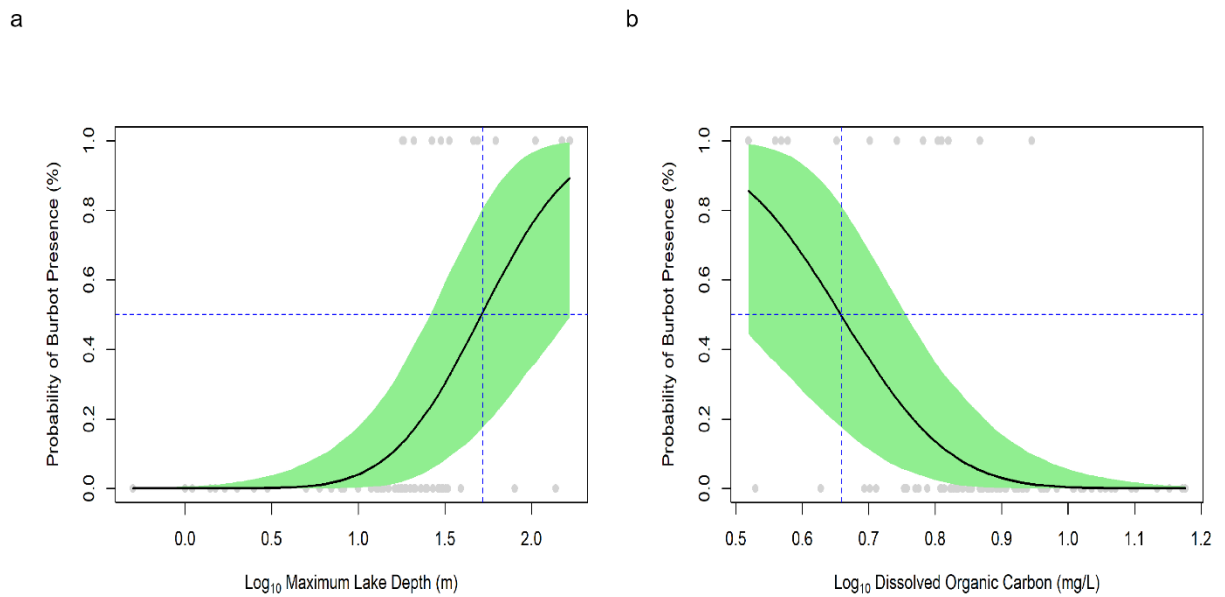


Figure 3.10. Response of burbot presence probability to (a) \log_{10} value of maximum lake depth (m) and (b) \log_{10} value of dissolved organic carbon (mg/L). Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

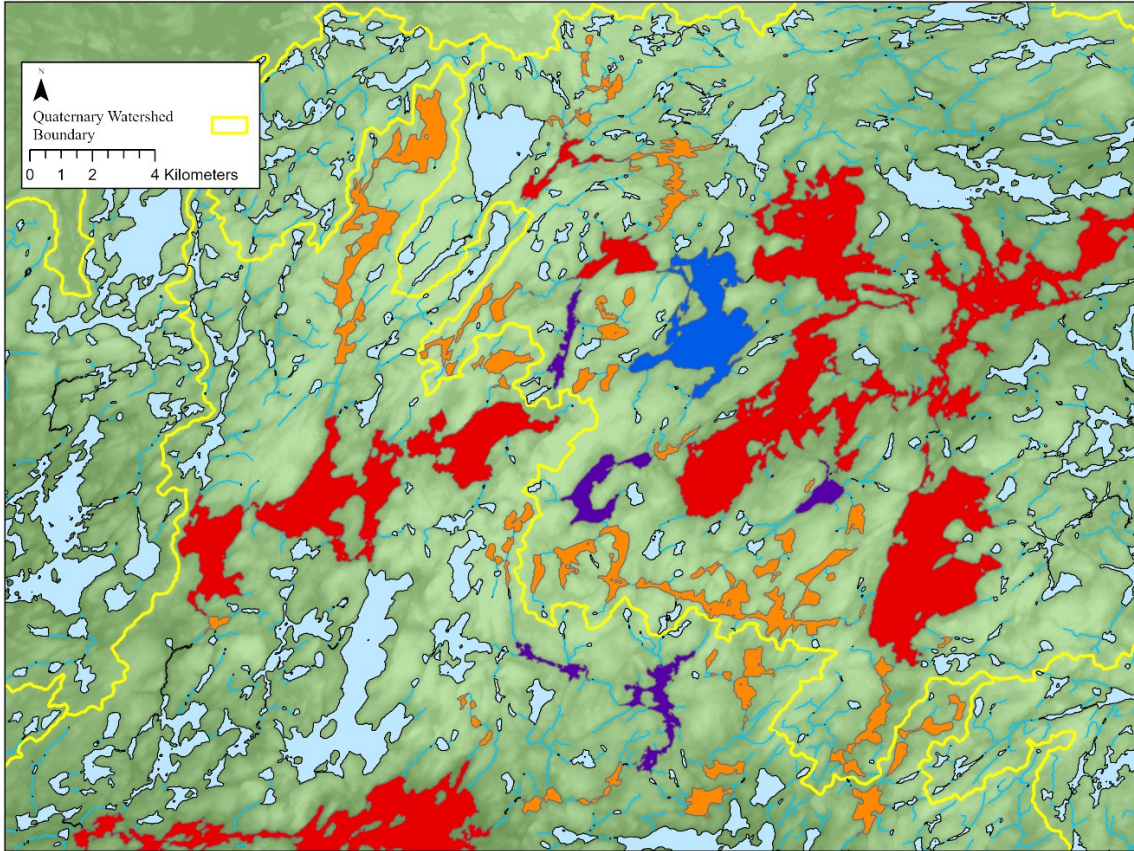


Figure 3.11. Actual and predicted distribution of burbot across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where burbot were predicted to be present but were not observed. Purple represents lakes where burbot were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

Slimy sculpin ($T_{jur} R^2 = 0.300$, 10-fold Cross Validation $T_{jur} R^2 = 0.107$) was primarily influenced by maximum lake depth (Figure 3.12). Presence-absence Model ESC shows that slimy sculpin is strongly positively associated with greater lake depths. The \log_{10} maximum lake depth at which the presence probability reaches 0.5 is 1.497 (95% CI: 0.645 to 2.222), which corresponds to an actual depth of 31.4 m (95% CI: 4.4 m to 167.0 m; Figure 3.12). Given the strong correlation between maximum lake depth and surface area (see Figure 3.1), the lake surface area at this probability is estimated to be

94.7 ha (95% CI: 9.0 ha to 703.5 ha). Figure 3.13 displays the actual and predicted distributions of slimy sculpin based on all model predictors. When applying a 0.5 probability threshold, Model ESC accurately predicted slimy sculpin presence and absence in 63 of the 81 study lakes.

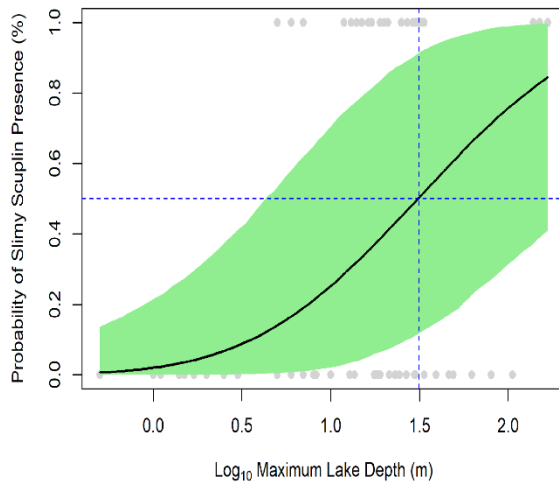


Figure 3.12. Response of slimy sculpin presence probability to \log_{10} value of maximum lake depth (m). Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

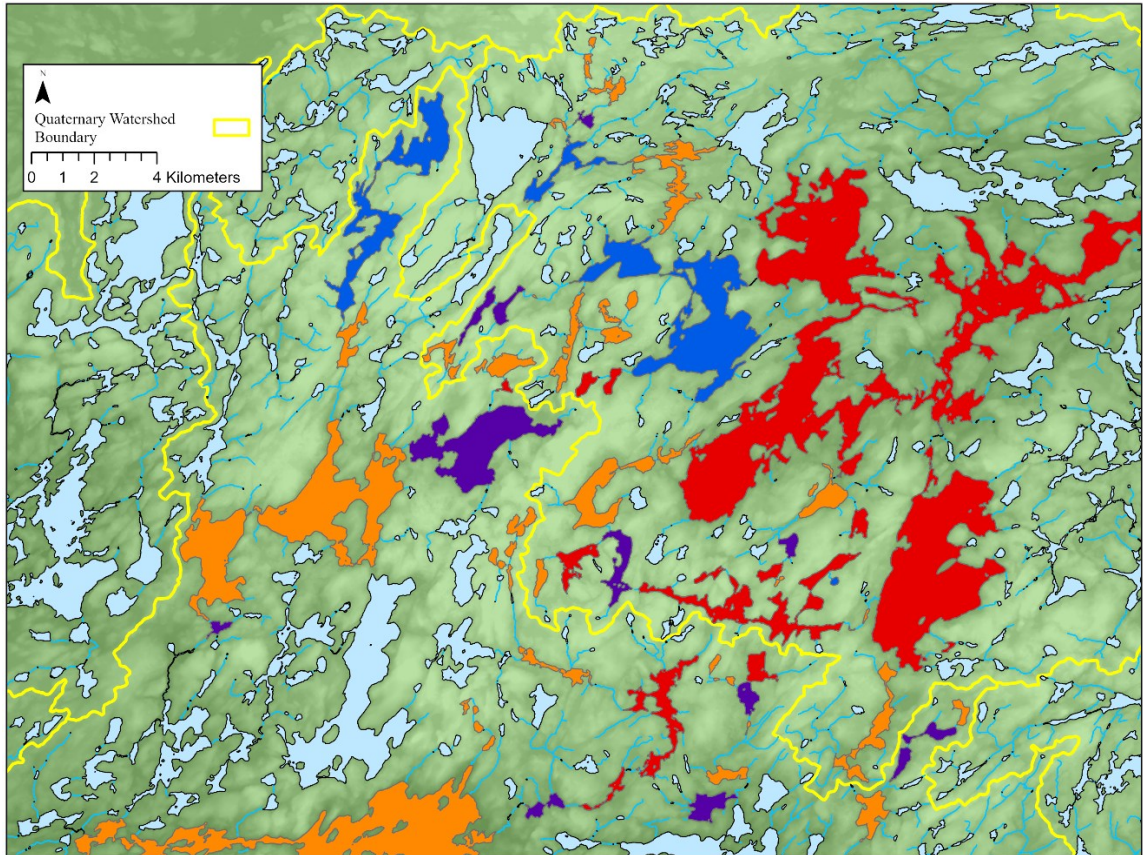


Figure 3.13. Actual and predicted distribution of slimy sculpin across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where slimy sculpin were predicted to be present but were not observed. Purple represents lakes where slimy sculpin were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

3.5.5 Spatial Drivers of Fish Presence

Spatial variables accounted for 52.6% of the explained variance across all species. Spatial eigenvector 2, which appears to differentiate between downstream lakes (high-order) and headwater lakes (low-order), accounted for 32.7% of the explained variance. Northern pike, white sucker, and finescale dace showed particularly strong associations with this variable. Spatial eigenvectors 1, 3, and 4 accounted for smaller portions of the explained variance, at 5.4%, 7.4%, and 7.1%, respectively. Eigenvector 1, which appears

to distinguish the Eagle and Dryberry watersheds, showed strong positive associations with fathead minnow and northern redbelly dace, suggesting both species are more likely to be present in the Eagle watershed.

Northern pike ($T_{jur} R^2 = 0.584$, 10-fold cross validation $T_{jur} R^2 = 0.473$) was best explained by spatial eigenvectors 2 (Figures 3.14). Based on presence-absence Model ESC, northern pike had a strong negative association with spatial eigenvectors 2. The spatial eigenvector 2 value where presence probability reaches 0.5 is -0.0497 (95% CI: -0.0937 to -0.0144), suggesting that northern pike are more likely to be found in further downstream lakes than headwater lakes (Figures 3.5, 3.6, and 3.14). The actual and predicted distribution of northern pike is displayed in Figure 3.15. If we use 0.5 probability as the cutoff for likely species presence, Model ESC correctly predicts the presence and absence of northern pike for 76 of the 81 study lakes.

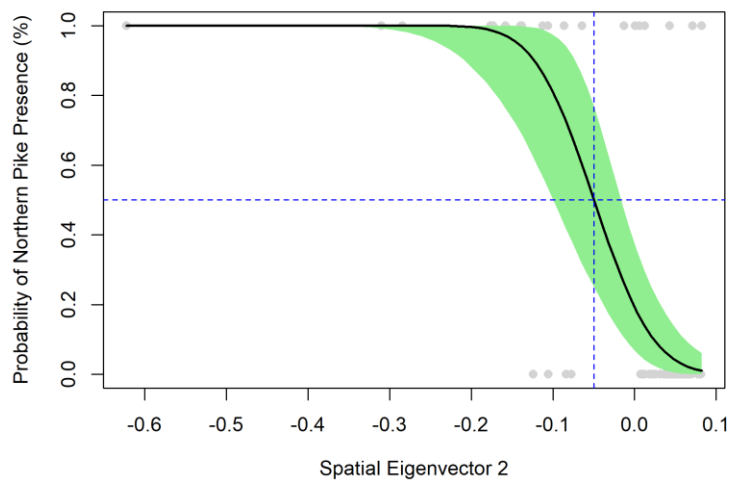


Figure 3.14. Response of northern pike presence probability to spatial eigenvector 2. Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

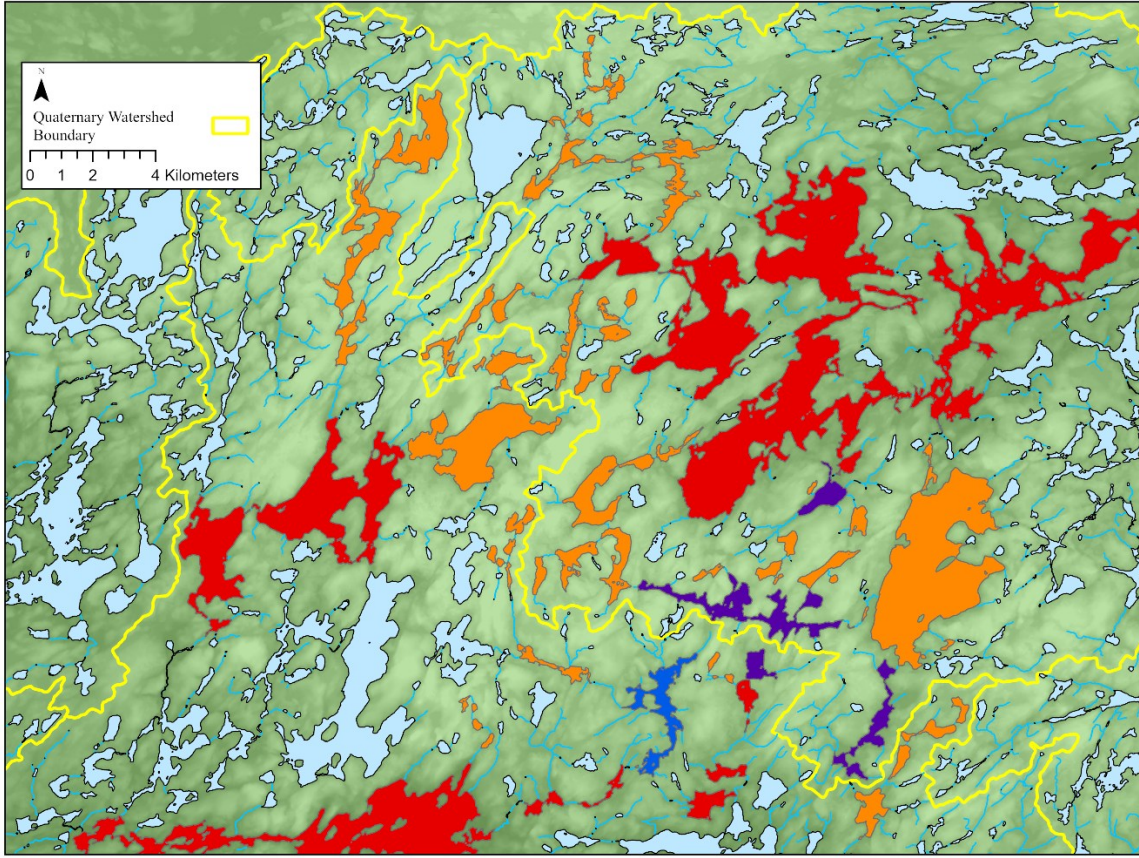


Figure 3.15. Actual and predicted distribution of northern pike across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where northern pike were predicted to be present but were not observed. Purple represents lakes where northern pike were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

White sucker ($T_{jur} R^2 = 0.482$, 10-fold cross validation $T_{jur} R^2 = 0.237$) was best explained by spatial eigenvectors 2 (Figures 3.16). Based on presence-absence Model ESC, white sucker has a strong negative association with spatial eigenvectors 2. The spatial eigenvector 2 value where presence probability reaches 0.5 is over 0.0826 (modeling limitation unable to determine exact value), suggesting that white sucker are more likely to be found in higher-order lakes than headwater lakes (Figure 3.16). The actual and predicted distribution of white sucker is displayed in Figure 3.17. If we use 0.5

probability as the cutoff for likelihood of species presence, Model ESC correctly predicts the presence and absence of white sucker for 72 of the 81 study lakes.

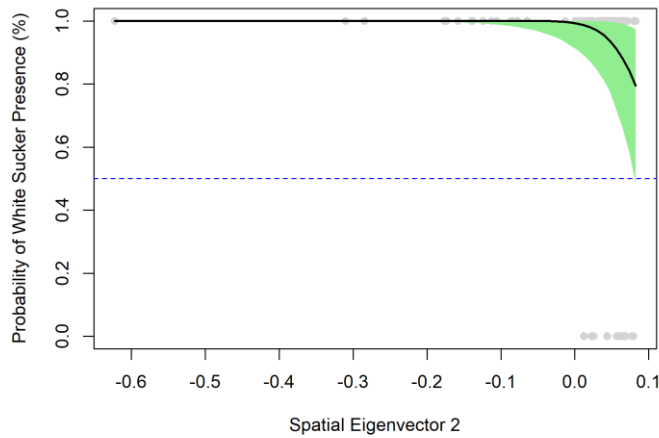


Figure 3.16. Response of white sucker presence probability to spatial eigenvector 2. Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

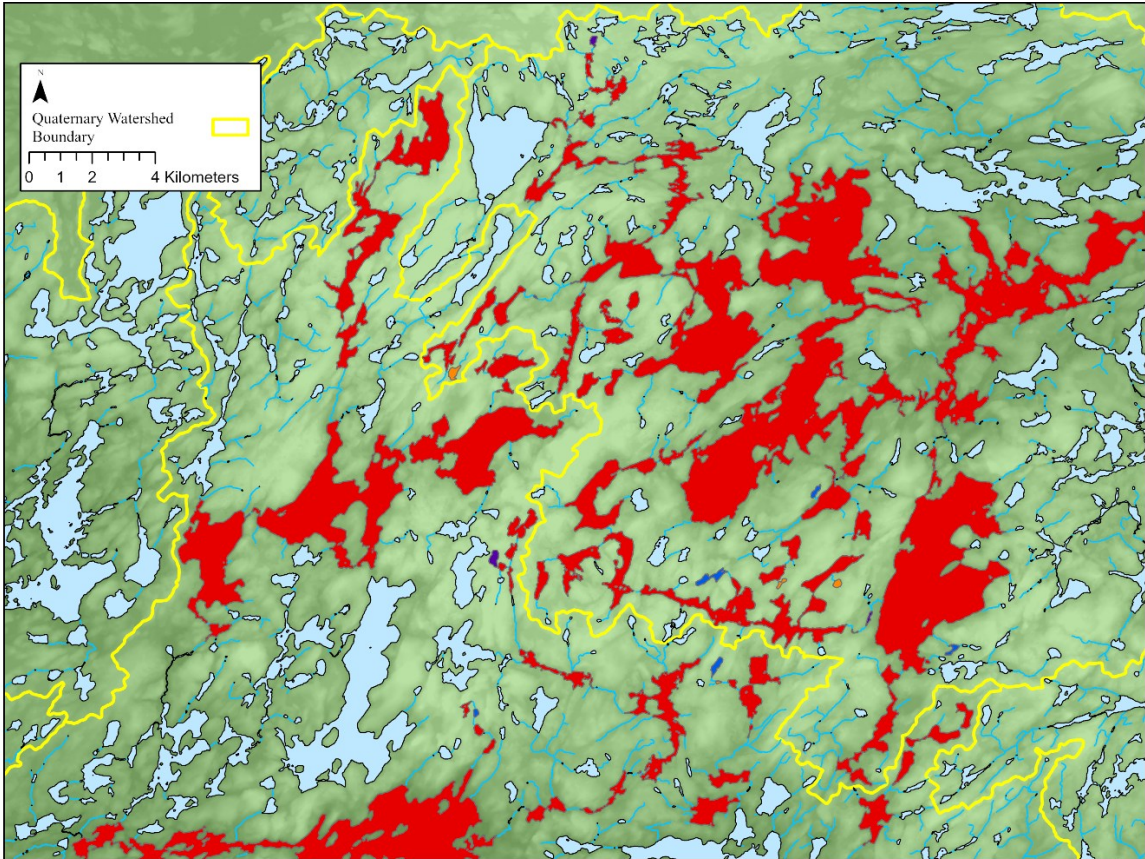


Figure 3.17. Actual and predicted distribution of white sucker across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where white sucker were predicted to be present but were not observed. Purple represents lakes where white sucker were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

Finescale dace (Tjur $R^2 = 0.369$, 10-fold cross validation Tjur $R^2 = 0.055$) was best explained by spatial eigenvectors 2 (Figures 3.18). Based on presence-absence Model ESC, finescale dace has strong positive association with spatial eigenvectors 2. The spatial eigenvector 2 value where presence probability reaches 0.5 suggests that finescale dace are more likely to be found in headwater lakes than downstream lakes (Figure 3.18). The actual and predicted distribution of finescale dace is displayed in Figure 3.19. If we use 0.5 probability as the cutoff for likely species presence, Model

ESC correctly predicts the presence and absence of finescale dace for 71 of the 81 study lakes.

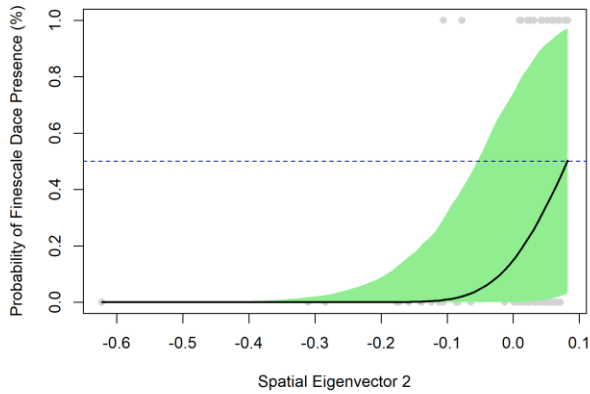


Figure 3.18. Response of finescale dace presence probability to spatial eigenvector 2. Plots are based on predictive ability of presence-absence Model ESC. Green shaded area represents 95% posterior probability.

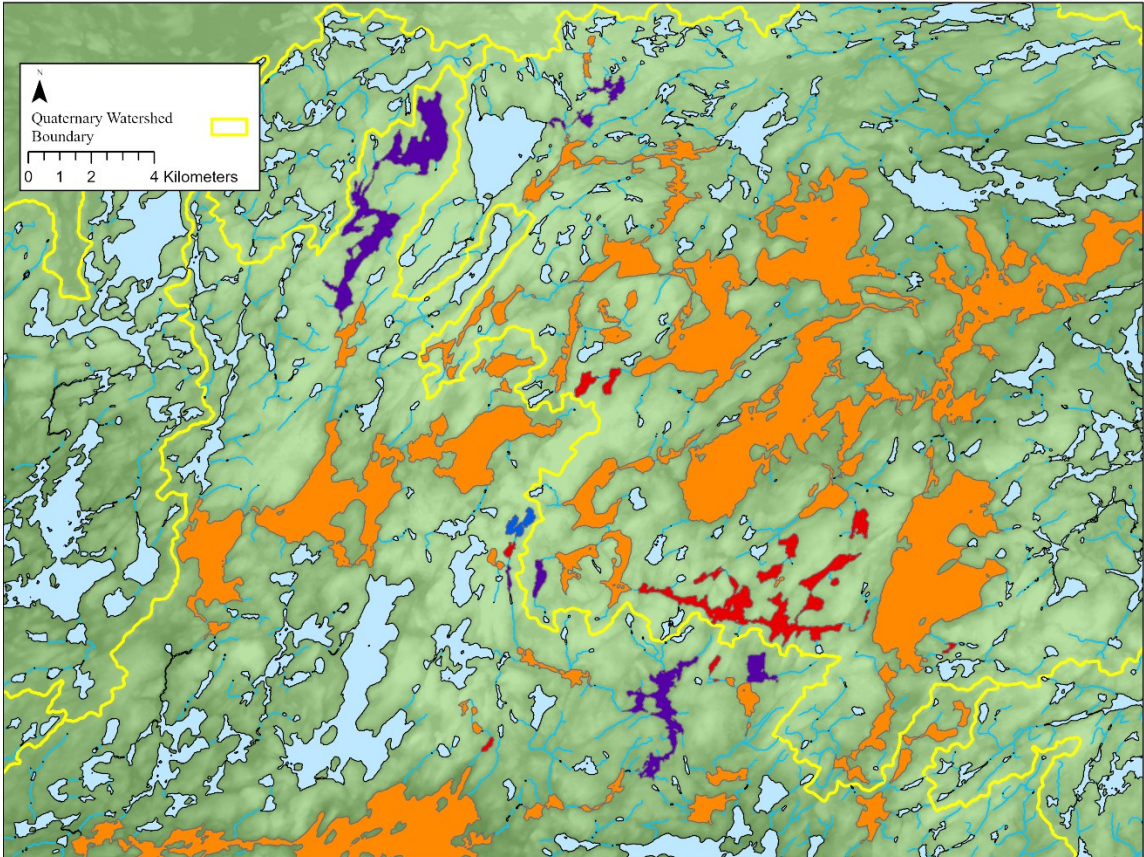


Figure 3.19. Actual and predicted distribution of finescale dace across study lakes. Red represents lakes where the predicted and actual presence align (and species is present). Orange represents lakes where the predicted and actual presence align (and species is absent). Dark blue indicates lakes where finescale dace were predicted to be present but were not observed. Purple represents lakes where finescale dace were observed but not predicted. The predicted distribution is based on the presence-absence Model ESC, with presence defined as a probability exceeding 0.5.

3.5.6 Species Associations (i.e., co-occurrences)

There are several strongly positive and negative species associations (>95% posterior probability) occurring in presence-absence Model ESC, based on species co-occurrence latent variables (Figure 3.20). Yellow perch appear to have a strong negative association with fathead minnow, finescale dace, northern redbelly dace, and slimy sculpin. By contrast, fathead minnow, finescale dace, northern redbelly dace, and slimy sculpin all have positive associations with one another.

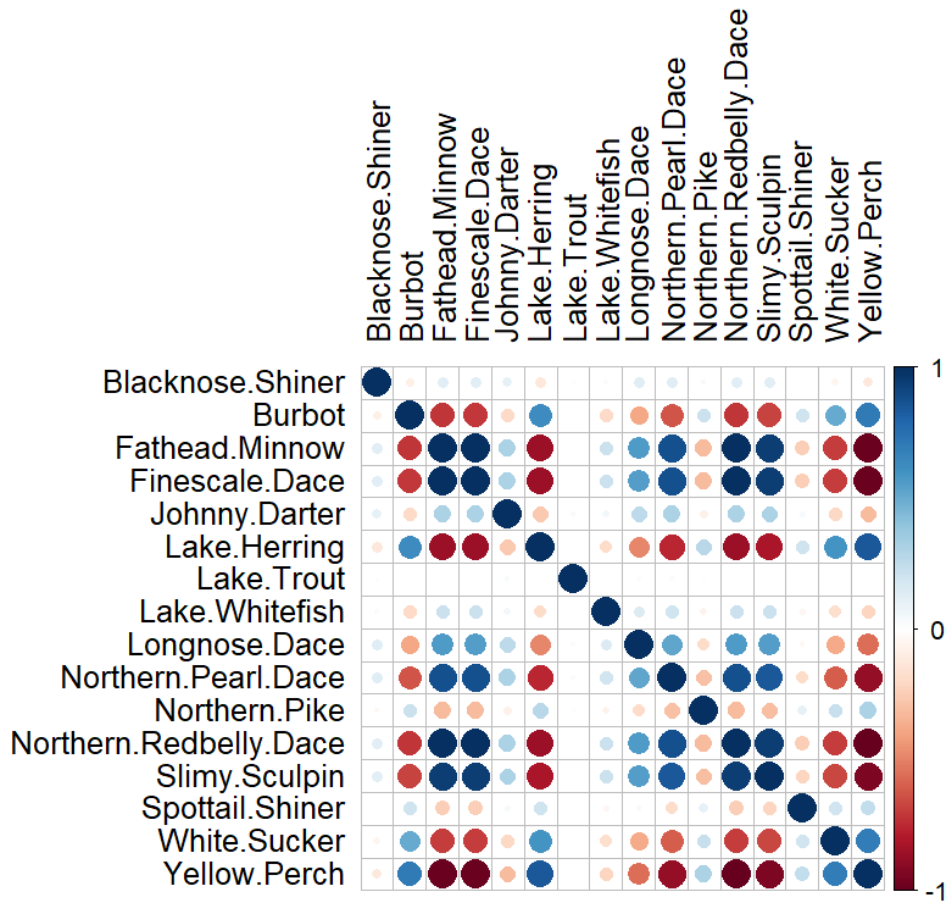


Figure 3.20. Species-to-species associations based on species co-occurrence latent variables produced in presence-absence Model ESC. Blue refers to positive and red to negative associations amongst species. Size of circles represent posterior probability.

Discussion

The study of fish metacommunity composition presents challenges specifically related to the integration of spatial patterns with physiochemical variables and biological drivers of species co-occurrence. These challenges have limited past research in their ability to provide a holistic framework for identifying the driving factors that influence fish distribution. Here, through the integration of spatial eigenvectors—which represent spatial connectivity variables among lakes—into a joint species distribution model

(JSDM) also incorporating physiochemical and biological factors, I have captured the key factors shaping fish community structure in boreal shield lakes at a level of spatial resolution not previously considered within freshwater communities.

The results demonstrate that spatial variables—specifically those related to lake connectivity, stream flow direction, and maximum stream gradient—are the primary factors influencing fish metacommunity composition in boreal shield lakes relative to the environmental and biological variables included in the analysis. These spatial variables, expressed as spatial eigenvectors, had the greatest explanatory power amongst all variable types in Model ESC (spatial: 52.6%, environmental: 36.0%, species co-occurrence: 11.4%) and ranked highest amongst spatial-only JSDM models. The integration of spatial eigenvectors with environmental variables and species co-occurrences (Model ESC) best explained and predicted fish community structure in terms of presence-absence, emphasizing the importance of a holistic, multivariate approach when studying metacommunities.

Of the spatial variables identified as most influential (i.e., lake connectivity, stream flow direction, and maximum stream gradient), flow direction combined with lake connectivity appears to have the greatest impact on fish community structure in boreal shield lakes. This is evident in the ranking of spatial JSDM models, where AEM-based models—incorporating flow direction— consistently outperformed MEM-based models, which unlike AEM-based models, assumes bidirectional movement of fish along streams (essentially ignoring flow direction or barriers to movement). This notable difference in explanatory power between AEM and MEM models highlights the importance of incorporating flow direction when studying aquatic species. While maximum stream

gradient contributed to developing the best performing spatial JSDM model for fish presence-absence in the lakes (SM1), its influence was secondary to flow direction and lake connectivity as its addition only slightly improved model fit (see difference between SM1 and SM2 Tjur R^2 values; Table 3.2).

While lake connectivity and stream flow direction appear to be the most influential amongst spatial variables in this research, maximum stream gradient was the most effective spatial weight investigated and included in AEM-based models. Maximum stream gradient essentially represents the largest fish movement barrier present along a stream. If, for example, a large waterfall is present along a stream, the sudden elevation change would be captured by the maximum stream gradient value for that watercourse. Alternatively, if no waterfalls are present, the gradient may reflect the presence of rapids, chutes, or cascades. Although other methods, such as fish barrier assessment indices and average stream gradient were tested as means to represent barriers to upstream movement, maximum stream gradient outperformed both of these alternatives. This is likely because fish barrier assessments are subject to in-field sampling bias and may overlook less obvious gradient-based barriers, while average stream gradient likely downweighs sudden elevational changes if the rest of the stream is relatively flat.

Despite the strong evidence I have found suggesting spatial connectivity strongly impacts fish community structure in lakes, many past and recent studies have reached the opposite conclusions. Traditionally, environmental factors have been emphasized in aquatic community studies, with fish communities often explained by environmental filters and species traits, while spatial variability has often been overlooked due to the difficulty of incorporating it into analyses (B. L. Brown et al., 2011; Jackson et al., 2001;

Leibold et al., 2004; Poff, 1997; Smith & Powell, 1971). Although the metacommunity framework has increased attention on spatial factors, their role is still often seen as secondary. For example, Henriques-Silva et al. (2013) concluded that environmental factors explained 16% of variation in Ontario lake fish communities, compared to just 1% for spatial factors (Henriques-Silva et al., 2013). Similarly, Magnuson et al. (1998) suggested that environmental variables were more influential than spatial variables in temperate lakes, possibly due to faster extinction rates reducing the impact of colonization events in boreal lakes (Magnuson et al., 1998).

A major limitation of these previous studies is the coarse spatial resolution at which their analyses were conducted. Many relied on broad spatial measures, such as geographic coordinates (Henriques-Silva et al., 2013) or lake connectivity metrics that failed to account for variables like flow direction and movement barriers (Kadoya et al., 2024). In contrast, my analysis used finer spatial scales by considering flow direction, stream gradient, and by treating individual lakes as spatial units. This approach provides a clearer understanding of spatial variation across landscapes at a more relevant level of spatial resolution (i.e. connections between adjacent lakes connected by waterways) and reveals the stronger influence of spatial factors on fish community composition among adjacent and connected boreal shield lakes.

The approach implemented here is unique in its application to understanding the influence of spatial connectivity on lake-based fish metacommunities on the boreal shield. Although AEM eigenvectors have been successfully applied in previous studies to assess the influence of connectivity and flow direction on aquatic organisms (Liu et al., 2013), their use in investigating lake-based fish metacommunities—particularly in boreal

shield lakes—has been limited. Research suggests that incorporating AEM eigenvectors, which can account for connectivity, flow direction, and spatial weights like stream gradient, can enhance the explanatory and predictive ability of stream-based metacommunity models (Mozzaquattro et al., 2020). Studies on freshwater fish, macroinvertebrates, and diatom metacommunities have also emphasized the importance of upstream movement barriers in shaping community structure in riverine environments (Dong et al., 2016; Mozzaquattro et al., 2020; Pollice et al., 2020). I have not found any studies that apply AEM eigenvectors to represent spatial variables for lake-based fish communities. Furthermore, no studies to date appear to have integrated this type of spatial eigenvector with JDSM into metacommunity models that simultaneously explore the impacts of environmental variables and species co-occurrence on community structure.

Although spatial eigenvectors used in my analysis broadly encompass lake connectivity, stream flow direction, and maximum stream gradient, each of the four individual spatial eigenvectors selected for JDSM modeling highlights key spatial patterns on the landscape. Of the four eigenvectors, spatial eigenvector 2 was the most important variable in explaining fish metacommunity composition. Although spatial eigenvector 2 does encompass flow direction and maximum stream gradients of connecting watercourses, a visualization of the sign and magnitude of this eigenvector makes it evident that this eigenvector also strongly highlights differences between high and low-order lakes. Along with accounting for 32.7% of the explained variation found in overall metacommunity composition, several fish species had particularly strong

relationships with spatial eigenvectors 2, including northern pike, white sucker, and finescale dace.

The positive association of northern pike with higher-order lakes indicates a preference for more downstream environments. These findings align with the known habitat preferences of northern pike. Although pike are relatively sedentary and territorial for most of the year, they can undertake substantial migrations during spawning, traveling tens to hundreds of kilometers to reach suitable tributary streams (Inskip, 1982). However, strong currents exceeding 1.5 m/sec can inhibit their movement (Harvey, 2009), as reflected in their preference for still or slow-moving water. This preference, combined with their limited ability to migrate upstream against strong water velocities or over natural barriers, may explain the spatial patterns observed in this study, particularly their reduced presence in headwater, low-order lakes (Harvey, 2009). This constraint is particularly relevant during spring spawning, when flow rates in streams and rivers are typically at their highest, further limiting pike's ability to access upstream habitats (Inskip, 1982).

Similar to northern pike, white sucker exhibit a positive association with higher-order lakes, although they are widely distributed, being present in 70 of the 81 lakes included in this study. This positive association reflects their absence from select headwater lakes, primarily those of lake orders 1 and 2, where white sucker are commonly absent. White sucker are known to migrate from lakes into streams during the spring spawning season, likely contributing to their widespread presence across lakes (Lucas & Baras, 2001; Reeb et al., 2008). However, their absence in certain lakes appears to be at least partially linked to extreme maximum stream gradients. White

sucker migration is known to be impacted by stream gradient, with previous research noting a substantial reduction in upstream movement ability between gradient of 10% and 20% (Bunt et al., 1999). These steep gradients likely act as fish movement barriers, preventing white sucker from colonizing some headwater lakes despite their migratory ability. In fact, in 8 of the 11 study lakes where white sucker were absent, no white sucker populations were found in directly connected upstream lakes. Among these 8 lakes, the average maximum stream gradient immediately downstream was 42.6% (SD = 24.8%), far exceeding the gradient threshold documented as a barrier to upstream movement for this species.

Unlike northern pike and white sucker, finescale dace demonstrated a positive relationship with spatial eigenvector 2, suggesting this species are present primarily in lower-order lakes (i.e., further upstream lakes). These findings align with research previous conducted by Booher and Walters (2021), who found an association of finescale dace with headwater streams, noting their preference for cool-water environments (Booher & Walters, 2021). Finescale dace are known to prefer small, shallow, spring-fed lakes with clear water, dense shoreline vegetation, and minimal presence of large predatory fish species (Stasiak & Cunningham, 2006). Although no negative species co-occurrences between finescale dace with other species were detected in the analysis, their preference for headwater lakes may be somewhat linked to predator avoidance (Cordero & Jackson, 2019). Other analyses of ELA lakes (including pike translocation experiments into lakes from which they are not native) have found evidence of negative associations between finescale dace and other minnow species with northern pike (Nicholson et al., 2015).

During the original ELA fish surveys, Beamish et al. (1976) noted several observations regarding differences between the two quaternary watersheds (Dryberry and Eagle watershed) included in this study, specifically regarding johnny darter, lake whitefish, and lake herring. By investigating species-specific relationships with spatial eigenvector 1, these observations were tested. Based on the model (Model ESC for presence-absence), only two species were suggested to be more or less likely to occur in either individual watershed (fathead minnow and northern redbelly dace were more likely to be found in Eagle watershed), neither of which were mentioned in Beamish et al. (1976).

Although the application of spatial eigenvectors, including MEM and AEM eigenvectors, has become more common in aquatic community analyses, the selection and justification of which eigenvectors to use and include in subsequent analyses varies among studies (Parreira et al., 2023; Sharma et al., 2011). Due to the potentially large number of eigenvectors produced through eigenvector mapping ($n-1$, where n represents the number of edges—i.e., rivers or streams connecting to lakes—in AEM mapping, and represents the number of sites—i.e., lakes—in MEM mapping), the process of how to select eigenvectors (variable selection) for analyses must be considered carefully. Typically, a process called forward selection has been applied to identify and choose the most relevant eigenvectors for modeling spatial patterns in ecological data (Blanchet et al., 2008b, 2008a, 2011), in which an initial model is run (e.g., a redundancy analysis or RDA) where no spatial eigenvectors are selected. Following this, eigenvectors are sequentially added to the model one at a time based on their ability to further explain the response variable (which, in the context of community ecology, often refers to patterns of

community composition). This process continues until a predefined stopping criterion is met, which, in the case of an RDA model, is based on an alpha value of 0.05. Once this is reached, the selected eigenvectors can be used as variables going forward. Although this process is effective in choosing specific eigenvectors that best explain a response variable, it does not base the decision to include certain eigenvectors on the spatial pattern it represents. Additionally, the process of forward selection also has the potential for introducing spurious or ‘nonsense’ spatial predictors when applied to large datasets, especially when eigenvectors are selected without visual validation (i.e., confirming that the eigenvector does actually represent a spatial pattern of interest). Although Dray et al., (2006) and Blanchet et al., (2008) both argue that eigenvectors with smaller eigenvalues represent finer-scale spatial patterns worth consideration, there is also an increasing likelihood that they do not actually reflect spatial patterns at all and are simply artefactual. By instead selecting a set number of eigenvectors for each analysis (4 in my case) and using the eigenvectors that explained the most spatial variation in the spatial matrix of interest (i.e., eigenvectors with the highest eigenvalues), the issues that come with forward selection are essentially avoided. For all analyses conducted in this study, the selected eigenvectors accounted for over 50% of the total variance within each spatial matrix.

In addition to the strong explanatory power of spatial variables on fish metacommunity composition, environmental variables also played an important role, especially in regard to coldwater fish species. Of the environmental variables, maximum lake depth, representing overall lake size due to its high correlation with lake surface area, explained 20.4% of the explained variation in metacommunity structure. DOC,

typically strongly associated with light attenuation in lakes, also played an important role for several species, although its overall impact was much smaller compared to maximum depth.

The strong preference of lake trout for deep lakes with low DOC levels (i.e., clearer water) aligns with the known preferences of the species. Being a coldwater species, lake trout favour deep, cold lakes, particularly in the southern extent of their range (Sellers et al., 1998). As surface temperatures rise during spring and summer, lake trout retreat to deeper, cooler waters, eventually residing in the hypolimnion (below the thermocline), during warmer months. In lake under 500 hectares, similar to many of the lakes included in this research, the thermocline typically forms above 6 meters, which maintains the cold temperatures that lake trout prefer, typically around 10° (Fee et al., 1996; Scott & Crossman, 1998).

With regards to DOC, my findings suggest that lake trout are most likely to be present in lakes with DOC concentrations of 6.6 mg/L or lower, which aligns with their known tolerance range (2.6-8.8 mg/L; Benoît et al., 2016). DOC can majorly influence light penetration and the thermal structure of lakes, both of which can impact lake trout population (Fee et al., 1996; Stasko et al., 2012). High DOC levels reduce light availability for photosynthesis, thereby decreasing primary production and limiting availability for benthic prey (Benoît et al., 2016; Sherbo et al., 2023). Such changes can also alter predator-prey dynamics, favouring species adapted to low-light conditions while hindering those that rely on clear water (Stasko et al., 2012). Additionally, higher DOC levels may lead to lower oxygen levels in deep lake layers, further stressing lake trout populations (Craig, 2016).

Similar to lake trout, slimy sculpin also displayed a strong preference for deeper lakes. Slimy sculpin are a cool-water, benthic species that prefer cobble substrate in both lake and stream habitats. Like other cottids, they lack swim bladders and are generally characterized by low mobility and strong site fidelity (Gray et al., 2004, 2018). Adult sculpin prefer boulders, while young-of-year sculpin prefer gravel and rubble substrates. They typically inhabit cold waters and are rarely found in temperatures above 19°C, with lethal temperatures estimated between 23 and 25°C (Gray et al., 2018). Given slimy sculpin's preference for cooler water, their presence in deeper lakes in the study area align with other observations elsewhere.

Burbot are a coldwater fish commonly found in large, deep lakes, particularly at the southern edge of their range, where they reside in cool rivers and reservoirs (McPhail & Paragamian, 2000). In lakes, adult burbot prefer bottom habitats, typically residing below the thermocline during the summer months at temperatures between 10-12°C (Hackney, 1973; McPhail & Paragamian, 2000). The results align with these past findings, predicting that the probability of burbot presence exceeds 0.5 at lake depths of 52.2 m (95% CI: 25.3 m to 167.0 m). Additionally, burbot tend to be more abundant in lakes with higher water clarity, specifically those with Secchi depths greater than 4.0 meters (Marshall & Ryan, 1987), which aligns with my findings related to DOC.

Even though environmental and spatial factors were the main explanatory variables in this analysis, several species associations were identified as not explained by either variable. It is important to recognize that species associations identified through JSDM modeling could reflect unknown or unspecified environmental or spatial variables, not otherwise included in the tested explanatory variables (Ovaskainen & Abrego, 2020).

This could lead to falsely assumed species interactions when taken at face value. For this reason, we cannot assume species interaction results are necessarily a reflection of true interactive effects without support from previous studies that have more directly focused on species-to-species relationships.

Northern redbelly dace, finescale dace, fathead minnow, and slimy sculpin were found to have strong positive associations with one another. Previous studies on predation and species co-occurrence indicate that northern redbelly dace, finescale dace, and fathead minnow show high segregation from top predators like northern pike (Nicholson et al., 2015). Although negative associations with northern pike were not detected in the current study, the positive associations among small-bodied fish observed here may be linked to predator avoidance (Cordero & Jackson, 2019). Cordero and Jackson (2019) found that northern redbelly dace, finescale dace, and fathead minnow exhibited strong segregation with top predators such as northern pike, suggesting that their co-occurrence tendencies in Ontario lakes may be related largely to predator avoidance (i.e., occurring in lakes together where predatory species are not present). The strong association between finescale dace and northern redbelly dace is expected, given their tendency to hybridize (Mee & Rowe, 2010) and their similar ecological niches (Stasiak & Cunningham, 2006).

Yellow perch were observed to have negative associations with finescale dace, northern redbelly dace, fathead minnows, and slimy sculpin. Studies have documented significant declines in prey species populations, such as fathead minnows and northern redbelly dace, due to yellow perch predation, with these populations rebounding following yellow perch extirpation (Soukup & Wisenden, 2023). Yellow perch, unless

controlled by predation, can dominate small lakes and exclude other species through predation and competition. They often prey on juvenile fish and can be the most abundant piscivores in small lakes, even exhibiting cannibalistic behaviour (T. G. Brown et al., 2009).

While several species associations were detected through JSMD modeling, it is important to note that spatial resolution is crucial when inferring interactions among species from empirical data. Biotic interactions typically occur at local scales, and when species are modeled across large spatial extents, these interactions may be diluted in the observed patterns. As a result, certain biological interactions may be more influential on within-lake distributions and abundance than at a whole-lake scale (Vallé et al., 2023).

Although environmental, spatial, and species co-occurrence latent variables were collectively able to effectively explain and predict the presence and absence of species within a metacommunity context, these same variables were ineffective in predicting relative abundance (conditional on presence). The substantial difference between all relative abundance full dataset and spatial-only JSMD models' explanatory and predictive power suggests that no combination of environmental, spatial, or species co-occurrence latent variables adequately modeled fish relative abundance. The poor predictive performance of these relative abundance models align with past studies that noted the challenges of modeling abundance, likely due to complex interactions and the need for large sample sizes (Pena et al., 2023). Although a hurdle model approach was necessary for this analysis due to the zero-inflated nature of the dataset, this approach reduced the sample size for relative abundance by excluding cases where species were absent (i.e., For each species, only lakes where the species is present are included in the

model). The reduced sample size likely contributed to the model's limitations, making cross-validation and model efficacy more challenging, emphasizing the need for substantially larger sample sizes when modeling abundance. Similar challenges have been reported in studies of both bird and fish communities with abundance metrics as a response variable using JSDMs and hurdle models, where poor model performance was common (Keppeler et al., 2022; Pena et al., 2023). Abundance models tend to struggle due to zero-inflated and long-tailed community data (Warton, 2005), making modeling difficult to model, especially in a multivariate context. In addition, it is also entirely likely that the scale used to classify relative abundance in the current study was too coarse to detect the potential influence of spatial, environmental or biological variables on fish abundance at a lake-level scale.

Despite the modeling approach being limited to explaining species presence and absence, the methods used and key findings are still widely applicable to the management of fish communities, lakes, and watersheds. While environmental conditions are often prioritized in habitat conservation, this research demonstrates the importance of spatial connectivity, specifically as it relates to the passage of species (i.e. specifically from one lake to another vs. across watersheds), in sustaining fish biodiversity. My findings suggest that effective management plans should consider the impacts of spatial variables, particularly understanding the role that stream gradients, flow direction, and lake connectivity play in shaping fish community structure and migration potential between waterbodies. Not only is this crucial for understanding why fish are where they are, but it can also be useful when investigating the spread of invasive fish species (Blanchet et al., 2008b). Although the potential influence of species associations on fish metacommunities

is likely limited (based on the current analysis), the ability to incorporate possible species interactions into ecological modeling allows for a more holistic approach to studying metacommunities.

Overall, this thesis demonstrates that the incorporation of spatial eigenvectors into JSDM models provides a clearer understanding of the role of spatial influences through connectivity on fish metacommunities, particularly at finer spatial scales. This framework not only lays the groundwork for future studies focused on exploring the role of spatial connectivity in shaping fish metacommunities but also offers insights for the management of aquatic systems by highlighting the key variables influencing species presence. By utilizing spatial eigenvectors, I was able to compare the effectiveness of different spatial variables in explaining fish species distribution. The integration of both environmental and spatial variables into JSDMs also enabled the identification of species co-occurrences that were not directly explained by these factors alone. Many of the co-occurrences identified align with previously observed species interactions, supporting the associations reported here. Despite limitations of this approach, such as its ability to effectively explain and predict species' relative abundance, JSDM modeling with the incorporation of spatial eigenvectors provides an effective way to investigate aquatic metacommunities.

References

- Baker, J. P., & Christensen, S. W. (1991). Effects of Acidification on Biological Communities in Aquatic Ecosystems. *Acidic Deposition and Aquatic Ecosystems*, 83–106. https://doi.org/10.1007/978-1-4613-9038-1_5
- Beamish, R., Blouw, L., & McFarlane, G. (1976). A fish and chemical study of 109 lakes in the Experimental Lakes Area (ELA), Northwestern Ontario, with appended reports on lake whitefish ageing errors and the Northwestern Ontario baitfish industry. In *Department of the Environment: Fisheries and Marine Service*.
- Beamish, R. J. (1976). Acidification of Lakes in Canada by Acid Precipitation and the Resulting Effects on Fishes. *Water, Air, and Soil Pollution*, 6(1976), 501–514.
- Benoit, P. O. (2014). *Influence of dissolved organic carbon (DOC) on fish production in north-temperate lakes*. McGill University, Macdonald Campus.
- Benoît, P. O., Beisner, B. E., & Solomon, C. T. (2016). Growth rate and abundance of common fishes is negatively related to dissolved organic carbon concentration in lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 73(8), 1230–1236. <https://doi.org/10.1139/cjfas-2015-0340>
- Blackburn, J., Hurkett, B., Redman, L., & Park, S. (n.d.). *Guide to Waterfall Fish Barrier Assessment*.
- Blanchet, F. G., Legendre, P., & Borcard, D. (2008a). Forward selection of explanatory variables. *Ecology*, 89(9), 2623–2632.
- Blanchet, F. G., Legendre, P., & Borcard, D. (2008b). Modelling directional spatial processes in ecological data. *Ecological Modelling*, 215(4), 325–336. <https://doi.org/10.1016/j.ecolmodel.2008.04.001>
- Blanchet, F. G., Legendre, P., Maranger, R., Monti, D., & Pepin, P. (2011). Modelling the effect of directional spatial ecological processes at different scales. *Oecologia*, 166(2), 357–368. <https://doi.org/10.1007/s00442-010-1867-y>
- Booher, E. C. J., & Walters, A. W. (2021). Biotic and abiotic determinants of finescale dace distribution at the southern edge of their range. *Diversity and Distributions*, 27(4), 696–709. <https://doi.org/10.1111/ddi.13227>
- Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations)? *Journal of Computational and Graphical Statistics*, 7(4), 434–455. <https://doi.org/10.1080/10618600.1998.10474787>
- Brown, B. L., Swan, C. M., Auerbach, D. A., Campbell Grant, E. H., Hitt, N. P., Maloney, K. O., & Patrick, C. (2011). Metacommunity theory as a multispecies, multiscale framework for studying the influence of river network structure on riverine communities and ecosystems. *Journal of the North American Benthological Society*, 30(1), 310–327. <https://doi.org/10.1899/10-129.1>
- Brown, T. G., Runciman, B., Bradford, M. J., & Pollard, S. (2009). A biological synopsis of yellow perch. *Canadian Manuscript Report of Fisheries and Aquatic Sciences*, 2883, 28.

- Brunskill, G. J., & Schindler, D. W. (1971). Geography and Bathymetry of Selected Lake Basins, Experimental Lakes Area, Northwestern Ontario. *Journal of the Fisheries Research Board of Canada*, 28(2), 139–155. <https://doi.org/10.1139/f71-028>
- Bunt, C. M., Katopodis, C., & Mckinley, R. S. (1999). Attraction and Passage Efficiency of White Suckers and Smallmouth Bass by Two Denil Fishways. In *North American Journal of Fisheries Management* (Vol. 19).
- Byström, P., Karlsson, J., Nilsson, P., Van Kooten, T., Ask, J., & Olofsson, F. (2007). Substitution of top predators: Effects of pike invasion in a subarctic lake. *Freshwater Biology*, 52(7), 1271–1280. <https://doi.org/10.1111/j.1365-2427.2007.01763.x>
- Chase, J. M., Jeliaskov, A., & Ladouceur, E. (2020). *Biodiversity conservation through the lens of metacommunity ecology*. 1469, 86–104. <https://doi.org/10.1111/nyas.14378>
- Chow-Fraser, P. (1991). Use of the morphoedaphic index to predict nutrient status and algal biomass in some Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 48(10), 1909–1918. <https://doi.org/10.1139/f91-227>
- Chu, C., Minns, C. K., Moore, J. E., & Millard, E. S. (2004). Impact of Oligotrophication, Temperature, and Water Levels on Walleye Habitat in the Bay of Quinte, Lake Ontario. *Transactions of the American Fisheries Society*, 133(4), 868–879. <https://doi.org/10.1577/t03-095.1>
- Cleugh, T. R., & Hauser, B. W. (1971). Results of the Initial Survey of the Experimental Lakes Area, Northwestern Ontario. *Journal of the Fisheries Research Board of Canada*, 28(2), 129–137. <https://doi.org/10.1139/f71-027>
- Cooke, S. J., Piczak, M. L., Nyboer, E. A., Michalski, F., Bennett, A., Koning, A. A., Hughes, K. A., Chen, Y., Wu, J., Cowx, I. G., Koehnken, L., Raghavan, R., Pompeu, P. S., Phang, S., Valbo-Jørgensen, J., Bendixen, M., Torres, A., Getahun, A., Kondolf, G. M., ... Taylor, W. W. (2023). Managing exploitation of freshwater species and aggregates to protect and restore freshwater biodiversity. *Environmental Reviews*, 24, 1–24. <https://doi.org/10.1139/er-2022-0118>
- Cordero, R. D., & Jackson, D. A. (2019). Species-pair associations, null models, and tests of mechanisms structuring ecological communities. *Ecosphere*, 10(7). <https://doi.org/10.1002/ecs2.2797>
- Correll, D. L. (1998). The Role of Phosphorus in the Eutrophication of Receiving Waters: A Review. *Journal of Environmental Quality*, 27(2), 261–266. <https://doi.org/10.2134/jeq1998.00472425002700020004x>
- COSEWIC. (2009). *COSEWIC assessment and update status report on the Roseate Tern *Sterna dougallii* in Canada*. Committee on the Status of Endangered Wildlife in Canada. www.sararegistry.gc.ca/status/status_e.cfm
- Craig, N. (2016). *Impacts of dissolved organic carbon on productivity of fish and benthic macroinvertebrates in north temperate lakes*.

- Davis, B. M., & Todd, T. N. (1998). Competition between larval lake herring (*Coregonus artedii*) and lake whitefish (*Coregonus clupeaformis*) for zooplankton. *Canadian Journal of Fisheries and Aquatic Sciences*, 55(5), 1140–1148. <https://doi.org/10.1139/f98-004>
- Davis, C. D., Epps, C. W., Flitcroft, R. L., & Banks, M. A. (2018). Refining and defining riverscape genetics: How rivers influence population genetic structure. *Wiley Interdisciplinary Reviews: Water*, 5(2), 1–20. <https://doi.org/10.1002/WAT2.1269>
- Desforges, J. E., Clarke, J., Harmsen, E. J., Jardine, A. M., Robichaud, J. A., Serré, S., Chakrabarty, P., Bennett, J. R., Hanna, D. E. L., Smol, J. P., Rytwinski, T., Taylor, J. J., Martel, A. L., Winegardner, A. K., Marty, J., Taylor, M. K., O’connor, C. M., Robinson, S. A., Reid, A. J., ... Cooke, S. J. (2022). The alarming state of freshwater biodiversity in Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, 79(2), 352–365. <https://doi.org/10.1139/cjfas-2021-0073>
- Dextrase, A. J., & Mandrak, N. E. (2006). Impacts of alien invasive species on freshwater fauna at risk in Canada. *Biological Invasions*, 8(1), 13–24. <https://doi.org/10.1007/s10530-005-0232-2>
- Dong, X., Li, B., He, F., Gu, Y., Sun, M., Zhang, H., Tan, L., Xiao, W., Liu, S., & Cai, Q. (2016). Flow directionality, mountain barriers and functional traits determine diatom metacommunity structuring of high mountain streams. *Scientific Reports*, 6(April), 1–11. <https://doi.org/10.1038/srep24711>
- Dray, S., Legendre, P., & Peres-neto, P. R. (2006). *Spatial modelling : a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM)*. 196, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Englund, G., Johansson, F., Olofsson, P., Salonsaari, J., & Öhman, J. (2009). Predation leads to assembly rules in fragmented fish communities. *Ecology Letters*, 12(7), 663–671. <https://doi.org/10.1111/j.1461-0248.2009.01322.x>
- Fang, X., Alam, S. R., Jacobson, P., Pereira, D., & Stefan, H. G. (2009). *Characteristics of Minnesota’s Cisco Lakes*. 532, 1–32.
- Fee, E. J., Hecky, R. E., Kasian, S. E. M., & Cruikshank, D. R. (1996). Effects of lake size, water clarity, and climatic variability on mixing depths in Canadian Shield lakes. *Limnology and Oceanography*, 41(5), 912–920. <https://doi.org/10.4319/lo.1996.41.5.0912>
- Findlay, D. L., Kasian, S. E. M., Turner, M. T., & Stainton, M. P. (1999). Responses of phytoplankton and epilithon during acidification and early recovery of a lake. *Freshwater Biology*, 42(1), 159–175. <https://doi.org/10.1046/j.1365-2427.1999.00458.x>
- Fullerton, A. H., Burnett, K. M., Steel, E. A., Flitcroft, R. L., Pess, G. R., Feist, B. E., Torgersen, C. E., Miller, D. J., & Sanderson, B. L. (2010). Hydrological connectivity for riverine fish: Measurement challenges and research opportunities. *Freshwater Biology*, 55(11), 2215–2237. <https://doi.org/10.1111/j.1365-2427.2010.02448.x>

- Gardunio, E. I. (2014). *Jumping and swimming performance of burbot and white sucker: implications for barrier design*. 92.
- Gelman, A., & Donald, Rubin. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4), 457–511.
- Gray, M. A., Allen Curry, R., Arciszewski, T. J., Munkittrick, K. R., & Brasfield, S. M. (2018). The biology and ecology of slimy sculpin: A recipe for effective environmental monitoring. *Facets*, 3(1), 103–127. <https://doi.org/10.1139/facets-2017-0069>
- Gray, M. A., Cunjak, R. A., & Munkittrick, K. R. (2004). Site fidelity of slimy sculpin (*Cottus cognatus*): Insights from stable carbon and nitrogen analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(9), 1717–1722. <https://doi.org/10.1139/F04-108>
- Guimarães, T. de F. R., Hartz, S. M., & Becker, F. G. (2014). Lake connectivity and fish species richness in southern Brazilian coastal lakes. *Hydrobiologia*, 740(1), 207–217. <https://doi.org/10.1007/s10750-014-1954-x>
- Hackney, P. A. (1973). *Ecology of the Burbot (Lota lota) with Special Reference to its Role in the Lake Opeongo Fish Community*. University of Toronto.
- Harvey, B. (2009). A Biological Synopsis of Northern Pike (*Esox Lucius*). *Fisheries (Bethesda)*, 31. <http://www.dfo-mpo.gc.ca/Library/337844.pdf>
- Heathcote, A. J., Giorgio, P. A., & Prairie, Y. T. (2015). Predicting bathymetric features of lakes from the topography of their surrounding landscape. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(5), 643–650. <https://doi.org/10.1139/cjfas-2014-0392>
- Heino, J., Siqueira, T., Landeiro, V. L., Kotanen, J., Bini, L. M., & Gr, M. (2013). *Metacommunity structuring in stream networks : roles of dispersal mode , distance type , and regional environmental context*. <https://doi.org/10.1002/ece3.834>
- Henriques-Silva, R., Lindo, Z., & Peres-Neto, P. R. (2013). A community of metacommunities: Exploring patterns in species distributions across large geographical areas. *Ecology*, 94(3), 627–639. <https://doi.org/10.1890/12-0683.1>
- Hossain, M., Arhonditsis, G. B., Hoyle, J. A., Randall, R. G., & Koops, M. A. (2019). Nutrient management and structural shifts in fish assemblages: Lessons learned from an Area of Concern in Lake Ontario. *Freshwater Biology*, 64(5), 967–983. <https://doi.org/10.1111/fwb.13278>
- Hulsman, M. F., Cahill, C. L., Erwin, A. C., Lunn, B. D., Tonn, W. M., Howland, K. L., & Bakix, A. B. M. (2016). Influence of potential fish competitors on Lake Trout trophic ecology in small lakes of the Barrenlands, N.W.T., Canada. *Journal of Great Lakes Research*, 42(2), 290–298. <https://doi.org/10.1016/j.jglr.2015.07.014>
- IISD. (2018). *IISD-ELA Broad-Scale Monitoring Protocol* (pp. 1–10).
- Inskip, P. D. (1982). *Habitat suitability index models: northern pike*. Western Energy and Land Use Team, Office of Biological Services, Fish and Wildlife Service, US Department of the Interior.

- Jackson, D. A., Peres-Neto, P. R., & Olden, J. D. (2001). What controls who is where in freshwater fish communities - The roles of biotic, abiotic, and spatial factors. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(1), 157–170. <https://doi.org/10.1139/cjfas-58-1-157>
- Johnson, M. G., Leach, J. H., Minns, C. K., & Olver, C. H. (1977). Limnological Characteristics of Ontario Lakes in Relation to Associations of Walleye (*Stizostedion vitreum vitreum*), Northern Pike (*Esox lucius*), Lake Trout (*Salvelinus namaycush*), and Smallmouth Bass (*Micropterus dolomieu*). *Journal of the Fisheries Research Board of Canada*, 34(10), 1592–1601. <https://doi.org/10.1139/f77-224>
- Kadoya, T., Nilsson, K. A., Kelly, J., Bartley, T. J., Säterberg, T., Guzzo, M. M., Esch, E., Koide, D., Matsuzaki, S. ichiro S., Terui, A., Akasaka, M., & MacDougall, A. S. (2024). Common processes drive metacommunity structure in freshwater fish. *Global Ecology and Biogeography*, 33(5). <https://doi.org/10.1111/geb.13822>
- Kaufmann, P. R., & Robison, E. G. (1998). Environmental Monitoring and Assessment Program Surface Waters : Field Operations and Methods for Measuring the Ecological Condition of Wadeable Streams. *Epa/620/R-94/00F, September*, 1–208.
- Keppeler, F. W., Andrade, M. C., Trindade, P. A. A., Sousa, L. M., Arantes, C. C., Winemiller, K. O., Jensen, O. P., & Giarrizzo, T. (2022). Early impacts of the largest Amazonian hydropower project on fish communities. *Science of the Total Environment*, 838(May), 155951. <https://doi.org/10.1016/j.scitotenv.2022.155951>
- King, K. B. S., Bremigan, M. T., Infante, D., & Cheruvelil, K. S. (2021). Surface water connectivity affects lake and stream fish species richness and composition. *Canadian Journal of Fisheries and Aquatic Sciences*, 78(4), 433–443. <https://doi.org/10.1139/cjfas-2020-0090>
- Laske, S. M., Haynes, T. B., Rosenberger, A. E., Koch, J. C., Wipfli, M. S., Whitman, M., & Zimmerman, C. E. (2016). Surface water connectivity drives richness and composition of Arctic lake fish assemblages. *Freshwater Biology*, 61(7), 1090–1104. <https://doi.org/10.1111/fwb.12769>
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M., & Gonzalez, A. (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecology Letters*, 7(7), 601–613. <https://doi.org/10.1111/j.1461-0248.2004.00608.x>
- Leibold, M. A., Rudolph, F. J., Blanchet, F. G., De Meester, L., Gravel, D., Hartig, F., Peres-Neto, P., Shoemaker, L., & Chase, J. M. (2021). The internal structure of metacommunities. *Oikos*, 1–13. <https://doi.org/10.1111/oik.08618>
- Lennox, P. A. (2019). *Examining the relationship between species richness of fish communities and their biomass in freshwater ecosystems*. University of Lethbridge.
- Littlefair, J., Hayhurst, L., Yates, M., Rennie, M. D., & Cristescu, M. (2024). A strong relationship between environmental DNA metabarcoding and rank-based abundance of fish. *Journal of Applied Ecology*.

- Liu, J., Soininen, J., Han, B. P., & Declerck, S. A. J. (2013). Effects of connectivity, dispersal directionality and functional traits on the metacommunity structure of river benthic diatoms. *Journal of Biogeography*, 40(12), 2238–2248. <https://doi.org/10.1111/jbi.12160>
- Lucas, M. C., & Baras, E. (2001). *Migration of Freshwater Fishes Ondrej Slavík*.
- Magnuson, J. J., Tonn, W. M., Banerjee, A., Toivonen, J., Sanchez, O., & Rask, M. (1998). ISOLATION VS. EXTINCTION IN THE ASSEMBLY OF FISHES IN SMALL NORTHERN LAKES. In *Ecology* (Vol. 79, Issue 8).
- Marshall, T. R., & Ryan, P. A. (1987). Abundance patterns and community attributes of fishes relative to environmental gradients. *Canadian Journal of Fisheries and Aquatic Sciences*, 44, 198–215. <https://doi.org/10.1139/f87-323>
- Matuszek, J. E., & Beggs, G. L. (1988). Fish pH. *Canadian Journal of Fisheries and Aquatic Sciences*, 45.
- McPhail, J. D., & Paragamian, V. L. (2000). Burbot biology and life history. *Burbot: Biology, Ecology and Management*, 11–23.
- Mee, J. A., & Rowe, L. (2010). Distribution of Phoxinus eos, Phoxinus neogaeus, and their asexually-reproducing hybrids (Pisces: Cyprinidae) in Algonquin provincial Park, Ontario. *PLoS ONE*, 5(10). <https://doi.org/10.1371/journal.pone.0013185>
- Mehner, T., Diekmann, M., Brämick, U., & Lemcke, R. (2005). Composition of fish communities in German lakes as related to lake morphology, trophic state, shore structure and human-use intensity. *Freshwater Biology*, 50(1), 70–85. <https://doi.org/10.1111/j.1365-2427.2004.01294.x>
- Milling, A. (2020). *Impacts of Mysis diluviana extirpation on a whole lake ecosystem*.
- Mills, K. H., Chalanchuk, S. M., Mohr, L. C., & Davies, I. J. (1987). Responses of fish populations in Lake 223 to 8 years of experimental acidification. *Canadian Journal of Fisheries and Aquatic Sciences*, 44(SUPPL. 1), 114–125. <https://doi.org/10.1139/f87-287>
- Molot, L. A., & Dillon, P. J. (1997). Colour - mass balances and colour - dissolved organic carbon relationships in lakes and streams in central Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(12), 2789–2795. <https://doi.org/10.1139/cjfas-54-12-2789>
- Moritz, C., Meynard, C. N., Devictor, V., Guizien, K., Labrune, C., Guarini, J. M., & Mouquet, N. (2013). Disentangling the role of connectivity, environmental filtering, and spatial structure on metacommunity dynamics. *Oikos*, 122(10), 1401–1410. <https://doi.org/10.1111/j.1600-0706.2013.00377.x>
- Mozzaquattro, L. B., Dala-Corte, R. B., Becker, F. G., & Melo, A. S. (2020). Effects of spatial distance, physical barriers, and habitat on a stream fish metacommunity. *Hydrobiologia*, 847(14), 3039–3054. <https://doi.org/10.1007/s10750-020-04309-8>
- Mushet, G. R., Yee, C., Bergson, S., & Blanchfield, P. J. (2023). Abundance and movement of fishes among nearshore habitat types in small boreal shield lakes. *Hydrobiologia*, 850(9), 1985–2003. <https://doi.org/10.1007/s10750-023-05205-7>

- Nicholson, M. E., Rennie, M. D., & Mills, K. H. (2015). Apparent extirpation of prey fish communities following the introduction of Northern Pike (*Esox lucius*). *Canadian Field-Naturalist*, *129*(2), 165–173. <https://doi.org/10.22621/cfn.v129i2.1697>
- Norberg, A., Abrego, N., Blanchet, F. G., Adler, F. R., Anderson, B. J., Anttila, J., Araújo, M. B., Dallas, T., Dunson, D., Elith, J., Foster, S. D., Fox, R., Franklin, J., Godsoe, W., Guisan, A., O'Hara, B., Hill, N. A., Holt, R. D., Hui, F. K. C., ... Ovaskainen, O. (2019). A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. *Ecological Monographs*, *89*(3), 1–24. <https://doi.org/10.1002/ecm.1370>
- Odrizola, I., Abrego, N., TlÁskal, V., ZrústovÁ, P., Morais, D., Větrovský, T., Ovaskainen, O., & Baldrian, P. (2021). Fungal Communities Are Important Determinants of Bacterial Community Composition in Deadwood. *MSystems*, *6*(1). <https://doi.org/10.1128/msystems.01017-20>
- Öhman, J., Buffam, I., Englund, G., Blom, A., Lindgren, E., & Laudon, H. (2006). Associations between water chemistry and fish community composition: A comparison between isolated and connected lakes in northern Sweden. *Freshwater Biology*, *51*(3), 510–522. <https://doi.org/10.1111/j.1365-2427.2006.01514.x>
- Olden, J. D., Jackson, D. A., & Peres-Neto, P. R. (2001). Spatial isolation and fish communities in drainage lakes. *Oecologia*, *127*(4), 572–585. <https://doi.org/10.1007/s004420000620>
- Orihe, D. M., Baulch, H. M., Casson, N. J., North, R. L., Parsons, C. T., Seckar, D. C. M., & Venkiteswaran, J. J. (2017). Internal phosphorus loading in canadian fresh waters: A critical review and data analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, *74*(12), 2005–2029. <https://doi.org/10.1139/cjfas-2016-0500>
- Ovaskainen, O., & Abrego, N. (2020). *Joint Species Distribution Modelling*. Cambridge University Press. <https://doi.org/10.1017/9781108591720>
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, *20*(5), 561–576. <https://doi.org/10.1111/ele.12757>
- Parreira, M. R., Tessarolo, G., & Nabout, J. C. (2023). Incorporating symmetrical and asymmetrical dispersal into Ecological Niche Models in freshwater environments. *Acta Limnologica Brasiliensia*, *35*. <https://doi.org/10.1590/S2179-975X2723>
- Pena, J. C., Ovaskainen, O., MacGregor-Fors, I., Teixeira, C. P., & Ribeiro, M. C. (2023). The relationships between urbanization and bird functional traits across the streetscape. *Landscape and Urban Planning*, *232*. <https://doi.org/10.1016/j.landurbplan.2023.104685>
- Pichler, M., & Hartig, F. (2021). A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods in Ecology and Evolution*, *12*(11), 2159–2173. <https://doi.org/10.1111/2041-210X.13687>

- Plumb, J. M., & Blanchfield, P. J. (2009). Performance of temperature and dissolved oxygen criteria to predict habitat use by lake trout (*Salvelinus namaycush*). *Canadian Journal of Fisheries and Aquatic Sciences*, 66(11), 2011–2023. <https://doi.org/10.1139/F09-129>
- Poff, N. L. R. (1997). Landscape filters and species traits: Towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society*, 16(2), 391–409. <https://doi.org/10.2307/1468026>
- Pollice, A., Jona-Lasinio, G., Gaglio, M., Blanchet, F. G., & Fano, E. A. (2020). Modelling the effect of directional spatial ecological processes for a river network in Northern Italy. *Ecological Indicators*, 112(January), 106144. <https://doi.org/10.1016/j.ecolind.2020.106144>
- Rahel, F. J. (1984). Factors Structuring Fish Assemblages Along a Bog Lake Successional Gradient. In *Source: Ecology* (Vol. 65, Issue 4).
- Reebs, S., LeBlanc, S., Fraser, A., Hardie, P., & Cunjak, R. A. (2008). Upstream and downstream movements of lake chub, *Couesius plumbeus*, and white sucker, *Catostomus commersoni*, at Catamaran Brook, 1990-2004. *Canadian Technical Report of Fisheries and Aquatic Sciences* 2791.
- Rennie, M. D., Kennedy, P. J., Mills, K. H., Podemski, C. L., Rodgers, C. M. C., Charles, C., Hrenchuk, L. E., Chalanchuk, S., Blanchfield, P. J., & Paterson, M. J. (2019). *Impacts of freshwater aquaculture on fish communities : A whole- ecosystem experimental approach*. January, 870–885. <https://doi.org/10.1111/fwb.13269>
- Robinson, C. L. K., & Tonn, W. M. (1989). Influence of Environmental Factors and Piscivory in Structuring Fish Assemblages of Small Alberta Lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, 46, 81–89.
- Sandstrom, S., Rawson, M., & Lester, N. (2013). Ministry of Natural Resources Ontario: Manual of instructions for broad-scale fish community monitoring using North American (NA1) and Ontario small mesh (ON2) gillnets. *Review Literature And Arts Of The Americas*, 1–90.
- Scheibel, N. C., Dembkowski, D. J., Davis, J. L., & Chipps, S. R. (2016). Impacts of Northern Pike on Stocked Rainbow Trout in Pactola Reservoir, South Dakota. *North American Journal of Fisheries Management*, 36(2), 230–240. <https://doi.org/10.1080/02755947.2015.1116472>
- Schindler, D. W., Armstrong, F. A. J., Holmgren, S. K., & Brunskill, G. J. (1971). Eutrophication of Lake 227, Experimental Lakes Area, Northwestern Ontario, by Addition of Phosphate and Nitrate. *Journal of the Fisheries Research Board of Canada*, 28(11), 1763–1782. <https://doi.org/10.1139/f71-261>
- Schindler, D. W., Bayley, S. E., Parker, B. R., Beaty, K. G., Cruikshank, D. R., Fee, E. J., Schindler, E. U., & Stainton, M. P. (1996). The effects of climatic warming on the properties of boreal lakes and streams at the Experimental Lakes Area, northwestern Ontario. *Limnology and Oceanography*, 41(5), 1004–1017. <https://doi.org/10.4319/lo.1996.41.5.1004>

- Schindler, D. W., Curtis, P. J., Bayley, S. E., Parker, B. R., Beaty, K. G., & Stainton, M. P. (1997). Climate-induced changes in the dissolved organic carbon budgets of boreal lakes. *Biogeochemistry*, *36*(1), 9–28. <https://doi.org/10.1023/A:1005792014547>
- Schindler, D. W., Fee, E. J., & Ruzsyczynski, T. (1978). Phosphorous input and its consequences for phytoplankton standing crop and production in the experimental lakes area and in similar lakes. *J. Fish. Res. Board Canada*, *35*(2), 190–196. <https://doi.org/10.1139/f78-031>
- Scott, W. B., & Crossman, E. J. (1998). *Freshwater Fishes of Canada*. Galt House Publications LTD.
- Sellers, T. J., Parker, B. R., Schindler, D. W., & Tonn, W. M. (1998). Pelagic distribution of lake trout (*Salvelinus namaycush*) in small Canadian Shield lakes with respect to temperature, dissolved oxygen, and light. *Canadian Journal of Fisheries and Aquatic Sciences*, *55*(1), 170–179. <https://doi.org/10.1139/f97-232>
- Sharma, S., Legendre, P., de Cáceres, M., & Boisclair, D. (2011). The role of environmental and spatial processes in structuring native and non-native fish communities across thousands of lakes. *Ecography*, *34*(5), 762–771. <https://doi.org/10.1111/j.1600-0587.2010.06811.x>
- Sherbo, B. A. H., Tonin, J., Paterson, M. J., Hann, B. J., Kozak, J., & Higgins, S. N. (2023). The effects of terrestrial dissolved organic matter on phytoplankton biomass and productivity in boreal lakes. *Freshwater Biology*, *68*(12), 2109–2119. <https://doi.org/10.1111/fwb.14178>
- Slongo, B. (2022). *The Effects of Climate Change on the Growth and Spawning Phenology of Fishes in Aquatic Boreal Environments*.
- Smith, C. L., & Powell, C. R. (1971). *The Summer Fish Communities of Brier Creek, Marshall County, Oklahoma*.
- Soukup, J. C., & Wisenden, B. D. (2023). Predator recognition learning by northern redbelly dace *Chrosomus eos* from a small kettle lake with a dynamic predator community. *Environmental Biology of Fishes*, *106*(12), 2193–2204. <https://doi.org/10.1007/s10641-023-01500-3>
- Spens, J., & Ball, J. P. (2008). Salmonid or nonsalmonid lakes: Predicting the fate of northern boreal fish communities with hierarchical filters relating to a keystone piscivore. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(9), 1945–1955. <https://doi.org/10.1139/F08-103>
- Stasiak, R., & Cunningham, G. R. (2006). *Prepared for the USDA Forest Service, Rocky Mountain Region, Species Conservation Project Finescale Dace (*Phoxinus neogaeus*): A Technical Conservation Assessment*. <http://www.fs.fed.us/r2/>
- Stasko, A. D., Gunn, J. M., & Johnston, T. A. (2012). Role of ambient light in structuring north-temperate fish communities: Potential effects of increasing dissolved organic carbon concentration with a changing climate. *Environmental Reviews*, *20*(3), 173–190. <https://doi.org/10.1139/a2012-010>

- Thompson, P. L., Guzman, L. M., De Meester, L., Horváth, Z., Ptacnik, R., Vanschoenwinkel, B., Viana, D. S., & Chase, J. M. (2020). A process-based metacommunity framework linking local and regional scale community ecology. *Ecology Letters*, 23(9), 1314–1329. <https://doi.org/10.1111/ele.13568>
- Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., Jonge, M. M. J. De, Oksanen, J., & Ovaskainen, O. (2020). *Joint species distribution modelling with the r - package H msc. 2020*(September 2019), 442–447. <https://doi.org/10.1111/2041-210X.13345>
- Urban, M. C. (2004). Disturbance heterogeneity determines freshwater metacommunity structure. *Ecology*, 85(11), 2971–2978. <https://doi.org/10.1890/03-0631>
- Vallé, C., Poggiato, G., Thuiller, W., Jiguet, F., Princé, K., & Le Viol, I. (2023). Species associations in joint species distribution models: from missing variables to conditional predictions. *Journal of Biogeography*, August, 1–14. <https://doi.org/10.1111/jbi.14752>
- Wagner, T., Hansen, G. J. A., Schliep, E. M., Bethke, B. J., Honsey, A. E., Jacobson, P. C., Kline, B. C., & White, S. L. (2020). Improved understanding and prediction of freshwater fish communities through the use of joint species distribution models. *Canadian Journal of Fisheries and Aquatic Sciences*, 77(9), 1540–1551. <https://doi.org/10.1139/cjfas-2019-0348>
- Warton, D. I. (2005). Many zeros does not mean zero inflation: Comparing the goodness-of-fit of parametric models to multivariate abundance data. *Environmetrics*, 16(3), 275–289. <https://doi.org/10.1002/env.702>
- Wilson, D. S. (1992). *Complex Interactions in Metacommunities, with Implications for Biodiversity and Higher Levels of Selection* (Vol. 73, Issue 6).
- Wudu, K., Abegaz, A., Ayele, L., & Ybabe, M. (2023). The impacts of climate change on biodiversity loss and its remedial measures using nature based conservation approach: a global perspective. *Biodiversity and Conservation*, 32(12), 3681–3701. <https://doi.org/10.1007/s10531-023-02656-1>

Appendix A

Table A. Summary of data sources for fish abundance, water chemistry, and lake morphology information used for JSDM modeling. AHI refers to Aquatic habitat inventory survey data.

* indicates lakes where only presence-absence data are available for fish abundance

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L92	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey
L93	1982	Gillnets and minnow traps	1982	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L100	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976
L101	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L102	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976
L104	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1991-1993, 1996	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L111	1982; 1990	Gillnets and minnow traps; Gillnets	1981, 1982, 1986-1990	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L128	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976
L130	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally	1973, 1978, 1982, 1986	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
		trapnets and rotenone						
L131	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1978, 1982, 1986	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L132	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1970, 1973, 1978, 1982, 1986, 1992, 2008	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L164	2019	Trap nets	1978, 1982, 1986-1994, 1997, 2008, 2016	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L165	2019	Trap nets	1986-1997, 2008	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L223	1973; 2019	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Gillnets and angling	1973-2004, 2008, 2012, 2015-2022	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L224	1973; 2019	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Gillnets and angling	1972, 1974-1978, 1981-2009, 2011-2022	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L225	1973; 2019	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Gillnets and minnow traps	1973, 1978, 1982, 1985-1989, 1992, 1996, 2004	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L226	1973; 2014	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Gillnets	1971, 1973-1983, 1985-1987, 1989-1997, 2001-2004	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L227	1973	Gillnets, wire minnow traps, dipnets, beach seines and	1970-2020	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
		occasionally trapnets and rotenone						
L228 (Teggau)	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1970, 1973, 1992, 1995	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L239	1973; 2019	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Trap nets, gillnets, angling, and seine netting	1973	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L240	1973; 1999-2019	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Gillnets, angling, and seine netting	1970-1972, 1974-1978, 1980-2008, 2016	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L254 (Ethelma)*	1975; Not specified	Not specified	2008	AHI; IISD ELA	AHI	IISD ELA	IISD ELA	DOC estimated based on water colour (Molot & Dillon, 1997)
L257	2014	Gillnets, minnow traps, seine netting	1978, 1982, 1986	IISD ELA; (Lennox, 2019)	Estimated based on lake surface area, elevation, and surrounding topography (Heathcote et al., 2015).	Total phosphorus estimated based on conductivity and lake depth (Chow-Fraser, 1991).	IISD ELA	IISD ELA
L258	1996	Gillnets	2021	IISD ELA	IISD ELA	IISD ELA	pH estimated using lake surface area (Beamish et al., 1976).	IISD ELA
L259	1973; 2002	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and	1973, 1978, 1982, 1986, 2021	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
		rotenone; Gillnets						
L261	1973; 2013	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Not specified	1968-1978, 1982, 1986	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L262 (Rogers)	1973; 1988	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Gillnets	1973, 1978, 1982, 1986, 1987, 1990	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L303	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1970-1978, 1980-1983, 1985-1987, 1989, 1992-1993, 1996, 2008-2011, 2015-2021	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L305	1973	Gillnets, wire minnow traps, dipnets, beach seines and	1970-1974, 1978, 1982-1983, 1985-1993, 1995, 1998,	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
		occasionally trapnets and rotenone	2003, 2008, 2014-2021					
L311	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976
L313	1997	Gillnets	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976
L315	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1978, 1982, 1986	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L316	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1978, 1982, 1986	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L317	1973	Gillnets, wire minnow traps, dipnets, beach seines and	1973, 1978, 1982, 1986	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
		occasionally trapnets and rotenone						
L322	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976
L373	2019	Gillnets and angling	1982, 1983, 1985-1987, 1989-1992, 1994-2009, 2011-2022	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L374 (Porcus)	1983; 2010	Gillnets; Gillnets	1983-1989, 2010	IISD ELA; OMNR	IISD ELA; OMNR	IISD ELA; OMNR	IISD ELA; OMNR	IISD ELA; OMNR
L375	1986-2019	Trap nets; Gillnets	1982, 1983, 1986-1994, 1997, 2001-2009, 2011, 2021, 2022	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L376 (Manomin)	1982; 2005	Not specified; Gillnets and seine nets	1984, 2007	AHI; IISD ELA	AHI	Total phosphorus estimated based on conductivity and lake depth (Chow-Fraser, 1991).	IISD ELA	IISD ELA
L377	1992	Gillnets and minnow traps	1978, 1982, 1986-1993, 2004, 2006-2007	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L379	1980	Not specified	1987, 2007	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L382	1973; 2003	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone; Gillnets	1973, 1977-1980, 1982, 1983, 1985-1998, 2001-2004	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L429	2017	Gillnets	2017	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L464 (Highwind)*	1967; 2023	Not specified; Angling and personal communications	1967	AHI; Personal communications	AHI	Total phosphorus estimated based on conductivity and lake depth (Chow-Fraser, 1991).	pH estimated using lake surface area (Beamish et al., 1976).	DOC estimated based on water colour (Molot & Dillon, 1997)
L466 (Dryberry)	1974; 2010; 2016; 2023	Not specified; Gillnets and minnow traps; Gillnets and minnow traps; Gillnets and minnow traps	2010, 2016	AHI; OMNR	AHI; OMNR	OMNR	OMNR	OMNR
L467 (MacDonald)	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally	1973, 1982, 1986	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
		trapnets and rotenone						
L468 (Roddy)	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1978, 1982, 1983, 1986, 1990, 1992, 1993, 1996	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L469 (Eagle)	1973; 2016; 2022	Not specified; Gillnets and minnow traps; Gillnets and minnow traps	2016, 2022	AHI; OMNR	AHI; OMNR	OMNR	OMNR	OMNR
L470	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1995-1997, 2008, 2016, 2018	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L589 (Silvery)*	1977	Not specified	1977	AHI	AHI	Total phosphorus estimated based on conductivity and lake depth (Chow-Fraser, 1991).	pH estimated using lake surface area (Beamish et al., 1976).	DOC estimated based on water colour (Molot & Dillon, 1997)

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L592	2022	Gillnets, minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey
L594	2022	Gillnets, minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey
L609 (Peterson)	1981	Gillnets and minnow traps	1981	OMNR; AHI	OMNR; AHI	Total phosphorus estimated based on conductivity and lake depth (Chow-Fraser, 1991).	OMNR	DOC estimated based on water colour (Molot & Dillon, 1997)
L611 (Lacourse)	1981	Gillnets and minnow traps	1981	OMNR; AHI	OMNR; AHI	Total phosphorus estimated based on conductivity and lake depth (Chow-Fraser, 1991).	OMNR	DOC estimated based on water colour (Molot & Dillon, 1997)
L613	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey
L614	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L615 (Windermere)	1976; 1990; 2010	Not specified; Gillnets; Gillnets	1982, 1987, 2010	OMNR	AHI	OMNR	OMNR	OMNR
L622	1983	Gillnets	1981-1984, 1986, 2021	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L623	1983	Gillnets and angling	1981, 1983-1986, 2021	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L624	2018	Gillnets and minnow traps	2018	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L625	2018	Gillnets	1987, 1996, 2010-2019	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L627	2016	Gillnets	1981, 1987, 2007, 2015	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L628	2016	Gillnets	1987	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L629	1982	Gillnets, minnow traps	1981-1986	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L631	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976
L634	Not specified	Not specified	2023	IISD ELA	Online resource	IISD ELA	IISD ELA	DOC estimated based on water colour (Molot &

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
								Dillon, 1997)
L649 (Lower Stewart)	1975; 2017; Not specified	Not specified; Gillnets; Not specified	1975, 1987, 2017	AHI; OMNR; IISD ELA	AHI; OMNR	AHI; OMNR; IISD ELA	AHI; OMNR; IISD ELA	AHI; OMNR; IISD ELA
L653 (Geejay)	1976; 1982	Rotenone; Not specified	1982, 1987	OMNR; AHI	AHI	AHI; IISD ELA	AHI; IISD ELA	AHI; IISD ELA
L659	Not specified	Gillnets and minnow traps	1981, 2007, 2019	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L660 (Winnange)	1985; 2010; 2015	Not specified; Gillnets; Gillnets	1985, 1990, 2010, 2015	AHI; IISD ELA	AHI; IISD ELA	IISD ELA	IISD ELA	IISD ELA
L661	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1978, 1981-1983, 1986, 1992, 1996, 2008	Beamish et al., 1976	Beamish et al., 1976	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA	Beamish et al., 1976; IISD ELA
L663	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1978, 1982, 1986, 1992	Beamish et al., 1976	Estimated based on lake surface area, elevation, and surrounding topography (Heathcote et al., 2015).	IISD ELA	IISD ELA	IISD ELA

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source					
					<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
L664	1973	Gillnets, wire minnow traps, dipnets, beach seines and occasionally trapnets and rotenone	1973, 1978, 1982, 1986	Beamish et al., 1976	Estimated based on lake surface area, elevation, and surrounding topography (Heathcote et al., 2015).	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L665	Not specified	Gillnets	2017	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L677	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey
L938	Not specified	Trap nets and gillnets	1986-1993, 1996, 2008	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L979	2015	Gillnets and minnow traps	1981, 1982, 1991, 1992, 1998, 2001-2004, 2006, 2008, 2009	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA	IISD ELA
L467OF (MacDonald Outflow)	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey
L468OF (Roddy Outflow)	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey
L625OF (625 Outflow)	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey

Lake	Fish Survey Year	Fish Survey Methods	Water Sampling Year	Data Source				
				<i>Fish Abundance</i>	<i>Maximum Lake Depth</i>	<i>Total phosphorus</i>	<i>pH</i>	<i>DOC</i>
Hornell	2022	Gillnets and minnow traps	2022	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey	Lakehead University Survey

Appendix B

Existing inaccuracies in the Ontario Integrated Hydrology Data

All GIS-related analyses were performed using ArcGIS Pro 3.1. The surface area of study lakes was determined using the Ontario Integrated Hydrology Data (OIHD) v1.0 (OMNRF, 2012), in conjunction with high-resolution imagery and digital elevation models (DEMs) from a 2017 aerial survey of the area, provided by the ELA. Any discrepancies in watercourse polylines and lake polygons from OIHD were corrected using both high-resolution imagery and field assessment observations.

Several corrections were made to watercourses within the study area. The OIHD inaccurately depicts a stream connecting L256 (Veronica Lake) to L101 in the Dryberry watershed, where no such connection exists. Likewise, it incorrectly shows a stream linking L259 to L468 (Roddy Lake) in the Eagle watershed, while omitting the actual stream connecting L259 and L379 in the same watershed. Additionally, in 2010, ELA researchers constructed a 200-meter diversion channel to redirect flow from L627 into L625 instead of L626, although the OIHD still depicts L627 as connected to L626.

Appendix C

Table C. Information on spatial weights used for AEM and MEM spatial eigenvectors.

Spatial Weight	Description	Calculation Method	Units of Measurement	Type of Spatial Eigenvector
None	Only using connectivity (MEM) and connectivity with directionality (AEM).	N/A	N/A	MEM and AEM
Hydrological Distance	The measured distance between connecting lakes following along their associated stream.	ArcGIS approach using the Ontario Integrated Hydrology Data (OIHD) v1.0 (OMNRF, 2012) lake polygons and stream polylines. The spatial analyst tool <i>Add Surface Information</i> was used to measure the distance between lake polygons of interest along stream polylines.	Meters (m)	MEM and AEM
Overland Distance	The measured overland distance between study lakes, regardless of hydrological connectivity to one another.	ArcGIS and <i>adespatial</i> (R-package) approach using the spatial analyst tool <i>Add Surface Information</i> to determine center points for lake polygons. Distances between center points of all lake polygons of interest were determined in R using <i>adespatial</i> .	Meters (m)	MEM only
Stream Sinuosity	The ratio of the stream length to straight line distance, determining how straight or meandering a stream between two lakes are.	ArcGIS approach using <i>Sinuosity Toolbox</i> developed by Dils and Yang 2015 which uses the Ontario Integrated Hydrology Data (OIHD) v1.0 (OMNRF, 2012) stream polylines to calculate the ratio of the stream length to straight line distance.	Ratio of the stream length to straight line distance (0-1); where 1 is equal to a completely straight line.	MEM and AEM

Spatial Weight	Description	Calculation Method	Units of Measurement	Type of Spatial Eigenvector
Average Stream Gradient	The average slope along a stream connecting two lakes of interest.	ArcGIS approach using the Ontario Integrated Hydrology Data (OIHD) v1.0 (OMNRF, 2012) lake polygons and stream polylines along with digital elevation models (DEMs) from a 2017 aerial survey of the area, provided by IISD-ELA. The spatial analyst tool <i>Add Surface Information</i> was used to calculate the average slope for a stream polyline by taking a weighted average of the slope from each line segment. The weighting is based on the 3D length of each segment. This results in longer segments having greater influence on the resulting calculation over shorter segments.	Percentage units (Grade)	AEM only
Maximum Stream Gradient	The maximum slope along a stream connecting two lakes of interest.	ArcGIS approach using the Ontario Integrated Hydrology Data (OIHD) v1.0 (OMNRF, 2012) lake polygons and stream polylines along with digital elevation models (DEMs) from a 2017 aerial survey of the area, provided by IISD-ELA. The spatial analyst tool <i>Add Surface Information</i> was used to calculate the maximum slope from the segment with the largest calculated value.	Percentage units (Grade)	AEM only
Fish Barrier Assessment Score	An assessment score to determine the difficulty for a fish to move upstream along a stream.	Calculated following the Alberta Conservation Association's (ACA) Guide to Waterfall Fish Barrier Assessment (Blackburn et al., n.d.).	Index (0-3); where 3 suggests a watercourse is unpassable while moving upstream.	AEM only

Appendix D

Table D. Summary of Spatial-only Joint Species Distribution models (JSDMs).

Spatial Model	Fish Data Type	Eigenvector Mapping¹	Spatial Weight	Spatial Weight Formula	Flow direction of watercourses
SM1	Presence-Absence	AEM	Maximum stream gradient along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM2	Presence-Absence	AEM	Hydrological connectivity only	-	Y
SM3	Presence-Absence	AEM	Average stream gradient along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM4	Presence-Absence	AEM	Hydrological watercourse length	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM5	Presence-Absence	AEM	Maximum fish barrier assessment score along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM6	Presence-Absence	AEM	Presence of fish barrier	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM7	Presence-Absence	AEM	Stream sinuosity along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM8	Presence-Absence	MEM	Hydrological watercourse length	$f= 1-d_{ij}/\max(d_{ij})$	N
SM9	Presence-Absence	MEM	Overland distance between lakes	$f= 1-d_{ij}/\max(d_{ij})$	N
SM10	Presence-Absence	MEM	Hydrological connectivity only	-	N
SM11	Presence-Absence	MEM	Stream sinuosity along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	N
SM12	Presence-Absence	MEM	Overland distance between lakes	$f= 1-d_{ij}/\max(d_{ij})$	N
SM13	Relative Abundance	AEM	Average stream gradient along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM14	Relative Abundance	AEM	Stream sinuosity along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM15	Relative Abundance	AEM	Hydrological connectivity only	-	Y
SM16	Relative Abundance	AEM	Maximum stream gradient along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM17	Relative Abundance	AEM	Hydrological watercourse length	$f= 1-d_{ij}/\max(d_{ij})$	Y
SM18	Relative Abundance	AEM	Maximum fish barrier assessment score along watercourse	$f= 1-d_{ij}/\max(d_{ij})$	Y

Spatial Model	Fish Data Type	Eigenvector Mapping¹	Spatial Weight	Spatial Weight Formula	Flow direction of watercourses
SM19	Relative Abundance	AEM	Presence of fish barrier	$f = 1 - d_{ij} / \max(d_{ij})$	Y
SM20	Relative Abundance	MEM	Hydrological connectivity only	-	N
SM21	Relative Abundance	MEM	Hydrological watercourse length	$f = 1 - d_{ij} / \max(d_{ij})$	N
SM22	Relative Abundance	MEM	Stream sinuosity along watercourse	$f = 1 - d_{ij} / \max(d_{ij})$	N

¹AEM refers to asymmetric eigenvector mapping (Blanchet et al., 2008). MEM refers to Moran's eigenvector mapping (Blanchet et al., 2011; Dray et al., 2006).

Appendix E

Table E. Summary of average environmental variables for each of the 81 lakes included in this study.

Lake	Watershed	Lake Surface Area (ha)	Maximum Lake Depth (m)	pH	Conductivity ($\mu\text{S}/\text{cm}$)	Total Nitrogen ($\mu\text{g}/\text{L}$)	Total Phosphorus ($\mu\text{g}/\text{L}$)	DOC (mg/L)
L92	Eagle	9.56	1.4	6.26	14.94	843.0	15.8	14.86
L93	Dryberry	6.38	8.0	6.57	22.24	430.8	10.7	8.04
L100	Dryberry	10.47	12.8	6.90	23.00	283.0	22.0	7.81
L101	Dryberry	11.14	0.5	6.90	17.00	414.0	21.0	5.96
L102	Dryberry	4.75	2.0	6.60	18.00	413.0	14.0	7.12
L104	Dryberry	7.88	5.0	6.72	24.00	375.0	10.7	9.30
L111	Dryberry	9.47	32.0	6.78	21.06	303.4	5.7	7.44
L128	Dryberry	8.47	2.0	6.80	20.00	295.0	8.0	8.67
L130	Dryberry	7.33	6.0	6.91	20.00	323.0	13.0	5.72
L131	Dryberry	3.32	2.0	6.26	16.60	257.0	11.0	7.56
L132	Dryberry	6.93	8.0	6.55	16.71	372.0	9.7	7.61
L164	Dryberry	21.39	7.0	6.70	21.14	395.5	9.5	11.19
L165	Dryberry	17.64	5.0	6.68	21.77	394.6	9.3	11.39
L227	Eagle	5.17	10.0	7.88	30.71	974.0	41.5	12.46
L228 (Teggau)	Eagle	1333.19	167.0	6.90	19.00	298.0	8.0	3.63
L239	Dryberry	54.62	30.4	6.70	25.00	324.0	32.0	8.36
L240	Dryberry	44.07	13.0	6.96	27.67	332.5	7.9	6.94
L254 (Ethelma)	Dryberry	394.99	33.6	7.00	30.00	255.0	4.0	5.03
L257	Eagle	28.42	23.1	6.62	23.00	397.1	8.8	10.21
L258	Eagle	71.98	138.0	7.34	35.00	340.9	6.4	8.32
L259	Eagle	96.82	20.0	6.62	21.40	278.7	12.0	6.45

Lake	Watershed	Lake Surface Area (ha)	Maximum Lake Depth (m)	pH	Conductivity (µS/cm)	Total Nitrogen (µg/L)	Total Phosphorus (µg/L)	DOC (mg/L)
L303	Dryberry	9.78	3.0	7.49	25.01	689.4	19.4	9.62
L305	Eagle	57.81	33.0	7.27	26.56	252.2	6.4	4.94
L311	Eagle	6.40	11.9	6.50	18.00	341.0	8.0	9.12
L313	Eagle	133.38	24.4	6.65	19.00	323.0	9.0	10.86
L315	Eagle	40.72	17.0	6.92	26.20	314.0	14.0	8.57
L316	Eagle	21.33	5.0	6.46	23.80	341.0	7.0	12.65
L317	Eagle	28.67	19.0	6.98	26.80	363.0	8.0	8.56
L322	Eagle	83.58	19.0	6.80	20.00	232.0	9.0	7.69
L374 (Porcus)	Dryberry	163.60	30.0	7.13	28.68	264.0	5.3	4.49
L376 (Manomin)	Eagle	528.18	80.1	7.16	38.00	242.5	3.5	6.73
L377	Eagle	27.54	18.0	6.96	24.27	259.0	6.1	6.37
L379	Eagle	163.60	26.7	7.17	25.00	256.0	7.0	5.52
L429	Eagle	15.69	8.3	6.80	16.90	516.5	6.2	7.17
L464 (Highwind)	Dryberry	777.19	62.2	7.00	30.00	263.4	6.1	5.03
L467 (MacDonald)	Dryberry	197.85	21.0	6.79	23.00	301.0	7.0	8.82
L468 (Roddy)	Eagle	297.27	25.0	7.72	28.45	342.7	9.2	6.14
L470	Dryberry	5.61	2.0	6.28	16.50	516.3	12.4	11.79
L589 (Silvery)	Dryberry	473.73	39.1	6.98	30.00	195.9	5.5	5.03
L592	Eagle	34.47	17.6	6.31	19.60	360.0	5.1	9.20
L594	Dryberry	66.71	13.6	6.44	18.46	366.0	5.4	5.03
L609 (Peterson)	Eagle	12.94	10.0	7.00	70.00	478.3	11.9	5.03
L611 (Lacourse)	Eagle	37.78	6.0	7.00	60.00	526.4	13.9	5.03
L613	Eagle	16.48	30.5	6.43	76.11	560.0	6.2	13.62
L614	Eagle	8.17	23.0	5.85	25.07	273.0	5.6	6.83
L615 (Windemere)	Eagle	96.62	49.0	7.18	41.30	264.0	4.0	3.30
L622	Eagle	40.19	31.0	6.91	25.17	324.7	5.9	7.44
L623	Eagle	37.13	21.0	6.90	28.63	369.6	6.5	6.52

Lake	Watershed	Lake Surface Area (ha)	Maximum Lake Depth (m)	pH	Conductivity (µS/cm)	Total Nitrogen (µg/L)	Total Phosphorus (µg/L)	DOC (mg/L)
L624	Eagle	0.76	1.0	6.50	23.00	493.2	10.0	8.36
L625	Eagle	82.07	30.0	7.86	25.07	378.4	8.8	7.37
L627	Eagle	35.28	8.0	6.99	18.50	310.5	6.3	5.89
L628	Eagle	20.08	18.6	6.72	23.00	290.5	6.0	8.53
L629	Dryberry	62.13	18.0	6.66	21.42	360.2	7.8	7.07
L631	Dryberry	39.39	8.0	7.00	23.00	315.0	15.0	6.80
L634	Dryberry	55.79	18.3	6.13	14.40	479.5	10.5	5.03
L649 (Lower Stewart)	Eagle	168.35	29.6	7.09	53.60	320.0	6.0	8.08
L653 (Geejay)	Eagle	188.00	46.3	7.31	43.00	271.4	3.0	6.61
L659	Eagle	27.25	7.0	6.86	20.30	258.5	10.4	6.43
L660 (Winnange)	Eagle	2421.43	150.0	7.10	35.63	181.4	4.7	3.78
L661	Dryberry	1.22	1.0	5.95	26.64	649.9	14.9	15.03
L663	Dryberry	41.69	21.4	6.68	25.17	398.0	23.0	10.46
L664	Dryberry	77.72	29.0	6.78	28.00	296.0	32.0	7.92
L665	Dryberry	29.42	16.3	6.78	15.70	347.2	4.8	8.42
L677	Eagle	3.30	1.5	6.45	17.93	315.0	7.4	7.37
L938	Dryberry	19.23	6.0	7.03	33.50	235.0	5.7	5.68
L979	Dryberry	2.26	1.0	6.34	23.97	594.9	16.2	11.62
L467OF (MacDonald Outflow)	Dryberry	0.56	1.7	6.18	18.60	358.0	3.3	10.56
L468OF (Roddy Outflow)	Eagle	1.23	1.1	6.49	16.48	334.0	3.9	6.65
L625OF (625 Outflow)	Eagle	0.38	1.1	5.74	14.60	783.0	11.2	8.31
Hornell	Eagle	5.45	2.5	5.81	13.95	653.0	15.1	14.21
L223	Eagle	27.30	14.0	6.90	22.40	294.6	6.4	5.15
L224	Eagle	26.24	27.0	6.96	18.95	238.2	5.8	3.38

Lake	Watershed	Lake Surface Area (ha)	Maximum Lake Depth (m)	pH	Conductivity ($\mu\text{S/cm}$)	Total Nitrogen ($\mu\text{g/L}$)	Total Phosphorus ($\mu\text{g/L}$)	DOC (mg/L)
L225	Eagle	3.71	2.0	5.15	22.65	570.9	13.0	10.17
L226	Eagle	16.62	15.0	6.94	22.84	427.9	14.1	6.66
L261	Eagle	5.75	10.0	6.00	16.78	416.8	17.9	7.90
L262 (Rogers)	Eagle	82.55	30.0	6.99	25.67	253.5	5.6	6.45
L373	Eagle	27.83	21.0	7.21	26.11	250.5	6.6	4.24
L375	Eagle	23.03	27.0	7.55	48.50	258.1	5.1	5.67
L382	Eagle	37.79	13.0	6.78	22.59	369.3	7.2	7.41
L466 (Dryberry)	Dryberry	9545.98	105.5	6.80	28.60	188.0	4.8	3.70
L469 (Eagle)	Eagle	2430.48	33.6	7.08	46.75	331.0	11.4	6.05

Appendix F

Table F. Species specific explanatory (Tjur R², AUC) and predictive (10-fold cross validation Tjur R² and AUC) power for Model ESC (presence-absence).

Common Name	Latin Name	Tjur R ² ¹	10-fold Cross Validation Tjur R ² ²	AUC ¹	10-fold Cross Validation AUC ²
Blacknose Shiner	<i>Notropis heterolepis</i>	0.1188	0.0015	0.6895	0.5231
Burbot	<i>Lota lota</i>	0.5061	0.3761	0.9606	0.9030
Fathead Minnow	<i>Pimephales promelas</i>	0.4693	0.0783	0.9712	0.6550
Finescale Dace	<i>Chrosomus neogaeus</i>	0.3687	0.0552	0.9379	0.5879
Johnny Darter	<i>Etheostoma nigrum</i>	0.4275	0.2930	0.9265	0.8133
Lake Herring (Cisco)	<i>Coregonus artedi</i>	0.4133	0.2846	0.9174	0.8004
Lake Trout	<i>Salvelinus namaycush</i>	0.6192	0.5207	0.9778	0.9265
Lake Whitefish	<i>Coregonus clupeaformis</i>	0.4535	0.3744	0.9270	0.8779
Longnose Dace	<i>Rhinichthys cataractae</i>	0.2031	0.0811	0.8294	0.6638
Northern Pearl Dace	<i>Margariscus margarita</i>	0.2233	0.0626	0.7918	0.5620
Northern Pike	<i>Esox lucius</i>	0.5838	0.4734	0.9474	0.8871
Northern Redbelly Dace	<i>Chrosomus eos</i>	0.6070	-0.0008	0.9982	0.5485

Common Name	Latin Name	Tjur R^{2 1}	10-fold Cross Validation Tjur R^{2 2}	AUC¹	10-fold Cross Validation AUC²
Slimy Sculpin	<i>Cottus cognatus</i>	0.2998	0.1074	0.8879	0.6545
Spottail Shiner	<i>Notropis hudsonius</i>	0.2227	0.1042	0.7889	0.6162
White Sucker	<i>Catostomus commersonii</i>	0.4817	0.2374	0.9688	0.8156
Yellow Perch	<i>Perca flavescens</i>	0.4208	0.0695	0.9505	0.6289

Appendix References

- Beamish, R., Blouw, L., & McFarlane, G. (1976). A fish and chemical study of 109 lakes in the Experimental Lakes Area (ELA), Northwestern Ontario, with appended reports on lake whitefish ageing errors and the Northwestern Ontario baitfish industry. In *Department of the Environment: Fisheries and Marine Service*.
- Blanchet, F. G., Legendre, P., Maranger, R., Monti, D., & Pepin, P. (2011). Modelling the effect of directional spatial ecological processes at different scales. *Oecologia*, *166*(2), 357–368. <https://doi.org/10.1007/s00442-010-1867-y>
- Chow-Fraser, P. (1991). Use of the morphoedaphic index to predict nutrient status and algal biomass in some Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, *48*(10), 1909–1918. <https://doi.org/10.1139/f91-227>
- Dray, S., Legendre, P., & Peres-neto, P. R. (2006). *Spatial modelling : a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM)*. 196, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Heathcote, A. J., Giorgio, P. A., & Prairie, Y. T. (2015). Predicting bathymetric features of lakes from the topography of their surrounding landscape. *Canadian Journal of Fisheries and Aquatic Sciences*, *72*(5), 643–650. <https://doi.org/10.1139/cjfas-2014-0392>
- Lennox, P. A. (2019). *Examining the relationship between species richness of fish communities and their biomass in freshwater ecosystems*. University of Lethbridge.
- Molot, L. A., & Dillon, P. J. (1997). Colour - mass balances and colour - dissolved organic carbon relationships in lakes and streams in central Ontario. *Canadian Journal of Fisheries and Aquatic Sciences*, *54*(12), 2789–2795. <https://doi.org/10.1139/cjfas-54-12-2789>