LAND COVER AND STREAM BIOLOGICAL INTEGRITY IN NORTH-CENTRAL INDIANA

by

Alex Adams

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science

Department of Forestry and Natural Resources

West Lafayette, Indiana

May 2024

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Mo Zhou, Co-Chair

Department of Forestry and Natural Resources

Dr. Reuben Goforth, Co-Chair

Department of Forestry and Natural Resources

Dr. Sara McMillan

Department of Agricultural and Biosystems Engineering

Iowa State University

Approved by:

Dr. Jessica P. Gurevitch

This study is dedicated to my mother, who is my first line of support, and the Purdue Department of Forestry and Natural Resources, who have treated me like family.

ACKNOWLEDGMENTS

The author gratefully acknowledges the Cropland Data Layer logistical and programming guidance from the Purdue University FACAI Lab (Dr. Jingjing Liang and Akane Ota-Abbasi). The author also gratefully acknowledges the assistance in field data collection from Macy Mahrenholz and Shelby Mathis. Lastly, the author gratefully acknowledges the statistical resources provided by Dr. Esteban Fernandez-Juricic of the Purdue University Department of Biological Sciences.

TABLE OF CONTENTS

LIST OF TABLES 7
LIST OF FIGURES 8
LIST OF ABBREVIATIONS 10
ABSTRACT 11
INTRODUCTION 12
CHAPTER 1. LITERATURE REVIEW 15
1.1 Role of Forests 15
1.2 Stream Health Metrics 16
1.3 Agriculture and Stream Integrity 18
1.4 Agricultural Drainage and the Role of Riparian Buffers 21
CHAPTER 2. LAND COVER CHANGE IN THE CROPLAND DATA LAYER 26
2.1 Data 26
2.2 Methods 26
2.2.1 Area of Interest Shapefile Export & Import 26
2.2.2 Land Cover Category Definition & Change Matrix 29
2.2.3 Visualizing Land Cover Change & Exporting a Raster 29
2.2.4 Technical Validation and Ground Truthing 30
2.3 Results 32
2.4 Discussion 34
2.5 Summary 37
CHAPTER 3. AGRICULTURAL AND FORESTED STREAM BIOLOGICAL INTEGRITY 38
3.1 Data 38
3.1.1 Site Descriptions 38
3.2 Methods and Materials 41
3.2.1 Site Sampling Procedures 41
3.2.2 Outlier Identification 42
3.2.3 Site Classification Metrics 44
3.2.4 Index of Biotic Integrity Calculation 45

3.2.5 Statistical Methods 46
3.3 Results 47
3.3.1 Site Classification Analysis 47
3.3.2 Summary Statistics 52
3.3.3 Statistical Tests 52
3.4 Discussion 58
3.5 Summary 61
APPENDIX A. LAND COVER CHANGE ANALYSIS R CODE 62
APPENDIX B. IBI DATA TABLE 64
REFERENCES 65

LIST OF TABLES

Table 3.1. Study sites sampled between June and July of 2022. Included are site code, stream name, latitude and longitude coordinates, Indiana county, and Level IV EPA Ecoregion. 40

Table 3.2. AICc model selection output table.. 46

Table 3.3. Parameters of sampled stream sites in north-central Indiana. QHEI scores and NDVI are also included. Land cover measure parameters indicating agricultural ("Ag") land are in bold; non-bolded items indicate forest ("Fo"). "QHEI Bank & Riparia" is the score from Section 4 of the Ohio EPA QHEI form. Riparian width is the average of the left and right riparia at the approximate center of the transect, perpendicular to the stream. CDL Ag:Forest is the ratio of pixel counts for agricultural and forest land cover taken from the USDA Cropland Data Layer. 50

Table 3.4. Summary statistics (mean ± standard deviation) of number of individuals, number of species, QHEI score, and IBI score for agricultural and forested streams. 52

Table 3.5. Assumptions and test selection for four dependent variables following land cover (IBI, number of individuals, number of species, QHEI). Normality was tested using the Shapiro-Wilk test. Homogeneity of variance was tested using mean-center and median-center (“med-center”) Levene’s tests.. 53

Table 3.6. Assumptions and test selection for five continuous variables (QHEI, IBI, number of intolerant species, percent tolerant species, percent cultivated crops in delineated study site drainage basin). QHEI was the independent variable to be used in analysis for all other variables except percent cultivated crops, which was analyzed with IBI. Normality was tested using the Shapiro-Wilk test. 54

Table B.1. Index of Biotic Integrity (IBI) component metrics for 21 sampled Indiana streams in Benton, Boone, Clinton, Howard, Montgomery, Tippecanoe, Warren, and White counties. Streams were sampled between 6/6/2023 and 7/20/2023. Fish samples were collected using a Smith-Root backpack electrofisher in one 100m or approximately 1800s transect. CPUE (catch per unit effort) is the total number of individuals captured at each site. DELTs are the number of fish displaying a deformity, erosion, lesion, or tumor. Total IBI scores and total species captured are also provided for sampled sites. Site 20 was omitted from further statistical analysis as an outlier.................. 64

LIST OF FIGURES

Figure 2.1. Selected Indiana counties in the study AOI. The extent of the created shapefile includes the above labeled 35 counties in northern Indiana... 28

Figure 2.2. Map of sixteen Indiana land cover sites visited for ground truthing between June and August 2022. Site types identified were afforested (n = 3), agriculture (n = 5), deforested (n = 3), and undisturbed forest (n = 5)... 31

Figure 2.3. Raster image of the Cropland Data Layer's measured landcover changes in Northern Indiana between 2010 and 2020. Lower right inset shows detail of Tippecanoe County, Indiana. 33

Figure 2.4. A comparison of CDL imagery for a selected area in northwestern Indiana during the years 2010, 2015, and 2020. Note increasing blue ("Woody wetlands") and tan ("Shrubland") areas with temporal progression.. 36

Figure 3.1. North-Central Indiana counties selected for fish sampling and QHEI assessment from June to July 2022. 39

Figure 3.2. Left: the author using an ABP-3 backpack electrofisher (ETS Electrofishing Systems, LLC; Madison, WI). Right: a forested study stream reach with downed woody debris. 42

Figure 3.3. Boxplots showing outliers in Indiana stream data for drainage area in km^2 (left; Figure 3.3a) and number of individual fishes captured (right; Figure 3.3b). Site 20 was an outlier for stream drainage (1041 km^2) and number of individuals (n = 716)... 43

Figure 3.4. A correlation matrix heatmap of the variables used in agriculture vs. forest land cover classification: Cropland Data Layer agriculture:forest ratio ("cdlrat"), average wooded riparian width ("ripav"), mean Normalized Difference Vegetation Index value ("ndvi_m"), and Qualitative Habitat Evaluation Index bank and riparia components ("qbr"). Mean NDVI values had a low correlation with other variables. 48

Figure 3.5. A biplot with square cosine values of variables used in agriculture vs. forest land cover classification. Higher square cosine values (cos2) indicate a more well-represented variable. CDL agriculture-to-forest pixel count ratio ("cdlrat") and QHEI Bank & Riparia subsection score ("qbr") are the two best-represented variables. Mean site NDVI ("ndvi_m") is very poorly represented. 49

Figure 3.6. Fish and QHEI stream sampling locations by land cover category (Agriculture, Forest). Sites were located in Benton, Boone, Clinton, Howard, Montgomery, Tippecanoe, Warren, and White counties in north-central Indiana.. 51

Figure 3.7. Boxplots of four variables by land cover for 20 sampled sites. Variables include number of individuals captured, number of species captured, Index of Biotic Integrity (IBI) score, and QHEI score. Outliers are labeled with site number.. 55

Figure 3.8. Boxplots of IBI score, QHEI score, number of intolerant species, and percent tolerant species for n = 20 sites. Outliers are labeled with site number; multiple outliers of the same value are labeled on one point. 56

Figure 3.9. Spearman Rank Correlation plots of biotic and abiotic measures for 20 stream sites. Clockwise from top left: QHEI score and Index of Biotic Integrity score (r = 0.27, p = 0.25); QHEI and percent tolerant species (r = -0.2, p > 0.30); IBI score and percent cultivated crops in drainage basin (r = -0.31, p = 0.178); QHEI and number of intolerant species (r = 0.23, p > 0.30). 57

LIST OF ABBREVIATIONS

AOI	Area of Interest
CDL	Cropland Data Layer
EPA	Environmental Protection Agency
EPT	Ephemeroptera Plecoptera Trichoptera
IBI	Index of Biotic Integrity
PCA	Principal Components Analysis
QHEI	Qualitative Habitat Evaluation Index
USDA	United States Department of Agriculture
USGS	United States Geological Survey

ABSTRACT

The Temperate Plains ecoregion of Indiana has experienced significant agricultural development since the 19th century, which has left streams vulnerable to impacts such as sedimentation and nutrient accumulation. This thesis describes first the accuracy of the USDA Cropland Data Layer (CDL) in land cover change, and second, the relationships between agricultural and forested land covers and stream biological integrity. I first employed the CDL to review land cover change, particularly relating to agriculture and forest, for the area of interest between 2010 and 2020. I determined that the CDL improved in accuracy for the area of interest in the chosen timeframe for non-agricultural and non-forest land cover. I concluded that the CDL was best used as a supplement to primary-source land cover measures. Next, I calculated the fish Index of Biotic Integrity (IBI) scores for 20 sampled agricultural and forested streams in North-Central Indiana. I also assessed the stream habitats at all sites using the Qualitative Habitat Evaluation Index (QHEI) and percent cultivated crops in drainage basin areas for all streams. Forested streams had significantly higher QHEI scores than agricultural streams (median = 62 and 40.4, respectively). No other relationships were statistically different, including IBI and land cover category, which may have been due to the small sample size (n = 20). I concluded that future studies may build on these findings by controlling for agricultural drainage types or using precise measures of forested land cover.

INTRODUCTION

Over half of all land in the United States is used for agriculture (Baker and Capel, 2011). Indiana is a heavily agricultural state, and over the last 200 years, approximately 75% of its forests have been removed (Carman, 2013), with 80% of all land cleared for crop and animal agriculture (Capel et al., 2018). Row crops such as corn and soybeans now dominate the central Indiana landscape, resulting in high vulnerability of streams to agricultural pollutants (Munn et al., 2018; U.S. EPA, 2020). Such farming practices can affect adjacent water bodies through erosion and sedimentation from tillage, nutrient loading associated with fertilizer application, and pesticide application (Meador and Frey, 2018; Munn et al., 2018; Nowell et al., 2018). Agricultural runoff, i.e., water flow over farm fields resulting from irrigation, precipitation, or snowmelt, is the primary source of these stressors. It can enter water bodies directly on the surface or through groundwater or drainage tile fallouts and impact stream habitats and biological integrity.

Fine sediment is also a major contributor to water quality impairment. For example, the National Rivers and Streams Assessment (NRSA) of 2013-2014 reported that 44% of all sampled stream miles had "Fair" or "Poor" levels of excess streambed fine sediments. Two-thirds of sampled stream miles in the Temperate Plains ecoregion, including Central Indiana, were rated "Fair" or "Poor" due to excess streambed fine sediments; one-third of the surveyed streams were "Poor" (U.S. EPA, 2020). Surface soil erosion contributes approximately one-third of fine sediments in Midwestern streams, while streambanks and channels comprise two-thirds (Gellis et al., 2017). Fine sediment can smother stream habitat and breeding areas, inhibit aquatic plant growth, and transport sorbed contaminants into water bodies (Meador and Frey, 2018; Munn et al., 2018; USGS, 2018). The Ohio River drainage basin, which covers most of Indiana, is estimated to yield 44.9 tons of fine sediments from agriculture each year (Robertson and Saad, 2019).

Another large source of agricultural pollution is nutrients associated with fertilizers. Each year in the United States, >10 million tons of nitrogen (N) and almost 2 million tons of phosphorus (P) fertilizers are used in agricultural operations (Munn et al., 2018). However, nutrient application is not always efficient, resulting in lost nutrients that do not contribute to agricultural production. For example, Potter et al. (2006) estimated that 28% of applied N and 16% of applied P are potentially lost from cropland annually. These lost nutrients often enter surface waters, causing eutrophication and increasing N and P availability. Elevated levels of these nutrients allow algae to bloom excessively, which can decrease dissolved oxygen concentrations and even create hypoxic (i.e., very low or no dissolved oxygen) conditions. Hypoxia is problematic because it often causes organism die-offs. The Ohio River drainage basin is estimated to yield 236 kg of N and 36.8 kg of P per square kilometer per year from non-manure fertilizer application (Robertson and Saad, 2019). The NRSA of 2013-2014 evaluated over half of the sampled rivers and streams in the Temperate Plains ecoregion as having very high ("Poor") levels of Total Phosphorus and Total Nitrogen when compared to reference sites (U.S. EPA, 2020). The USGS conducted a National Water Quality Assessment (NAWQA) of selected areas in the United States and found that the sampled areas in Central Indiana had elevated nutrient levels (Munn et al., 2018). Nitrate and ammonia, in particular, accumulate predominantly in water, including subsurface drains and open streams. Indiana is one of three Mississippi River drainage basin states that provide 40% of the nitrogen entering the Gulf of Mexico (Capel et al., 2018).

Like fertilizers, pesticides and their byproducts are present throughout field-stream systems. Dozens of pesticide compounds may exist at a single stream site (Meador and Frey, 2018; Nowell et al., 2018). Pesticides are not always directly toxic to fish. For example, insecticides can indirectly affect fishes through chronic toxicity to aquatic macroinvertebrates that serve as fish prey (Nowell et al., 2018). A model simulation also indicated that north-central Indiana streams had a 5-50% probability that the annual mean concentration of atrazine, an herbicide, would exceed the benchmark for drinking water (Capel et al., 2018). Agriculture has numerous and widespread effects on Midwestern streams and rivers.

The goals of this paper are twofold. My first goal is to evaluate the United States Department of Agriculture National Agricultural Statistics Service Cropland Data Layer as a means to compare differences in agricultural and forested land cover over time (Chapter 3). My second goal is to investigate the relationship between surrounding land cover (i.e., row crop versus forest) and the biological integrity of streams in the Temperate Plains ecoregion of Indiana (Chapter 4).

To achieve my first goal, I extracted land cover information for northern Indiana from the USDA NASS Cropland Data Layer and classified it using relevant categories. I then used the land cover change raster to select sites for ground truthing the projected CDL land cover type. To accomplish my second goal, I sampled fishes at stream sites in north-central Indiana to use them to calculate an Index of Biotic Integrity (IBI; Karr, 1981; Karr et al., 1986; Ohio EPA, 2006) for each site. I also assessed local stream habitats using the Qualitative Habitat Evaluation Index (QHEI; Rankin, 1989), as well as CDL pixel counts and riparian measures to categorize sites as agricultural or forested. Finally, I used models to determine any significant correlations between surrounding land use and IBI scores.

CHAPTER 1. LITERATURE REVIEW

1.1 Role of Forests

The Temperate Plains ecoregion in central Indiana includes the Illinois/Indiana Prairies, (EPA Ecoregion 54a, hereafter "Indiana Prairies") and the Loamy, High Lime Till Plains, (EPA Ecoregion 55b, hereafter "Till Plains") (Woods, 1998). The Indiana Prairies and Till Plains were flattened from Ice Age glaciation, later resulting in areas of wetlands, forests, and tallgrass prairie. In particular, the Indiana Prairies were predominantly prairie with oak-hickory forest, and the Till Plains were predominantly forest. Beginning in the 19th century the land was gradually converted to agriculture, degrading streams with chemical and fine sediment pollution, bank erosion, and higher water temperatures (Woods, 1998). In addition, widespread and intensifying agriculture within a flat landscape with poor drainage eventually required extensive ditching and subsurface drain installation (Capel et al., 2018). The Midwest accounts for approximately 30% of the 1 million square kilometers of wetlands in the contiguous United States that have been converted to agriculture through drainage. As a result, extensive land clearing and wetland draining confined most of the remaining forests in the Indiana Temperate Plains ecoregion to isolated woodlots and riparian zones along streams and rivers (Woods, 1998).

At the local level, stream health is controlled by the presence and quality of a riparian forest buffer. A riparian forest buffer is an area of perennial plants bordering a water body that is managed for conservation purposes (MacFarland et al., 2017). Riparian forests benefit adjacent- and downstream water bodies and associated stakeholders. Among these benefits are nutrient uptake, filtration of fine sediments in overland flow, pesticide sequestration, streambank stabilization, shade, instream cover, and food sources (e.g., leaf fall) for aquatic biota (e.g., MacFarland et al., 2017).

Riparian vegetation has a "baffling effect" on overland flow, weakening stormwater surface flow and facilitating fine sediment deposition in the riparian zone. Fine sediment deposition is also coupled with nutrification because phosphorus and certain forms of nitrogen are sorbed by soil particles that are mobilized by surface flow. Dosskey et al. (2010) found that root networks remaining after aboveground vegetation removal could continue to protect stream health for decades after removal. The ability of riparian zones to control and sequester N, P, and other nutrients also depends on many factors, including soil type, nutrient load, vegetation composition and nutrient uptake, and stream geomorphology. However, riparian zones generally control nutrient and sediment deposition in streams (Naiman et al., 2005). Indeed, because two-thirds of sedimentation is from bank and channel erosion and sedimentation can degrade algal, invertebrate, and fish communities, maintaining a healthy root system in the riparian zone is key (USGS, 2018).

The River Continuum Concept (RCC) posits that headwater streams are typically forested, and therefore have photosynthesis-limiting shade that creates dependence on allochthonous (i.e., imported organic) material by resident aquatic biota (Vannote et al., 1980). Woody vegetation also filters solar radiation, regulating stream temperature and summarily increasing a stream's O_2 capacity, which decreases respiratory stress in aquatic organisms (NRCS & Wildlife Habitat Council, 2007). While woody vegetation provides the greatest shade for streams, fully-grown grasses and forbs can sufficiently shade streams narrower than 2.5 m (Blann et al., 2002).

1.2 Stream Health Metrics

Headwater streams represent at least 50% of the overall length of all stream systems on earth (Richardson, 2020). Also known as upper reaches, they are the source of water for larger streams and rivers. Headwater streams are defined as 1st, 2nd, and 3rd order streams in the River Continuum Concept, which is based on Strahler stream order, a system for classifying streams by their number of tributaries (Strahler, 1952, 1957; Vannote et al., 1980). These headwater streams average 0.8 to 3.7 m wide, with size increasing exponentially with Strahler stream order (Downing, 2012). Headwater streams are ranked as critically important for climate regulation, food web dynamics, nutrient cycles, and recreation, while the larger rivers they influence are critically important for water consumption, food sources, and flood control (Yeakley et al., 2016). They influence connected higher-order streams by transporting nutrients, pollutants, organic matter, and sediment downstream. However, as small and often unprotected water bodies, they are highly vulnerable to changes in the environment (Richardson, 2019).

Several indicators are used to determine stream health, including benthic macroinvertebrates, fish assemblages, and physical habitat metrics. Benthic macroinvertebrates are macroscopic invertebrate animals that live in (i.e., infauna) or on the surface of (i.e., epibenthic) stream substrates. They are used as bioindicators because they vary in pollution tolerance by species and cannot migrate away from polluted conditions, meaning that only a few tolerant species will be found in streams with suboptimal water quality. For example, agricultural insecticides such as permethrin impact insect reproduction and development (Capel et al., 2018). Macroinvertebrates are also more likely to be present than fish in very small, low-order streams, and are easily captured and identified (U.S. EPA, 2013). There are several indices used in benthic macroinvertebrate sampling, including the Ephemeroptera Plecoptera Trichoptera (EPT) Index, which provides a rapid assessment of water quality based on the presence of the generally pollution-intolerant insect orders Ephemeroptera, Plecoptera, and Trichoptera (Barbour et al., 1999).

Fish assemblages reflect the diversity and numbers of fish species present in a water body. They are used as bioindicators because a diverse assemblage requires a variety of food sources, complex habitat for shelter, and specific spawning conditions (Barbour et al., 1999). Because fish can move away from degraded areas, the composition of a fish assemblage is an indicator of water quality and stream health. For example, fertilizers such as ammonia (NH_3) impact the reproductive, respiratory, and nervous systems of fishes (Capel et al., 2018). Like benthic macroinvertebrates, there are many ways to use fishes as indicators, which can be chosen based on region, stream size, and existing pollution level. Among the indices appropriate for small Midwestern streams are the number and identity of darter species, headwater species, sucker species, and intolerant species; % omnivores, insectivores, and carnivores; and total number of individuals (Karr, 1981; Karr et al., 1986).

Stream physical habitat is an important component in determining the effects of anthropogenic land uses on streams (Frissell et al., 1986). This is typically done by evaluating physical habitats within, adjacent to, and along streams. This can be done to quickly estimate stream health by using a metric-driven assessment, the Qualitative Habitat Evaluation Index (QHEI; Rankin, 1989). Evidence of human disturbance (e.g., developed surrounding land, dams), habitat complexity and cover, and sediment type are all components of physical habitat quality and are easy to evaluate. In particular, riparian vegetation composition is a useful indicator because it reduces sediment and overland nutrient runoff entering the stream, stabilizes riparian sediment, and provides organic material for instream cover and food (MacFarland et al., 2017; Rankin, 1989).

The NRSA of 2013-2014 found that Indiana stream health indicators rated poorly (U.S. EPA, 2020). Macroinvertebrate composition in Temperate Plains streams was rated as 46% Poor and 30% Fair, and fish assemblages were rated as 34% Poor and 31% Fair. Notably, for all streams, a Poor sedimentation rating made a Poor benthic macroinvertebrate condition twice as likely. The USGS NAWQA supported this with the conclusion that macroinvertebrate composition in the sampled region of Indiana was influenced by physical habitat, not nutrient levels. That is, the nutrients were sufficiently abundant not to limit the growth of biota (Munn et al., 2018).

1.3 Agriculture and Stream Integrity

Alexander et al. (2008) created an improved water-quality model to evaluate and predict how total N and total P move from sources in the Mississippi and Atchafalaya River Basins into the Gulf of Mexico. They found that ~50% of the N and ~25% of the P entering the Gulf of Mexico came from corn and soybean row crop agriculture. Indiana delivered the greatest N yield (1806.6 kg/km/yr) and had the third largest share of the total flux (10.1%). Illinois, Iowa, and Indiana together accounted for 38.2% of all N flux entering the Gulf of Mexico.

A study by Kladivko et al. (2004) assessed the nitrate N concentrations from subsurface drains in southeastern Indiana over 15 years of row crop agriculture. The area studied was the Southeast Purdue Agricultural Center, within the Eastern Corn Belt Plains. The tile drains were spaced variously at 5, 10, and 20 m apart. The authors determined that nitrate concentrations were unrelated to drain spacing, but instead decreased by more than half as a result of lower fertilizer application, the use of winter cover crops, and rotations of no-till corn and soybeans.

Stream sedimentation is problematic not only because of the accelerated changes to stream morphology, but also because sediment is able to retain sorbed pollutants from agriculture. Wolf et al. (2020) studied benthic macroinvertebrate responses to agricultural stream sediments and found that intolerant species experienced mortality where highly-tolerant species did not. However, the mortality levels were not proportional to the levels of agricultural development between sites. This may have been due to a large decrease in biotic integrity with a small increase in development, followed by a “ceiling” where increased development yielded almost no response.

Wang et al. (2007) sampled Wisconsin streams and found that those in high-agriculture, low-forest regions had the greatest median N and P concentrations, fewer EPT taxa, lower counts of EPT macroinvertebrates, and fewer overall taxa present than less developed forested regions. However, their measurements also showed highly variable biological integrity even at low nutrient levels, implying that non-nutrient factors impact assemblages without N and P deposition. The authors concluded that indirect non-nutrient factors, nutrient effects, and interactions explained much of the variation in fish and macroinvertebrate assemblages in the selected streams.

To develop nutrient criteria specific to nutrient-heavy Indiana streams, Caskey et al. (2010) measured the relationship between nutrient-based stressors and biotic communities. The study streams' basins were overwhelmingly agricultural (77%), with forested basins only comprising 16% of the study area. The two most abundant aquatic insect families were Chironomidae (41.7%) and Hydropsychidae (17.3%), which are considered tolerant of poor stream conditions. Tolerant species such as central stoneroller *Campostoma anomalum*, creek chub *Semotilus atromaculatus*, and bluntnose minnow *Pimephales notatus* were abundant in the sampled fish communities (13.3%, 9.9%, and 9.3% total relative abundance, respectively).

Caskey and Frey (2009) assessed agricultural stream fish community composition in the Indiana-Ohio Eastern Corn Belt Plains ecoregion. The two most abundant fish species were central stoneroller and bluntnose minnow (25.7% and 11.1%, respectively), accounting for 36.8% of the fishes captured. Differences between similar biological communities were attributed to environmental factors other than nutrient levels; a canonical correspondence analysis suggested that an increase in mean bankfull depth (i.e., maximum possible depth of the stream channel before stream overflow) increased the number of fish taxa present.

One of the potential sources of N and P that can affect fish and macroinvertebrate assemblages in streams is eutrophication. Much attention has been given to the effects and mechanisms of eutrophication in lakes. However, there is evidence that streams also exhibit eutrophication that can impact biological integrity. In a literature review, Dodds and Smith (2016) found that P and N control phytoplankton biomass. High levels of either nutrient can increase benthic algal biomass, contradicting the widely-held assumption that P is the limiting nutrient for freshwater ecosystems. They also cite several papers finding that high N and P levels jointly decreased fish and aquatic macroinvertebrate abundances. Hypoxia is the main concern with lacustrine eutrophication, but N and P can also cause changes in herbivory and detritus consumption by influencing primary productivity and primary consumption.

Another way N and P pollution influence stream health is through nutrient-limited growth and trophic changes. Evans-White et al. (2009) suggested that slight N and P increases can cause a change in food quality, shifting trophic structure and therefore negatively impacting assemblage diversity. Shredders are a trophic guild of macroinvertebrate species that consume plant-based detritus by chewing or boring. Collector-gatherers are a trophic guild of macroinvertebrate species that consume detritus. The authors found that fast-growing shredder and collector-gatherer taxa utilized high-P food sources and could out-compete other invertebrate taxa under eutrophic conditions. Predators, which have a better balance between body C:N and C:P than their prey, were less affected. There was a negative relationship between macroinvertebrate taxa diversity and nutrient levels. Shredder taxa also decreased in diversity with increasing P, and chironomid species dominated the community. Broadly, macroinvertebrate species with relatively high N or P requirements are limited by those nutrients, and could grow without constraint and out-compete other species in enriched conditions. This change in macroinvertebrate assemblage composition can then affect other parts of the trophic system. Their results found variability in indicator taxa even under low-nutrient conditions, which corroborated the findings of Wang et al. (2007).

Camargo and Alonso (2006) conducted a global literature review of aquatic nitrogen pollution and identified a rise in inorganic N concentrations in water bodies everywhere. They also listed the drastic effects of anthropogenic N eutrophication, such as hypoxia, lowered light penetration, and trophic shifts in macroinvertebrates and fish, including the decline of salmonids and EPT species. N pollution is therefore of critical importance for aquatic ecosystem health worldwide. Meador and Frey (2018) studied streams across the Midwest for predictors of fish community composition, measuring N, P, streambed sedimentation, dissolved oxygen (DO), riparian vegetative cover, riparian disturbance, bed sediment contaminants, streamflow variability, pesticides, and instream habitat cover. Total N was of the highest importance; however, the factors controlling N pollution are complex and involve precipitation, tile drainage, and tillage practices. Sedimentation and P were also important influences for the Temperate Plains ecoregion, followed by DO, riparian vegetative cover, and riparian disturbance.

The USGS Midwest Stream Quality Assessment (MSQA), conducted in 2013, presented a cross-section of biological impacts of various anthropogenic stressors (USGS, 2018). It determined that sensitive forms of algae require hard substrates, cool water, and low levels of herbicides, particularly triazines, such as atrazine. Similarly, aquatic invertebrate diversity decreases with the presence of excessive soft substrates, such as silt and muck. Notably, mean nitrate concentration increased exponentially with the percentage of a watershed planted in corn. Ammonia and pesticides also degraded aquatic macroinvertebrate diversity. Fish diversity decreased with total nitrogen and soft substrates, which cover breeding habitat and prevent reproduction. Pollutants like pesticides were found at higher-than-tolerable levels for invertebrates and algae. Indeed, because riparian zones inhibit runoff, high-quality forested riparia were associated with robust macroinvertebrate communities even within heavily agricultural watersheds.

1.4 Agricultural Drainage and the Role of Riparian Buffers

Riis et. al (2020) conducted a meta-analysis that ranked ecosystem services provided by different riparian vegetation types. They determined that forested riparian zones were of medium- or high importance for N and P removal, sediment removal, pesticide removal, erosion control, flow regulation, habitat provision, regulation of microclimate, pollination, and standing woody biomass. Other types of riparia (herbs/grass, wet forest, and wetlands) were variously ranked as medium- or high importance for some, but not all, of the same services. For example, wetlands were of 'medium' importance for standing crop of non-woody biomass, while all other types of riparia were 'low' importance. Overall, dry forest riparian zones were of the highest importance for the greatest portion of ecosystem services.

Simon and Collison (2002) observed the root numbers and strengths of young trees (<10 years on average) as well as herbaceous vegetation. They determined that riparian vegetation influences bank stability through the number of roots per unit area and the strength of those roots. They found that larger (>5 mm) roots typical of woody vegetation are stronger and better at bank reinforcement than a large number of small roots per unit area, typical of herbaceous species. However, this was based on observations during an unusually dry period. In contrast, riparian trees came in second to switch grass *Panicum virgatum* in terms of bank reinforcement during a high rainfall period. The authors recommended that riparian management should also consider other factors, including mechanical and hydrologic properties, tree canopy cover during dormancy, rooting depth, and transpiration, and that a mixture of woody and herbaceous species would be most beneficial.

Pollen-Bankhead and Simon (2010) corroborated the findings of Simon and Collison (2002. They accounted for seasons and included more riparian tree species, finding that soil cohesion varied with season. Root volume was the critical factor in streambank stability in winter and spring, rather than the suction force resulting from evapotranspiration. When evapotranspiration was the most critical factor in soil cohesion, longleaf pine *Pinus palustris* and river birch *Betula nigra* provided similar suction to switch grass. It was suggested that a combination of trees and switch grass is ideal for bank stability. Polvi et al. (2014) quantified riparian species' root tensile strengths and found that tree roots were stronger than any other taxonomic group. While they did not account for the effects of different sediment textures on roots, they nonetheless advised that woody and non-woody species should be used together.

It has been long accepted that riparian vegetation reduces agricultural sediment runoff into streams (Naiman et al., 2005). The primary vehicle for sediment trapping is the physical resistance of above-ground vegetation and its ability to slow down overland flow. Notably, it has been shown that vegetated buffer zones can reduce sediment runoff into streams from 60% to 90% (Daniels and Gilliam, 1996; Cooper et al., 1987). This research necessarily preceded studies of nutrient transport, because N and P sorb to soil particles; therefore, preventing or reducing sedimentation also prevents nutrient runoff. On the other hand, there is much debate on the effectiveness of vegetated buffers for nutrient removal. One confounding factor is the seasonality of vegetation versus precipitation. Liu et al. (2014) argued that plants are dormant for most of the Midwestern winter, spring, and early summer, and the water table is too far below the root zone during the dry season for plants to uptake N. However, subsurface flow is irregular and often heaviest during spring when plants are exiting dormancy. A study by Stauffer et al. (2000) also indicated that wooded riparian zones effectively protect fish and macroinvertebrate communities from intensive agriculture in the Midwest. Indeed, their conclusion was that the quality of the riparian zone had a greater impact on fish assemblages than runoff potential.

Osborne and Kovacic (1993) conducted a literature review to assess the potential effectiveness of vegetated buffers. They found that forested buffer strips were generally effective for N reduction (40-100%). These results were applicable to forested buffer strips as narrow as 16 m. Grass buffer strips were also effective to a smaller degree, reducing 10-60% of N from tile drainage and 54-84% from surface drainage. They also studied the nutrient-reducing effects of a vegetated buffer strip in a central Illinois field with tile drainage. They determined that forested buffers removed nitrate from shallow tile drains better than grass buffers. However, they were more efficient at nutrient removal from surface flow than subsurface drains.

Structural best management practices are also subject to degradation and become less effective over time. Bracmort et al (2006) assessed grassed waterways, grade stabilization structures, field borders, and parallel terraces that had been implemented in two highly agricultural drainage areas of Black Creek watershed in Allen County, Indiana. Using the Soil and Water

Assessment Tool (SWAT) the authors estimated that structure installation decreased sediment deposition by 16-32% and decreased P deposition by 10-24% when implemented in the 1970's and 1980's. After 20 or more years post-installation, the structures reduced sediment yield by only 7-10% and P yield by 7-17%.

While there is consensus that riparian vegetation reduces sedimentation in streams, studies disagree on whether including trees is less effective, or as equally effective as grass-only buffers (Yuan et al., 2009; Hughes and Quinn, 2014). Regardless, bank stability is improved by the presence of trees in a riparian zone (Simon and Collison, 2002; Pollen-Bankhead and Simon, 2010; Polvi et al., 2014), and bank erosion is the source of two-thirds of stream sedimentation (USGS, 2018).

There is similar disagreement about N and P runoff potential related to riparian vegetation type (Liu et al., 2014; Stauffer et al., 2000). A model simulation indicated that buffer zones did not protect invertebrate assemblages in regions where crop fields were tile-drained, including Central Indiana (Munn et al., 2018). However, because of nutrient sorption to soil, riparian vegetation may prevent a considerable amount of runoff to surface waters in the absence of subsurface drainage (Carpenter et al., 1998). A literature review by Feld et al. (2018) concluded that wooded riparian buffers of sufficient width and length could retain up to 100% of nutrients and sediment from runoff.

Baker et al. (2006) conducted a study within the Temperate Plains region of Indiana to investigate agricultural chemical transport routes. The study site was highly agricultural (87% of land use) and tile drained. The two largest sources of water and chemical pollutants to Leary Weber Ditch were overland flow and tile drains. Pesticide parent compounds were more prevalent than degradates in overland flow than in tile drain water. Overland flow contributed less water than tile drain flow when rainfall was <0.5 in/hr (10% and 90%, respectively) and when rainfall was >0.75 in/hr (40% and 60%, respectively). Overland flow became a significant vehicle for chemical pollutants during extreme precipitation events, but tile drains contributed the majority of pollution to the Leary Weber Ditch under all precipitation levels.

Stone and Wilson (2006) also studied the Leary Weber Ditch to compare pollutant flow paths during storms. Agricultural chemicals can appear more quickly in water from tile drains than from typical soil matrix flow due to the presence of preferential flow pathways (e.g., tunnels from invertebrates). This faster transport prevents pollutants from adsorbing to soil particles. This study used chloride concentrations to test the flow contributions of both flow pathways. During two storm events, preferential flow initially contributed little to overall flow (11% of total flow, 40% of peak flow), although its contribution increased substantially following the initial period (51% of total flow, 81% of peak flow). Therefore, there is potential for agricultural pollutants to bypass adsorption by the soil matrix and be transported rapidly into tile drain effluent during heavy rainfall events.

One of the sources of disagreement in estimates of agricultural pollutants entering streams may be due to low resolution of stream maps used in analyses. Baker et al. (2007) evaluated the impacts of stream map resolution on estimates and predictions of riparian buffer effectiveness. Buffers were defined by contiguous and stream-adjacent forest- and wetland pixels from the National Land Cover Dataset (NLCD). Buffer width was defined as the span of these forest-wetland pixels between cropland pixels and the stream. Nutrient loading was simulated from cropland pixels using different levels of buffer retentiveness. The authors determined that improved estimates of buffer width, gaps, and variability from increased stream map resolution led to generally decreased estimates of retentiveness in riparian buffers. The improved stream map resolution improved estimates of agricultural pollutants entering streams.

Effert-Fanta et al. (2019) studied row crop agriculture in Eastern Illinois and found increased abundances of herbivorous- and omnivorous fish in streams with intensive agriculture and poor buffering zones. Areas with high agricultural development and little buffering also had the highest fish and macroinvertebrate abundances, due to increased nutrients and light within tolerable levels. The abundance of pollution-tolerant macroinvertebrates was similar across all sites regardless of buffer quality, possibly due to the homogeneous stream substrate across sites. However, forested buffers provided woody debris and created deeper pools in streams, creating ideal habitat for carnivorous fish and a more balanced trophic structure. Intolerant and endangered fish species also occupied the well-buffered, low-agriculture streams. Overall, high buffering led to similarity between streams regardless of agricultural intensity, indicating that forested buffer zones could control for the negative effects of agriculture. A literature review by the EPA indicated that vegetated riparian zones >25 m wide effectively removed N by 75% or more, and that forested buffers were more effective than grass alone (Mayer et al., 2005).

Richardson and Beraud (2014) conducted a meta-analysis of riparian forest harvesting. Like many others, they found increased N in harvested riparian streams, but a low P response.

They theorized this could be because of efficient uptake by P-limited primary producers. The effects of riparian harvest overlapped the zero-effect line in a regression analysis, which the authors suggested resulted from the differences in environments across studies. The authors also found that aquatic macroinvertebrate density generally increased after riparian forest harvest. They used potential evapotranspiration (PET) as a proxy for stream temperature. A higher PET (a warmer stream) induced a greater negative response from shredders and EPT, suggesting their sensitivity to logging under high temperatures.

The presence of riparian canopy influences the stream microclimate by decreasing sunlight penetration, particularly in small streams with dense canopies. Canopy density is, therefore, a key determinant of stream temperature. Streams without sufficient riparian canopies can exhibit elevated water temperatures that create stressful conditions for fish (Evans-White, 2009). According to Gregory et al. (1991), downed woody debris from well-developed riparian canopies provides habitat for invertebrates and influences stream morphology, leading to greater habitat diversity, such as deep pools, backwaters, and side channels, all of which also create cover for fish.

Riparian zone vegetation types can drastically differ in total biomass. While material from trees results mainly from senescence (i.e., seasonal leaf fall), most material from herbaceous plants enters a stream via flooding. Plant matter is a primary food source for aquatic invertebrates, and while trees provide much higher biomass than herbaceous plants, this material is typically nutrient-poor and time-consuming to break down. As a result, streams with dense canopies do not typically support high numbers of herbivorous invertebrates. Shredders are a dominant group of invertebrates in shaded streams that rely on autochthonous material provided from terrestrial sources (e.g., tree leaves). Otherwise, unshaded streams have a greater abundance of invertebrates, which creates higher food availability for fish (Gregory et al., 1991).

CHAPTER 2. LAND COVER CHANGE IN THE CROPLAND DATA LAYER

2.1 Data

The USDA National Agricultural Statistics Service (NASS) Cropland Data Layer (CDL) is one of the most extensive tools for visualizing and measuring land cover in the United States. The CDL is in the U.S. Public Domain, can be easily viewed and downloaded online, has a 30m resolution, and has annual data for the contiguous U.S. every year since 2008 (USDA NASS, n.d.-b). The current recommended online interface for visualizing, selecting, and downloading CDL data is CroplandCROS (USDA National Agricultural Statistics Service Cropland Data Layer, 2024). During this study, only its predecessor, "CropScape," was available via the George Mason University Center for Spatial Information Science and Systems (GMU CSISS, n.d.; USDA National Agricultural Statistics Service Cropland Data Layer, 2020). CropScape is a strong candidate for use in land cover analysis because it is easy to access and offers a comprehensive range of land cover types, particularly row crops such as corn and soybeans. This study aimed to determine the CDL's applicability for measuring differences in landcover between years. I used CDL data from the years 2010 and 2020, as 2010 was the first year with consistent 30m nationwide coverage, and 2020 was the most recent year available at the time of analysis. Thirty-five counties in the northern portion of Indiana were selected and analyzed in change visualization due to the presence of row crop agriculture and forest.

2.2 Methods

2.2.1 *Area of Interest Shapefile Export & Import*

Nationwide data can be downloaded by year from the Cropland National CDL's webpage (USDA NASS, n.d.-a). For more narrow surveys, the data requires a defined area of interest (AOI). This can be done two ways, either by creating a zipped ESRI shapefile or GML file, then importing it into the CropScape program, or by directly defining an AOI using the CropScape interactive data layer. CropScape's built-in AOI drawing tool only allows for drawing rectangles, circles, and polygons, or selecting AOIs based on region, state, county, or Agricultural Statistics District

(ASD). Therefore, defining a more complex or narrow AOI must be done externally and then imported into CropScape as either a GML file or a zipped ESRI shapefile including .shp, .shx, .dbf, and .prj files. The “tigris” package in R allows AOI definition using state and county names or FIPS codes. The “sf” package enables shapefile writing; shp, .shx, .dbf, and .prj files must be zipped for import into the CropScape program. Alternatively, a shapefile can be created and exported using queries in a GIS application.

To download AOI data from CropScape, the AOI cannot be more than 4,000,000 square kilometers. Data can be downloaded as a raster, PDF, or CSV file. For manipulation in RStudio, a raster or CSV is required. Depending on the AOI size, it may be necessary to manipulate the data on a high-performance computing system. For this study, an AOI was created as a shapefile in R including 35 northern Indiana counties (Fig. 2.1), and then imported to CropScape. CSVs of the data were then downloaded in the default Albers projection for the years of interest. R code used for this study, written in R version 4.1.2 (“Bird Hippie”), is included in Appendix A. Raster manipulation in Rstudio was done using the Purdue University Scholar Cluster, a high-performance computer cluster.

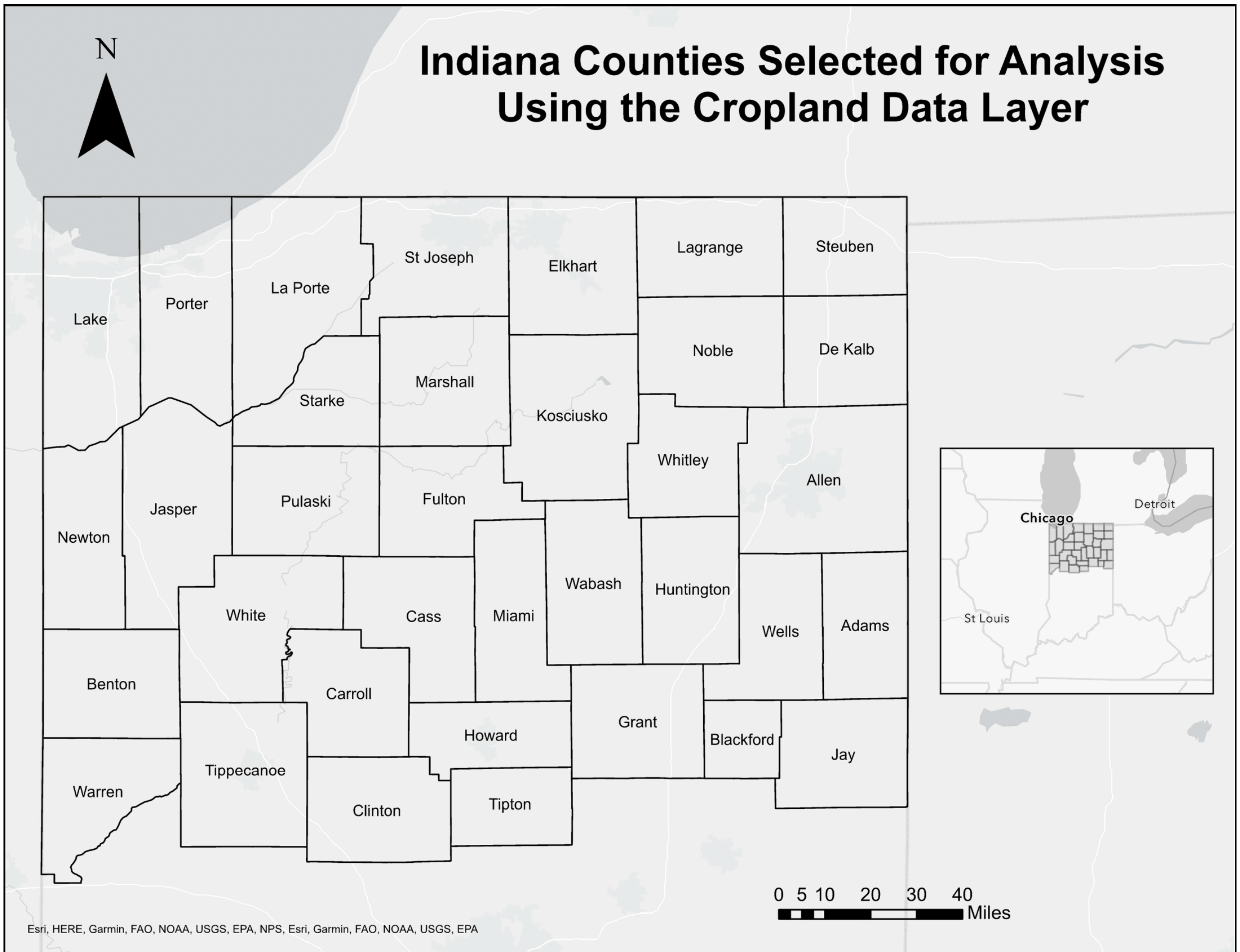

Figure 2.1. Selected Indiana counties in the study AOI. The extent of the created shapefile includes the above labeled 35 counties in northern Indiana.

2.2.2 *Land Cover Category Definition & Change Matrix*

It is necessary to confirm which land cover categories are present in the AOI; this can be done on a state- and year-by-year basis by downloading AOI-specific data from the visual CropScape interface (GMU CSISS, n.d.). For AOIs spanning all 48 states, the pixel counts and acreages by category are available on the Cropland FAQs page (USDA NASS, n.d.-b). Each category also has a unique numeric value, which can be used to collapse or redefine land cover categories. The code used in this study aggregated values from similar categories into larger categories. It should be noted that the CDL is based on satellite imagery and is therefore meant to be used as land cover data, not land-use. Lark et al. (2017) recommend combining CDL classes into larger categories to reduce the error from satellite imagery; in this study, land cover change types were collapsed into the following categories:

Afforestation: any non-forest land use category to "forest" between 2010 and 2020

Deforestation: "forest" to any non-forest land use category between 2010 and 2020

Undisturbed Forest: "forest" to "forest" between 2010 and 2020 (no change)

Agriculture: "cropland" to "cropland" between 2010 and 2020 (no change)

Other: all other categories and changes (e.g., "barren" to "developed", "open water" to "open water", etc.)

For the purposes of this visualization, the category "Agriculture" also includes fallow or idle cropland (i.e., land with nothing growing at time of measurement). Tree crops were also classified as "Agriculture", under the assumption that farming operations would create an environment more similar to row crops than to natural forests. The study category "Forest" includes evergreen, deciduous, and mixed forests; the typical forest type in the study region is mixed hardwoods dominated by oak, hickory, and maple species.

2.2.3 *Visualizing Land Cover Change & Exporting a Raster*

To effectively visualize land cover change, a function must be created to reclassify each pixel value into an aggregated category value. Two new raster images can then be drawn using this function and the TIFs of the chosen AOIs. These can be used to create a single raster with collapsed change categories (e.g., "A" to "B", "B" to "A"), which is then saved as a TIF. Importing the TIF into ArcMap or ArcGIS enables further analysis and manipulation.

A new raster was created from the existing CDL raster images of land use in 2010 and 2020. The raster was projected onto a satellite imagery basemap in ArcMap and ArcGIS and visually evaluated for projection accuracy. Because of the high resolution of the data, the agricultural or forested areas had to be approximately 20 pixels or larger in size and reasonably contiguous to be considered of significance.

2.2.4 *Technical Validation and Ground Truthing*

Technical validation was performed first by visually comparing contemporary satellite imagery to the change raster. I gathered parcel numbers from a GIS layer of property boundaries (Indiana Geographic Information Office, 2019). I then retrieved landowner contacts from publicly available assessed value information from the Indiana Department of Local Government Finance (DLGF, 2021). Landowners were contacted via physical or electronic mail for permission to visit sites. I received permission to visit sixteen sites: three afforested, three deforested, five agricultural and five undisturbed forest. All deforested sites were located in La Porte County, while the agricultural, afforested, and undisturbed forest sites were located in Benton, Clinton, Howard, Tippecanoe, Warren, and White counties (Fig. 2.2). Ground truthing was conducted between June and August of 2022, in which I and at least one field crew member visually surveyed the site and, if the CDL-projected change could not be verified visually, communicated with the landowners about the site history.

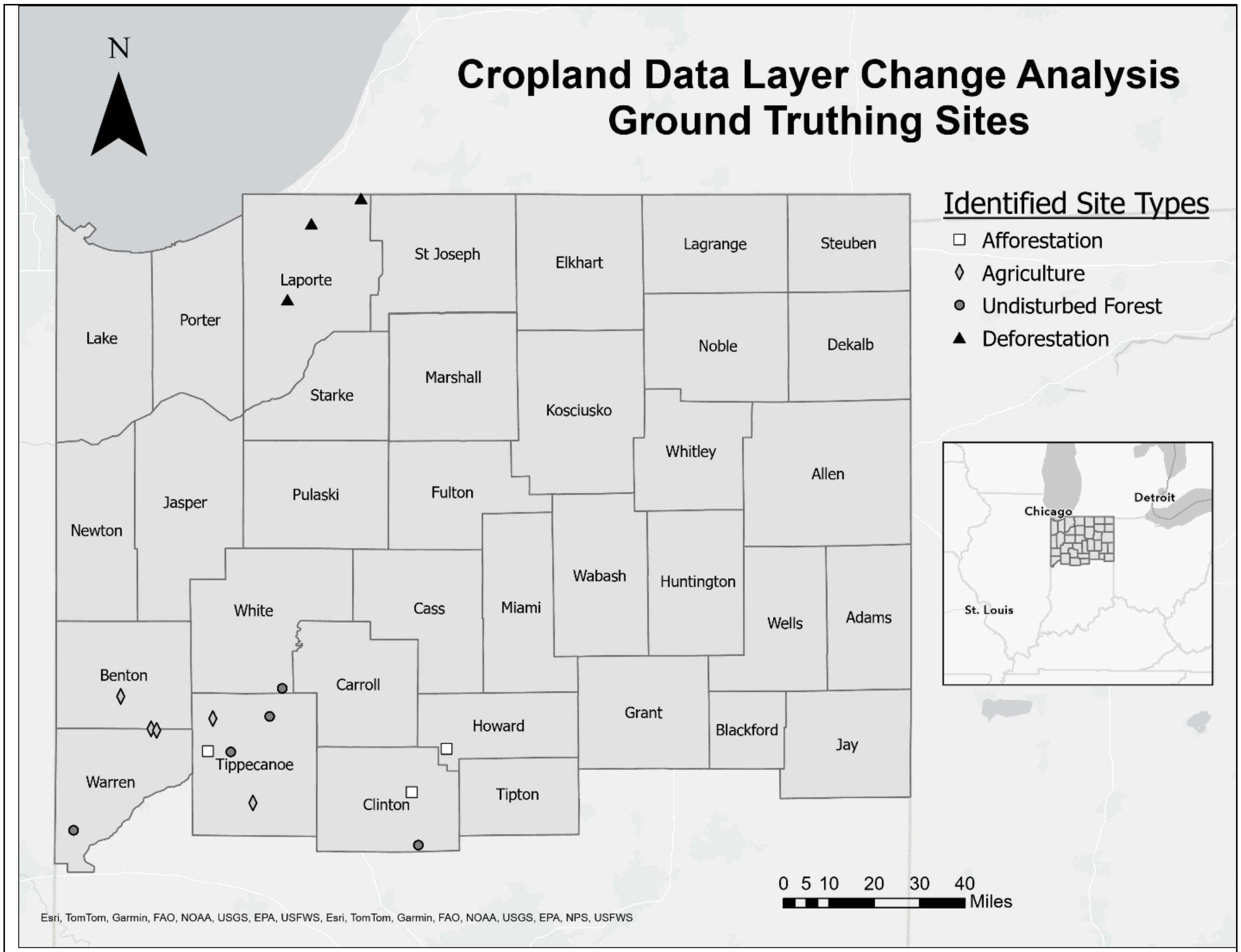

Figure 2.2. Map of sixteen Indiana land cover sites visited for ground truthing between June and August 2022. Site types identified were afforested (n = 3), agriculture (n = 5), deforested (n = 3), and undisturbed forest (n = 5).

2.3 Results

In terms of resolution, the total combined area of the 35 counties is 14,661.75 square miles, or 9,383,520 acres (US Census Bureau, 2011). The same AOI in the USDA Cropland Data Layer was 9,233,581 acres. At a resolution of 30 m, there is a loss of 149,939 acres or 1.6% of the recorded area, considered negligible.

The land use change raster displayed several patterns (Fig. 2.3). First, that the northwest quarter of the study area, including the Kankakee River and Calumet Region, seemingly had the highest proportion of deforestation as identified by the CDL. Second, the northeastern quarter of the study area seemingly had the highest proportion of afforestation. Third, that the greatest concentration of undisturbed forest was along large rivers and major tributaries. Lastly, the overwhelming majority of the total landcover was agriculture, distantly followed by the “Other” category.

Of the six “change” sites visited, none were correctly identified by the CDL. One “afforested” area was a planted forest older than the study timeframe of 10 years; one was a small patch of natural forest adjacent to alternating grassy and agricultural areas; and the last was a misidentified grassy area. The three “deforested” areas in the northwestern quadrant of the region were determined through site reconnaissance to be long-standing bogs and wetlands, with only one site confirmed to have experienced selective logging upstream. The sites containing agricultural and undisturbed forest land cover (n = 10) were confirmed to have been accurately identified by the CDL, both through ground truthing and existing literature describing land cover type and distribution in Indiana (Carman, 2013; Munn et al., 2018; Woods, 1998).

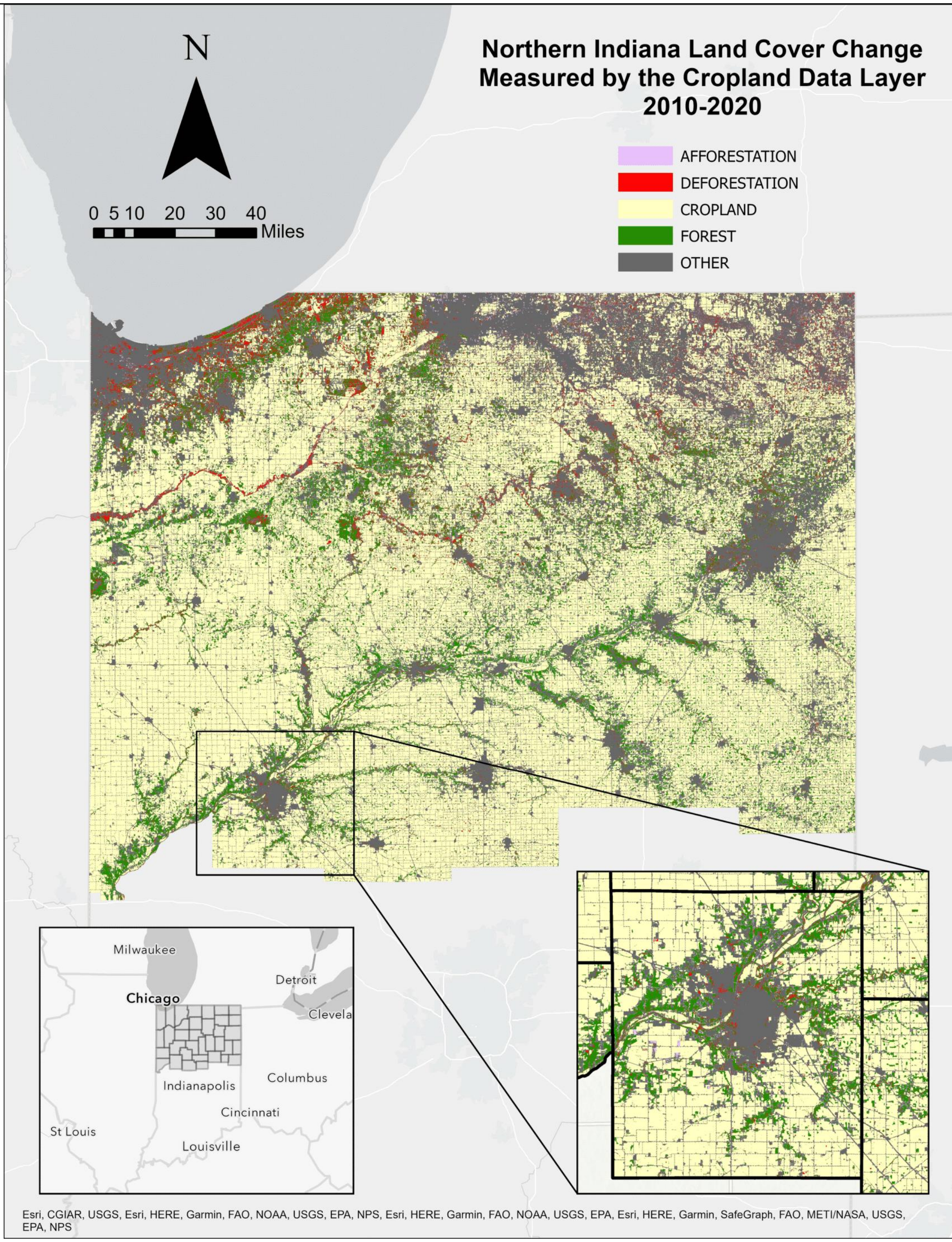

Figure 2.3. Raster image of the Cropland Data Layer's measured landcover changes in Northern Indiana between 2010 and 2020. Lower right inset shows detail of Tippecanoe County, Indiana.

2.4 Discussion

I determined that the CDL was accurate for cropland and undisturbed forest in the AOI. However, it was much less accurate for areas of afforestation and deforestation. Specifically, the sites identified as deforested were misidentified in the reference year, 2010, and were correctly identified by 2020; the same was true for two afforested sites. These findings indicate that the CDL experienced an improvement in accuracy for non-agricultural, non-forest landcover in northern Indiana between 2010 and 2020 (Fig. 2.4). In a study conducted for several Midwestern farm bureaus, the CDL showed a decrease in deciduous forest alongside an increase in woody wetlands across multiple Midwestern states (Decision Innovation Solutions, 2013). This is similar to my finding in northwestern Indiana, and may be due to improvements in, or improved integration with, the USGS National Land Cover Database, which provides non-agricultural land cover information to the CDL (USDA NASS, n.d.-b).

Lark et al. (2015) suggested analyzing all years of CDL data within the temporal window of interest to understand the trajectory of change in land cover over time and account for classification errors between years. This is recommended for analyzing crops that may undergo yearly rotation. As demonstrated here, it also applies to non-crop land cover change analysis, especially when early years of CDL data are used. In quantifying CDL accuracy for South Dakota between 2006 and 2012, Reitsma et al. (2016) also found that land use change accuracy was consistent for cropland and grassland. However, for categories including forests, wetlands, and non-agricultural developed areas, they found low accuracy which decreased with time. This contrasts with my finding that the CDL has experienced an improvement in accuracy for non-agricultural, non-forest categories since 2010, but this discrepancy may be due to differences in study area, years of interest, and category collapse, a general challenge also noted by Lark et al. (2017).

Dunn et al. (2015) questioned the accuracy of the grassland area calculations in Lark et al. (2015), partly because the CDL struggles to differentiate between grassland-type land covers (e.g., alfalfa and other hay). However, they concluded that ground truthing and satellite imagery should be used, rather than the CDL alone. Dunn et al. (2017) emphasized the need for ground truthing as a direct source of landcover information. It is interesting to note that one of the six “change” sites studied here was misidentified in 2020, and the site was a grassy area. While not misidentified as a similar category (e.g., hay), it does support the grassland-related identification issues identified by other studies (Dunn et al., 2015; Dunn et al., 2017; Lark et al., 2015; Reitsma et al., 2016).

All of the studies cited concluded that the CDL poses accuracy-based challenges for land cover change analysis. This agrees with my results and supports the idea that the CDL should be used for land cover change quantification only with consideration for the accuracy of the years of interest and extensive verification from ground truthing and satellite imagery. The accuracy of the CDL in broadly measuring agriculture, and to a lesser extent forest, in northern Indiana has been reinforced by the ground truthing in this study

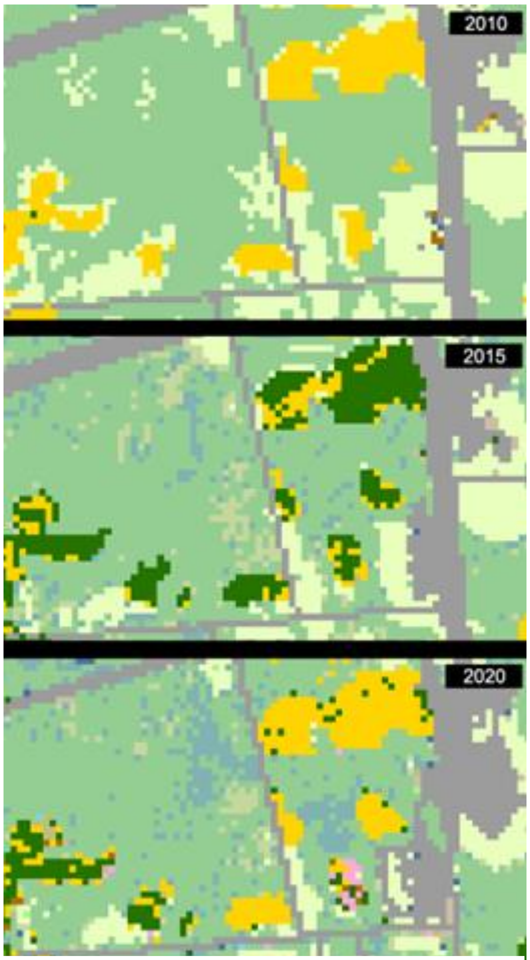

Figure 2.4. A comparison of CDL imagery for a selected area in northwestern Indiana during the years 2010, 2015, and 2020. Note increasing blue (“Woody wetlands”) and tan (“Shrubland”) areas with temporal progression.

2.5 Summary

Land cover is a subject of much interest to environmental science, whether for preliminary sampling site selection or as data analysis. The CDL is an extensive and easily accessible collection of land cover data for the contiguous United States. At the same time, the accuracy of the CDL is constantly improving, which poses challenges when using it in land cover change analysis. The goal of this study was to determine whether differences in land cover between years could be accurately measured using the CDL. The years chosen were 2010 and 2020, and the area measured was northern Indiana. After data manipulation and visualization in R and ArcGIS and site reconnaissance at areas of “afforestation” and “deforestation,” it was concluded that the CDL was unreliable in identifying land cover change based on aggregated forest types for northern Indiana between the years 2010 and 2020. Further research into the CDL as an analysis tool for land use changes may focus on developments in its accuracy over time in identifying and differentiating similar categories of non-agricultural land use, such as woody wetlands and deciduous forest.

In the case of studies using the CDL for land cover change analysis, it is advisable for researchers and data analysts to seek supplemental or alternative means of land cover change measurement. Additionally, any usage of the CDL for contemporary land cover classification should also include site reconnaissance and visual verification through current satellite imagery. In the case of CDL usage as a means of site selection for land cover change-based study, communication with landowners or managers about the history of their land should be used in addition to ground truthing and satellite imagery.

CHAPTER 3. AGRICULTURAL AND FORESTED STREAM BIOLOGICAL INTEGRITY

3.1 Data

3.1.1 *Site Descriptions*

My study focused on small streams in the Central Corn Belt Plains and Eastern Corn Belt Plains ecoregions of Indiana. I selected study streams based on the following criteria: a wetted width <10m, a maximum depth <1m, and accessible with permission by landowners or public domain when sampled. The original objective was to sample at least 30 streams. However, site selection was constrained by time limitations, inclement weather, low flow conditions, personnel logistics, and a limited number of participating landowners. Within these constraints, I sampled 21 sites over a single season to avoid year-to-year variation.

Four sample sites were in the Central Corn Belt Plains ecoregion and 17 were in the Eastern Corn Belt Plains ecoregion (Table 3.1). Both ecoregions occur within the EPA Level II Central U.S. Plains ecoregion. The sites were typically current- or past agricultural ditches or stream reaches in the Wildcat Creek, Sugar Creek, and Big Pine Creek watersheds. A few exceptions to the <10m stream width criterion were made for Sugar Creek sites that were wadable due to low water conditions and could be sampled using a backpack electrofisher. Sites were located throughout Benton, Boone, Clinton, Howard, Montgomery, Tippecanoe, Warren, and White counties in Indiana (Fig. 3.1).

Study sites were primarily rural, with a few urban or exurban stream reaches. All sites were characterized by adjacent land uses on a continuum from intensive agriculture to secondary forest. At least two of the privately owned forested sites were in recovery after intensive agriculture in previous decades. Two forested sites were properties managed by NICHES Land Trust, a local land stewardship organization. Two sites were adjacent to active agriculture but maintained a forested buffer zone. Three sites were owned by Purdue University: the Animal Sciences Resource Education Center, the Purdue University Forestry and Natural Resources Farm, and Throckmorton Agricultural Center. Another site was at Tippecanoe Battlefield and Museum, a forested public park and national historic landmark. The remaining 11 sites were privately owned stream reaches across the intensive agricultural-forested continuum.

All study sites were on properties with ≥100m stream reaches that met the above selection criteria. Where possible, I conducted site reconnaissance to determine site suitability for inclusion in the study and access points for sampling. This was especially critical for selecting entry and exit points for heavily forested streams. A few sites had obstacles (e.g., thickets adjacent to the stream or severe channel narrowing) that made access throughout the 100m study reach difficult or impassable. For impassable sites, modifications to the sampling protocol were made as described in Section 3.2.1. However, most streams were fully accessible throughout the 100m study reach.

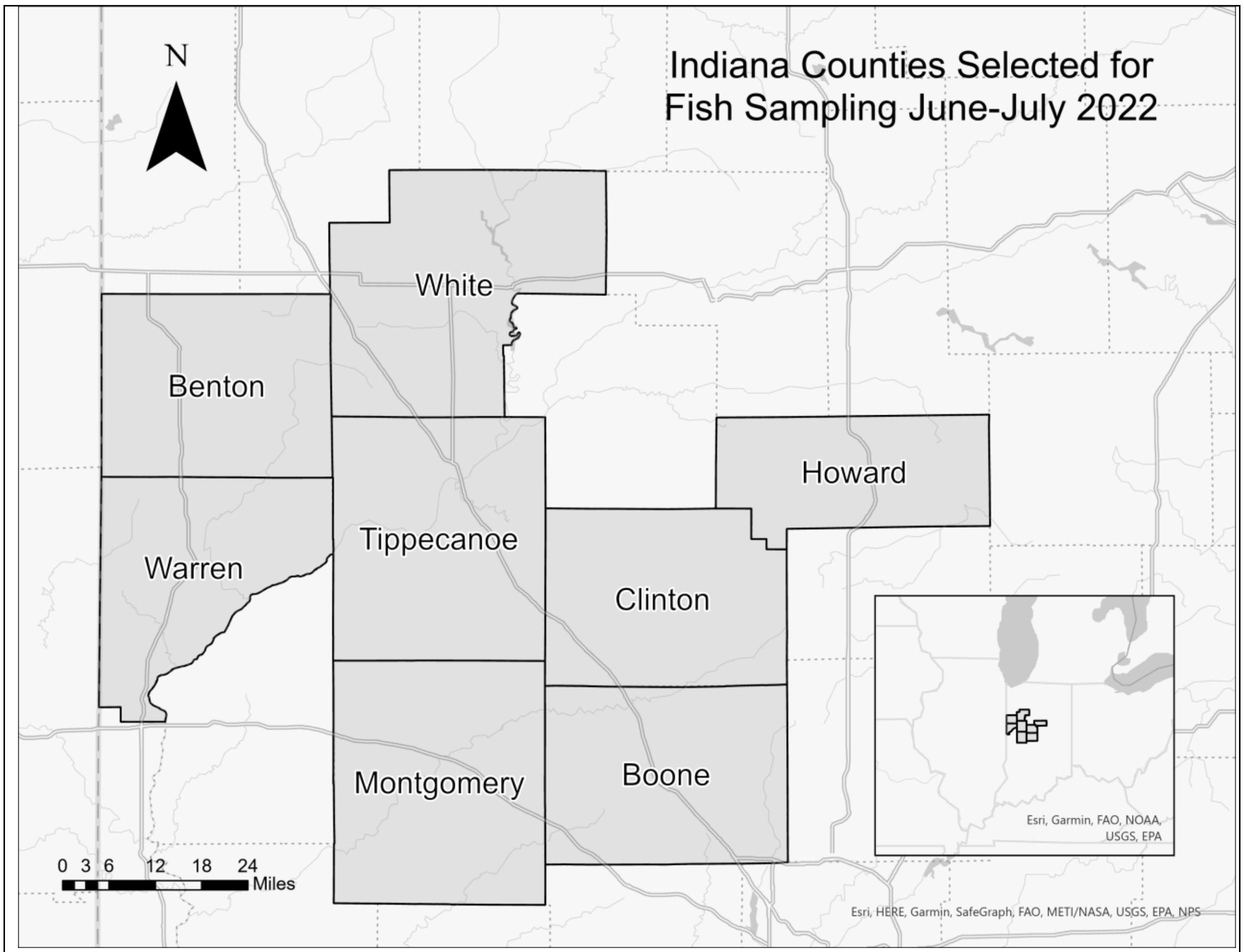

Figure 3.1. North-Central Indiana counties selected for fish sampling and QHEI assessment from June to July 2022.

Table 3.1. Study sites sampled between June and July of 2022. Included are site code, stream name, latitude and longitude coordinates, Indiana county, and Level IV EPA Ecoregion.

Sampled Stream Site	Stream Name	Lat.	Lon.	County	Level IV Ecoregion
S1	Marshall Ditch	40.50097	-87.02756	Tippecanoe	54a Illinois/Indiana Prairies
S2	Jordan Creek	40.42003	-86.97051	Tippecanoe	55b Loamy High Lime Till Plains
S3	Indian Creek	40.42165	-87.04330	Tippecanoe	55b Loamy High Lime Till Plains
S4	Opossum Run	40.22898	-87.47116	Warren	54a Illinois/Indiana Prairies
S5	Unnamed Stream 1	40.57376	-86.80646	Tippecanoe	55b Loamy High Lime Till Plains
S6	Throckmorton Ditch	40.29589	-86.89978	Tippecanoe	55b Loamy High Lime Till Plains
S7	McClamrock Ditch	40.19302	-86.37290	Clinton	55b Loamy High Lime Till Plains
S8	Mud Pine Creek	40.55466	-87.32108	Benton	54a Illinois/Indiana Prairies
S9	Brown Ditch	40.47611	-87.22434	Warren	54a Illinois/Indiana Prairies
S10	Unnamed Stream 2	40.47215	-87.20654	Warren	54a Illinois/Indiana Prairies
S11	South Fork Wildcat Creek	40.32210	-86.39366	Clinton	55b Loamy High Lime Till Plains
S12	West Honey Creek	40.42748	-86.28267	Howard	55b Loamy High Lime Till Plains
S13	Burnett Creek	40.50628	-86.84590	Tippecanoe	55b Loamy High Lime Till Plains
S14	Little Potato Creek	40.22745	-86.65612	Montgomery	55b Loamy High Lime Till Plains
S15	Goldsberry Creek	40.14837	-86.67031	Boone	55b Loamy High Lime Till Plains
S16	Wolf Creek	40.13009	-86.64146	Boone	55b Loamy High Lime Till Plains
S17	Prairie Creek	40.13796	-86.60414	Boone	55b Loamy High Lime Till Plains
S18	Bowers Creek	40.18752	-86.77072	Montgomery	55b Loamy High Lime Till Plains
S19	Lye Creek Drain	40.15062	-86.81981	Montgomery	55b Loamy High Lime Till Plains
S20	Sugar Creek	40.10744	-86.82320	Montgomery	55b Loamy High Lime Till Plains
S21	Little Sugar Creek	40.05064	-86.82392	Montgomery	55b Loamy High Lime Till Plains

3.2 Methods and Materials

3.2.1 *Site Sampling Procedures*

I sampled fish communities in the study reaches with a field crew (i.e., the author plus 1-3 field personnel) between June 1st and July 31st, 2022. Each sampling event occurred between 09:00 and 17:00. The field crew completed sample collection within two hours at most sites. We sampled 2-4 sites daily depending on weather conditions, stream morphology, and stream size. Fish were sampled using a DC backpack electrofisher (Model ABP-3; ETS Electrofishing Systems, LLC; Madison, WI; Figure 3.2) with the following settings: 160-190 V, a frequency of 120 Hz, and a 25% duty cycle. Small adjustments in these settings were necessary based on study reach water chemistry. I determined the appropriate electrofisher settings for each study site based on fish recovery time after briefly operating the electrofisher downstream of the study reach.

To sample fish in each of the study reaches, I operated the backpack electrofisher while the remaining crew members used dip nets to collect the stunned fishes. Stunned fishes were kept in a 5-gallon (18.9L) bucket or 20L flow-through livewell filled with stream water. The electrofisher was operated for ≈30 minutes (1800 s), or until a 100m reach was covered at each site. Exceptions were made for impassible terrain (e.g., stream too narrow and banks too steep to sample) or thick large woody debris accumulations in the stream channel. The field crew sampled all available stream habitats, including stream shallows and a variety of instream cover types (e.g., rootwads, undercut banks, aquatic vegetation, and boulders).

The field crew identified captured fishes to species, counted the number of individuals of each species in the sample, and then released them after sampling the study reach. When fishes were highly abundant in a study reach, we processed the samples at the approximate halfway point in the reach to prevent oxygen stress and high mortality. We then sampled and processed the fish captured from the midpoint to the end of the study reach. All fish identifications and counts were recorded on a site-specific field datasheet. For consistency, one field crew member recorded data for each site, while the other field personnel identified the fish species in the sample.

I completed Qualitative Habitat Evaluation Index (QHEI; Rankin, 1989; Ohio EPA, 2006) forms for each study site after processing the fish sample. Some bias is inherent in QHEI scoring due to the study design. Specifically, backpack electrofishing precludes sampling very deep or fast streams, or streams with excessive amounts of instream cover (e.g., fallen trees) that would prevent movement throughout the study reach.

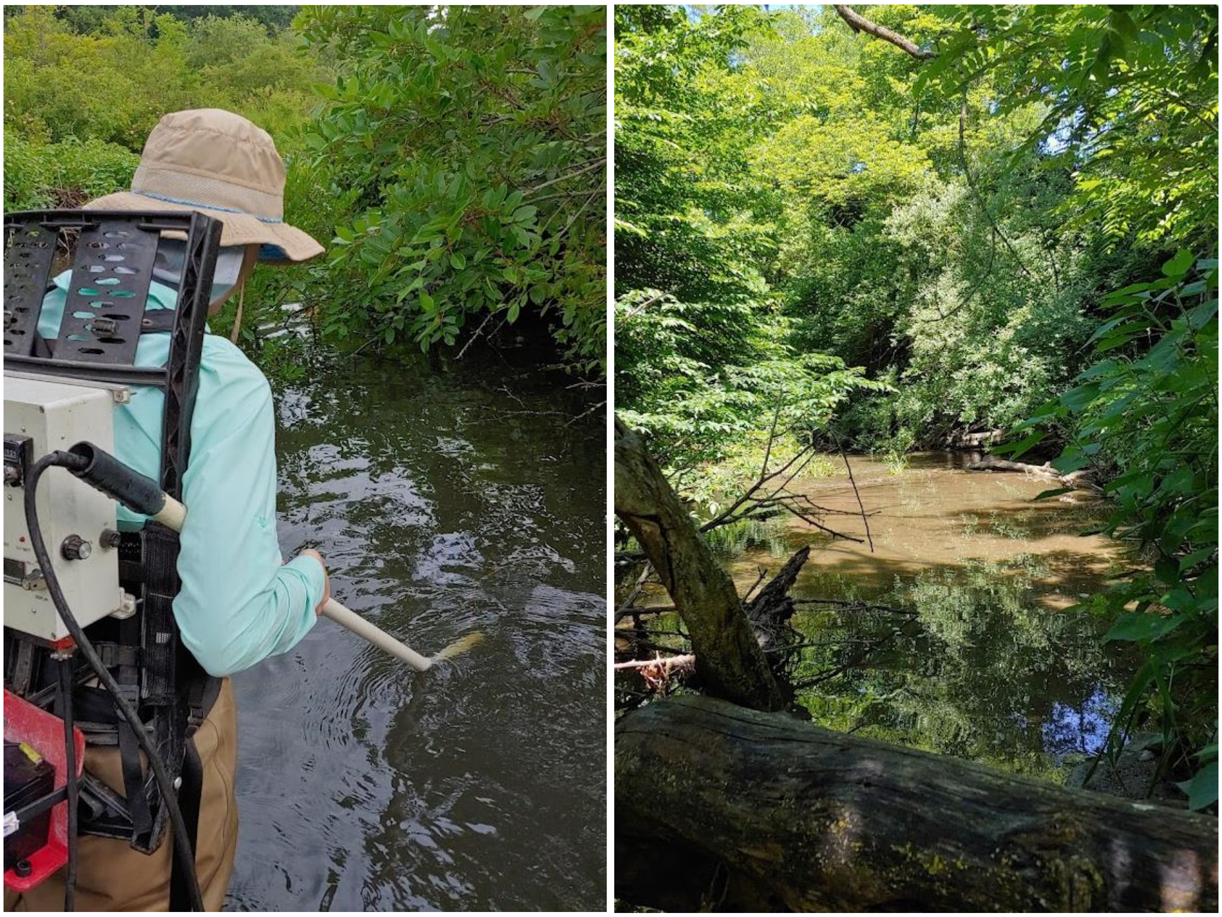

Figure 3.2. Left: the author using an ABP-3 backpack electrofisher (ETS Electrofishing Systems, LLC; Madison, WI). Right: a forested study stream reach with downed woody debris.

3.2.2 *Outlier Identification*

In a preliminary analysis of data normality, I determined that S20, Sugar Creek, was an outlier. S20 had a drainage area of 1041.2 km^2 (402 mi^2); the average drainage area of all other sites was 39.1 km^2 (24.3 mi^2) (Fig. 3.3a). Similarly, S20's number of individuals was 716; the average of all other sites was 204.9 (Fig. 3.3b). Therefore, while S20 was sampleable in low-water conditions, it did not meet the stream criteria and was excluded from the data analysis.

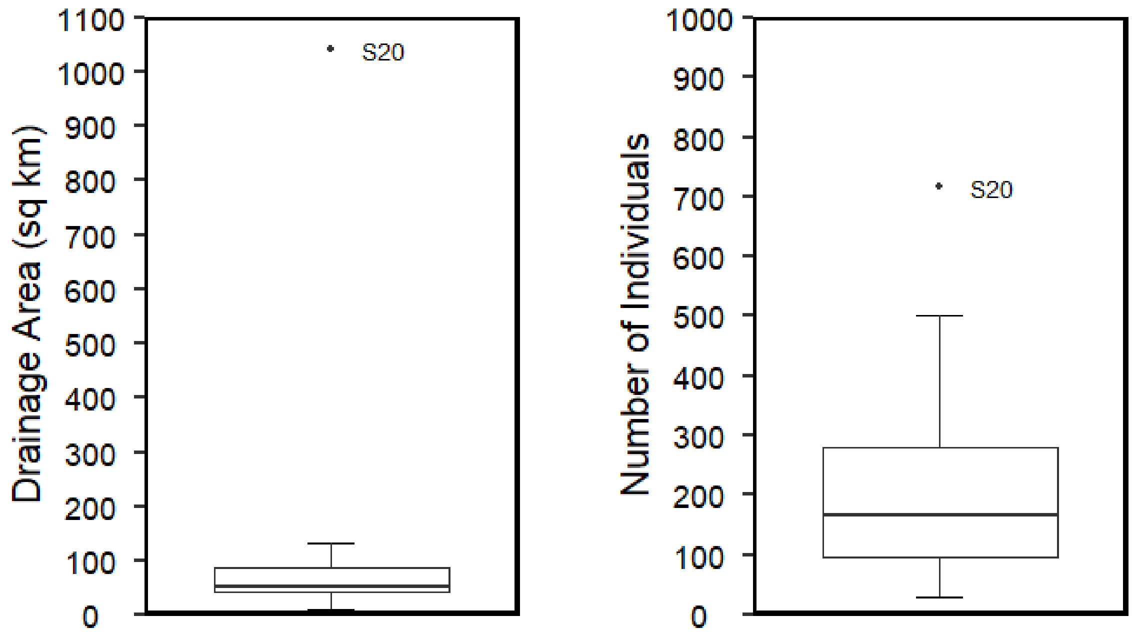

Figure 3.3. Boxplots showing outliers in Indiana stream data for drainage area in km^2 (left; Figure 3.3a) and number of individual fishes captured (right; Figure 3.3b). Site 20 was an outlier for stream drainage (1041 km^2) and number of individuals (n = 716).

3.2.3 *Site Classification Metrics*

I delineated the watersheds of all 20 study sites and measured the agricultural land cover using ModelMyWatershed software (Dewitz, 2021; Stroud Water Research Center, 2017). Cultivated crops covered at least 50% of the drainage basin area in 19 of 20 sites, at least 75% of the drainage basin area in 16 of 20 sites, and at least 90% of the drainage basin area in 11 of 20 sites.

Before comparing site metrics, sites were categorized as agricultural or forested based on riparian characteristics. To do this, I chose four variables related to land cover to use in a principal components analysis (PCA): Cropland Data Layer (CDL) agriculture:forest ratio, riparian width, the QHEI Bank Erosion and Riparian Zone metric score, and the Normalized Difference Vegetation Index (NDVI) (Masek et al., 2006; Vermote et al., 2016). I chose a cutoff value between agricultural and forested for each variable, and classified sites based on agreement between variables.

I created buffers for each site to measure study site riparian landcover using the CDL. These were written in RStudio using the sf package and centered on the site coordinates with an approximate 1km radius. I then imported shapefiles of the buffers to CropScape, the visual interface for the CDL. Data was downloaded in .csv format for the study year (2022) and imported into RStudio. Landcover types were collapsed into eight categories: Cropland, Herbaceous, Water, Developed, Barren, Forest, Pasture, and Wetlands, following the same procedure described in Chapter 3. I used the median ratio of agriculture-to-forest pixel counts for site classification (median = 3.3). Sites below the median were categorized as forest, and values greater than the median were categorized as agriculture.

I measured riparian width using the measurement tool in Google Earth. Measurements in meters were taken at the site coordinates, approximately perpendicular to the stream's angle of meander at the coordinates. The left and right banks' riparia widths were averaged for a single stream riparian measurement. Because multiple streams had no riparian zone (riparian average = 0 m), the median of the dataset could not be used for categorization. Instead, the category threshold was based on the proposed minimum width for a functional riparian zone (20 m) proposed by Feld et al. (2018).

Metric 4 of the Ohio EPA QHEI is Bank Erosion and Riparian Zone. The maximum possible score for this metric is 10 and the minimum is 1. This metric includes three components:

erosion, riparian width, and flood plain quality. Erosion is scored from 1 to 3 (none/little); riparian width is scored from 0 (none) to 4 (wide >50m); and flood plain quality is scored based on the land use >100m away from the stream (maximum score of 3). The median sum value of these components was 5.6; sites below the median were classified as agriculture while sites greater than the median were classified as forest.

The Normalized Difference Vegetation Index (NDVI) is a widely used indicator of vegetation greenness based on satellite imagery (Remote Sensing Phenology, 2018). I used the MODIS/Terra+Aqua Land Cover Type (LC) Yearly L3 Global 500 m SIN Grid (product ID: MOD13A3) at the 16-day 250-m NDVI band (Masek et al., 2006; Vermote et al., 2016). This provided lower spatial resolution and much higher temporal resolution than the CDL. The coordinates of each site's approximated transect center were entered into the MODISTools package in R, along with the 16-day window that included the sampling date. The area covered was 4 km^2 centered on each of the 21 transects. The median NDVI was 0.7; sites below the median were classified as agricultural and sites greater than the median were classified as forested.

3.2.4 *Index of Biotic Integrity Calculation*

I calculated fish Index of Biotic Integrity (Karr, 1981; Karr et al., 1986; Simon and Dufour, 1997; hereafter, IBI) for each site using Microsoft Excel. Using this approach, fish species' characteristics and community composition at the sites were used to score metrics that sum to provide the IBI. The metrics include the number of darters, madtoms, and sculpin; number of darters; %headwater species; number of sunfish; number of minnow species; number of suckers; %pioneer species; number of sensitive species; %tolerant species; %omnivores; %insectivores; %carnivores; CPUE; %simple lithophiles; and %DELTs (i.e., fish exhibiting external deformities, erosions, lesions, or tumors). For sites with <50 individuals, I used the IBI score modifications recommended by Simon and Dufour (1997). DELTs were not counted in sampling, so all study sites with 50 or more individuals were given a score of 5. A table of the IBI component scores is included in Appendix B.

The protocols used to calculate IBI differ based on the drainage area of the sampled stream (i.e., ≤51.8 km^2 [10 sites] and >51.8 km^2 [10 sites]). I used the drainage areas of Indiana streams provided by Hoggatt (1975). The average drainage area included in the study was 62.9 km^2 (min = 11.5 km^2, max = 132.6 km^2). Site S10 did not have an appropriate measurement point on the

USGS drainage map (i.e., drainage at closest measurement point = 380.7 km^2). Therefore, I approximated the S10 drainage area by averaging the drainage at the three most comparable ditches, S1 (Marshall Ditch, 20.6 km^2), S6 (Little Wea Creek Tributary, 86.5 km^2), and S9 (Brown Ditch, 48.2 km^2).

3.2.5 *Statistical Methods*

The variables included in statistical analysis were: land cover category (agricultural or forested; hereafter, category), QHEI score, IBI score, number of individuals, number of species, number of sensitive species, and number of tolerant species. As QHEI score includes land cover components, it was necessary to determine the validity of using it as an independent variable with IBI as a dependent variable. Therefore, I ran four different linear models for comparison: IBI following category, IBI following QHEI, IBI following the cross of category and QHEI, and IBI following category and QHEI (hereafter models 1, 2, 3, and 4, respectively).

I developed a corrected Akaike Information Criteria (AICc) model selection table using the four linear models. Model 2 was the best model, with an AICc of 156.15 and a weight of 0.61. The model 1 AICc was 158.05 with a weight of 0.24. The model 4 AICc was 159.30 with a weight of 0.13. Model 3 was the weakest model identified, with an AICc of 162.86 and a weight of 0.02. (Table 3.2). Models 1 and 2 (Category and QHEI, respectively) were used in the analysis.

Table 3.2. AICc model selection output table.

Model selection based on AICc:

	K	AICc	Delta_AICc	AICcWt	Cum.Wt	LL
QHEI	3	156.15	0.00	0.61	0.61	-74.33
Category	3	158.05	1.90	0.24	0.85	-75.28
Category + QHEI	4	159.30	3.14	0.13	0.98	-74.31
Category * QHEI	5	162.86	6.71	0.02	1.00	-74.29

IBI following QHEI, number of sensitive species following QHEI, and percent tolerant species following QHEI required use of the Pearson Product Moment Correlation coefficient or Spearman Rank Correlation coefficient because they were all continuous variables. IBI following category, QHEI following category, number of individuals following category, and number of species following category were tested using independent t-tests or nonparametric equivalents. A p-value ≤ 0.05 was considered significant for all tests.

3.3 Results

3.3.1 *Site Classification Analysis*

I conducted a principal components analysis (PCA) using the site categorization metrics, including CDL ratio, riparian width, QHEI subsection, and NDVI. A heatmap of the correlation matrix showed little correlation of NDVI values with any other metric. Additionally, the correlation with the CDL agriculture-to-forest pixel ratio was close to zero (Fig. 3.4). A biplot with square cosine values was used to determine similarity between variables, impact of each variable on the two most significant principal components, and the amount of representation of each variable within a component. This plot showed that CDL agriculture-to-forest ratio and QHEI subsection were the two most well-represented variables in the two most significant principal components; average riparian width was less well-represented; and NDVI was very poorly represented (Fig. 3.5).

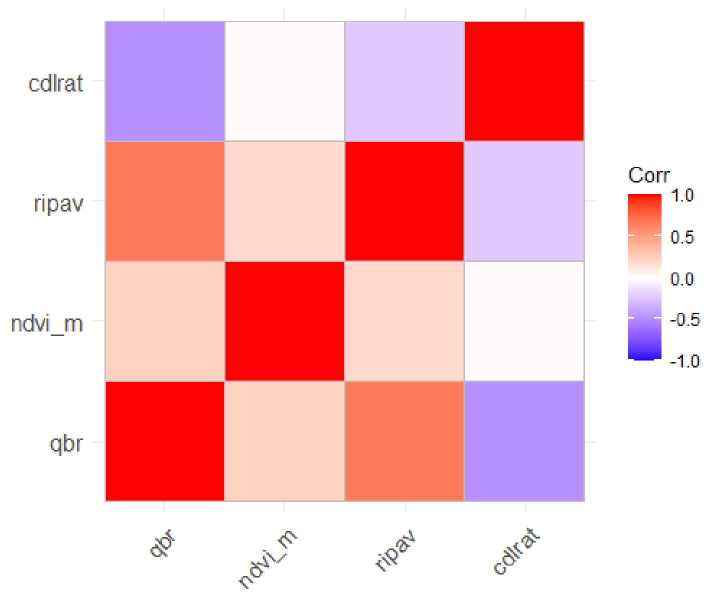

Figure 3.4. A correlation matrix heatmap of the variables used in agriculture vs. forest land cover classification: Cropland Data Layer agriculture:forest ratio (“cdlrat”), average wooded riparian width (“ripav”), mean Normalized Difference Vegetation Index value (“ndvi_m”), and Qualitative Habitat Evaluation Index bank and riparia components (“qbr”). Mean NDVI values had a low correlation with other variables.

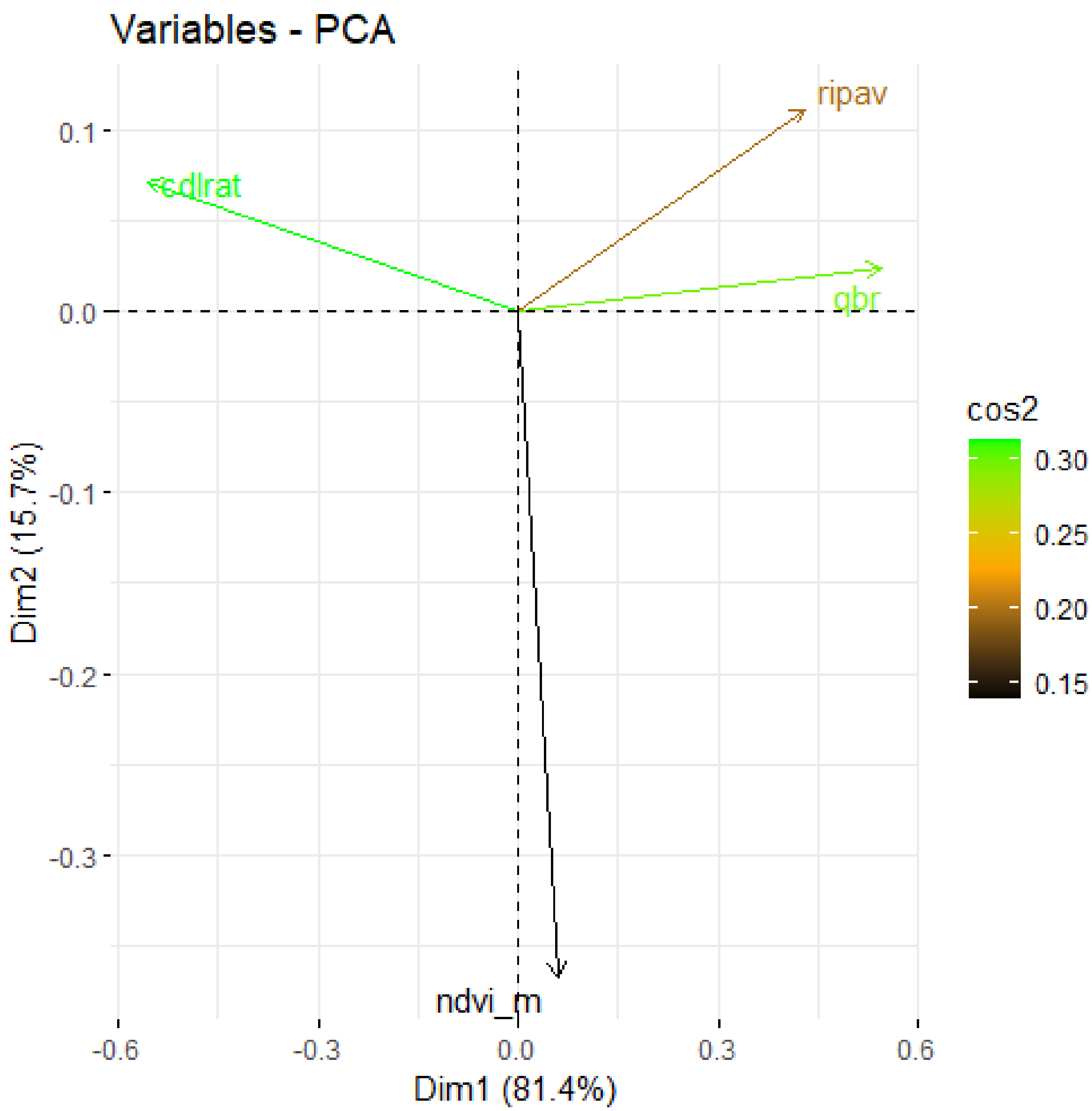

Figure 3.5. A biplot with square cosine values of variables used in agriculture vs. forest land cover classification. Higher square cosine values (cos2) indicate a more well-represented variable. CDL agriculture-to-forest pixel count ratio ("cdlrat") and QHEI Bank & Riparia subsection score ("qbr") are the two best-represented variables. Mean site NDVI ("ndvi_m") is very poorly represented.

As supported by the PCA, NDVI proved unreliable as a means of categorizing landcover (Table 3.3). For example, a site with no wooded riparian zone had an NDVI of 0.82, while a site with an extensive riparian zone had an NDVI of 0.65. This is due to the NDVI's inability to differentiate between tree leaf greenness and other forms of vegetation, including row crops. For this reason, NDVI was not included in the categorization of study sites.

Table 3.3. Parameters of sampled stream sites in north-central Indiana. QHEI scores and NDVI are also included. Land cover measure parameters indicating agricultural ("Ag") land are in bold; non-bolded items indicate forest ("Fo"). "QHEI Bank & Riparia" is the score from Section 4 of the Ohio EPA QHEI form. Riparian width is the average of the left and right riparia at the approximate center of the transect, perpendicular to the stream. CDL Ag:Forest is the ratio of pixel counts for agricultural and forest land cover taken from the USDA Cropland Data Layer.

Sampled Stream Site	QHEI	NDVI	QHEI Bank & Riparian	Riparian Width (m)	CDL Ag:Forest	Land Cover Category
S1	29.0	**0.53**	**3.0**	**6.5**	**623.60**	**Ag**
S2	63.0	0.77	7.0	199.5	0.23	Fo
S3	68.0	0.76	8.5	520.0	0.05	Fo
S4	51.0	**0.65**	9.0	209.5	1.63	Fo
S5	66.5	0.71	9.5	606.0	1.08	Fo
S6	42.0	**0.52**	**4.0**	**0.0**	**18.06**	**Ag**
S7	54.5	**0.51**	**5.5**	36.5	**3.73**	**Ag**
S8	31.0	**0.65**	**3.0**	42.0	**272.42**	**Ag**
S9	41.0	**0.56**	**3.0**	**0.0**	**3.40**	**Ag**
S10	24.0	**0.59**	**3.0**	**0.0**	**5.94**	**Ag**
S11	72.5	**0.48**	**5.0**	**19.0**	**8.71**	**Ag**
S12	71.0	**0.57**	6.5	75.0	2.99	Fo
S13	62.5	0.77	10.0	46.0	0.43	Fo
S14	47.0	0.82	**3.0**	**0.0**	**18.01**	**Ag**
S15	67.5	0.80	6.5	108.0	2.92	Fo
S16	73.0	0.76	6.5	80.0	3.17	Fo
S17	62.0	**0.69**	7.0	32.5	2.36	Fo
S18	40.5	0.86	**3.0**	**0.0**	**649.40**	**Ag**
S19	49.8	0.87	**4.8**	**9.5**	**13.04**	**Ag**
S21	56.8	0.76	5.8	57.5	1.91	Fo

The CDL, average riparian width, and QHEI subsection were in strong agreement for 18 out of 20 sites. The two remaining sites, S8 and S9, were ambiguous due to higher-than-expected riparian width averages (36.5m and 42.0m, respectively) compared to the QHEI subsection and CDL values. Site S8 was a stream in recovery from intensive agriculture, with a highly variable upstream riparian width. Site S9 had a wide but sparsely wooded riparian buffer on the right bank and no wooded riparian zone on the left bank. Both were also classified as agricultural by the QHEI subsection and CDL values, which were the best-represented variables in the PCA. Therefore, both were categorized as agricultural sites. There were 10 agricultural sites and 10 forested sites after classification (Fig. 3.6).

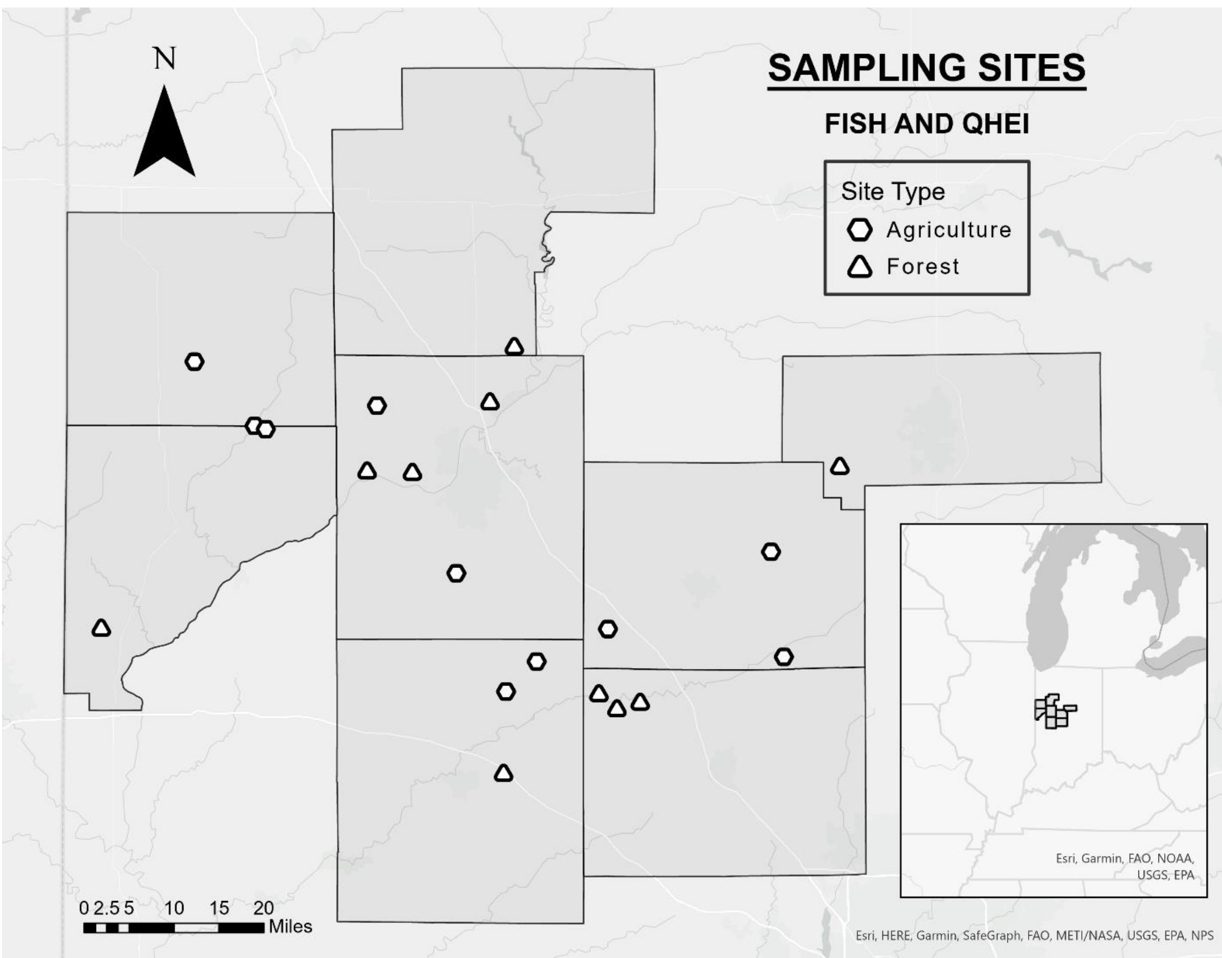

Figure 3.6. Fish and QHEI stream sampling locations by land cover category (Agriculture, Forest). Sites were located in Benton, Boone, Clinton, Howard, Montgomery, Tippecanoe, Warren, and White counties in north-central Indiana.

3.3.2 *Summary Statistics*

For all sites, the mean number of individuals captured per site was 204.9 (median = 163.5, SD = 148.4). The mean number of species identified at each site was 12.2 (median = 12, SD = 6.1). The mean site QHEI score was 53.6 (median = 55.6, SD = 15.2). The mean site IBI score was 36.7 (median = 40, SD = 11.1).

For agricultural sites, the mean number of individuals captured was 147.1 (median = 102.5, SD = 120.5). The mean number of species identified at each site was 10.4 (median = 12, SD = 4.5). The mean site QHEI score was 43.1 (median = 41.5, SD = 14.1). The mean site IBI score was 34.0 (median = 40, SD = 12.8).

For forested sites, the mean number of individuals captured was 262.7 (median = 248.5, SD = 156.8). The mean number of species identified at each site was 13.9 (median = 13, SD = 7.1). The mean site QHEI score was 64.1 (median = 64.8, SD = 6.6). The mean site IBI score was 39.4 (median = 40, SD = 8.9). Mean and standard deviation values for all four metrics across both categories are provided in Table 3.4.

Table 3.4. Summary statistics (mean ± standard deviation) of number of individuals, number of species, QHEI score, and IBI score for agricultural and forested streams.

Category	Mean Individuals	Mean Species	Mean QHEI	Mean IBI
Agricultural	147.1 (±120.5)	10.4 (±4.5)	43.1 (±14.1)	34.0 (±12.8)
Forested	262.7 (±156.8)	13.9 (±7.1)	64.1 (±6.6)	39.4 (±8.9)

3.3.3 *Statistical Tests*

A Shapiro-Wilk test was used to test the data for normality. Outliers were detected by drawing boxplots. Homogeneity of variance was tested using mean-center and median-center Levene's tests. The Kruskal-Wallis test was used when assumptions for an independent t-test (i.e., "Welch's t-test") were not met or when the data contained outliers.

The Shapiro-Wilk test p-values (>0.05) indicated the data were normally distributed for number of individuals, number of species, and QHEI score for both agricultural and forested streams (Table 3.5). The Shapiro-Wilk test indicated that the IBI score data were not normally distributed for agricultural sites ($p = 0.011$). Boxplots revealed that the number of individuals found at agricultural site S14 was an outlier ($n = 431$). The boxplot of QHEI scores indicated that forested site S4 was also an outlier (n = 51.0) (Fig. 3.7). Kruskal-Wallis tests were used for number of individuals, QHEI score, and IBI score, while Welch's t-test and Kruskal-Wallis test were run for the number of species. Because the Kruskal-Wallis test requires all factors to be numeric, agricultural was coded as 1 and forested was coded as 2.

A Shapiro-Wilk test was run for QHEI score, IBI score, percent tolerant species, number of intolerant species, and percent cultivated crops in delineated study site watersheds. The Shapiro-Wilk test indicated that IBI score, number of intolerant species, and percent cultivated crops had non-normal distributions ($p = 0.035$, $p = 0.001$, and $p < 0.001$, respectively) (Table 3.6). Using boxplots, the percent tolerant species at S6 (86%) was revealed to be an outlier. The number of intolerant species at sites S21 (n = 10), S16, and S17 (n = 11 at both sites) were also determined to be outliers (Fig. 3.8). As mentioned in Section 3.2.3, cultivated crops accounted for at least 90% of drainage basin area for 11 of 20 sites, which created outliers. The Spearman Rank Correlation Coefficient was used for all correlation analyses due to violation of one or more assumptions.

Table 3.5. Assumptions and test selection for four dependent variables following land cover (IBI, number of individuals, number of species, QHEI). Normality was tested using the Shapiro-Wilk test. Homogeneity of variance was tested using mean-center and median-center ("med-center") Levene's tests.

Dependent Variable	Data Normality		Outliers	Homogeneity of Variance	Test Used
	Agriculture	Forest			
IBI	Non-normal ($p = 0.011$)	Normal ($p = 0.801$)	No	Heterogeneous (mean-center $p = 0.029$) (med-center $p = 0.277$)	Kruskal-Wallis
No. of Individuals	Normal ($p = 0.054$)	Normal ($p = 0.561$)	Yes	-	Kruskal-Wallis
No. of Species	Normal ($p = 0.156$)	Normal ($p = 0.530$)	No	Homogeneous (mean-center $p = 0.121$) (med-center $p = 0.132$)	Welch's; Kruskal-Wallis
QHEI	Normal ($p = 0.665$)	Normal ($p = 0.696$)	Yes	-	Kruskal-Wallis

Table 3.6. Assumptions and test selection for five continuous variables (QHEI, IBI, number of intolerant species, percent tolerant species, percent cultivated crops in delineated study site drainage basin). QHEI was the independent variable to be used in analysis for all other variables except percent cultivated crops, which was analyzed with IBI. Normality was tested using the Shapiro-Wilk test.

Variable	Data Normality	Outliers	Test Used
QHEI	Normal ($p = 0.184$)	No	-
IBI	Non-normal ($p = 0.035$)	No	Spearman Rank Sum Correlation
% Tolerant Spp.	Normal ($p = 0.424$)	Yes	Spearman Rank Sum Correlation
No. of Intolerant Spp.	Non-normal ($p = 0.001$)	Yes	Spearman Rank Sum Correlation
% Cultivated Crops	Non-normal ($p < 0.001$)	Yes	Spearman Rank Sum Correlation

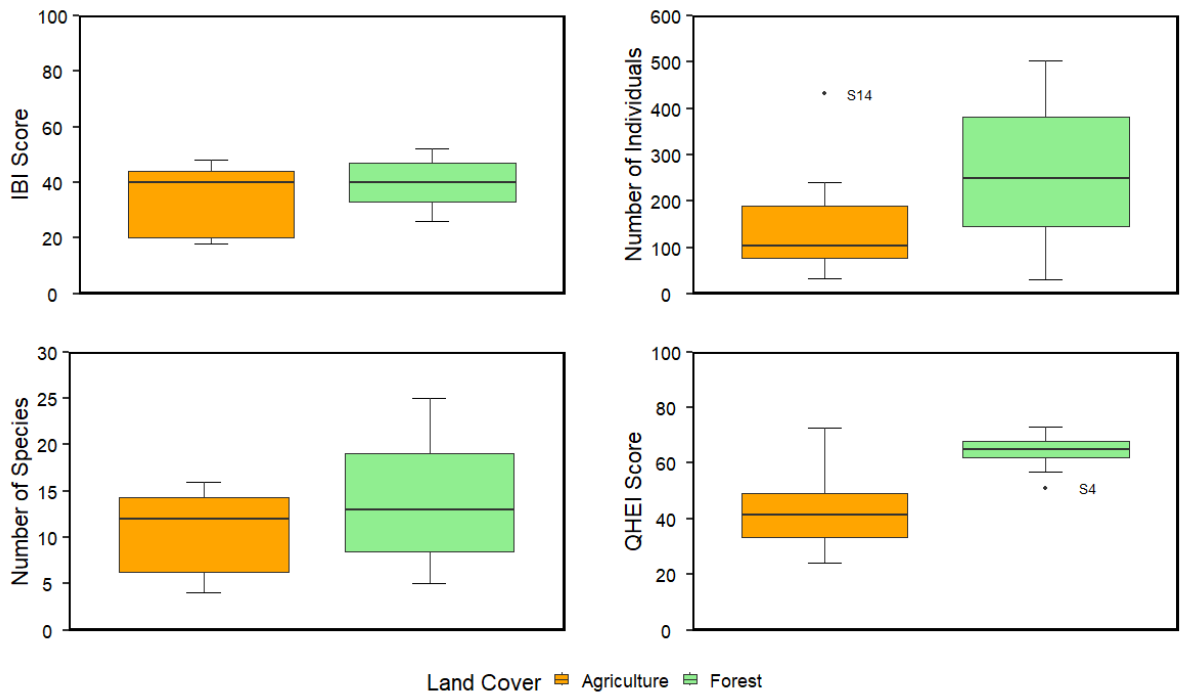

Figure 3.7. Boxplots of four variables by land cover for 20 sampled sites. Variables include number of individuals captured, number of species captured, Index of Biotic Integrity (IBI) score, and QHEI score. Outliers are labeled with site number.

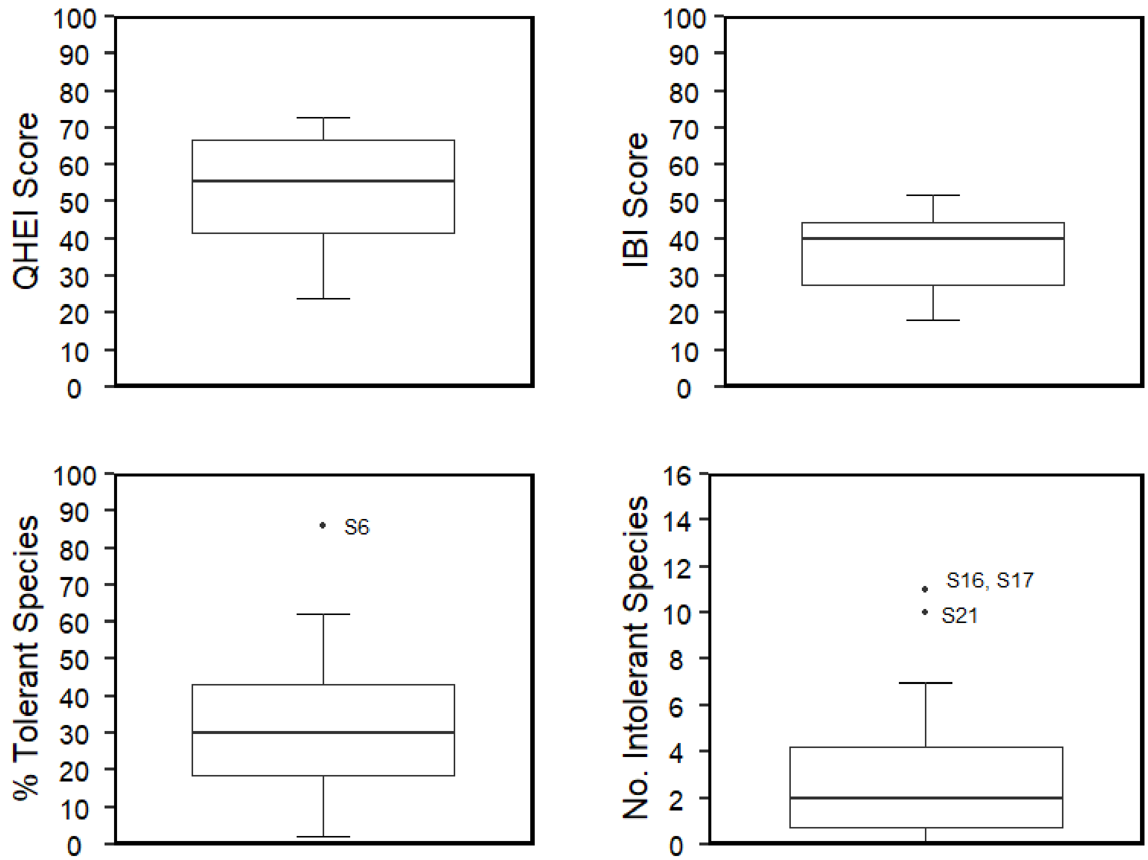

Figure 3.8. Boxplots of IBI score, QHEI score, number of intolerant species, and percent tolerant species for n = 20 sites. Outliers are labeled with site number; multiple outliers of the same value are labeled on one point.

The nonparametric Kruskal-Wallis rank sum test showed no significant difference in median IBI scores between agricultural and forested streams ($\chi 2 = 0.83$, df = 1, $p > 0.36$). The Kruskal-Wallis test showed no significant difference in median number of individuals between agricultural and forested streams ($\chi 2 = 3.3$, df = 1, $p = 0.070$). Welch's t-test showed no significant difference in mean number of species between agricultural and forested streams (df = 15.24, $p > 0.2$). The Kruskal-Wallis test likewise showed no significant difference in median number of species between agricultural and forested streams ($\chi 2 = 1.05$, df = 1, $p > 0.3$). The Kruskal-Wallis test showed a significant difference in median QHEI scores between agricultural and forested streams ($\chi 2 = 9.14$, df = 1, $p = 0.002$).

QHEI and IBI were not significantly correlated based on Spearman Rank Correlation analysis ($r = 0.27$, $df = 18$, $p = 0.25$; Figure 3.9). There was also no significant correlation between QHEI and percent tolerant species using the Spearman Rank Correlation coefficient ($r = -0.2$, $df = 18$, $p > 0.30$; Figure 3.9). QHEI and number of intolerant species were not significantly correlated based on Spearman Rank Correlation analysis ($r = 0.23$, $df = 18$, $p > 0.30$; Figure 3.9). Finally, IBI and percent cultivated crops were not significantly correlated based on Spearman Rank Correlation analysis ($r = -0.31$, $df = 18$, $p = 0.178$; Figure 3.9).

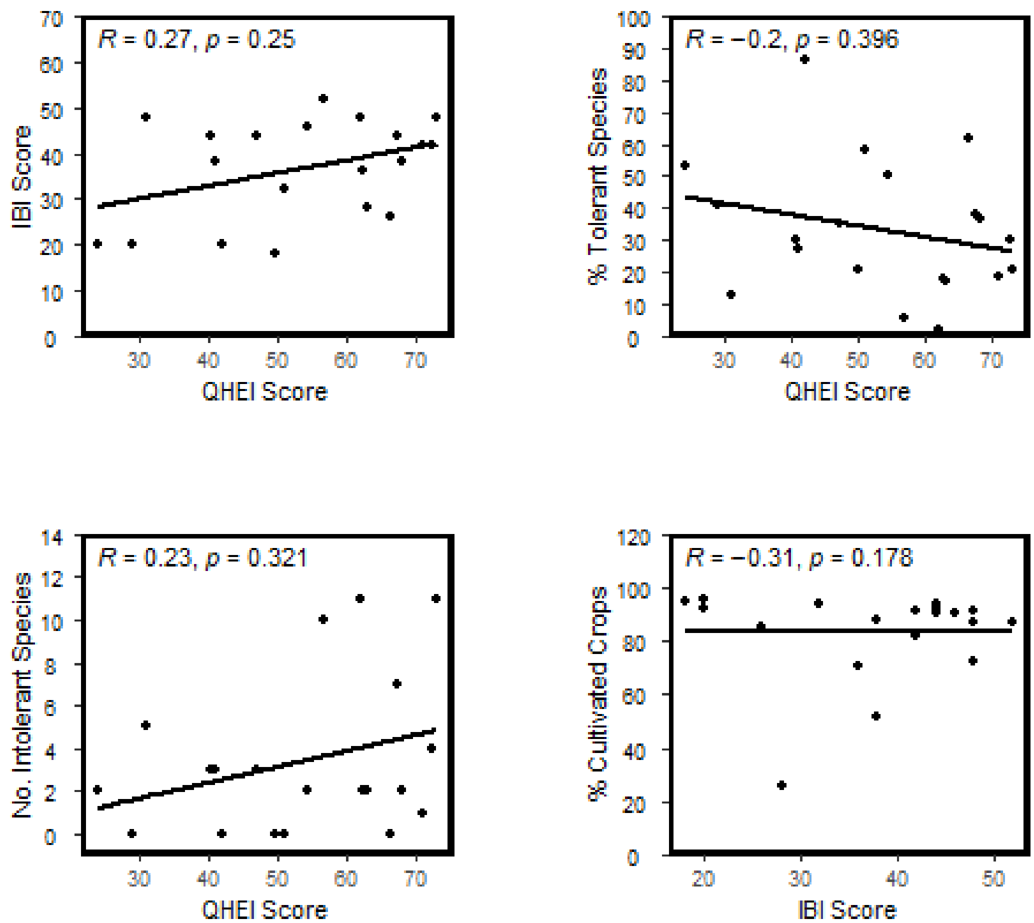

Figure 3.9. Spearman Rank Correlation plots of biotic and abiotic measures for 20 stream sites. Clockwise from top left: QHEI score and Index of Biotic Integrity score ($r = 0.27$, $p = 0.25$); QHEI and percent tolerant species ($r = -0.2$, $p > 0.30$); IBI score and percent cultivated crops in drainage basin ($r = -0.31$, $p = 0.178$); QHEI and number of intolerant species ($r = 0.23$, $p > 0.30$).

3.4 Discussion

My study findings suggest that fish IBIs, site QHEI scores, and component scores for both protocols were largely uncorrelated with local land use. This may have been due, in part, to the logistical challenges encountered during the study. In particular, the number of study sites was smaller than originally planned, which resulted in underpowered tests. The weather conditions during sampling were also very dry (NOAA, 2023). This may have affected local fish assemblage compositions due to fish movement away from suboptimal conditions (Ross et al., 1985). Limitations in transect size and location may have also affected species representation in fish samples (Karr et al., 1986). It must also be noted that the IBI calculations in my study were based on historical drainage area measurements from Hoggatt (1975). The measurements used in my study classify 10 of the 20 sites as “large”, while contemporary measurements of delineated drainage area based on site coordinates classify seven of the 20 sites as “large” (Stroud Water Research Center, 2017; USGS, 2019). Hoggatt (1975) was used by Sullivan et al. (2003) and Lau et al. (2006) for drainage area measurement; the methods used in other studies cited in this discussion are unknown.

Stream IBI scores for both agricultural ($\bar{x}$ = 34.0 ± 12.8) and forested ($\bar{x}$ = 39.4 ± 8.9) study sample sites fell within the range of scores corresponding to Poor and Fair biotic conditions, respectively, but were not significantly different (Doll, 2011; Karr et al., 1986). Hrodey et al. (2009) also found that fish IBI scores of first- to fourth-order Wabash River tributaries were not significantly different between agricultural, forested, and fallow field land uses. They suggested this was because <400 individuals were captured at nearly half of the study sites. Shields et al. (1995) reported higher variability in IBI scores for streams with total fish catches <400 individuals, and Simon and Dufour (1997) recommended that sites with total catches <50 should be excluded in IBI calculation, or the calculation must be modified. Sites S1, S2, and S19 had fewer than 50 captured individuals, and all sites except S14, S15, S16, and S17 had fewer than 400 captured individuals, which could be the cause of the increased variability in overall IBI scores (Fore et al., 1994).

It is possible that land cover categories used in this study were not representative in scale to the extent that there would be measurable differences in IBIs between forested and agricultural streams and ditches. Studies by Sullivan et al. (2003), Lau et al. (2006), and Clark-Kolaks (2022) found no significant correlation between the riparian zone scores taken from QHEI site assessments and IBI scores. Similarly, Roth et al. (1996) found that catchment-scale agricultural land cover was a strong predictor of IBI, but that reach-scale (1500 m) median vegetation width was a weak predictor. This contradicts my lack of significant correlation between IBI score and percent cultivated crops in study area drainage basins, further indicating a possible issue due to sample size.

The lack of significant differences between IBI scores of forested and agricultural streams sampled in my study may also be due to tile drainage in agricultural areas beyond the forested riparian buffer. Previous research has suggested that forested riparian areas may be poor at mitigating agricultural runoff to streams in the presence of tile drainage (Baker et al., 2006; Stone and Wilson, 2006). While tile drainage was not a factor in stream classification for this study, it is a very common feature of agricultural lands in Indiana (Pavelis, 1987), and most drainage basins in my study were strongly agricultural (Dewitz, 2021; Stroud Water Research Center, 2017). Alternatively, the areas adjacent to streams and ditches lacking forested riparian buffers were often grassed. This may indicate the efficacy of grassed riparian strips in mitigating the addition of fine sediments and nutrients to ditches and streams (Hughes and Quinn, 2014; Osborne and Kovacic, 1993; Yuan et al., 2009).

I found no significant correlation between IBI and QHEI. Regression models developed by Moerke and Lamberti (2006) indicated that IBI was related to local and instream scale metrics, including large woody debris volume, dissolved oxygen, specific conductivity, and turbidity, which were not measured in my study. In a study of 29 Minnesota streams, Talmage et al. (2002) likewise found a positive correlation between fish IBI and metrics such as %woody debris and %overhanging vegetation. Sullivan et al. (2003) and Lau et al. (2006) both found significant positive correlations between IBI and QHEI scores across east-central Indiana stream transects ($n = 42$ and 40, respectively). The streams sampled by Lau et al. (2006) had comparable QHEI (i.e., 26 – 83) and fish IBI (i.e., 20 – 52) scores to those in my study. This suggests that the absence of a significant correlation between these variables in my study may have been related to sample size. Shields et al. (1995) studied Mississippi streams and found no significant relationship between IBI and QHEI and attributed this (in part) to small capture numbers (i.e., $n < 400$ fishes) and a lack of pristine reference sites, both of which were issues in my study.

There were no significant correlations between land cover type and the number of fish captured or fish species richness. These relationships may have been masked by the comparably small number of study sites, the small physical size of the streams sampled, or the low numbers of fish in the samples under very low water conditions. However, landscape factors related to fish abundance and fish species richness have not been consistently identified in other studies. For example, Hrodey et al. (2009) found that the total number of fish in a sample was best explained by QHEI score, followed by watershed area and % upstream forest. They also found that fish species richness was significantly higher in forested than agricultural streams. On the other hand, Moerke and Lamberti (2006) found that species richness was most strongly predicted by drainage area, low-flow yield, urban land cover, and lacustrine geology, while forested- and agricultural land covers were not strong predictors of species richness. It is clear that the relationship between landscape factors and fish abundance and fish species richness is complex and not always directly correlated with land cover type.

There were no significant correlations between QHEI and the number of intolerant fish species or the percentage of tolerant fish species in a sample. Lau et al. (2006) found a higher number of intolerant fishes in natural than channelized streams, although there was no difference in the number of tolerant fishes across these stream types. Roth et al. (1996) anecdotally observed higher numbers of intolerant fishes at sites with Habitat Index (HI) values above 60 (n = 23) but the relationship was not statistically quantified. Shields et al. (1995) tested correlations between fish IBI components and 10 physical habitat measures across four categories, including riparian conditions; severity of channel incision; substrate and habitat heterogeneity; and cover and pool formation. They found that the number of intolerant fishes was negatively correlated with channel depth and the presence of kudzu (*Pueraria montana*), an invasive vine that outcompetes shading- and bank-stabilizing riparian plants native to the study area in Mississippi. The number of intolerant fishes was positively correlated with the availability of pool habitat. The opposite was true for the proportion of tolerant fishes, which was positively correlated with channel depth and kudzu presence and negatively correlated with the availability of pool habitat. The physical properties included in my study were based on qualitative assessments that may not have been sufficient to capture the associations between abiotic factors and the fish community.

I found that median QHEI scores were significantly higher for forested streams (median = 62.0, IQR = 5.75) compared to agricultural streams (median = 40.4, IQR = 15.7). Hrodey et al. (2009) also found that QHEI scores were significantly higher for forested riparian streams than agricultural streams. Moerke and Lamberti (2006) also found that QHEI scores were positively correlated with the amount of local forest land cover. This makes sense, because QHEI scores include components that score higher due to the presence of a forested riparian zone (e.g., riparian land cover, presence of woody debris; Rankin, 1989).

3.5 Summary

The aim of this study was to investigate the relationship between surrounding land usage (i.e., row crop versus forest status) and the biological integrity of streams in the Temperate Plains ecoregion of Indiana. QHEI, land cover, and fish species data were collected for 20 sampled stream sites in north-central Indiana. Mean IBI scores were not significantly different between ditches and streams with agricultural vs. forested riparian areas, and QHEI was not significantly correlated with IBI score in the sampled Indiana streams. These results suggest that there may be no association between land cover and biotic integrity in Indiana Temperate Plains streams. However, due to time and resource constraints and outlier exclusion, only 20 sites were used in analysis. Future research on the correlations between land cover and stream biotic integrity may build on these findings by using larger sample sizes, controlling for drainage types in agricultural fields or the usage of non-riparian management practices, or employing precise measures of forest area rather than binary land cover categories. The dynamics between agriculture, forests, and stream integrity are complex and highly interconnected. A better understanding of the intricacies of landscape-stream interactions mediated by riparia is necessary to create and improve management strategies.

APPENDIX A. LAND COVER CHANGE ANALYSIS R CODE

```
library(base)
library(utils)
library(graphics)
library(sp)
library(raster)

# packages below are only necessary if creating and exporting a shapefile in Rstudio:
library(dplyr)
library(magrittr)
library(sf)
library(tigris)

# define study area county cluster for shapefile
study_area <- tigris::counties(state = "IN", cb = TRUE) %>%
  st_as_sf() %>% filter(NAME %in% c(
        "Adams", "Allen", "Benton", "Blackford", "Carroll", "Cass", "Clinton",
        "DeKalb", "Elkhart", "Fulton", "Grant", "Howard", "Huntington", "Jasper",
        "Jay", "Kosciusko", "LaGrange", "Lake", "LaPorte", "Marshall", "Miami",
        "Newton", "Noble", "Porter", "Pulaski", "St. Joseph", "Starke", "Steuben",
         "Tippecanoe", "Tipton", "Wabash", "Warren", "Wells", "White", "Whitley"
          ))

#create shapefile of study area
   st_write(study_area, paste0("studyarea.shp"))

#load the TIFs downloaded from Cropscape for years 0 and 1:
tmp1 <- paste("year1raster.tif", sep = "/")
tmp0 <- paste("year0raster.tif", sep = "/")

#define as raster data:
tif1 <- raster(tmp1)
tif0 <- raster(tmp0)

# Change raster value to 1 = "forest", 2 = "cropland", and 3 = other:
myfun <- function(x){
  z <- rep(NA, length(x))
  z[which(x == 0)] <- 0
  z[which(x >= 141 & x <= 143)] <- 1
  z[which((x >= 1 & x <= 57) | (x >= 59 & x <= 77) | (x >= 205 & x <= 254))] <- 2
  z[which(x == 58 | (x >= 78 & x <= 140) | (x >= 144 & x <= 204) | x >= 255)] <- 3
  return(z)
}
```

```
#calculate new raster:
CDL_yr0 <- calc(tif0, fun = myfun)
CDL_yr1 <- calc(tif1, fun = myfun)

# create a new raster to display the change
# 0 = outside the map (original 0)
# 1 = anything to "forest"
# 2 = "forest" to anything
# 3 = "forest" to "forest"
# 4 = "cropland" to "cropland"
# 5 = all other categorical changes

mydat <- overlay(CDL_yr0, CDL_yr1, fun = function(x,y){
  z <- rep(NA, length(x)) # create a vector of NAs
  z[(x == 0 & y == 0)] <- 0
  z[(x != 1 & y == 1)] <- 1
  z[(x == 1 & y != 1)] <- 2
  z[(x == 1 & y == 1)] <- 3
  z[(x == 2 & y == 2)] <- 4
  z[is.na(z)] <- 5
  return(z)
})

# save as a tif file
writeRaster(mydat, "mydat.tif")
```

APPENDIX B. IBI DATA TABLE

Table B.1. Index of Biotic Integrity (IBI) component metrics for 21 sampled Indiana streams in Benton, Boone, Clinton, Howard, Montgomery, Tippecanoe, Warren, and White counties. Streams were sampled between 6/6/2023 and 7/20/2023. Fish samples were collected using a Smith-Root backpack electrofisher in one 100m or approximately 1800s transect. CPUE (catch per unit effort) is the total number of individuals captured at each site. DELTs are the number of fish displaying a deformity, erosion, lesion, or tumor. Total IBI scores and total species captured are also provided for sampled sites. Site 20 was omitted from further statistical analysis as an outlier.

	Sampled Stream Site																				
IBI Components	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	S17	S18	S19	S20	S21
Darter/madtom/sculpin species	1	2	4	3	2	1	5	2	0	1	3	3	3	4	3	5	5	5	1	6	4
Darter species	1	2	3	2	1	0	5	2	0	1	3	2	2	3	3	5	4	3	0	4	3
Headwater species	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
Sunfish species	2	0	1	1	0	0	0	3	3	1	2	0	1	2	1	3	3	2	0	2	3
Minnow species	1	6	9	6	3	4	6	4	5	3	5	2	6	6	9	10	8	7	0	10	9
Sucker species	1	0	0	0	0	0	1	2	1	0	1	0	2	2	3	3	1	0	0	1	4
% Pioneer species	69	33	29	59	45	46	38	13	4	71	55	24	38	74	63	57	45	41	0	57	47
Sensitive species	0	2	2	0	0	0	2	5	3	2	4	1	2	3	7	11	11	3	0	11	10
% Tolerant species	41	17	37	58	62	86	50	13	27	53	30	19	18	35	38	21	2	30	21	5	6
% Omnivores	0	0	10	19	15	13	8	8	8	0	6	0	9	19	17	7	0	22	0	5	2
% Insectivores	91	77	60	39	38	13	33	84	77	46	59	81	82	38	29	43	60	58	27	43	52
% Carnivores	3	0	0	0	0	0	0	7	15	1	10	0	0	4	0	2	5	10	73	2	8
CPUE	32	30	279	252	131	70	240	166	107	98	98	161	141	431	415	471	502	196	33	716	250
% Simple lithophiles	25	37	26	54	60	41	54	4	7	19	36	42	40	11	23	21	19	8	0	22	33
% DELTS	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total IBI Score	20	28	38	32	26	20	46	48	38	20	42	42	36	44	44	48	48	44	18	46	52
Total Species	7	8	14	10	5	5	12	15	12	6	12	5	12	15	19	25	19	16	4	21	22

REFERENCES

Alexander, R. B., Smith, R. A., Schwarz, G. E., Boyer, E. W., Nolan, J. V., & Brakebill, J. W. (2008). Differences in phosphorus and nitrogen delivery to the Gulf of Mexico from the Mississippi River Basin. Environmental Science & Technology, 42(3), 822–830. https://doi.org/10.1021/es0716103

Baker, M. E., Weller, D. E., & Jordan, T. E. (2007). Effects of stream map resolution on measures of riparian buffer distribution and nutrient retention potential. Landscape Ecology, 22(7), 973–992. https://doi.org/10.1007/s10980-007-9080-z

Baker, N. T., & Capel, P. D. (2011). Environmental factors that influence the location of crop agriculture in the conterminous United States. In Scientific Investigations Report (2011–5108). U.S. Geological Survey. https://doi.org/10.3133/sir20115108

Baker, N. T., Stone, W. W., Wilson, J. T., & Meyer, M. T. (2006). Occurrence and transport of agricultural chemicals in Leary Weber Ditch Basin, Hancock County, Indiana, 2003–04 (U.S. Geological Survey Scientific Investigations Report 2006–5251; p. 44).

Barbour, M. T., Gerritsen, J., Snyder, B. D., & Stribling, J. B. (1999). Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, benthic macroinvertebrates, and fish, second edition (EPA 841-B-99-002). U.S. Environmental Protection Agency; Office of Water. https://www3.epa.gov/region1/npdes/merrimackstation/pdfs/ar/AR-1164.pdf

Blann, K., Frost Nerbonne, J., & Vondracek, B. (2002). Relationship of riparian buffer type to water temperature in the Driftless Area ecoregion of Minnesota. North American Journal of Fisheries Management, 22(2), 441–451. https://doi.org/10.1577/1548-8675(2002)022<0441:RORBTT>2.0.CO;2

Bracmort, K. S., Arabi, M., Frankenberger, J. R., Engel, B. A., & Arnold, J. G. (2006). Modeling long-term water quality impact of structural BMPs. Transactions of the ASABE, 49(2), 367–374. https://doi.org/10.13031/2013.20411

Camargo, J., & Alonso, A. (2006). Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: A global assessment. Environment International, 32, 831–849. https://doi.org/10.1016/j.envint.2006.05.002

Capel, P. D., McCarthy, K. A., Coupe, R. H., Grey, K. M., Amenumey, S. E., Baker, N. T., & Johnson, R. L. (2018). Agriculture—A river runs through it—The connections between agriculture and water quality. In Agriculture—A river runs through it—The connections between agriculture and water quality (USGS Numbered Series 1433; Circular, Vol. 1433, p. 216). U.S. Geological Survey. https://doi.org/10.3133/cir1433

Carman, S. F. (2013). Indiana forest management history and practices. In: Swihart, Robert K.; Saunders, Michael R.; Kalb, Rebecca A.; Haulton, G. Scott; Michler, Charles H., Eds. 2013. The Hardwood Ecosystem Experiment: A Framework for Studying Responses to Forest Management. Gen. Tech. Rep. NRS-P-108. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station: 12-23. http://www.fs.usda.gov/treesearch/pubs/42897

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., & Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecological Applications, 8(3), 559–568. https://doi.org/10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2

Caskey, B. J., & Frey, J. W. (2009). Biological-community composition in small streams and its relations to habitat, nutrients, and land use in agriculturally dominated landscapes in Indiana and Ohio, 2004, and implications for assessing nutrient conditions in Midwest streams (Scientific Investigations Report 2009–5055). USGS NAWQA.

Caskey, B. J., Frey, J. W., & Selvaratnam, S. (2010). Breakpoint analysis and assessment of selected stressor variables on benthic macroinvertebrate and fish communities in Indiana streams: Implications for developing nutrient criteria (Scientific Investigations Report 2010–5026). USGS NAWQA.

Clark-Kolaks, S., Gaston, K., & Holloway, D. (2022). White River mainstem 2020: A detailed analysis of the fish community and aquatic habitat. Proceedings of the Indiana Academy of Science, 130(2), 144–155.

Cooper, J. R., Gilliam, J. W., Daniels, R. B., & Robarge, W. P. (1987). Riparian areas as filters for agricultural sediment. Soil Science Society of America Journal, 51(2), 416–420. https://doi.org/10.2136/sssaj1987.03615995005100020029x

Daniels, R. B., & Gilliam, J. W. (1996). Sediment and chemical load reduction by grass and riparian filters. Soil Science Society of America Journal, 60, 246–251.

Decision Innovation Solutions. (2013). 2013 multi-state land use study.

Dewitz, J. (2021). National land cover database (NLCD) 2019 products [dataset]. U.S. Geological Survey. https://doi.org/10.5066/P9KZCM54

DLGF. (2021, May 24). Assessed value search. Indiana Department of Local Government Finance. https://www.in.gov/dlgf/understanding-your-tax-bill/assessed-value-search/

Dodds, W. K., & Smith, V. H. (2016). Nitrogen, phosphorus, and eutrophication in streams. Inland Waters (Print), 6(2), 155–164. https://doi.org/10.5268/IW-6.2.909

Doll, J. C. (2011). Predicting biological impairment from habitat assessments. Environmental Monitoring and Assessment, 182(1), 259–277. https://doi.org/10.1007/s10661-011-1874-4

Dosskey, M. G., Vidon, P., Gurwick, N. P., Allan, C. J., Duval, T. P., & Lowrance, R. (2010). The role of riparian vegetation in protecting and improving chemical water quality in streams. JAWRA Journal of the American Water Resources Association, 46(2), 261–277. https://doi.org/10.1111/j.1752-1688.2010.00419.x

Downing, J. (2012). Global abundance and size distribution of streams and rivers. Inland Waters, 2, 229–236. https://doi.org/10.5268/IW-2.4.502

Dunn, J. B., Merz, D., Copenhaver, K. L., & Mueller, S. (2017). Measured extent of agricultural expansion depends on analysis technique. Biofuels, Bioproducts and Biorefining, 11(2), 247–257. https://doi.org/10.1002/bbb.1750

Dunn, J. B., Mueller, S., & Eaton, L. (2015, April 29). Comments on cropland expansion outpaces agricultural and biofuel policies in the United States. https://greet.es.anl.gov/files/comments-cropland-expansion

Effert-Fanta, E. L., Fischer, R. U., & Wahl, D. H. (2019). Effects of riparian forest buffers and agricultural land use on macroinvertebrate and fish community structure. Hydrobiologia, 841(1), 45–64. https://doi.org/10.1007/s10750-019-04006-1

Evans-White, M. A., Dodds, W. K., Huggins, D. G., & Baker, D. S. (2009). Thresholds in macroinvertebrate biodiversity and stoichiometry across water-quality gradients in Central Plains (USA) streams. Journal of the North American Benthological Society, 28(4), 855–868. https://doi.org/10.1899/08-113.1

Feld, C. K., Fernandes, M. R., Ferreira, M. T., Hering, D., Ormerod, S. J., Venohr, M., & Gutiérrez-Cánovas, C. (2018). Evaluating riparian solutions to multiple stressor problems in river ecosystems—A conceptual study. Water Research, 139, 381–394. https://doi.org/10.1016/j.watres.2018.04.014

Fore, L. S., Karr, J. W., & Conquest, L. L. (1994). Statistical properties of an index of biological integrity used to evaluate water resources. Canadian Journal of Fisheries and Aquatic Sciences, 51(5), 1077–1087. https://doi.org/10.1139/f94-107

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

Gellis, A. C., Fuller, C. C., & Van Metre, P. C. (2017). Sources and ages of fine-grained sediment to streams using fallout radionuclides in the Midwestern United States. Journal of Environmental Management, 194, 73–85. https://doi.org/10.1016/j.jenvman.2016.06.018

GMU CSISS. (n.d.). CropScape—NASS CDL program. Retrieved August 31, 2023, from https://nassgeodata.gmu.edu/CropScape/

Gregory, S., Swanson, F., McKee, W., & Cummins, K. (1991). An ecosystem perspective of riparian zones. Bioscience, 41. https://doi.org/10.2307/1311607

Hoggatt, R. E. (1975). Drainage areas of Indiana streams. United States Geological Survey. https://www.in.gov/dnr/water/publications/drainage-areas-of-indiana-streams/

Hrodey, P. J., Sutton, T. M., Frimpong, E. A., & Simon, T. P. (2009). Land-use impacts on watershed health and integrity in Indiana warmwater streams. The American Midland Naturalist, 161(1), 76–95. https://doi.org/10.1674/0003-0031-161.1.76

Hughes, A. O., & Quinn, J. M. (2014). Before and after integrated catchment management in a headwater catchment: Changes in water quality. Environmental Management, 54(6), 1288–1305. https://doi.org/10.1007/s00267-014-0369-9

Indiana Geographic Information Office. (2019). Parcel boundaries of Indiana 2019: Parcel boundaries maintained by county agencies in Indiana [File Geodatabase; Polygon Feature Class].

Indiana Geological Survey. (2019, April 26). HYDROGRAPHY_LOCALRES_FLOWLINE_UNCLASSIFIED_NHD_IN: Unclassified drainage flowlines in watersheds of Indiana (U. S. Geological Survey, 1:2,400, line shapefile). https://maps.indiana.edu/metadata/Hydrology/Water_Bodies_Flowlines_Unclassified_LocalRes.html

Karr, J., Fausch, K., Angermeier, P., Yant, P., & Schlosser, I. (1986). Assessing biological integrity in running waters. A method and its rationale. III Nat Hist Surv Spec Publ, 5.

Karr, J. R. (1981). Assessment of biotic integrity using fish communities. Fisheries, 6(6), 21–27. https://doi.org/10.1577/1548-8446(1981)006<0021:AOBIUF>2.0.CO;2

Kladivko, E. J., Frankenberger, J. R., Jaynes, D. B., Meek, D. W., Jenkinson, B. J., & Fausey, N. R. (2004). Nitrate leaching to subsurface drains as affected by drain spacing and changes in crop production system. Journal of Environmental Quality, 33(5), 1803–1813. https://doi.org/10.2134/jeq2004.1803

Lark, T. J., Mueller, R. M., Johnson, D. M., & Gibbs, H. K. (2017). Measuring land-use and land-cover change using the U.S. Department of Agriculture's cropland data layer: Cautions and recommendations. International Journal of Applied Earth Observation and Geoinformation, 62, 224–235. https://doi.org/10.1016/j.jag.2017.06.007

Lau, J. K., Lauer, T. E., & Weinman, M. L. (2006). Impacts of channelization on stream habitats and associated fish assemblages in east central Indiana. The American Midland Naturalist, 156(2), 319–330. https://doi.org/10.1674/0003-0031(2006)156[319:IOCOSH]2.0.CO;2

Liu, X., Vidon, P., Jacinthe, P.-A., Fisher, K., & Baker, M. (2014). Seasonal and geomorphic controls on N and P removal in riparian zones of the US Midwest. Biogeochemistry, 119(1), 245–257. https://doi.org/10.1007/s10533-014-9963-4

MacFarland, K., Straight, R., & Dosskey, M. (2017). Riparian forest buffers: An agroforestry practice. USDA. https://www.fs.usda.gov/nac/assets/documents/agroforestrynotes/an49rfb01.pdf

Masek, J., Vermote, E., Saleous, N., Wolfe, R., Hall, F., Huemmrich, K., Gao, F., Kutler, J., & Lim, T.-K. (2006). A landsat surface reflectance data set for North America, 1990–2000. Geoscience and Remote Sensing Letters, IEEE, 3, 68–72. https://doi.org/10.1109/LGRS.2005.857030

Mayer, P. M., Reynolds Jr., S. K., & Canfield, T. J. (2005). Riparian buffer width, vegetative cover, and nitrogen removal effectiveness: A review of current science and regulations. EPA/600/R-05/118.

Meador, M. R., & Frey, J. W. (2018). Relative importance of water-quality stressors in predicting fish community responses in Midwestern streams. JAWRA Journal of the American Water Resources Association, 54(3), 708–723. https://doi.org/10.1111/1752-1688.12646

Moerke, A. H., & Lamberti, G. A. (2006). Scale-dependent influences on water quality, habitat, and fish communities in streams of the Kalamazoo River Basin, Michigan (USA). Aquatic Sciences, 68(2), 193–205. https://doi.org/10.1007/s00027-006-0837-2

Munn, M. D., Frey, J. W., Tesoriero, A. J., Black, R. W., Duff, J. H., Lee, K., Maret, T. R., Mebane, C. A., Waite, I. R., & Zelt, R. B. (2018). Understanding the influence of nutrients on stream ecosystems in agricultural landscapes. In Understanding the influence of nutrients on stream ecosystems in agricultural landscapes (USGS Numbered Series 1437; Circular, Vol. 1437, p. 92). U.S. Geological Survey. https://doi.org/10.3133/cir1437

Naiman, R. J., Décamps, H., McClain, M. E., & Likens, G. E. (2005). Riparia: Ecology, conservation, and management of streamside communities. Elsevier. https://doi.org/10.1016/B978-012663315-3/50008-3

NOAA National Centers for Environmental Information, monthly drought report for annual 2022. (2023, January). https://www.ncei.noaa.gov/access/monitoring/monthly-report/drought/202213

Nowell, L. H., Moran, P. W., Schmidt, T. S., Norman, J. E., Nakagaki, N., Shoda, M. E., Mahler, B. J., Van Metre, P. C., Stone, W. W., Sandstrom, M. W., & Hladik, M. L. (2018). Complex mixtures of dissolved pesticides show potential aquatic toxicity in a synoptic study of Midwestern U.S. streams. Science of The Total Environment, 613–614, 1469–1488. https://doi.org/10.1016/j.scitotenv.2017.06.156

NRCS & Wildlife Habitat Council. (2007). Riparian systems (Fish and Wildlife Habitat Management Leaflet 45). https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs143_010137.pdf

Osborne, L. L., & Kovacic, D. A. (1993). Riparian vegetated buffer strips in water-quality restoration and stream management. Freshwater Biology, 29(2), 243–258. https://doi.org/10.1111/j.1365-2427.1993.tb00761.x

Pavelis, G. A. (1987). Farm drainage in the United States. History, status, and prospects. Miscellaneous publication number 1455. Superintendent of Documents, U. https://eric.ed.gov/?id=ED295043

Pollen-Bankhead, N., & Simon, A. (2010). Hydrologic and hydraulic effects of riparian root networks on streambank stability: Is mechanical root-reinforcement the whole story? Geomorphology, 116(3–4), 353–362. https://doi.org/10.1016/j.geomorph.2009.11.013

Polvi, L. E., Wohl, E., & Merritt, D. M. (2014). Modeling the functional influence of vegetation type on streambank cohesion. Earth Surface Processes and Landforms, 39(9), 1245–1258. https://doi.org/10.1002/esp.3577

Potter, S. R., Andrews, S., Atwood, J. D., Kellogg, R. L., Lemunyon, J., Norfleet, L., & Oman, D. (2006). Model simulation of soil loss, nutrient loss and soil organic carbon. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/ceap/na/?cid=nrcs143_014128

Rankin, E. T. (1989). The qualitative habitat evaluation index [QHEI]: Rationale, methods, and application. Ohio Environmental Protection Agency. https://epa.ohio.gov/static/Portals/35/documents/QHEI_1989.pdf

Reitsma, K. D., Clay, D. E., Clay, S. A., Dunn, B. H., & Reese, C. (2016). Does the U.S. cropland data layer provide an accurate benchmark for land-use change estimates? Agronomy Journal, 108(1), 266–272. https://doi.org/10.2134/agronj2015.0288

Remote Sensing Phenology. (2018, November 27). NDVI, the foundation for remote sensing phenology | U.S. Geological Survey. https://www.usgs.gov/special-topics/remote-sensing-phenology/science/ndvi-foundation-remote-sensing-phenology#overview

Richardson, J. S. (2019). Biological diversity in headwater streams. Water, 11(2), Article 2. https://doi.org/10.3390/w11020366

Richardson, J. S. (2020). Headwater streams. In M. I. Goldstein & D. A. DellaSala (Eds.), Encyclopedia of the World's Biomes (pp. 371–378). Elsevier. https://doi.org/10.1016/B978-0-12-409548-9.11957-8

Richardson, J. S., & Beraud, S. (2014). Effects of riparian forest harvest on streams: A meta-analysis. The Journal of Applied Ecology, 51(6), 1712–1721. https://doi.org/10.1111/1365-2664.12332

Riis, T., Kelly-Quinn, M., Aguiar, F. C., Manolaki, P., Bruno, D., Bejarano, M. D., Clerici, N., Fernandes, M. R., Franco, J. C., Pettit, N., Portela, A. P., Tammeorg, O., Tammeorg, P., Rodríguez-González, P. M., & Dufour, S. (2020). Global overview of ecosystem services provided by riparian vegetation. BioScience, 70(6), 501–514. https://doi.org/10.1093/biosci/biaa041

Robertson, D. M., & Saad, D. A. (2019). Spatially referenced models of streamflow and nitrogen, phosphorus, and suspended-sediment loads in streams of the Midwestern United States (Scientific Investigations Report 2019–5114; Scientific Investigations Report, p. 74).

Ross, S. T., Matthews, W. J., & Echelle, A. A. (1985). Persistence of stream fish assemblages: Effects of environmental change. The American Naturalist, 126(1), 24–40.

Roth, N. E., Allan, J. D., & Erickson, D. L. (1996). Landscape influences on stream biotic integrity assessed at multiple spatial scales. Landscape Ecology, 11(3), 141–156. https://doi.org/10.1007/BF02447513

Shields, F. D., Knight, S. S., & Cooper, C. M. (1995). Use of the index of biotic integrity to assess physical habitat degradation in warmwater streams. Hydrobiologia, 312(3), 191–208. https://doi.org/10.1007/BF00015512

Simon, A., & Collison, A. J. C. (2002). Quantifying the mechanical and hydrologic effects of riparian vegetation on streambank stability. Earth Surface Processes and Landforms, 27(5), 527–546. https://doi.org/10.1002/esp.325

Simon, T. P., & Dufour, R. (1997). Development of index of biotic integrity expectations for the ecoregions of Indiana. V. Eastern Corn Belt Plain (EPA 905/R-96/002). U.S. Environmental Protection Agency.

State of Ohio Environmental Protection Agency. (2006). Methods for assessing habitat in flowing waters: Using the qualitative habitat evaluation index (QHEI). OHIO EPA Technical Bulletin EAS/2006-06-1.

Stauffer, J. C., Goldstein, R. M., & Newman, R. M. (2000). Relationship of wooded riparian zones and runoff potential to fish community composition in agricultural streams. Canadian Journal of Fisheries and Aquatic Sciences, 57(2), 307–316. https://doi.org/10.1139/f99-197

Stone, W. W., & Wilson, J. T. (2006). Preferential flow estimates to an agricultural tile drain with implications for glyphosate transport. Journal of Environmental Quality, 35(5), 1825–1835. https://doi.org/10.2134/jeq2006.0068

Strahler, A. N. (1952). Hypsometric (area-altitude) analysis of erosional topography. GSA Bulletin, 63(11), 1117–1142. https://doi.org/10.1130/0016-7606(1952)63[1117:HAAOET]2.0.CO;2

Strahler, A. N. (1957). Quantitative analysis of watershed geomorphology. Eos, Transactions American Geophysical Union, 38(6), 913–920. https://doi.org/10.1029/TR038i006p00913

Stroud Water Research Center. (2017). Model my watershed [Computer software]. https://wikiwatershed.org/

Sullivan, B. E., Rigsby, L. S., Berndt, A., Jones-Wuellner, M., Simon, T. P., Lauer, T., & Pyron, M. (2004). Habitat influence on fish community assemblage in an agricultural landscape in four east central Indiana streams. Journal of Freshwater Ecology, 19(1), 141–148. https://doi.org/10.1080/02705060.2004.9664521

Talmage, P. J., Perry, J. A., & Goldstein, R. M. (2002). Relation of instream habitat and physical conditions to fish communities of agricultural streams in the northern Midwest. North American Journal of Fisheries Management, 22(3), 825–833. https://doi.org/10.1577/1548-8675(2002)022<0825:ROIHAP>2.0.CO;2

U.S. EPA. (2013, September 3). Indicators used in the national aquatic resource surveys [Collections and Lists]. https://www.epa.gov/national-aquatic-resource-surveys/indicators-used-national-aquatic-resource-surveys

U.S. EPA. (2020). National rivers and streams assessment 2013-2014 report: A collaborative survey (Reports and Assessments EPA 841-R-19-001; National Rivers and Streams Assessment). https://www.epa.gov/national-aquatic-resource-surveys/nrsa

USDA NASS. (n.d.-a). Research and science—Cropland data layer releases. Retrieved March 2, 2022, from https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php

USDA NASS. (n.d.-b). USDA - National Agricultural Statistics Service—Research and science—CropScape and cropland data layers. Retrieved April 28, 2022, from https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#Section1_10.0

USDA National Agricultural Statistics Service Cropland Data Layer. (2020). Published crop-specific data layer [online]. https://nassgeodata.gmu.edu/CropScape/

USDA National Agricultural Statistics Service Cropland Data Layer. (2024). Published crop-specific data layer [online]. https://croplandcros.scinet.usda.gov/

USGS. (2018). The Midwest stream quality assessment—Influences of human activities on streams (Fact Sheet) [Fact Sheet]. USGS.

USGS. (2019). The StreamStats program. https://streamstats.usgs.gov/ss/

USGS. (20190624). CONTOURS_24K_USGS_IN [Vector digital data]. https://gisdata.in.gov/server/rest/services/Hosted/CONTOURS_24K_USGS_IN/FeatureServer/53/metadata?format=default&f=html

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., & Cushing, C. E. (1980). The river continuum concept. Can. J. Fish. Aquat. Sci., 37, 130–137.

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sensing of Environment, 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008

Wang, L., Robertson, D. M., & Garrison, P. J. (2007). Linkages between nutrients and assemblages of macroinvertebrates and fish in wadeable streams: Implication to nutrient criteria development. Environmental Management (New York), 39(2), 194–212. https://doi.org/10.1007/s00267-006-0135-8

Wolf, J. F., Prosser, R. S., Champagne, E. J., & McCann, K. S. (2020). Variation in response of laboratory-cultured freshwater macroinvertebrates to sediment from streams with differential exposure to agriculture. Water, Air, & Soil Pollution, 231(1), 13. https://doi.org/10.1007/s11270-019-4376-6

Woods, A. J. (1998). Ecoregions of Indiana and Ohio [Map]. USEPA, Office of Research and Development ; For sale by U.S. Geological Survey.

Yeakley, A. J., Ervin, D., Chang, H., Granek, E. F., Dujon, V., Shandas, V., & Brown, D. (2016). Ecosystem services of streams and rivers. In D. J. Gilvear, M. T. Greenwood, M. C. Thoms, & P. J. Wood (Eds.), River Science (pp. 335–352). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118643525.ch17

Yuan, Y., Bingner, R. L., & Locke, M. A. (2009). A review of effectiveness of vegetative buffers on sediment trapping in agricultural areas. Ecohydrology, 2(3), 321–336. https://doi.org/10.1002/eco.82