4-2020

Creating a Distribution Model of Invasive Rusty Crayfish (Faxonius Rusticus) in Michigan Streams Using Publically Accessible Data

Robert C. Homan
Western Michigan University, robertcarlhoman@gmail.com

Recommended Citation

Homan, Robert C., "Creating a Distribution Model of Invasive Rusty Crayfish (Faxonius Rusticus) in Michigan Streams Using Publically Accessible Data" (2020). *Master's Theses*. 5136.
https://scholarworks.wmich.edu/masters_theses/5136

CREATING A DISTRIBUTION MODEL OF INVASIVE RUSTY CRAYFISH (*FAXONIUS RUSTICUS*) IN MICHIGAN STREAMS USING PUBLICALLY ACCESSIBLE DATA

by

Robert C. Homan

A thesis submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Master of Science
Geography
Western Michigan University
April 2020

Thesis Committee:

Kathleen Baker, Ph.D., Chair
Brian Roth, Ph.D.
Gregory Veeck, Ph.D.

CREATING A DISTRIBUTION MODEL OF INVASIVE RUSTY CRAYFISH (*FAXONIUS RUSTICUS*) IN MICHIGAN STREAMS USING PUBLICALLY ACCESSIBLE DATA

Robert C. Homan, M.S.

Western Michigan University, 2020

The purpose of this thesis is to create a predictive model of habitat suitability for the invasive rusty crayfish (*Faxonius rusticus*) throughout the state of Michigan. *F. rusticus* often outcompete and extirpate native crayfish species, so understanding their habitats of success is instrumental in monitoring vulnerable ecosystems. Michigan State University and the Michigan DNR conducted extensive field surveys across 461 streams sites from 2014-2016. This project compares this field data set to data from publicly available national datasets with the purpose of revealing the ecosystems most vulnerable to the introduction of *F. rusticus*. The pattern of *F. rusticus* habitat at a local (100 acres) scale and landscape (1000 acres) scale are determined by comparing the current locations of the species in Michigan against a number of variables quantifying the physical geography of the locations that may affect the spread, growth and survivability of these crayfish. The presence of *F. rusticus* is also compared to the presence/absence of other species at each surveyed site. An Artificial Neural Network (ANN) model using variables from Soil Survey Geographic Database (SSURGO) and National Land Cover Database (NLCD) datasets found 45 stream locations vulnerable to *F. rusticus* invasion. This model also determines the variables that have the greatest influence on the model at the center of this research.

ACKNOWLEDGEMENTS

I would like to begin by thanking Dr. Brian Roth and Mr. Kelley Smith of Michigan State University's Department of Fisheries and Wildlife. I would literally not have a thesis if these two men were not generous enough to let me build upon the foundations of their previous research. I am extraordinarily grateful for both of you and hope that this research is beneficial to your work on researching Michigan's invasive crayfish.

Secondly, I would like to thank all the people at Western Michigan University who helped me, directly and indirectly with my work. First up is the chair of my committee, Dr. Kathleen Baker. You taught me how to program, how to handle each step of the thesis procedure and kept me focused and on track. I would never have completed this without you. Next is Dr. Gregory Veeck, also of my thesis committee for his valuable instructions on formulating a thesis and his general support of all graduate students within WMU's Geography Department. Next is Dr. Dave Lemberg who suggested I focus my aquatic invasive species interest around *Faxonius rusticus*. I would also like to thank Dr. Benjamin Ofori-Amoah, Mary Lou Brooks, Dr. Adam Mathews, Dr. Lisa Dechano-Cook, Dr. Charles Emerson, Dr. Li Yang, Dr. Laiyin Zhu, Leesa Jaquays, and the rest of the WMU Geography Faculty for their advice, teachings, aid, and general support. I would also like to thank the other Department of Geography graduate students for their support, advice, and comradery. Thank you Esther Akoto Amoaka, Sami Abdulkarim A. Alwulayi, Nolan Bergstrom, Natalie Bond, Erik Breidinger, Hailey Cantrell, Dennis Donkor, Claire Gilbert, Kevin Haines, Mike Hayosh, Ian Kelley, Eugene Kujo Opare Ahwireng, Connor Landeck, Skye Leake, Graeme McCormick Timmeney, Lydia Serwaah Yeboah, April Shirey, Rajesh Sigdel, Emily Szymanski, Natalie Woloszyn, and anyone else I may have missed.

Acknowledgements – Continued

Third, I would like to thank the Michigan Department of Natural Resources Fisheries Division for their employment, support and aid. They allowed me to find my passion, develop my research GIS skills, and gave me support to complete this endeavor. Thank you, Gary Whelan, Seth Herbst, Jim Dexter, Lynne Thoma, Chris Tom Goniea, Todd Grischke, Ellen Grove, Joel Nohner, Suzanne Stone, Terri Smith, Christian Lesage, Cristel Davis, Cheryl Benson, Christina Baugher, Marissa Lay, Kesiree Thiamkeelakul, and everyone else in the Department. I would also like to thank Jan-Michael Hessenauer, Ph.D., for providing data on several streams of Michigan. I would also like to thank Allen Kei for his video tutorial on the neuralnet tool in RStudio.

Fourth, Dr. Dana Infante of Michigan State University's Department of Fisheries and Wildlife, Kyle Herreman, Arthur Cooper, Jared Ross, Sam Betances, Emily Dean, Erin Tracy, and the rest of and the rest of the ALE (Aquatic Landscape Ecology) Lab. Thank you for providing stream flow information, for helping me learn GIS software, and for helping me apply to graduate school. Their support is invaluable. I would also like to thank Dr. Kim Scribner and Dr. Jeanette Kanefsky of MSU's Dept. of Fisheries and Wildlife's Genetics Laboratory for the laboratory experience and support of my research and further education.

Finally, I would like to thank my parents, David and Cathy Homan, my brother, Steven Homan and his wife, Anna Homan. I would not be the person I am without their love and support.

Robert C. Homan

TABLE OF CONTENTS

CHAPTER

Table of Contents – Continued

CHAPTER

Table of Contents – Continued

APPENDICES

LIST OF TABLES

LIST OF FIGURES

List of Figures – Continued

CHAPTER I

INTRODUCTION – HOW TO COMBAT THE SPREAD OF RUSTY CRAYFISH (*FAXONIUS RUSTICUS*[1])

Rusty Crayfish (*Faxonius rusticus*) are a significant threat to Michigan's freshwater ecosystems. Like the more infamous sea lamprey (*Petromyzon marinus*) and zebra mussel (*Dreissena polymorpha*), *F. rusticus* are an invasive non-native species that was introduced into the ecosystem in an uncontrolled manner. Large, aggressive, and extremely adaptable, *F. rusticus* are difficult to eliminate and capable of extirpating native crayfish species (Olden et al., 2011). Preventative measures need to be taken to protect Michigan's native crayfish populations.

Although the invasive rusty crayfish (*Faxonius rusticus*) has been present in Michigan's waters for over 130 years, little has been done to predict what sites are suitable for the species, and thus, where the species might invade next. The locations of *F. rusticus* populations in Michigan have been collected through state crayfish surveys conducted by Lippson (1975) and Smith et al. (2018), providing data on their population densities, but this research has not yet produced any predictive models. Olden et al. (2011) modeled the vulnerability of Wisconsin's lakes and streams to *F. rusticus* invasions, but as stated, this model was exclusively focused on the water bodies of Wisconsin. The aim of this thesis is to create a predictive model of suitable habitats and potential spread of *F. rusticus* in the State of Michigan.

---

[1] Note: The rusty crayfish (*Faxonius rusticus*), virile crayfish (*Faxonius virilis*), northern clearwater crayfish (*Faxonius propinquus*), and the calico (papershell) crayfish (*Faxonius immunis*), were part of the *Orconectes* genus until they were reclassified as members of the *Faxonius* genus in August 2017 (Crandall & De Grave, 2017). Most of the cited reports and surveys refer to these species by their former names: *Orconectes rusticus*, *Orconectes virilis*, *Orconectes propinquus*, and *Orconectes immunis*, respectively.

A model of site suitability will potentially yield two major research benefits. Similar to Olden et al. (2011), this model will identify the types of waterbodies that are particularly vulnerable to *F. rusticus* invasion. With this information, preventative actions can be taken to help identify ecosystems vulnerable to invasion. The second major benefit is that this model can theoretically identify waterbodies that are less vulnerable to *F. rusticus*. By identifying these waterbodies, these sites can be further explored to determine what factors are present that inhibit the dispersion of *F. rusticus*, potentially highlighting the limiting factors necessary to block their spread.

As Figure 1a illustrates, in the years between 1975 and 2016, *F. rusticus* have expanded their range throughout the state of Michigan (Smith et al., 2018). Before a predictive model can be created, it is necessary to determine the factors that are related to dispersion and growth of *F. rusticus* populations. By analyzing a number of biological, physiological and chemical factors essential to their presence and propagation, this research outlined in this thesis endeavors to determine patterns in the presence of *F. rusticus* and what factors are associated with presence of the species.

The objectives of this project are the following:

- To compare Smith et al.'s (2018) crayfish survey data to the data derived from the publicly available Soil Survey Geographic Database (SSURGO) and National Land Cover Database (NLCD) databases in order to reveal the ecosystems most vulnerable to the introduction of *F. rusticus*.
- To determine habitation characteristics for *F. rusticus* at a local scale (40.4678 hectares or 100 acres) and landscape (404.678 hectares or 1000 acres) scale.

- To compare the presence/absence of other native species of crayfish (*F. propinquus* and *F. virilis*) against these same variables and the presence/absence of *F. rusticus*.

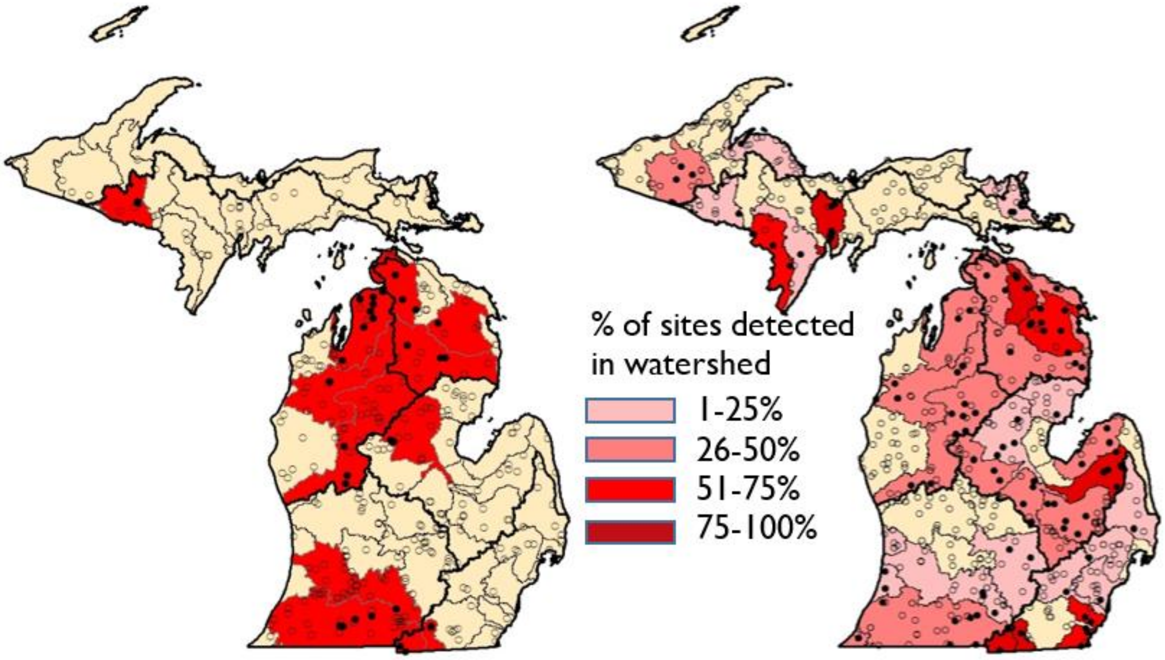

Figure 1.1 – Distribution of *Faxonius rusticus* in 1975 (L) and 2014-2016 (R)
(Detections are not available for the 1975 data due to unknown sites where crayfish were not captured)
Source: Smith et al. (2018)

This thesis will consist of five sections. The first section provides background information on the ecology of *F. rusticus* and Michigan's native crayfish species. The second section is a literature review which will detail the invasion speed of *F. rusticus*, the impact the species has on their adopted ecosystems, and a summary of biological, physiological and chemical factors that may influence their grown and diffusion rates. The third section explains the methods that will be used in this study to determine the ideal habitat of *F. rusticus* and to create a predictive model of their expansion into new habitats and locations. The fourth section is the results of the research and modelling. Finally, the fifth section is a discussion of the implications this research could have on determining the presence/absence of *F. rusticus* and where future research is needed.

Background – Michigan's Crayfish: The Old and the New

Rusty crayfish (*Faxonius rusticus*) (Figure 1b) is a species of crayfish native to western Ohio, Indiana, Kentucky and northern Tennessee, predominantly found in the Ohio River Basin (see Figure 1c) (Minnesota Sea Grant, 2016; United States Geological Survey (USGS) – NAS - Nonindigenous Aquatic Species, 2019). *F. rusticus* require permanent bodies of water for survival, since they live in open water and only create burrows during periods of temperature extremes (Minnesota Sea Grant, 2016; USGS – NAS - Nonindigenous Aquatic Species, 2019). They are opportunistic, omnivorous feeders that subsist on a variety of food sources including, but not limited to: detritus, macrophytes (aquatic vegetation), benthic invertebrates, and fish eggs (Roth et al., 2007). With average carapace lengths greater than 60mm, they are a relatively large species of crayfish and have particularly sizable chelae (claws) (Minnesota Sea Grant, 2016). Combining their body and chelae sizes with their particularly aggressive behavior, *F. rusticus* frequently fight other crayfish species and often force the losing crayfish to relocate to less desirable environments (Garvey et al., 2004). *F. rusticus* are believed to have first passed from Ohio into the waters of Michigan over 130 years ago through the shipping canals connecting watersheds of the Maumee River and the Ohio River (Smith et al. 2018). Given that popular sport fish like smallmouth bass (*Micropterus dolomieu*) prey on *F. rusticus*, it is widely assumed that *F. rusticus* were released into Michigan by fishermen using this species as bait (Roth et al., 2016).

Figure 1.2 – Adult Rusty Crayfish (*Faxonius rusticus*)
(Approximate Size)
Source: Tennessee Wildlife Resources Agency (n.d.)

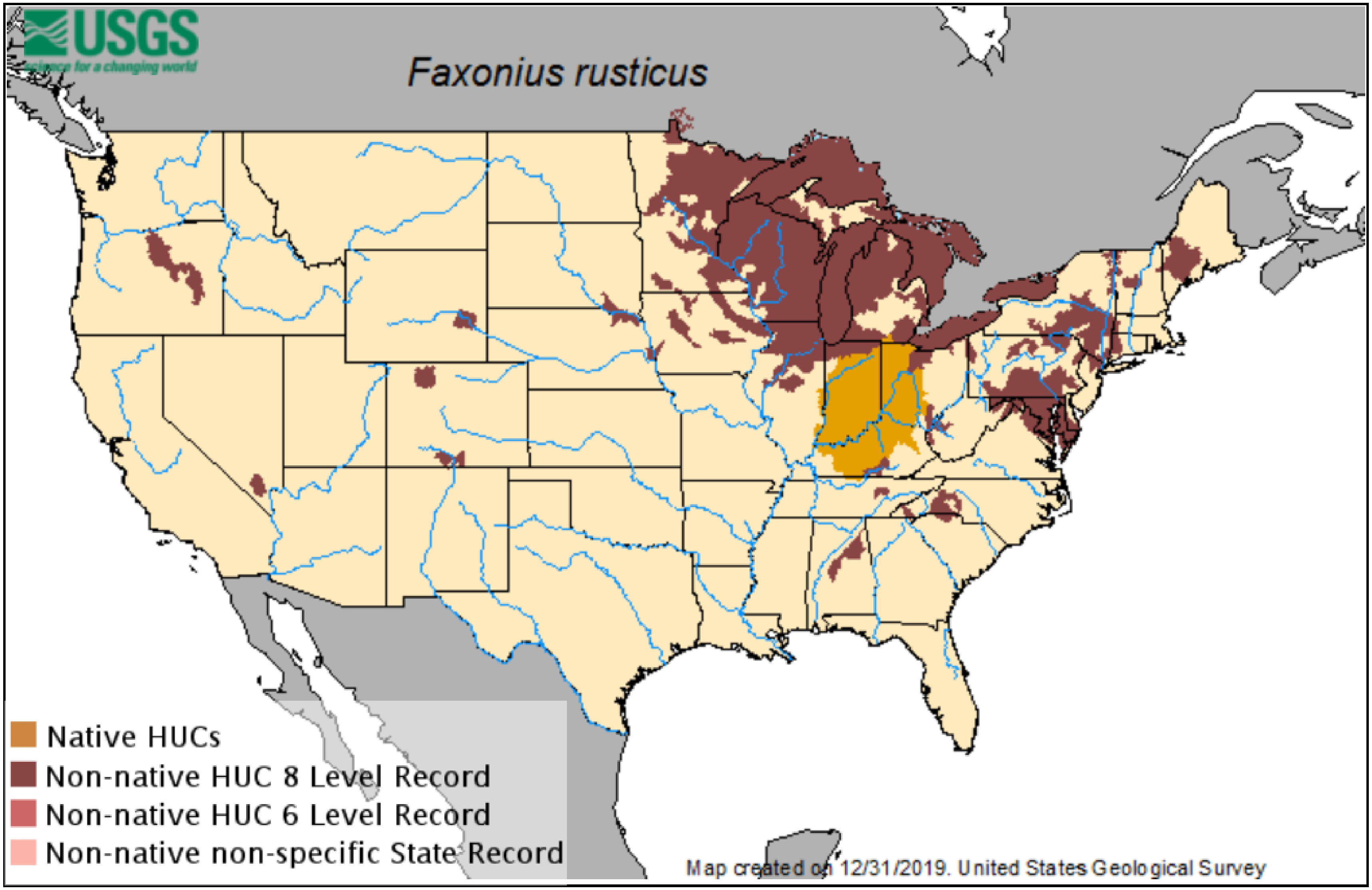

Figure 1.3 – Native Range and Non-native Range of *Faxonius rusticus*
Source: USGS – NAS - Nonindigenous Aquatic Species (2019)

Michigan is home to eight native species of crayfish: the virile crayfish (*Faxonius virilis*), the northern clearwater crayfish (*Faxonius propinquus*), the calico (or papershell) crayfish (*Faxonius immunis*), the big water crayfish (*Cambarus robustus*), the devil crayfish (*Cambarus diogenes*), the paintedhand mudbug (*Cambarus polychromatus*), the digger crayfish (*Creaserinus fodiens*) (often referred to by the synonymized name *Fallicambarus fodiens* (Crandall & De Grave, 2017)), and the white river crayfish (*Procambarus acutus acutus*). Of these eight species, *F. virilis* and *F. propinquus* are the most commonly occurring species in the state, and along with *C. robustus*, typically reside in open water areas of lakes and streams (Roth et al., 2016). *C. diogenes*, *C. polychromatus* and *Creaserinus fodiens* are burrowing species of crayfish, normally residing in the chimney-shaped burrows they dig (Roth et al., 2016). Due to their reclusive lifestyle, they are rarely observed or captured by humans. *F. immunis* and *P. acutus* reside in open water, but will also dig and reside in burrows during certain environmental periods (i.e. such periods where their waterbody of residence dries out) (Missouri Department of Conservation, n.d.; Roth et al., 2016). *F. virilis* and *F. propinquus* are the species of crayfish most commonly found competing with *F. rusticus*, since all three species prefer open water habitats.

CHAPTER II

LITERATURE REVIEW: SPREAD, IMPACT, ADAPTABILITY AND LIMITING FACTORS OF RUSTY CRAYFISH (FAXONIUS RUSTICUS)

2.1: Introduction

This literature review summarizes scholarship for five areas of research pertaining to the proliferation, physical properties, ecological limits, and behavior of the invasive rusty crayfish (*Faxonius rusticus*), along with their impacts on non-native aquatic communities. The first section will detail the spread and invasion patterns of *F. rusticus*. The second section will highlight the changes *F. rusticus* inflict on aquatic ecosystems. The third will report on both the interactions between *F. rusticus* and native crayfish species and the habitat preferences of *F. rusticus* and native crayfish species, including information of their preferred substrates. In previous research, these topics were frequently covered together so it makes sense to discuss them in the same section. The fourth section will summarize the thermal preferences and tolerances of *F. rusticus*. Finally, the fifth section will examine the effects that non-temperature related physiological and ecological properties of aquatic ecosystems (calcium, water velocity, etc.) have on *F. rusticus*.

2.2: The Spread of Rusty Crayfish (*Faxonius rusticus*)

Rusty crayfish (*Faxonius rusticus*) are believed to have been originally introduced into Michigan by anglers as bait. Anglers either deliberately or unintentionally released them into Michigan waters (Roth et al., 2016). After *F. rusticus* established themselves in their new habitat, they proceeded to diffuse into surrounding ecosystems. The majority of the research on the spread of *F. rusticus* has been performed in Wisconsin and thus this section will be primarily focused on Wisconsin-based studies, though the first literature examined will pertain to Michigan, the second portion pertains to Illinois, and the final section pertains to the Laurentian Great Lakes region. This section will present some of the most salient research analyzing the distribution and potential spread of *F. rusticus*.

From June 1994 through October 1995, 102 sites in Illinois were sampled for *F. rusticus* using 3.1-m minnow seines to determine their range and population density so these findings could be compared to a previous survey performed from 1972-1982 (Taylor & Redmer, 1996). *F. rusticus* were found in thirty-nine of the Illinois sites, including thirty-two where they were previously absent, and were found primarily in isolated segments of the waterbodies they inhabited, with population densities of 0.4-6.2 individuals/m^2 (Taylor & Redmer, 1996).

A data set containing 2,775 crayfish locality records for the state of Wisconsin was created, with archived information gathered from various historical records and 2004 data obtained from live sampling at 251 stream and lake segments (Olden et al., 2006). All survey locations were mapped using ArcGIS Version 8.3 and assigned to one of three time frames: 1870-1964 (pre-invasion), 1965-1984 (early post-invasion), and 1985-2004 (extant) (Olden et al., 2006). The number of recorded sightings of *F. rusticus* were found to have increased from 7% of the total crayfish recorded in the early post-invasion time period to 36% during the extant time period (Olden et al., 2006). By the last period, *F. rusticus* were present in 86% of Wisconsin's watersheds and most frequently recorded in the Northern Highland Lake-District (Olden et al., 2006). The ability of *F. rusticus* to increase their population and range by such a large extent in such a short time period demonstrates the necessity for determining their invasion pathways.

A prediction model for *F. rusticus* occurrence was created for the State of Wisconsin. Olden et al. (2011) evaluated ecosystem sensitivity of Wisconsin's freshwater ecosystems to rusty crayfish invasions, creating a database of Wisconsin's crayfish occurrences as a first step. Variables influencing *F. rusticus* introductions and success were assessed and a multi-response artificial neural network (MANN) was used to model crayfish occurrences for *F. rusticus*, virile crayfish (*Faxonius virilis*) and northern clearwater crayfish (*Faxonius propinquus*) (Olden et al., 2011). Vulnerability was measured as the product of the risk of exposure to *F. rusticus*, the probability of native species occurrence and sensitivity. Through this, the MANN predicted that 388 of the 4,200 surveyed lakes and 23,523 or the 90,360 stream kilometers of Wisconsin were vulnerable to the introduction and establishment of *F. rusticus* and 115 lakes and ~5,000 stream kilometers were vulnerable to the both of these factors as well as the extirpation of *F. virilis* and/or *F. propinquus* (Olden et al., 2011).

Nine (9) variables likely determining crayfish occurrences were selected for the lakes in the survey. Since human activity has been determined to be a primary facilitating variable enabling the spread of rusty crayfish (Capelli & Magnuson, 1983), the first two variables selected were the number of boat landings in the lakes and the lakes' relative human accessibility. Johnson et al. (2008) found a strong positive relationship ($r = 0.69$, $P < 0.01$) between the number of boat landings in the Wisconsin lakes and the amount of human visitation, so that was a well selected variable. The lakes were divided into three groups to determine their accessibility: easily accessible due to public boat landings or easy to navigate waters ("boat access"), only accessible via wilderness trails ("trail access") or inaccessible ("no access") (Olden et al., 2011). The other seven (7) variables all relate to the physical and hydrological properties of the lakes in the survey. These variables are; the lake's surface area (km^2); its shoreline perimeter (km); its maximum depth (m); the lake's mean summer transparency (m); its specific conductance (μmho/cm), which indicates the interconnectivity of lakes; hydrologic type (seepage with no connecting streams (0) or drainage with connecting streams (1)); and waterbody type (natural lake (0) or impoundment (1)) (Olden et al., 2011). Olden et al. (2011) choose to ignore the critical thresholds (the lowest/highest amount of the chemical/property that is necessary is necessary for an organism to survive in the environment) of dissolved calcium (>2-3 mg/L) and pH (>5.5), theorizing that the majority of Wisconsin's lakes are above these values (Olden et al., 2011). That decision may be worth revisiting, since not every lake exceeds these thresholds and lakes falling below these numbers cannot sustain crayfish life, which potentially could lead to incorrectly predicting *F. rusticus* vulnerability in lakes where *F. rusticus* literally cannot survive.

Eleven (11) variables were chosen for streams as predictors for rusty crayfish occurrence. These variables are: the channel order; the channel slope or gradient (in degrees); the stream baseflow (m^3/s); the amount of agricultural row crops in the upstream riparian area (URA); the amount of urban land use in the URA; the amount of forested wetlands in the URA; the amount of calcium rich carbonate bedrock (i.e. limestone, dolomite, etc.) in the URA; the amount of lacustrine clay and silt in the URA; the mean soil permeability (mm/hour); mean annual air temperature (°C); and mean precipitation from 1960-1990 (mm). The variables occurring in the URA were selected due to being strong predictors of calcium and pH concentrations.

Only four of the predictor variables for determining the likelihood of rusty crayfish occurrence in lakes (the number of boat landings, the hydrologic type (drainage with connecting streams), the waterbody type (impoundment), and maximum depth) and five for streams (the channel gradient, the stream baseflow, the amount of agricultural row crops in the URA, the amount of urban land use in the URA, and the amount of carbonate bedrock in the URA) were found to be significant (P value of less than 0.05). The channel gradient was found to be a negative contribution to rusty crayfish presence, while the other eight significant predictor variables all had positive contributions. Positive contribution for hydrologic type means that rusty crayfish are more likely to occur in drainage with connecting streams than seepage with no connecting streams. Positive contribution for waterbody type means that rusty crayfish are more likely to occur in impoundments than natural lakes. The relative importance of the significant predictor variables for lakes (from highest to lowest) were: landings, hydrologic type, waterbody type, and maximum depth. For streams, the relative importance was as follows: the amount of agricultural row crops in the URA, the channel gradient, the amount of urban land use in the URA, the amount of carbonate bedrock in the URA, and the stream baseflow. To put it another way, drainage impoundments with deep water, with a high number of boat landings and low-gradient channels, with a high baseflow, draining watersheds containing high concentrations of agricultural row crops, urban land use (increasing the likelihood of crayfish release due to bait, pet trade, etc.) and carbonate bedrock (providing calcium) were the habitats most vulnerable to rusty crayfish invasions (Olden et al., 2011).

Given that human actions are primarily responsible for the introduction of *F. rusticus*, it stands to reason that *F. rusticus* would have an easier time establishing themselves at impoundments, due to the relative amount of human activity. In addition to Olden et al. (2011) this theory has been supported by the work of Johnson et al. (2008) and Havel et al. (2015). Johnson et al. (2015) found dam impoundments to be 2.5 times more likely to have *F. rusticus* populations than natural lakes. This conclusion was reached after an experiment was conducted using information on water physiochemistry (i.e. conductance, depth, surface area), boating activity, and distance in regards to five aquatic invasive species (AIS) for 4,200 lakes and 1,081 impoundments in Wisconsin and the Upper Peninsula of Michigan to determine whether dams help invasive species get a foothold in freshwater environments (Johnson et al., 2008). Two-tailed Pearson X^2 and logistic regression were used predict the probability of invasion at any given location (Johnson et al., 2008). This finding is also corroborated by the findings of Havel et al. (2015) who warn about the vulnerability of reservoirs to aquatic invasive species (AIS), including *F. rusticus*, while highlighting how the high connectivity of reservoirs to other waterbodies undeniably aids in the spread of AIS.

Detecting the spread of *F. rusticus* (and other invasive crayfish species) was greatly improved through environmental DNA (eDNA) detection techniques developed by Dougherty et al. (2016) and Larson et al. (2017). Crayfish samples were collected at twelve lakes in Vilas County, Wisconsin, and Gogebic County, Michigan, to determine the population density of *F. rusticus*. Water samples were also taken from each location at the same time. After developing a primer for detecting the eDNA of *F. rusticus* and testing it against a control group of samples, the collected water samples from the twelve lakes were tested for rusty crayfish eDNA (Dougherty et al., 2016). eDNA for *F. rusticus* was detected at eleven of the twelve sites, including two sites where *F. rusticus* were not collected during the sampling (Dougherty et al. 2016. When rusty crayfish are less abundant, eDNA can be more easily detected in clearer water (Dougherty et al., 2016). In a later replication of this work, Larson et al. (2017) correctly detected eDNA of *F. rusticus* in the Laurentian Great Lakes.

Peters et al. (2014) looked at the spread of *F. rusticus* throughout the Laurentian Great Lakes from 1882-2008 by using inclusion and proportion curves. These expansion rates were compared to native crayfish species (such as *F. virilis*) in both Lake Michigan, Lake Huron and

Lake Erie (where *F. rusticus* have resided for over 100 years and may be a native species). Peters et al. (2014) found that *F. rusticus* started their Great Lakes expansion (excluding Lake Erie) after 1980 and spread more rapidly in Lake Huron and Lake Michigan than Lake Erie, despite having resided in Lake Erie for a longer period of time.

The most comprehensive research on the change of *F. rusticus* species distribution in Michigan (and the source of Michigan crayfish information that this thesis will be examining in depth) was performed by Smith et al. (2018). From 2014-2016, the researchers examined 966 stream sites, with 2 sites per stream section throughout the entire state of Michigan to determine the distribution of native and non-native crayfish species and how these locations compare to Michigan’s 1975 crayfish sampling records. The substrate composition of each site was also recorded (Smith et al., 2018). Smith et al. (2018) used the following occupancy model to estimate the species residing in the location:

$$L(\varphi,p) = (\varphi^{n.}\Pi p_t^{nt}(1 - p_t)^{n.-nt}) \text{ x } (\varphi\Pi(1 - p_t) + (1 - \varphi))^{N-n.}$$

where t is the number of searchers at a site, N is the total number of sites surveyed, and n is the number of sites where at least one detection occurred, φ is the probability of occupancy, p is the detection probability for a single searcher, and n_t is the number of detections on tth survey.

This model assumes that if a given species was present at one sub site, it will be present at the other; that the species were correctly identified; and that no species moved into or out of the sampling site between surveys. *F. rusticus* were found to have expanded their range from 12 of Michigan’s HUC 8 watersheds to 34 (including the previous 12) and were found in 20% of the sampled locations.

2.3: The Effects of Rusty Crayfish (*Faxonius rusticus*) on Aquatic Ecosystems

Roth et al. (2006) created a bioenergetic model to determine if and how the growth of *F. rusticus* constrain feeding rates and prey preference while determining trophic levels of *F. rusticus* and their prey by using stable carbon and nitrogen isotopes. Results from over 6,000 bioenergetic model simulations, along with stable isotope predictors, showed that *F. rusticus* were primarily predators that target energy-rich prey such as Ephemeroptera (mayflies), but will supplement their diet with large amounts of energy-poor prey if their preferred prey is unavailable (Roth et al., 2006). This predation is indicative of the powerful changes *F. rusticus* can inflict on indigenous species, primarily macroinvertebrates.

To determine the effects *F. rusticus* have on the abundance of macroinvertebrates and periphytons, an enclosure-exclosure experiment was performed in July and August 1992, when fifteen enclosures were placed in a site downstream of Bond Falls in the Middle Branch Ontonagon River and given twenty-two days to grow algae. On the twenty-second day (July 24), either one, two or no *F. rusticus* specimens were added to each enclosure (five each) to simulate low, high and non-existent *F. rusticus* densities (Charlebois & Lamberti, 1996). On August 10, the crayfish stomach contents were analyzed and MANOVA was used to analyze and compare populations of macroinvertebrates, periphyton chlorophyll ɑ, and the periphyton biomass found in the enclosures. During June and July, 1992, a longitudinal study was performed at four sites downstream of Bond Falls, each with different *F. rusticus* densities so as to simulate densities naturally decreasing with distance downstream (Charlebois & Lamberti, 1996). Crayfish density was determined by counting all *F. rusticus* within 4 replicate 0.5- m-wide belt transects aligned perpendicular to stream flow (Charlebois & Lamberti, 1996). Fifteen rocks were chosen from each site and the periphyton chlorophyll ɑ and biomass, macroinvertebrate abundance and taxa richness were analyzed for each rock using multiple ANOVAs (Charlebois & Lamberti, 1996). *F. rusticus* densities were found to have a direct and significant negative effect invertebrate taxa and densities, while periphyton responded positively to higher rusty crayfish densities (Charlebois & Lamberti, 1996).

Bobeldyk and Lamberti (2008) followed up on the research and research design of Charlebois and Lamberti (1996) by conducting a longitudinal study in 2004 at three of the previously examined sites (one with high *F. rusticus* density, one with intermediate *F. rusticus* density and one with no *F. rusticus*). To determine the effects of *F. rusticus* on benthic invertebrate densities, two packs of leaves, one that crayfish could access and one they could not, were placed at each site and allowed to decompose for 28 days, after which the samples, along with periphyton taken from the cobble, were analyzed using two-way ANCOVA and one-way ANOVA (Bobeldyk & Lamberti, 2008). As with the Charlebois and Lamberti (1996), the results suggest that the presence of *F. rusticus* substantially reduces invertebrate densities in both cobble and leaf packs, as the site lacking crayfish had 2.5 times more benthic invertebrates than either site where *F. rusticus* were present, although periphyton biomass remained unchanged in the Bobeldyk and Lamberti (2008) experiment.

Similarly, to determine the effects that rusty crayfish densities have on macroinvertebrate populations, six enclosures were placed in Charlotte Creek, in central New York, during the summers of 2007 and 2010, with two enclosures with no crayfish, two enclosures with a low *F. rusticus* density (three crayfish), and the final two having a high density of *F. rusticus* (eight crayfish). After three weeks, macroinvertebrate samples were taken from each enclosure, along with three samples taken directly from the stream (no enclosure) as a control. The macroinvertebrate taxa were sorted into feeding groups and species diversity after collection, and the number of taxa, the proportion and density of each taxa, and the total densities of macroinvertebrates was calculated and compared with crayfish density using MANOVA. An experiment with the same procedures was also performed in the summer of 2012 with two tanks containing no crayfish, two containing four *F. rusticus*, and four containing four clearwater crayfish (*Faxonius propinquus*). For the *F. rusticus* specific density experiments, no crayfish-density effect was found, but an enclosure effect where the median macroinvertebrate density decreased as the *F. rusticus* density increased was found to be statistically significant. *F. propinquus* and *F. rusticus* were found to have similar effects on macroinvertebrate populations (Kuhlmann, 2016). This is backed up by McCarthy et al. (2006) which found that populations densities of total zooplankton and the orders of Gastropoda, Diptera and Ephemeroptera were significantly reduced by invasive *F. rusticus*.

Macrophytes are not the only organisms effected by the invasive *F. rusticus*. Wilson et al. (2004) estimated *F. rusticus* diffusion in Trout Lake, Wisconsin, which was measured by trapping crayfish at approximately 73 sites throughout the lake one per year from 1983-1999 and recording their abundance and spread. The population densities of crayfish (all species), fish, macroinvertebrates, snails and macrophytes were also recorded from Lodge-Kratz datasets and through samplings made by North Temperate Lakes Long-Term Ecological Research (NTL-LTER) (fish and macroinvertebrates were measured once a year at six specific sites from 1981-1999 (possibly 1982 for macroinvertebrates), macrophytes were sampled once a year at four specific sites from 1982-1999, and snails were sampled in 1984 and 1995-2000) (Wilson et al., 2004). Results indicate that since arriving in 1979, rusty crayfish spread along the entire shoreline of Trout Lake by 1997, and the absolute population of all crayfish (all species combined) increased yearly despite the *F. rusticus* nearly eradicating the native species (Wilson et al., 2004). This was because the rusty crayfish population exceeded the numbers of native crayfish lost (Wilson et al. 2004). Due to the population increase of *F. rusticus*, snail abundance fell to catastrophic numbers and macrophyte species richness dramatically declined (Wilson et al., 2004). Bluegill and pumpkinseed sunfish populations also declined, possibly due to reduced macrophyte presence or predation on their eggs by *F. rusticus* (Wilson et al., 2004). Through ANOVA, it was determined that macroinvertebrate abundance did not significantly correlate with *F. rusticus* population increases (Wilson et al., 2004).

Roth et al. (2007) ran three studies to determine the relationship between *F. rusticus*, macrophytes, and *Lepomis* species (bluegill and pumpkinseed fish), continuing Wilson et al.'s (2004) long-term study for the years 2001 to 2004, by conducting a comparative study from the summers of 2001 to 2004 assessing the relative abundance of crayfish, fish and macrophytes in fifty seven lakes in Vilas County, Wisconsin, and comparing the predation rates of *Lepomis* species on *F. rusticus* populations in Wisconsin's Arrowhead, Big, Trout and Wild Rice Lakes. Both the long-term and comparative studies show that as rusty crayfish populations increase, macrophyte cover and dry mass (mass of dried macrophytes) declined as did catches of all *Lepomis* species (Roth et al., 2007). The results of the predation study indicated that the abundance of *F. rusticus* was inversely proportional to the abundance of *Lepomis* species in many lakes (Roth et al. 2007). Although the dominant theory is that the decline of cover occupied by macrophytes leads to the predation of *Lepomis* eggs, the causes for the decline of *Lepomis* in areas with high *F. rusticus* populations is unclear and more research is needed (Roth et al., 2007). As *Lepomis* are considered predators of juvenile *F. rusticus*, it would stand to reason that high abundance of *Lepomis* would negatively affect *F. rusticus* populations (Hein et al., 2006; Roth et al., 2007).

Predation has been shown to actively decrease *F. rusticus* populations under certain conditions. From 2001-2003, Hein et al. (2006) depleted the *F. rusticus* population in Sparkling Lake, Wisconsin, by 55% through trapping in tandem with predation by smallmouth bass (*Micropterus dolomieu*) and rock bass (*Ambloplites rupestris*). On average, the fish removed ~247 kg more *F. rusticus* beyond just trapping, but predation by fish primarily targeted large crayfish, so trapping was more effective at removing all crayfish of all ages and at all reproductive states.

2.4: Interactions with Native Species and Habitat Preference of Rusty Crayfish (*Faxonius rusticus*)

A considerable portion of previously identified research is focused on the displacement and extirpation of native crayfish by *F. rusticus*. Reid and Nocera (2015) studied the effects that *F. rusticus* were having on native crayfish assemblages in Ontario. This study incorporated sampling crayfish across ninety-nine sites in thirteen southeastern Ontario watersheds from July 25 to October 5, 2011 and again from May 30 to October 16, 2012, with 200 minutes of crayfish sampling taking place at each site. For each site, water temperature, conductivity, water clarity and channel width were recorded, along with a habitat quality assessment determined by the Qualitative Habitat Evaluation Index (QHEI). The non-parametric Mann-Whitney test was used to compare native crayfish catches from sites where *F. rusticus* were present or, alternatively, absent. In the same study, *F. rusticus* and northern clearwater crayfish (*Faxonius propinquus*) co-occurrence patterns were analyzed using the C-score index, and MANOVA to identify tested habitat differences in sites with and without *F. propinquus*. *F. propinquus* were the most abundant species making up 56% of the collected crayfish and occurring at 64% of sampled sites compared to *F. rusticus,* which constituted 39% of collected crayfish, but eres found at 41% of sites. However, at sites where both species were present, *F. propinquus* comprised only 12.2% of the captured crayfish compared to 72.5% when *F. rusticus* were not present.

Crayfish diffusion and abundance from 1972-77 for sixty-seven lakes in Vilas County, Wisconsin, were recorded and compared with a series of independent variables including: substrates, human activity, geographic isolation, lake size and interspecific competition using multiple stepwise multiple regression. Larger lakes are more likely to have crayfish than smaller lakes, but lake size does not influence the "species richness" (number of different species). Rusty crayfish were the only species that were found to have a strong, significant relationship to human interaction and geographic isolation, yet this study also registered an insignificant correlation between substrate types and abundance (Capelli & Magnuson, 1983).

Smith et al. (2018) showed the changes in the ranges of native crayfish population in Michigan from 1975 to 2016. As previously stated, *F. rusticus* expanded their range from 12 HUC 8 watersheds to 34 and were found in 20% of the total samples. During the same period, *F. propinquus* range expanded from 42 HUC 8 watersheds to 45, but were only found in 42% of samples as opposed to 72% of samples surveyed in in 1975 (Smith et al., 2018). *F. virilis* increased its range to from 33 watersheds to 43 and maintained a consistent 27% presence. Only 23% of sites containing *F. rusticus* also contained *F. propinquus* (as opposed to 43% in 1975). The cooccurrence of *F. rusticus* and *F. virilis* was relatively stable with 16% cooccurrence in 1975 and 15% in 2016 (Smith et al., 2018).

In 2006, estimates of the abundance and habitat association of rusty (*F. rusticus*), virile (*F. virilis*) and papershell crayfish (*Faxonius immunis*) within Lake of the Woods, Ontario, were made by catching samples, while at the same time noting temperature, substrate composition, population densities, depth and presence of macrophytes. These results were compared to historical records of distribution going back as far as 1963 (Jansen et al., 2009). Findings indicate that *F. virilis* remain the most abundant species, but *F. rusticus* drastically increased population densities between 1963 and 2006, expanding their range by an average of 2.1 km per year (Jansen et al., 2009). For each species, populations were not consistently found on any particular substrate, but increasing depth, corresponding with decreased temperature, acted as a natural barrier for the invasive *F. rusticus* and *F. immunis* (Jansen et al., 2009).

F. rusticus may not always display a substrate preference, but does influence the substrate preference of co-occurring crayfish. When *F. rusticus*, which have a positive association with cobble and a negative association with sand, co-occurred with native crayfish species, the indigenous crayfish associated with different substrates than in areas where rusty crayfish were absent (Smith et al., 2018). The cobble-favoring *Cambarus robustus* shifted to woody debris and *F. immunis*, which previously were positively associated with clay and silt, was now positively associated with live vegetation and negatively associated with sand and pebbles (Smith et al., 2018). *F. propinquus* went from associating with cobble and sand to pebbles and live vegetation (and gained a negative association with silt), while *F. virilis* inhabited sand despite having a negative association to it when *F. rusticus* were not present (Smith et al., 2018). This corresponds with findings from the work of Taylor and Redmer (1996), who note a strong positive association between the dominance of *F. rusticus* and cobble habitats.

Substrate preference was not the only aspect of habitat favorability examined by researchers when assessing the interactions between native crayfish and *F. rusticus*. Northern clearwater (*F. propinquus*), virile (*F. virilis*) and *F. rusticus* were captured in northern Michigan waters and, from July 4 through August 12, 2007. Keller & Hazlett (2010) tested their thermal preferences by placing each crayfish in the middle of an aquarium where the ends were heated and cooled to 22°C and 18°C respectively. The sections of the tanks were measured as one of five sections and each crayfish's preferred thermal section was recorded and later compared using ANCOVA. *F. rusticus* were found to prefer 22°C water, which is 1°C higher than the other two species which interestingly correlates with the average summer temperature of its habitat (Keller and Hazlett, 2010).

2.5: The Effects of Temperature on Rusty Crayfish (*Faxonius rusticus*)

Temperature may be a key component in facilitating the invasion of *F. rusticus*. Keller and Hazlett (2010) determined preferred temperature of *F. rusticus* in northern Michigan waters, and follow-up questions included how *F. rusticus* is affected by non-preferred water temperature and if temperature can be a limiting factor?

In order to establish if temperature can be a potential limiting factor for *F. rusticus*, the critical thermal minimum and maximum (CTMin and CTMax, respectively), or highest and lowest temperatures that they can reside in before death, needed to be established. Thankfully, ample research has been published on this topic. The preferred method of determining CTMax and CTMin is to place *F. rusticus* into a bowl of water with the same temperature as they are acclimated to and either heating (for CTMax) or cooling (for CTMin) the water at a rate of 0.5-0.8°C per minute until the crayfish are unable to right themselves if they are flipped over (Layne et al., 1987; Mundahl, 1989; Mundahl & Benton, 1990).

Layne et al. (1987) pioneered this experiment, collecting *F. rusticus* in Indian Creek, Butler County, Ohio, throughout the year of 1984, either testing temperatures directly from the stream or acclimating the crayfish to a temperature of either 5°C or 25°C, and then performing the CTMax and CTMin determination. They concluded this project by analyzing the results using ANOVA to compare temperature tolerances throughout the year. Both tests were performed on the crayfish in 25°C water at different times of the day to account for the influence of diel cycle on the critical thresholds. Crayfish collected in the summer were found to have a higher CTMax than crayfish collected in the winter even with temperature acclimations, whereas the CTMin was adjusted seasonally (Layne et al., 1987).

Mundahl (1989) studied changes in the CTMax in adult *F. rusticus* from the previously mentioned Dicks Creek, Harker's Run and Indian Creek collected during the spring, summer and winter of 1988. The crayfish were placed in water the same temperature as the creek they were taken from. A 1.0°C in water temperature corresponded to 0.38°C change in CTMax for a mean increase of 10.8°C (Mundahl, 1989).

Mundahl (1989) also investigated the effect the diel changes (changes throughout a 24-hour period) had on CTMax of juvenile *F. rusticus*. For thirty hours in July 1988, juvenile rusty crayfish were collected from a stream and placed in a tank (with sheltering rocks) which fluctuated in temperature throughout the day while their reactions were recorded in order to observe the effects of diel changes in juvenile rusty crayfish (Mundahl, 1989). The CTMax for juveniles is lowest in the morning and highest in the afternoon and evening, altering their mean CTMax by 1.2°C (Mundahl, 1989). Juveniles also burrow themselves in the sand and gravel when temperatures are dangerously high (Mundahl, 1989). Layne et al. (1987) found no relationship between the diel cycle and either critical threshold.

Mundahl and Benton (1990) performed a CTMax and CTMin comparison between adult and juvenile *F. rusticus* collected from three creeks in Butler County, Ohio (Dicks Creek, Harker's Run and Indian Creek), during June, July and August, 1988. Juveniles had a CTMax

0.9 to 2.6°C higher than the adults and a CTMin 0.5 to 1.5°C higher than the adults (Mundahl & Benton, 1990).

Westhoff and Rosemberger (2016) reviewed 56 pertinent studies to determine the temperature tolerance (upper limit, lower limit and breadth), temperature preference and optimal growth temperature of freshwater crayfish, specifically focusing on these three parameters. They summarize the information found in the studies (methodology, temporal trends, taxa covered, etc.), estimate the highest and lowest mean observed temperatures for each species and each parameter, and use linear-regression to examine the relationship between absolute latitude and each temperature metric and between range extent and temperature tolerance breadth (Westhoff & Rosemberger, 2016). Of all species, *F. rusticus* was found to have the highest CTMax range with 27.4-41.5°C of any species. The CTMin of *F. rusticus* was 0.9°C (Westhoff & Rosemberger, 2016).

F. rusticus have a high CTMax, but that does not necessarily mean that they thrive at the upper and lower points of their temperature tolerances. In another study, the effect of temperature on the growth rate of juvenile *F. rusticus* was determined by collecting samples in November and December 1988, when the water was 2.5-5.0°C. The samples were divided into four groups. Each group was placed in a tank of 16, 20, 25 and 29°C water, respectively, for twenty-nine days, and the amount of growth, as compared to when they were first captured, was recorded (Mundahl & Benton, 1990). Temperature preferences for adults acclimated in a lab and juveniles acclimated to a stream were determined by dividing a tank into twelve equal sections of increasing temperature, placing the crayfish in the tank in the temperature section they were acclimated to, then observing the movement patterns of the crayfish for 10 minutes, and recording the section they selected at 15 second intervals (Mundahl & Benton, 1990). Growth rates (length grown and weight gained) increase as the temperature increases, with the lowest at 16°C and highest at 25 and 29°C, but mortality following molting also increases with higher temperatures, especially over ≥25°C. The lab-acclimated crayfish preferred higher temperatures than the field-acclimated crayfish (Mundahl & Benton, 1990).

Higher temperatures may lead to higher mortality, but low temperatures may also prove a key to limiting the spread of *F. rusticus*. All crayfish, including *F. rusticus*, are coldblooded and thus are less active in colder temperatures. The previously mentioned Jansen et al. study (2009) captured no *F. rusticus* in water deeper than 8.5m, indicating that depth, corresponding with decreased temperature, acted as a natural barrier for the invasive *F. rusticus* and *F. immunis* (Jansen et al., 2009). Unfortunately, Jansen et al. (2009) conclude that this water depth barrier is more likely to simply slow the spread of *F. rusticus* than to stop it.

2.6: Physiological and Biological Factors Impacting the Proliferation and Growth of Rusty Crayfish (*Faxonius rusticus*)

One of the reasons that rusty crayfish (*F. rusticus*) diffusion has been so successful is that the species appears highly adaptable. Sargent and Lodge (2014) demonstrate this hardiness in their research to determine if invasive *F. rusticus* have higher growth and survival rates than *F. rusticus* in their native habitats. In the summer of 2011, young-of-the-year (YOY) crayfish from mothers born in both native and invaded habitats were housed within three invaded Wisconsin lakes with varying densities of *F. rusticus*. Their growth and mortality rates were recorded and analyzed using ANOVA (Sargent & Lodge, 2014). In the summer of 2012, rusty crayfish from mothers born in both native and invaded habitats were housed in mesocosms where their growth and survival rates were recorded and measured with a linear mixed effect model, to determine what factors influence growth and survival rates and to determine if those rates are based on genetics (Sargent & Lodge, 2014). The invading *F. rusticus* were found to have faster growth rates than the native *F. rusticus* as well as higher survival rates in all locations except those with poor food quality and no fish present (Sargent & Lodge, 2014).

The increased growth rates and survivability rates reported by Sargent & Lodge (2014) are alarming, but do not mean that *F. rusticus* expansion is "unstoppable". Temperature and depth are two factors shown to have potential in stopping the invasion, but they are not the only factors to consider. A great deal of research has been performed to determine what other physiological and biological factors potentially impede the growth, survivability and spread of *F. rusticus*. This section will highlight the impact of several of these key factors on *F. rusticus* populations.

The most intriguing factor in analyzing the spread of *F. rusticus* relates to calcium. Calcium (Ca) is an essential component of the exoskeleton composition of all crayfish species. In order to grow, *F. rusti*cus must regularly molt and regrow their exoskeletons. Growing a new exoskeleton requires a high amount of Ca to facilitate calcification, which means that *F. rusticus* could be potentially hindered by freshwater environments with low Ca concentrations (Cairns & Yan, 2009; Edwards et al., 2013).

Crayfish dispersal and abundance from 1972-77 for sixty-seven lakes in Vilas County, Wisconsin, were recorded and compared to a series of independent variables including calcium (Ca) using multiple stepwise multiple regression (Capelli & Magnuson, 1983). The data suggested that the *Orconectes* (*Faxonius*) crayfish require a Ca level of 2.5 ppm at the absolute minimum, since none were found in lakes containing Ca levels <2.5 ppm (Capelli & Magnuson, 1983).

Cairns and Yan (2009) performed a metadata analysis to summarize the effects that low aqueous calcium (Ca) has on freshwater crustaceans and their Ca saturation points, minimum Ca thresholds to allow for survival and the toll that suboptimal Ca concentrations exert on the organisms. Results indicate that for crayfish living in areas of low Ca, the lethal Ca threshold is likely between 1.0 and 2.5 mg of Ca L^{-1}, whereas crayfish living in areas with standard or higher amounts of Ca have a lethal Ca threshold of between 5 and 10 mg of Ca L^{-1}. Crayfish living in harder waters had a higher saturation points (the concentration of external Ca above which there is no measurable increase in Ca uptake) as compared to those living in Ca poor environments.

Edwards et al. (2015) examined the effects that calcium concentrations have on the carapaces (dorsal shells) of *Faxonius virilis*, *Faxonius rusticus*, and *Cambarus bartonii* (Appalachian brook crayfish/common crayfish) across 24 lakes in south-central Ontario. The primary objectives of the experiment were to determine if there was a significant correlation between the amount of calcification of the carapaces and the calcium present in the lakes and to compare the previously mentioned correlation between the three species. The calcium concentration, pH and alkalinity of the 24 lakes were recorded and the crayfish were collected via passive overnight traps. *F. virilis* were caught in 19 of the lakes, *F. rusticus* were caught in 5 of the lakes, and *C. bartonnii* were also caught in 5 of the lakes. The carapaces of the crayfish were measured for length, and the calcium percentage of the carapace was determined using spectrophotometry. The relationship between the calcium concentrations in the carapaces and in the lakes was tested using simple linear regression. Segmented regression was used to identify when a saturation point was reached where amount of carapace calcium did not increase with increasing calcium concentrations in the lake. The results showed that the carapace calcium concentration increased logarithmically with increasing lake calcium concentrations up unto the saturation point, which is greater than 8 mg l^{-1} (Edwards et al., 2015). This significant increase was not found for *F. rusticus* and *C. bartonii*, likely due to the fact that they had much smaller sample sizes than *F. virilis*, and were only found in lakes with specific calcium concentrations (*F. rusticus* $\geq$ 8 mg l^{-1}, *C. bartonii* 9.16-56.8 mg l^{-1}) (Edwards et al., 2015).

Calcium concentrations may also influence specific behaviors of *F. rusticus*. To test the effects that low dissolved Ca levels in the waters of the Canadian Shield have on rusty crayfish, forty (40) one-year old *F. rusticus* were taken from the Little Rogue River (LRR), Ontario, Canada, on October 5 of 2010 and the samples were divided into three tanks each containing a different level of dissolved Ca: Thirteen represented a control group and were given Ca equal to that of the LRR (34.5 mg*L^{-1}), fourteen were given dissolved Ca equal to the mean dissolved Ca of waters in the Canadian Shield (2.5 mg*L^{-1}), and thirteen were given the lowest amount of dissolved Ca found in the Canadian Shield (1.0 mg*L^{-1}) (Edwards et al., 2013). The rusty crayfish are given a specified amount of dissolved Ca until they regrow their carapace following molting during which time their behavior, including their reaction to predators, was recorded in 21-minute trials and analyzed with repeated-measures ANOVA (Edwards et al., 2013). The crayfish provided with the lowest amount of dissolved Ca had a significant increases in mortality, which led Edwards et al. (2013) to conclude that the limiting lower threshold for dissolved calcium is conservatively between 1 and 2.5 mg*L^{-1}. *F. rusticus* found in waters with less dissolved Ca than the control groomed less frequently and were found to be vigilant for longer periods after sensing a predator than the control group (Edwards et al., 2013).

A pair of factors related to calcium concentration are alkalinity and pH. Alkalinity is the ability of the water to neutralize acidity and is usually measured in the amount mg/L CaCO3 present. pH is the unit of measurement used to determine the amount of “acidic” hydrogen atoms in a liquid. The scale ranges from 0 (most acidic) to 14 (most basic) and water generally has a pH of 7, which is neutral. More alkaline water will require a greater degree of acid to lower its pH than water with less alkalinity (Cox, 1995).

Both pH and alkalinity are potentially valuable variables to examine when analyzing the spread of *F. rusticus*. Edwards et al. (2015) found that Berrill et al. (1985) indicated that habitats with a pH range of 5.4-6.1 were unsuitable for juvenile *F. rusticus*. This low pH has been found to impede crayfish's calcium absorption and thus weaken their carapaces, which increases their mortality rates (Edwards et al., 2015). Wheatley and Gannon (1995) proclaimed the importance of alkalinity in calcium reabsorption. Post-molt crayfish (specifically *Procambarus clarkii*) in aquatic ecosystems where sodium (NA) have been removed from the water reabsorbed calcium at a 50% reduced rate than those in a sodium present environment (Wheatley and Gannon, 1995).

Water velocity in streams also impacts the growth of *F. rusticus*, making it a factor that is potentially worth looking into. In June and July 2009, Form I (sexually active) and Form II (sexually inactive) male *F. rusticus* were collected from four high velocity streams (HVS) (43-55 cm/s), 5 low velocity streams (LVS) (24-27 cm/s) and 7 lakes in Northern Wisconsin and Michigan's Upper Peninsula (Perry et al. 2013). The crayfish had measurements of their left and right chelae (claws) and their bodies size was measured and this data was compared with water velocity using MANOVA. Crayfish in HVS were found to exhibit smaller right and left chelae and bodies than those in LVS and lakes, and no *F. rusticus* greater than 45 mm were found in HVS (Perry et al., 2013).

One final factor that may be worth consideration in examining the diffusion of *F. rusticus* is the role of dams and other impoundments. Adams (2013) detailed the effects downstream impoundments could have on *Orconectes* species (although not *O. rusticus* (as it was known at the time)). The effects of small dams on crayfish populations in the Yazoo River basin of northern Mississippi was explored by comparing crayfish assemblages at four outlet sites and four undammed sites in August, September and November 2004 and with seven undammed sites, four outlet sites, and three sites affected by extremely small impoundments in 2010 (Adams, 2013). Using catch-per-unit-effort (CPUE) data, the experiment showed that small dams and impoundments generally favored *Procambarus* species over *Orconectes* species (Adams, 2013). Each small impoundment was estimated to influence crayfish assemblages over 2 km downstream (Adams, 2013). It needs to be stated again that *F. rusticus* was not among the species examined in Adams (2013) work. However, Olden et al. (2011) did assess whether or not a lake was an impoundment or a natural lake AS one of the variables used to measure a Wisconsin waterbody's vulnerability to be invaded by *F. rusticus*.

Looking at the scale and speed of the *F. rusticus* invasion, it is unlikely that one specific factor can potentially combat the threat of future incursion. Multiple factors need to be analyzed to determine the exact extent that any of these variables can aid or impede the spread of *F. rusticus*. Calcium, water velocity, substrate, Alkalinity and pH seem like the most promising conditions for investigation

CHAPTER III

DATA AND METHODS – DEVELOPING A PREDICTIVE MODEL

The purpose of this study is to analyze the pattern of *F. rusticus* (rusty crayfish) in Michigan and evaluate the utility of publicly available data for modelling distributions of *F. rusticus* and other common crayfish species in the state. The study has three main objectives:

- To compare Smith et al.'s (2018) crayfish survey data to the data derived from the publicly available Soil Survey Geographic Database (SSURGO) and National Land Cover Database (NLCD) databases in order to reveal the ecosystems most vulnerable to the introduction of *F. rusticus*.
- To determine habitation characteristics for *F. rusticus* at a local (40.4678 hectares or 100 acres) and landscape (404.678 hectares or 1000 acres) scale.
- To compare the presence/absence of other native species of crayfish (*F. propinquus*, and *F. virilis*) against these same variables vis-a-vis the presence/absence of *F. rusticus*.

3.1: Overview of Data Sources

In order to accomplish these objectives, the Smith et al. (2018) crayfish survey data was obtained courtesy of Dr. Brian Roth and Kelley Smith. The Smith et al. (2018) data included the known geographical locations where *F. rusticus* have been detected in Michigan streams. The information was gathered by Kelley Smith, a graduate student at Michigan State University, and Dr. Brian Roth, associate professor in Michigan State University's Department of Fisheries and Wildlife. Dr. Roth and Mr. Smith kindly agreed to provide the crayfish survey data for this analysis. The data set contains the following relevant variables: an identification number for each site; the latitude and longitude of every sample site; the nearest Great Lake, watershed, sub-shed and tributary to the sample site; how many crayfish of each of 8 species (*Cambarus diogenes*, *Cambarus robustus*, *Fallicambarus fodiens*, *Faxonius immunis*, *Faxonius propinquus*, *Faxonius rusticus*, *Faxonius virilis*, and *Procambarus acutus acutus*) were found at each survey location; the substrate type characterizing each site; and additional notes on the site and substrate. Two surveys were conducted at each waterbody, so these survey locations were arranged in pairs by unique site (i.e. Survey locations 1 and 2 both took place at Hayward Creek, but each sample has unique latitude/longitude coordinates and site characteristics observed *in situ* independently from one another). In total, there were 966 survey locations included in the study.

The three most prevalent species of crayfish found were *F. rusticus*, *F. propinquus* and *F. virilis*. *F. propinquus* and *F. virilis* are the two most common native species of crayfish in Michigan and compete for the same habitat and resources as *F. rusticus.* While the main focus of this study is *F. rusticus*, these common native species were also considered during modeling. The other native crayfish species were present at too few sites to be able to use their presence for statistical analysis (*C. diogenes* were present at ~5.487% of the total sites surveyed, *C. robustus* at ~5.797%, *F. fodiens* at ~1.346%, *F. immunis* at ~4.865%, and *P. acutus acutus* at ~0.578%). Coordinate pairs identifying crayfish survey locations where *F. rusticus, F. propinquus* or *F. virilis* were present were converted into a point-based shapefile in ArcMap 10.17.1 (Environmental Systems Research Institute (ESRI), 2019), allowing spatial overlay with other datasets.

The SSURGO dataset was chosen because it contained soil data for the entire State of Michigan and each county’s information was updated quite recently (updates occurred in September 2018). Soil data layers contain information on calcium, substrate, and pH, all variables identified in the Introduction as significant to the problem. Since this research considered crayfish invasion on a macro-level, it was necessary to supplement the soil data with databases that provided insight into the land surface in areas beyond just the examined waterbodies themselves. The National Land Cover Dataset (NLCD) (Multi-Resolution Land Characteristics (MRLC) Consortium, n.d.a) provides land cover class descriptions at 30m x 30m resolution that are widely accepted by researchers, encompass the entire State of Michigan, and included classes that correspond to land cover variables examined in previous research (Olden et al. 2011; etc.). The NLCD also has the added benefit of being processed in raster format and therefore the land-use/land-cover (LULC) can be easily quantified at various levels of scale. These qualities made the NLCD data layer the ideal choice for analyzing LULC information. Interestingly, both SSURGO and NLCD are national models, so this form of research could be applied to the rest of the continental United States should the methods prove useful.

Finally, the baseflow data for Michigan's streams were used in conjunction with the soil and land cover datasets. Low stream baseflows have been shown to relate to *F. rusticus* expansion and were a significant variable in the prediction model of Olden et al. (2011). Dr. Dana Infante of Michigan State University's Department of Fisheries and Wildlife, generously provided the projected stream discharge data (in cubic feet per second) for nearly all of Michigan's streams (Stewart et al., 2016). Base flow data is currently also available on Michigan Open GIS (2019). As with the soil and NLCD variables, a detailed description of the stream baseflow can be found in Table 3.1.

Table 3.1 - Variables Examined and Used in the Artificial Neural Network

Variable Examined	Source	Original Column Name	Created Name	Normalized Name	Description
F. rusticus Presence	Smith et al. (2018)	O_rusticus	RusPres		Is *F. rusticus* present in the examined survey location
F. propinquus Presence	Smith et al. (2018)	O_propinqu	NCCPres		Is *F. propinquus* present in the examined survey location
F. virilis Presence	Smith et al. (2018)	O_virilis	VirPres		Is *F. virilis* present in the examined survey location
Total Sand	SSURGO	sandtotal_r		Norm_sand_r	Sand Particles between 0.05-2.0 mm in diameter as a weight percentage of the less than 2.0 mm fraction. (Relative Value (likely the 50th percentile))
Total Silt	SSURGO	silttotal_r		Norm_silt_r	Silt particles 0.002 mm to 0.05 mm in diameter as a weight percentage of the less than 2.0 mm fraction. (Relative Value)
Total Clay	SSURGO	claytotal_r		Norm_clay_r	Clay particles less than 0.002 mm in diameter as a weight percentage of the less than 2.0 mm fraction. (Relative Value)
Organic Material	SSURGO	om_r		Norm_om_r	The amount of decomposing organic matter as a weight percentage of the less than 2.0 mm fraction. (Relative Value)
$CaCO_3$	SSURGO	caco3_h		Norm_caco3_h	The amount of calcium carbonate as a weight percentage of the less than 2.0 mm fraction. (High Value (likely the 90th-95th percentile))

Table 3.1 - Contintued

Variable Examined	Source	Original Column Name	Created Name	Normalized Name	Description
Extract Acidity	SSURGO	extracid_r		Norm_extracid_h	The amount of Hydrogen atoms in the soil whose activation could lead to acidification. (Relative Value)
Average pH	SSURGO	ph1to1h2o_h; ph01mcac2_h	Ave_pH_h	Norm_pH_Ave_h	The average of 2 pH measurements or the non-zero pH if one measurement was 0. (High Value)
Total Pebble	SSURGO	fragvol_h; fragkind; fragsize_h	pebble_all_h	Norm_Peb_All	The percent volume (fragvol_h) of the soil horizon containing fragments (fragsize_h) ≥4 mm and <64 mm. (High Value)
Total Cobble	SSURGO	fragvol_h; fragkind; fragsize_h	cobble_all_h	Norm_Cob_All	The percent volume (fragvol_h) of the soil horizon containing fragments (fragsize_h) ≥64 mm and <256 mm. (High Value)
Open Water	NLCD		LULCWater, LULCWater1k	Norm_Water, Norm_Water1k	Locations where less than <25% of the total area consists of soil and/or vegetation and the rest is open water. (At local (100 acres) and landscape (1000 acres) scales)
Woody Wetland	NLCD		LULCWW, LULCWW1k	Norm_WW, Norm_WW1k	Areas that are periodically flooded or saturated, and which trees and shrubs comprise >20% of the total vegetative cover. (At local and landscape scales)
Emergent Herbaceous Wetland	NLCD		LULCEmWet, LULCEmWet1k	Norm_EmWet, Norm_EmWet1k	Periodically flooded or saturated areas whose vegetative cover is comprised of >80% herbaceous plants. (At local and landscape scales)
Cultivated Crops	NLCD		LULCCrops, LULCCrops	Norm_Crops, Norm_Crops1k	Areas where human planted crops (such as corn and vegetables) make up >20% of the total vegetation. (At local and landscape scales)

Table 3.1 - Continued

Variable Examined	Source	Original Column Name	Created Name	Normalized Name	Description
Developed, Open Space	NLCD		LULCDevOS, LULCDevOS1k	Norm_DevOS, Norm_DevOS1k	Areas that contain a mixture of vegetation and human-made developments. <20% of the land is made up of impervious surfaces. (At local and landscape scales)
Developed, Low Intensity	NLCD		LULCDevL, LULCDevL1k	Norm_DevL	Areas that contain a mixture of vegetation and human-made developments. 21-49% of the land is made up of impervious surfaces. (At local and landscape scales)
Developed, Medium Intensity	NLCD		LULCDevM, LULCDevM1k	Norm_DevM, Norm_DevM1k	Areas that contain a mixture of vegetation and human-made developments. 50-79% of the land is made up of impervious surfaces. (At local and landscape scales)
Developed, High Intensity	NLCD		LULCDevH, LULCDevH1k	Norm_DevH, Norm_DevH1k	Areas that contain human-made developments and (except for extreme cases) vegetation. 80-100% of the land is made up of impervious surfaces. (At local and landscape scales)
Deciduous Forest	NLCD		LULCDecFor, LULCDecFor1k	Norm_DecFor, Norm_DecFor1k	Areas where trees greater than >5m account for >20% of the total vegetative cover. >75% of the trees here lose their leaves in the fall and winter. (At local and landscape scales)
Evergreen Forest	NLCD		LULCConFor, LULCConFor1k	Norm_ConFor, Norm_ConFor1k	Areas where trees greater than >5m account for >20% of the total vegetative cover. >75% of the trees here maintain their foliage year-round. (At local and landscape scales)

Table 3.1 - Continued

Variable Examined	**Source**	**Original Column Name**	**Created Name**	**Normalized Name**	**Description**
Mixed Forest	NLCD		LULCMixFor, LULCMixFor1k	Norm_MixFor, Norm_MixFor1k	Areas where trees greater than >5m account for >20% of the total vegetative cover. Neither deciduous or evergreen species account for >75% of the total tree cover. (At local and landscape scales)
Pasture/Hay	NLCD		LULCPast, LULCPast1k	Norm_Past, Norm_Past1k	Areas where human-planted pasture/hay vegetation accounts for >20% of the total vegetation. (At local and landscape scales)
Grasslands or Herbaceous	NLCD		LULCGrass, LULCGrass1k	Norm_Grass, Norm_Grass1k	Areas where >80% of the vegetative cover is comprised of graminoid (grass-like) or herbaceous (non-woody) plants. (At local and landscape scales)
Shrub/Scrub	NLCD		LULCShrub, LULCShrub1k	Norm_Shrub, Norm_Shrub1k	Areas where vegetation <5m tall comprise >20% of the total vegetative cover. (At local and landscape scales)
Barren Land	NLCD		LULCBarren, LULCBarren1k	Norm_Barren, Norm_Barren1k	Areas consisting of high accumulations of earthen materials (dirt, sand, etc.) and where vegetation comprises <15% of the total cover. (At local and landscape scales)
Longitude	Smith et al. (2018)	Lon		Norm_Long	Longitude where the sample was taken
Latitude	Smith et al. (2018)	Lat		Norm_Lat	Latitude where the sample was taken
Projected Baseflow	Stewart et al. (2016)	Proj_baseflow_cps		Norm_BF	Projected baseflow of the stream (in cubic feet per second)

3.1.i: SSURGO Data

The soil information used for this analysis is available to the public through the SSURGO (Soil Survey Geographic Database), which contains information on soil data at a sub-county level (USDA - Natural Resources Conservation Service (NRCS) 2014). The Web Soil Survey, provided by the Natural Resource Conservation Service (NRCS) of the United States Department of Agriculture (USDA), has soil data mapped across each of Michigan's counties (see Figure 3.1), with mapping scales ranging from 1:12,000 to 1:24,000 (United States Department of Agriculture (USDA) – Natural Resources Conservation Service (NRCS), n.d.a; USDA – NRCS, n.d.b). Soil data for every county in Michigan was downloaded separately, linked to Microsoft Access and then opened and manipulated in ArcMap 10.7.1 (Environmental Systems Research Institute (ESRI), 2019).

Due the fact that *Faxonius rusticus* reside in permanent water bodies, the data was cleaned to exclude all soil units that did not overlap with a permanent body of water (in this case, streams). The State of Michigan's GIS Open Data website houses and provides "Stream River Assessment Units", which allow for the creation of a file layer containing only the permanent water bodies (Department of Natural Resources (DNR), 2019) [2]. This shapefile was opened in ArcMap and a 30-meter buffer was created around it to convert it from vectors into polygons. 30-meters was selected as the buffer size due to Olden et al. (2011) using a 30-meter buffer width to classify the upstream riparian area in a similar project. The buffer data was joined to the soil maps. All soil sections that did not intersect with this buffer were removed.

---

[2] https://gis-midnr.opendata.arcgis.com/datasets/c0f0eef5603e43d5a9864f8fdf6ddfe2_5

Figure 3.1 - The Counties of Michigan
Source: GIS Open Data (2019a)[3].

After reducing the total number of soil sections examined, it became necessary to determine which variable attached to the soil data would be joined to the crayfish location coordinates for modelling as the SSURGO data has a myriad of categories. For example, the chorizon table alone has 159 variables. After examining the variables available in the SSURGO data set and comparing them to the substrates examined and chemical factors examined in

---

[3] http://gis-michigan.opendata.arcgis.com/datasets/67a8ff23b5f54f15b7133b8c30981441_0

previous research studies (Cairns & Yan, 2009; Olden et al., 2011; Edwards et al., 2013; Smith et al., 2018; etc.), as well as the consistency and availability of the SSURGO variables across Michigan watersheds, a set of variables was extracted for use in further modeling. The selected variables were: total clay, total silt, total sand, total pebbles (rocks $\geq$4 mm and <64 mm), total cobble (rocks $\geq$64 mm and <256 mm), total organic material (om), $CaCO_3$ quantity (calcium carbonate), pH, and alkalinity (in the form of available hydrogen atoms (labeled 'extracid')). The $CaCO_3$ quantity corresponded to the calcium bedrock variable examined by Olden et al. (2011) in their prediction model. Additionally, pH and alkalinity corresponded to lacustrine clay variable used by Olden et al. (2011) in that same model. This extraction was performed for all counties of Michigan and the county data were merged into one shapefile encompassing the permanent streams in the State of Michigan. A description of these variables, along with any variables that were modified, can be found in Table 3.1.

After examining the values for soil variables for each sample, it became apparent that certain soil variables were not measured within each soil unit and, therefore, in each survey location. A number of the examined sites had values of zero at most, if not all, of the evaluated SSURGO variables. It is possible that the value of these variables, such as the amount of gypsum present, represent actual measured quantities of 0, but it is also just as possible that zero values represent units without measured data. As it is impossible for pH to have a value of 0, it can be assumed that any site with a pH of 0 simply did not have any soil data recorded for it. Since measurements of 0 can completely throw off the statistical results, all survey locations with a reported pH of 0 were removed. This removed 136 of the original 966 survey locations from the analysis. The remaining 830 survey locations were the examined targets.

Additionally, this merged data was joined with a map Michigan's HUC 8 watershed boundaries (see Figures 3.2 and 3.3). These boundaries will allow for future potential comparisons and integration with governmental databases. The HUC 8 boundaries shapefile were obtained from the State of Michigan's GIS Open Data Website (GIS Open Data, 2019b). This combined data was merged with Smith et al. (2018) survey location coordinates.

Figure 3.2 – HUC 8 Watersheds of Michigan
Source: GIS Open Data (2019b)[4]

---

[4] https://gis-michigan.opendata.arcgis.com/datasets/watershed-boundary-8-digit

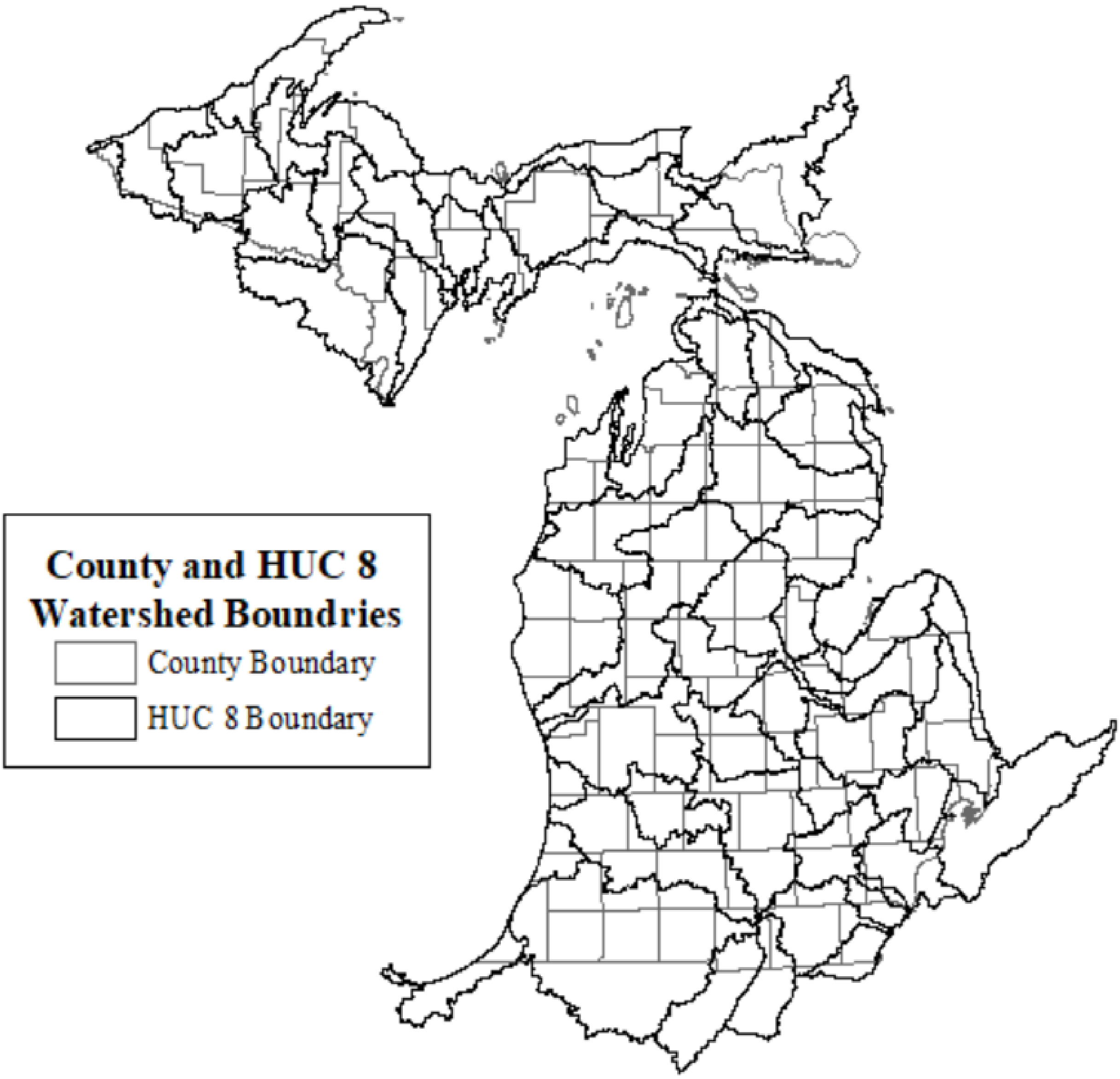

Figure 3.3 - Map of Michigan’s County and HUC 8 Watershed Boundaries

3.1.ii: NLCD

The National Land Cover Database (NLCD) is used to describe land-use land-cover classes (LULCs) in Michigan. Olden et al. (2011) found that *F. rusticus* presence corresponded with land taken up by row crops (crops that can be planted in rows for tilling) and urban environments. For this study, these factors, along with the other LULCs in Michigan were investigated at both a local scale and landscape scale. Using Anderson et al. (2018) wetlands density classification scales, the local scale is defined at a 40.4678-hectare (100 acre) (404678 m^2) circle centered on each survey location and the landscape scale is defined as a 404.678-hectare (1000 acre) (4046780 m^2) circle centered around each survey location.

The land cover classes, as defined by the National Land Cover Database 2016 (NLCD2016) legend, found in the state of Michigan were determined to be the following (the number is the official NLCD2016 value assigned to the class): Open Water (11); Developed, Open Space (21); Developed, Low Intensity (22); Developed, Medium Intensity (23); Developed, High Intensity (24); Barren Land (31); Deciduous Forest (41); Evergreen Forest (42); Mixed Forest (43); Shrub/Scrub (52); Grasslands/Herbaceous (71); Pasture/Hay (81); Cultivated Crops (82); Woody Wetlands (90); and Emergent Herbaceous Wetlands (95) (see Figure 3.4) (Multi-Resolution Land Characteristics (MRLC) Consortium, n.d.b). A detailed explanation of each of these habitat types can be found in Table 3.1. Cultivated Crops and Woody Wetlands correspond to the row crops and forested wetland classes (respectively) investigated by Olden et al. (2011). Olden et al.'s (2011) urban land use class corresponds with the 4 "Developed" land use classes found in the 2016 dataset.

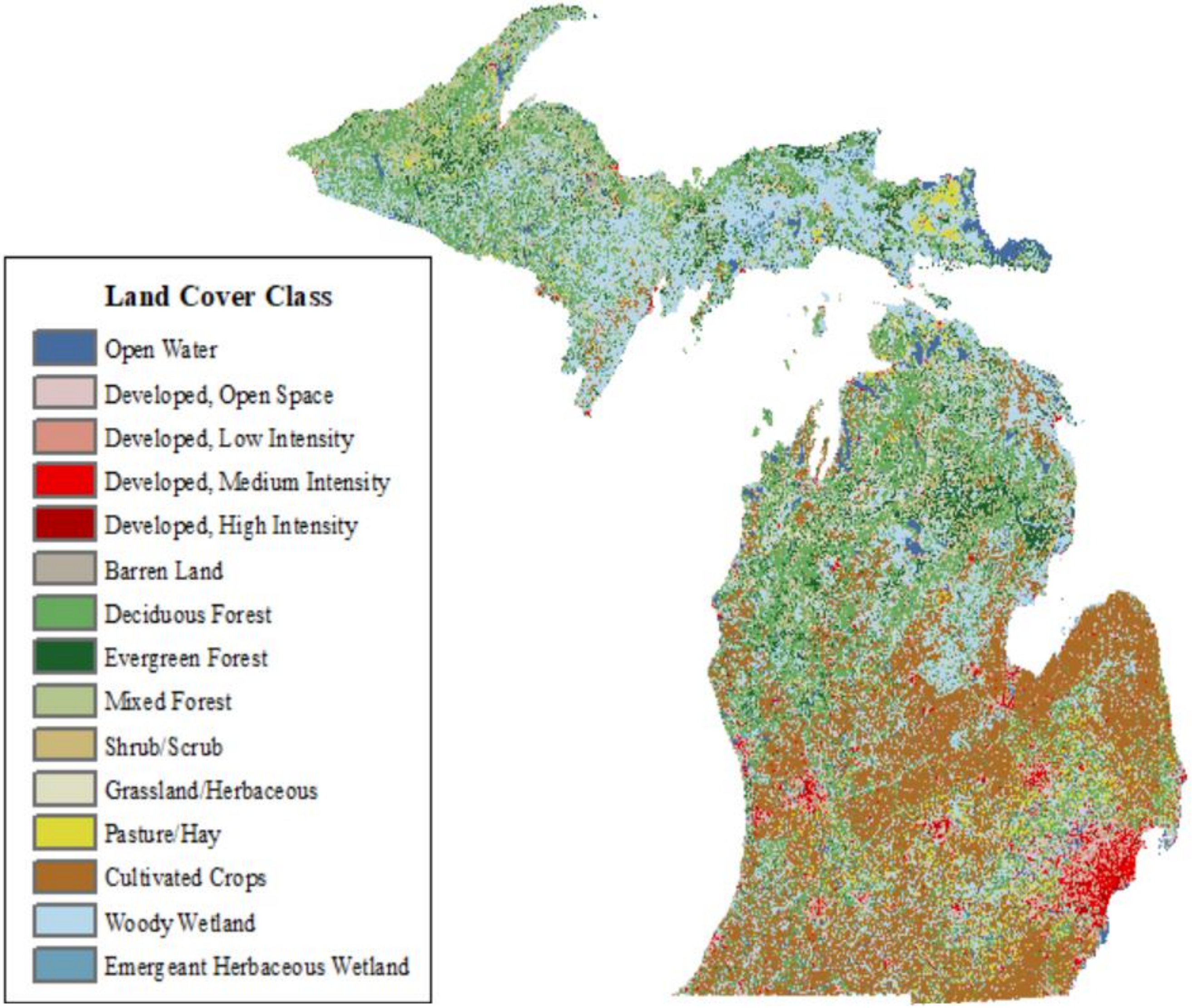

Figure 3.4 - 2016 National Land Cover Database (NLCD) Land Cover Classifications of Michigan (Source: MRLC Consortium, n.d.b)

To determine percentages of each LULC, a pixel value reflecting the presence/absence of each of the classes at both the local and landscape scales was calculated in ArcMap 10.7.1. These LULC percentages were linked to each of the crayfish survey locations. The fact that each pixel has a different value generally meant that each survey location corresponded to a slightly different value. Even when two samples were taken in close proximity, the latitude and longitudes were usually not located within the same pixel. A description of these local and landscape LLC classes, including the names used in the ANN model and means tables, can be found in Table 3.1.

3.2: Statistical Methods

The primary statistical software used for this project are IBM SPSS Statistics 26 (SPSS) and RStudio (RStudio Team, 2016). Due to its ease of use and ease of manipulation of data tables, SPSS was chosen as the program with which the majority of statistical tests were performed.

The average pH, the average percent volume of pebbles (high value), and the average percent volume of cobble (high value) needed to be derived from other variables using SPSS, as they were unable to be accurately modelled in the form available in the SSURGO dataset. In SSURGO, there were two methods used for determining pH: the 1:1 soil-water ratio method and the 0.01M $CaCl_2$ method (USDA – NRCS, 2014). Typically, only one method was used for each site, though some used both, which meant that a new pH field was needed in order to prevent pH recordings of 0 from entering the model. The new variable field (dubbed Ave_pH_h) was derived by the following rules: If one pH method had a value of 0, the two pH (high value) values would be added together; and if both pH's were greater than 0, the average of the two values would be used.

The (high values) of the total pebble and cobble found at each survey location had to be derived from 3 separate variables: the fragment kind (so as to exclude wood and shell fragments), the fragment size (to determine whether the fragments fall into the pebble ($\geq$4mm and <64mm) or cobble ($\geq$64mm and <256mm) classifications) and the fragment volume (the value that will be used for the research).

For modeling purposes, the *F. rusticus* count data from the original survey was simplified to a binary variable denote the species as being either present or absent at each examined coordinate. This classification allowed for *F. rusticus* presence (or RustPres as the variable was named) to be used as an independent variable, with 0 representing absence and 1 representing presence. This reclassification was also performed for *F. propinquus* (NCCPres) and *F. virilis* (VirilePres).

An artificial neural network (ANN) algorithm was selected to model presence/absence of crayfish species with regard to soil, LULC and streamflow variables. ANNs are effective at determining modeling data associations between multiple large number of dependent variables that may interact with each other. Olden et al. (2011) had success with modelling crayfish occurrences in Wisconsin using multi-response artificial neural network (MANN), so it stood to reason that ANNs would be suitable for this research. The ANNs are able to create predictions on crayfish presence based on a sample of the survey variables and so results could be compared to the actual values. RStudio was selected as the researcher has experience with this program and it has the neuralnet package, which allows for the creation of ANNs. Additionally, the NeuralNetTools package (Beck, 2018), was used since it can specifically determine the importance (or weight) of each variable in the network through its Olden function (named after

Julian D. Olden and specifically based around Olden et al. 2004). This importance was measured in terms of positive/negative impact and degree of intensity.

All non-crayfish variables used as independent variables in the ANN were normalized between 0 and 1. The normalization equation was as follows:

$$N = (V - V_{min})/(V_{max} - V_{min})$$

Where N is the normalized final variable, V is the variable being examined, V_{max} is the maximum value of the variable being used in the experiment, and V_{min} is the minimum value of the variable in being used in the experiment.

The normalization process for the adjusted baseflows was modified. The largest projected baseflow was 3932.72 cfs (cubic feet per second); however, all but 58 of survey locations had a projected baseflow of <100 cfs. Retaining all values made the majority of the normalized values far closer to 0 than is optimal for being useful for the model. To address this issue, all baseflows greater than 737.88 cfs were given a value of 1, with baseflows from 0-737.88 normalized as described above. The names given to the normalized variables are listed in Table 3.1.

Using the neuralnet and NeuralNetTools packages, the presence of *F. rusticus* was modeled with respect to the normalized soil variables, LULC variables at both local scale and landscape scale, and baseflow. To determine the accuracy of the model, approximately 50% of survey locations where the *F. rusticus* were present (79 survey locations) and where they were absent (336 survey locations) were randomly selected for model training and the other locations were used for assigning accuracy. Ten (10) hidden nodes were used in the ANN, as it was found to yield the best results after experimenting with multiple quantities. The number of steps used in the neural network was the default of 1e+05. A confusion matrix showed the accuracy of the predictions, with specificity being looked at in depth. These results were exported back into SPSS, where the results were compared to the actual presence (measured as absence, presence, false positive and false negative) and then converted into geographic coordinates so that the results can be mapped. The full code could be found in Appendix A.

F. propinquus and *F. virilis* were also tested separately with the same variables and procedures as those for *F. rusticus*. The number of randomly selected survey locations where *F. propinquus* were present was 170 and the number when they were absent was 245. The number of randomly selected survey locations where *F. virilis* were present was 81 and the number where they were absent was 334. Multiple tests were performed for each species as the results can be dependent on which survey locations were randomly selected as the testing variables. These full codes can also be found in Appendix A.

The importance of each input variable for predicting *F. rusticus* presence was determined in a similar manner. As before, an identical number of randomly selected coordinates were used to predict the importance of each soil variable, LULC variable, baseflow, and temperature. Additionally, the Olden function determines the importance (or weight) of each variable. The goal was to find variables with a high degree of importance. Importance is quantified such that the larger the number, the greater the variable's importance, regardless of its sign (positive or negative). The exact values of importance assigned by the Olden function have little meaning apart from the magnitude and sign. These results were posted in the Results Section.

CHAPTER IV

RESULTS

4.1: Current Distribution

Data provided by Smith et al. (2018) shows the current distribution of *F. rusticus*, along with *F. propinquus* and *F. virilis*, in the State of Michigan (see Figure 4.1 and Table 4.1). *F. rusticus* was the sole crayfish species present at 118 of the 830 sites, making it the second most common crayfish in Michigan, in terms of total sites occupied. It co-occurred only with *F. propinquus* (the most prevalent species), at 17 sites, with *F. virilis* (the third most prevalent species) at 18 sites, and with both of these species at 8 sites (see Table 4.1). Other crayfish species in the state occupied different habitat types and/or were found at too few sites to be statistically relevant and so are not included in this analysis.

4.2: *Faxonius rusticus* Habitat Predictions

Using 79 sample locations containing *F. rusticus* and 336 without, the ANN was run 10 times. The results were converted back to shapefiles (R to SPSS to Excel to ArcMap). The confusion matrix for Test 1 of *F. rusticus* predictions (Table 4.2) has been provided as an example. The rest of the confusion matrices of the 10 selections and their specificities can be found in Appendix B. The accuracy of the ANN model for the presence of *F. rusticus* was determined by the total percentage of correct predictions (*F. rusticus* is present in sites where it is present and absent in sites it is absent). The accuracy of various model runs with randomly selected training sites ranged from 0.7880 (Test 9) to 0.8386 (Test 5) with an average of 0.8181. However, if we make the assumption that predictions of *F. rusticus* presence in locations where

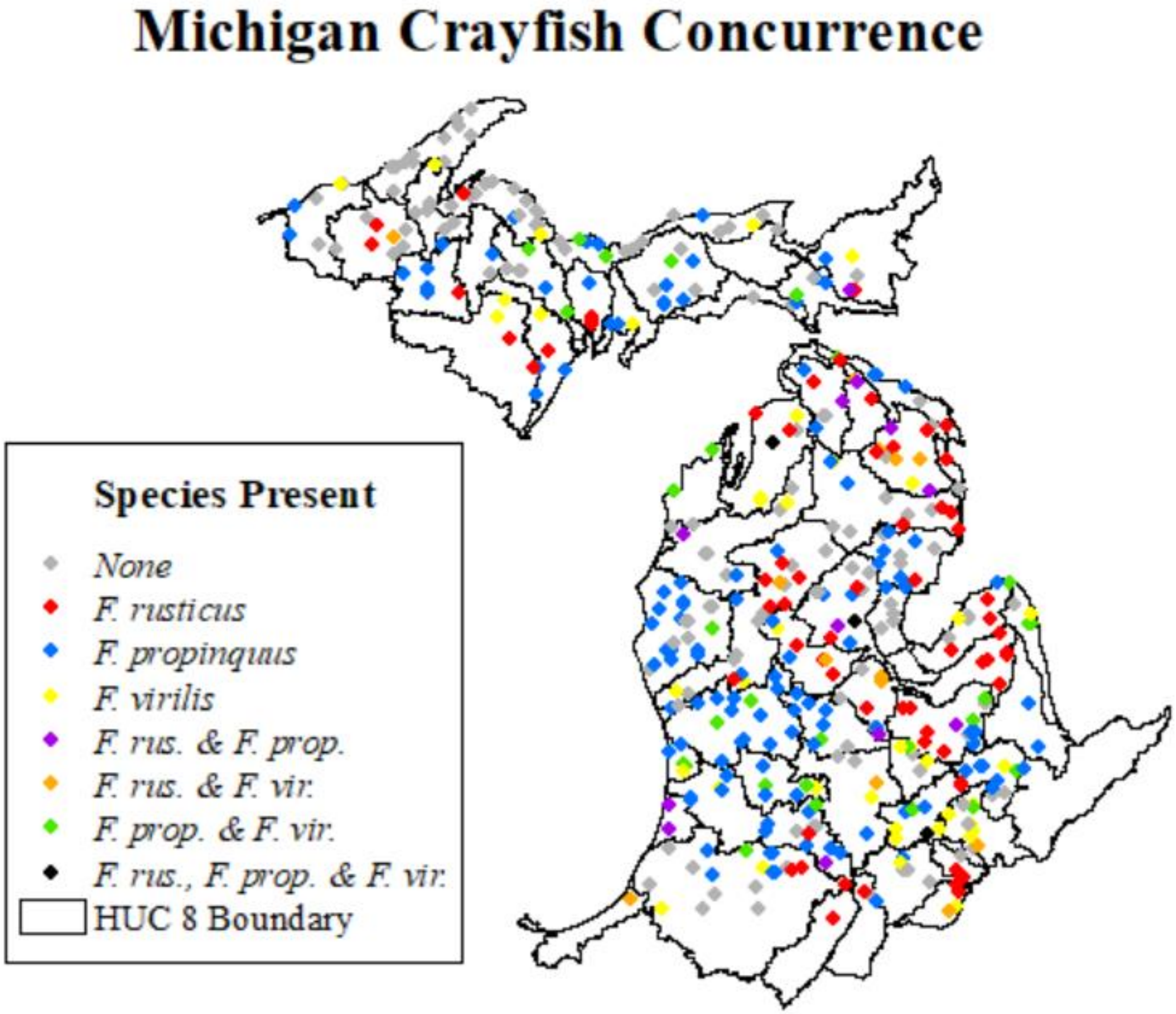

Figure 4.1 – Concurrence of *Faxonius rusticus*, *Faxonius propinquus* and *Faxonius virilis* in Michigan
Source: Smith et al. (2018); GIS Open Data (2019b)

Table 4.1 – Total number of Each Species Concurrence

Crayfish Species Combination	Number of Sites Combination Occurs
No Crayfish[1]	272
F. rusticus only	118
F. propinquus only	261
F. virilis only	81
F. rusticus & *F. propinquus*	17
F. rusticus & *F. virilis*	18
F. propinquus & *F. virilis*	57
F. rusticus, *F. propinquus* & *F. virilis*	6

[1] *F. rusticus*, *F. propinquus*, & *F. virilis* are the only species being looked at in this table

Source: Smith et al. (2018)

Table 4.2 – Confusion Matrix of Test 1 for *Faxonius rusticus* Prediction

Confusion Matrix (*Faxonius rusticus*) (Test 1)			
		Prediction	
		0	1
Actual	0	301	34
	1	46	34

they are absent are also correct in that they reflect potential suitable habitats; the accuracies range from 0.8843 (Test 9) to 0.9181 (Test 8) with an average of 0.9007. The specificities ranged from 0.4571 (Test 6) to 0.6066 (Test 5) with an average of 0.5353 (see Table 4.3).

Due to the fact that the sample locations selected for the ANNs were chosen at random, a few sites were selected in all 10 training processes, and thus do not have predictions associating with these survey locations. Those sample locations were Rows 65 (FID 88), 545 (FID 789), and 599 (FID 877). The number of times each site was predicted also varied from 1 test to all 10 for the other rows, depending on the random selection among model runs.

Table 4.3 – Accuracy and Specificity for Each ANN Test Performed on *Faxonius rusticus*

Test	Accuracy	Specificity
1	0.8072	0.5000
2	0.8337	0.5902
3	0.8193	0.5342
4	0.8217	0.5455
5	0.8386	0.6066
6	0.7928	0.4571
7	0.8337	0.5797
8	0.8313	0.5625
9	0.7880	0.4583
10	0.8145	0.5185

Note: Adjusted accuracy assumes that predicting *Faxonius rusticus* presence in an area it is currently absent from is a correct result and represents suitable habitats

In the results of the 10 model runs, *F. rusticus* were predicted to appear at survey sites where they were not found 340 times at 215 different locations (see Figure 4.2). These locations were spread throughout the state, but three particularly large clusters appeared. The first cluster forms almost a line, running up the southeastern coast (and coast adjacent counties) of Michigan, starting from Monroe County and moving up to the tip of the Thumb in Huron County (see Figure 4.3). This area includes the following counties: Monroe, Lenawee, Wayne, Washtenaw, Macomb, Livingston, Oakland, St. Clair, Lapeer, Sanilac, and Huron; and the following watersheds: Raisin, Ottawa-Stony, Detroit, Huron, Clinton, Flint, St. Clair, Cass, and Birch-Willow. The majority of the predictions in this area were focused on in the bodies, branches and tributaries of the Flint River, Cass River, and the Clinton River (including Red Run, a human created run flowing into the Clinton River).

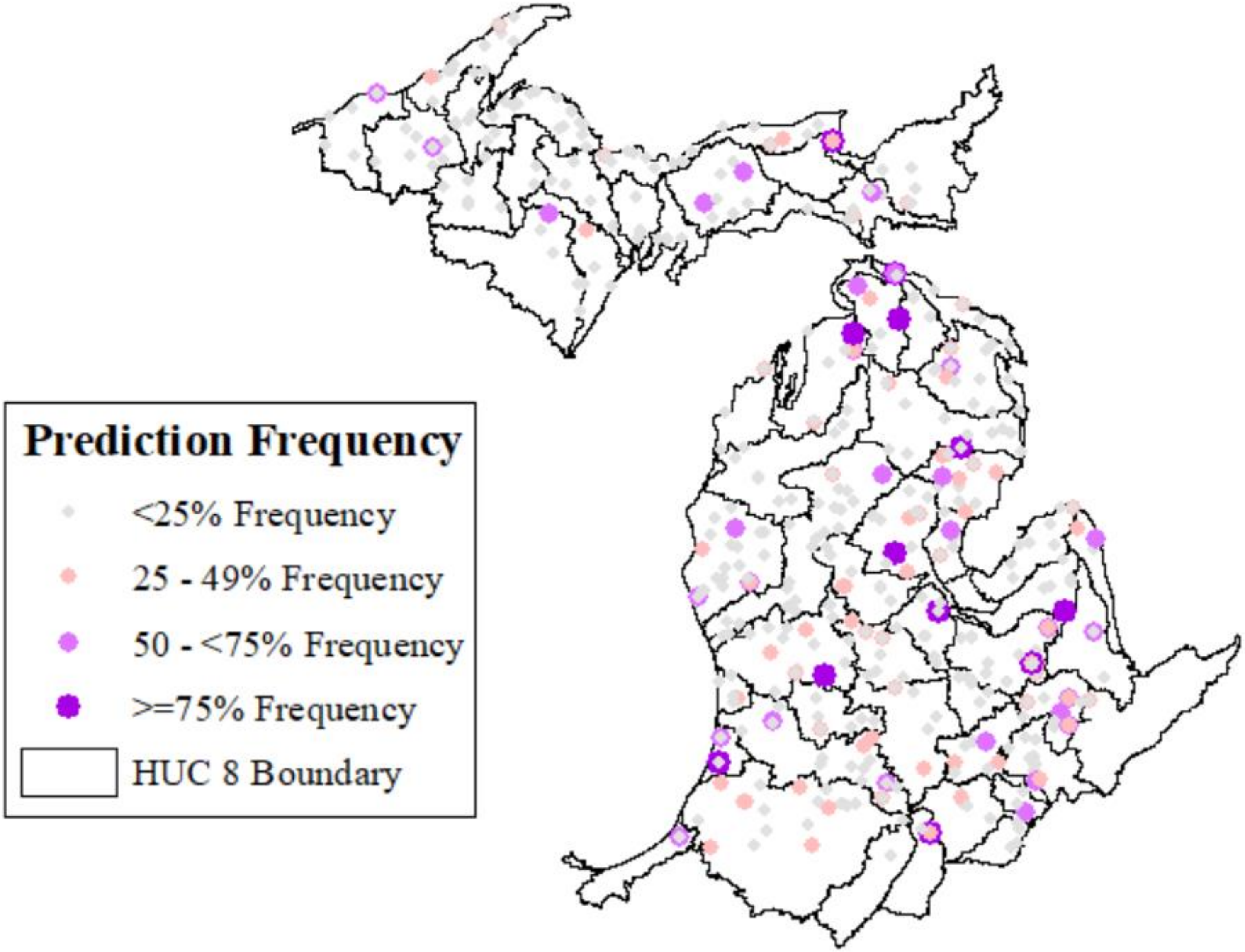

Figure 4.2 – Frequency of *Faxonius rusticus* Predictions at Sites Where They are Currently Absent

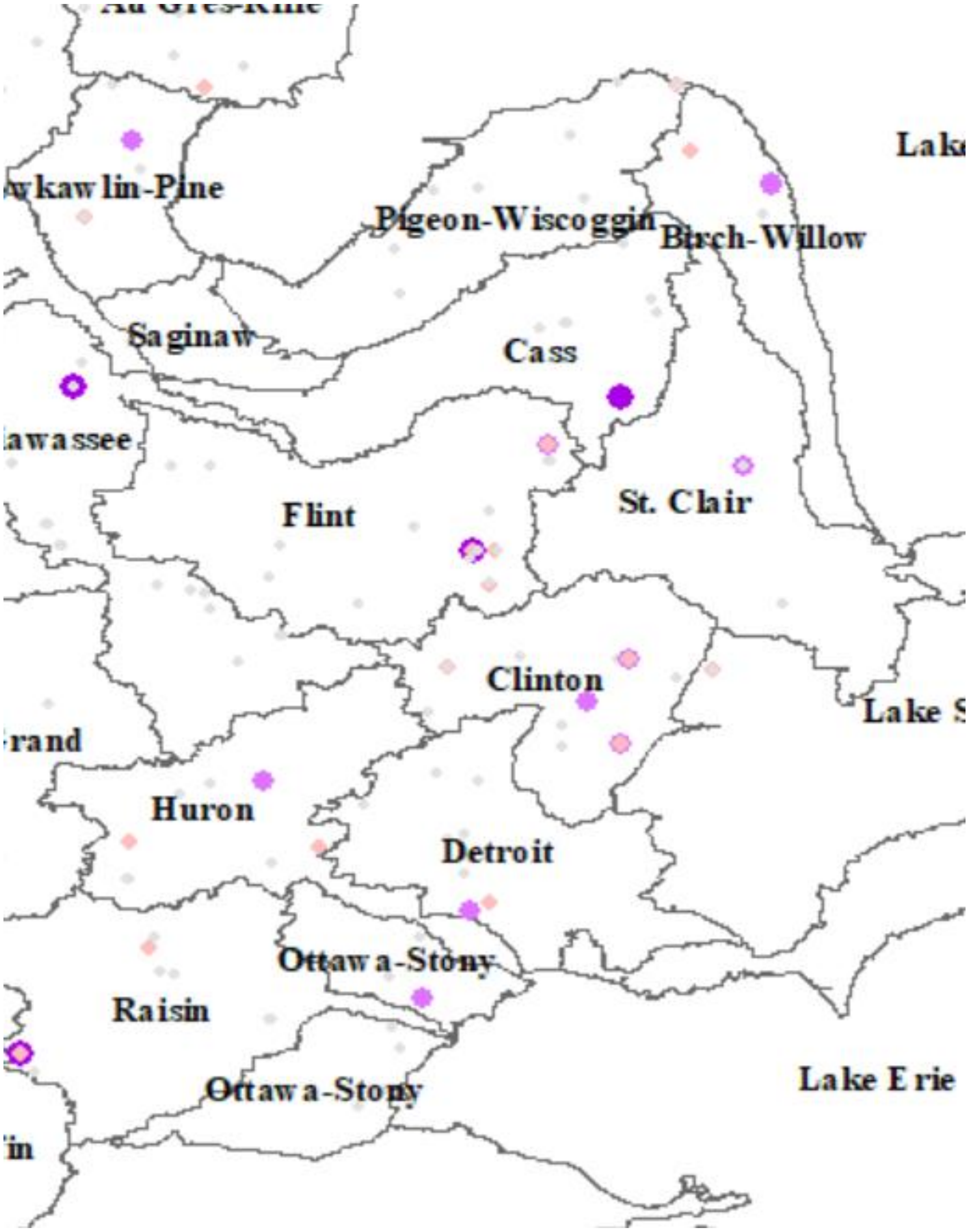

Figure 4.3 – Southeast Michigan *Faxonius rusticus* Prediction Cluster

Other waterbodies in this region containing high numbers of predictions included East Branch Willow Creek, Rock Falls Creek, New River, Frank and Poet Drain, Blakely Drain, Black Creek (River Raisin Watershed), Sandy Creek and Davis Creek. This area also contains the state's highest concentration of the "Developed" LULCs while the northern section contains a high level of the cultivated cropland (see Figure 3.4).

The second large concentration occurs in the Northern Lower Peninsula within Charlevoix, Emmet and Cheboygan Counties (see Figure 4.4). This includes the Boardman-Charlevoix, Cheboygan and Long Lake-Ocqueoc watersheds. *F. rusticus* was predicted for Hay Marsh Creek, South Branch Boyne River, Maple River, and Little Black River. At the landscape scale, the Maple River survey locations' percentage of LULC occupied by pasture and grassland classes far exceeded the state's average for both classes (0.14724 and 0.180145, respectively). The Little Black River selections showed higher than average pasture at a landscape scale and higher than average Developed - Open Space at both local and landscape scales. Over 75% of the total LULC at a local scale of the Hay Marsh Creek sites was classified as Woody Wetland in the NLCD dataset (MRLC Consortium, n.d.a).

The third major cluster of predicted *F. rusticus* presence is found in the counties surrounding the west side of Saginaw Bay (see Figure 4.5). This includes the following counties: Saginaw, Midland, Bay, Arenac, Ogemaw and Roscommon. The watersheds in this region are: Shiawassee, Tittabawassee, Kawkawlin-Pine, Muskegon, Au Gres-Rifle, and Au Sable. The waterbodies predicted were Knappen Creek, Harper Creek, Rifle River, Saganing River, Bluff Creek, and Marsh Creek. This region's land cover overwhelmingly consisted of the Woody Wetland and Cultivated Crops Classes.

The Black River in the Black-Macatawa Watershed of Allegan and Van Buren Counties also contained several highly predicted sites. The primary land cover in this area is cultivated cropland. The only other waterbody that contained a site with a 75% or greater prediction occurrence rate was Cheney Creek, in the Tahquamenon Watershed in Chippewa County.

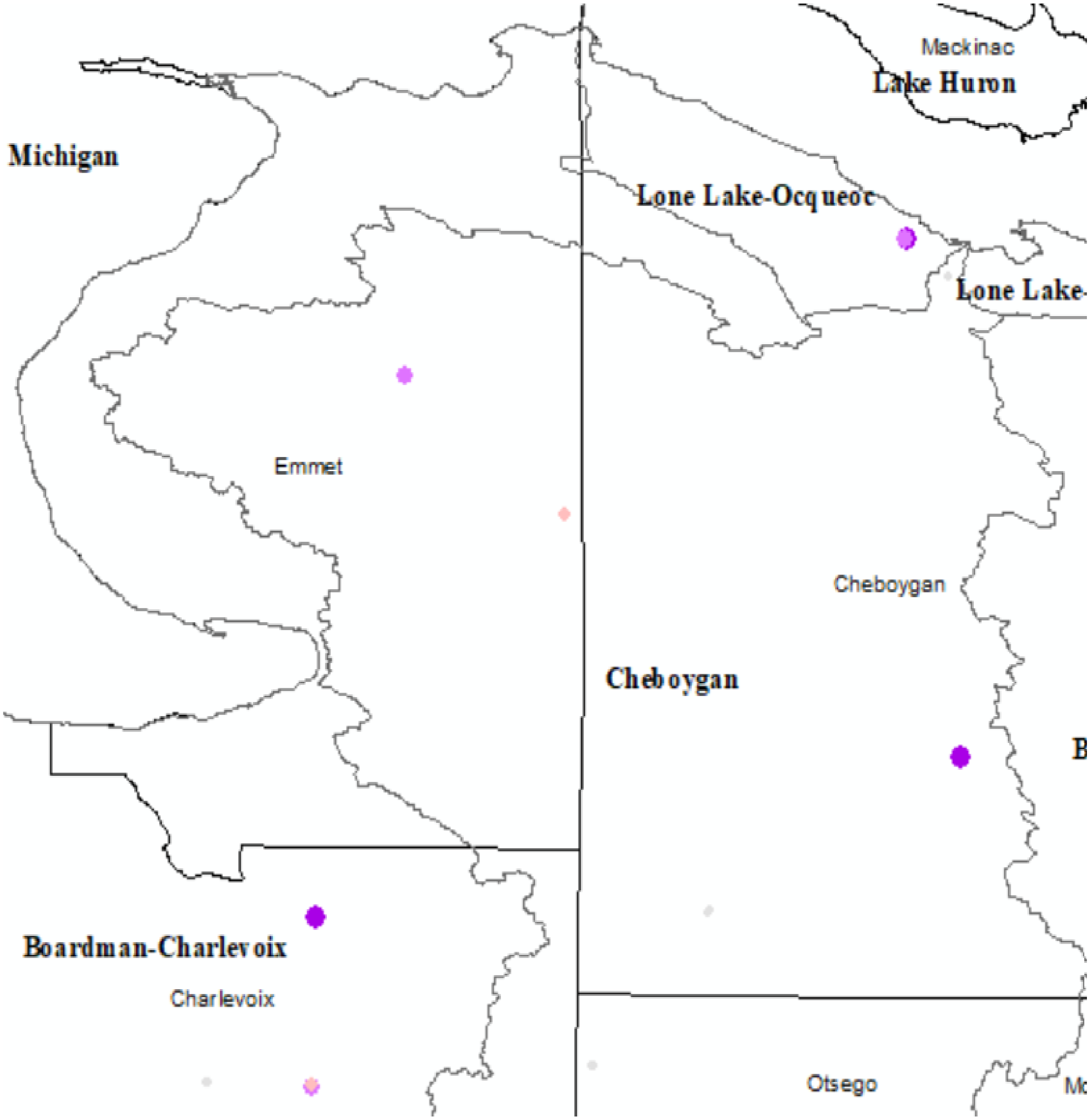

Figure 4.4 – Northern Lower Peninsula *Faxonius rusticus* Prediction Cluster

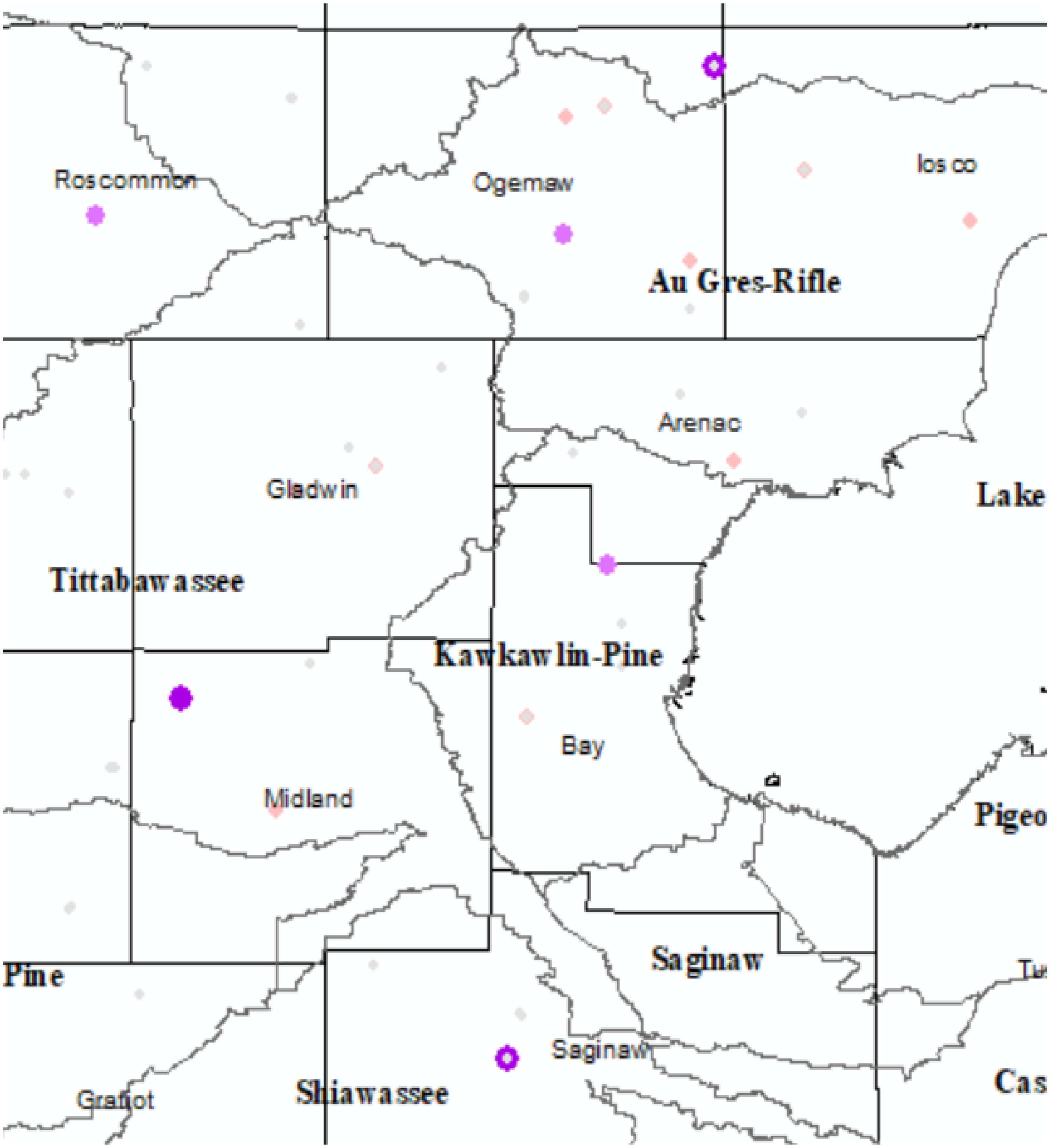

Figure 4.5 – Central East Lower Peninsula *Faxonius rusticus* Prediction Cluster

In this location Woody Wetlands constituted over 53% of the land cover at a local scale and over 75% at the landscape scale there.

Figure 4.6 shows the current locations of *F. rusticus* and 45 vulnerable sites to *F. rusticus* invasion (sites currently unoccupied by *F. rusticus* where the number of times *F. rusticus* was predicted to occur in at least 50% of the predictions). Fourteen of these predicted survey locations contain *F. propinquus*: Clinton River (FID 230), Grand River (FID 695), two sites on the Pere Marquette River (FIDs 707 & 708), Rifle River (FID 532), Maple River (FID 609), Kintz Creek (FID 226), Bluff Creek (FID 355), North Branch Black River (FID 126), Black Creek (FID 672), Wilkes Creek (FID 631), Fox River (FID 875), Davis Creek (FID 149), and North Branch Kalamazoo River (FID 26); eight contain *F. virilis* (Pierce Drain (FID 91), Healy Drain (FID 248), Trout Creek (FID 760), East Branch Sturgeon River (FID 804), Hay Marsh Creek (FID 569), Little Iron River (FID 506), and two sites on Rock Falls Creek (FIDs 282 & 283); and four contain both *F. propinquus* and *F. virilis* (Black Creek (FID 324), two sites on Little Black River (FIDs 612 & 613), and Grand River (FID 694) (see Figure 4.7). FIDs 355, 126, 631, and 760 are the predicted sites closes to current *F. rusticus* populations.

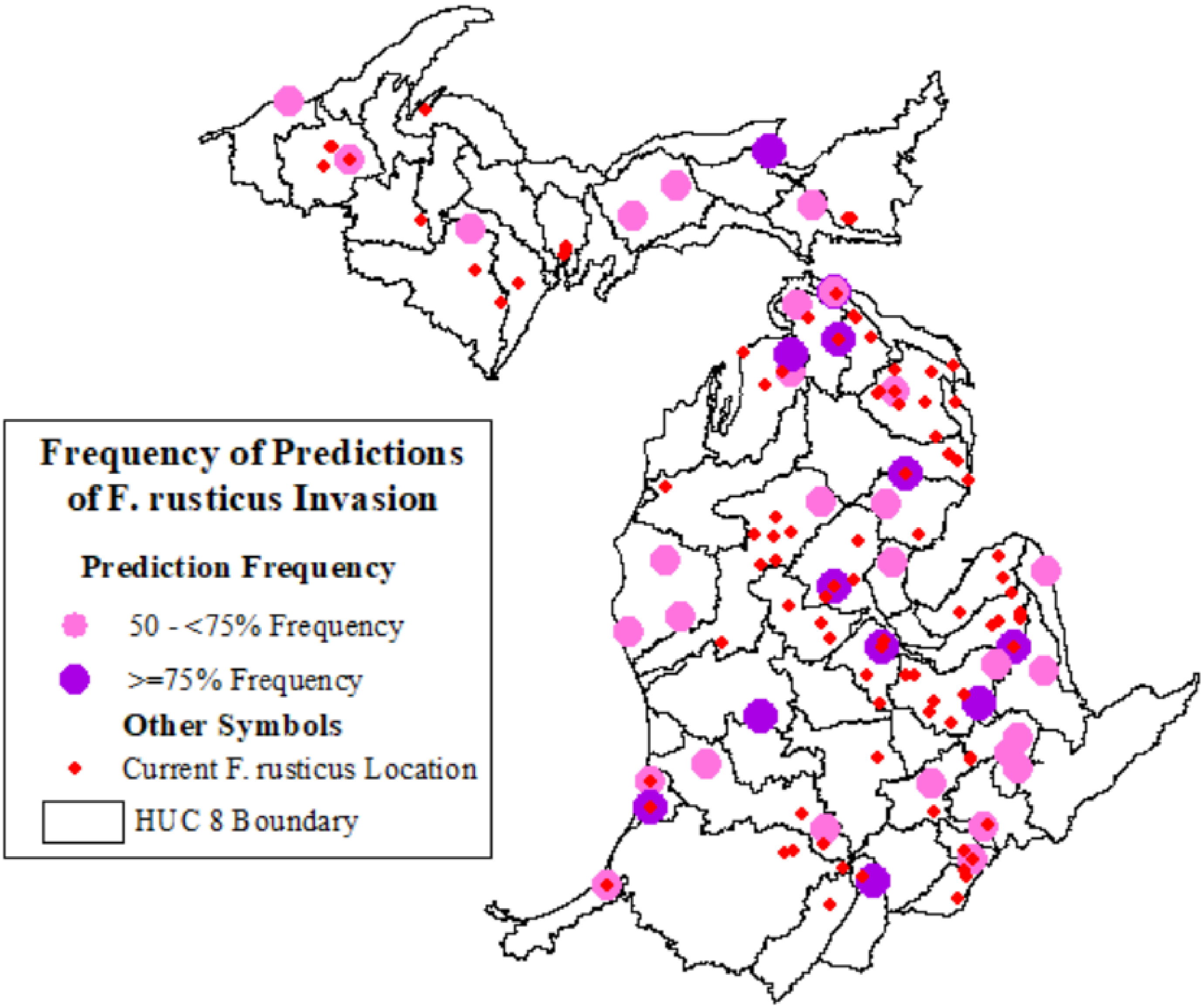

Figure 4.6 – Frequency of Predictions for *Faxonius rusticus* Invasion with the Current Locations of *Faxonius rusticus*

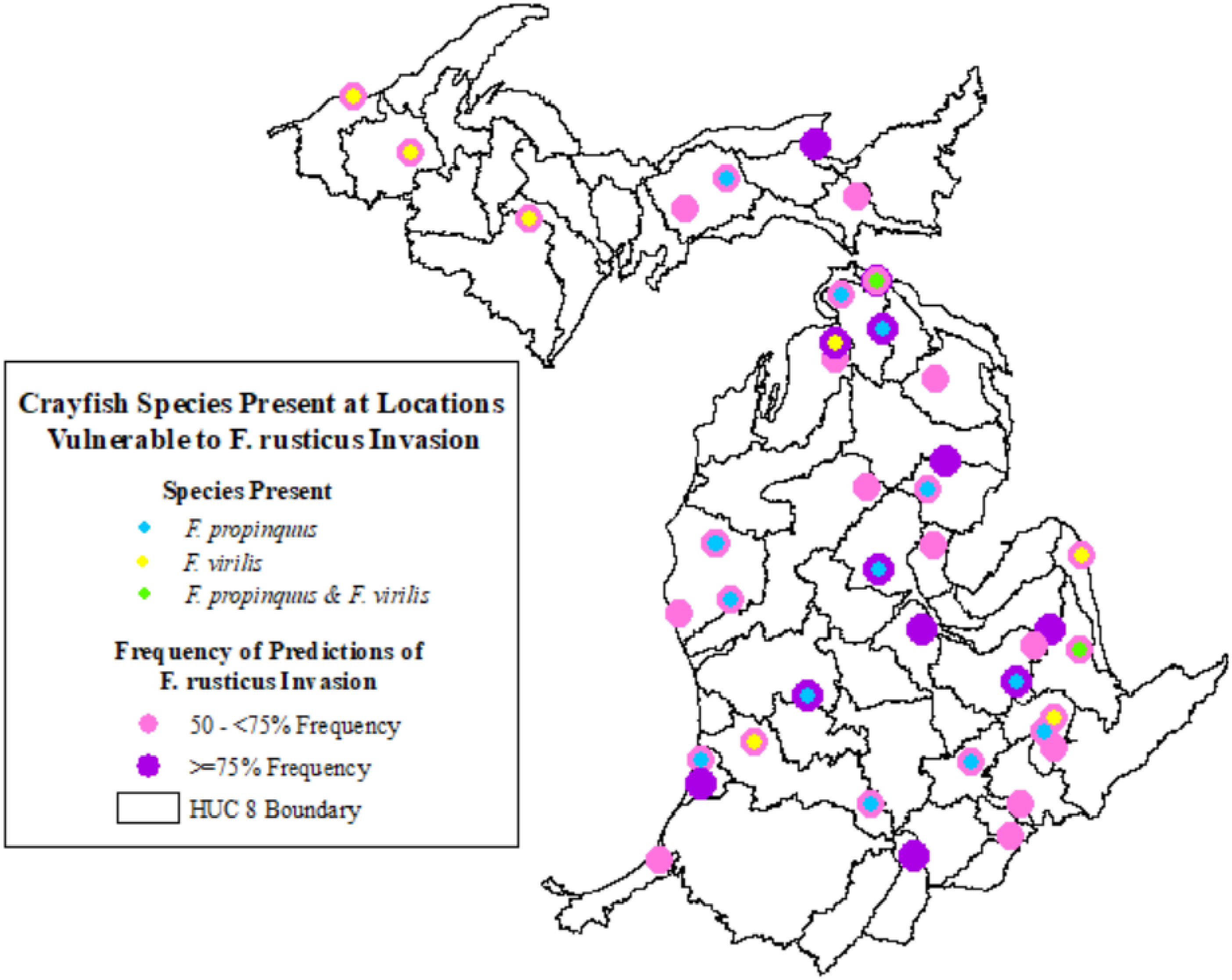

Figure 4.7 – Crayfish Species Present at the Likely *Faxonius rusticus* Invasion Locations

4.3: Selected Patterns in the Accuracy of the ANN for *Faxonius rusticus*

Figures 4.8 shows the accuracy for each survey location across model runs. Figure 4.9 shows the accuracy for sites where *F. rusticus* is present. The false negatives (sites where the ANN predicted that *F. rusticus* were absent, but they were actually present) were primarily concentrated in several areas. The first region consists of the four watersheds bordering Little Bay de Noc in Delta County, in the Tacoosh-Whitefish, two locations in the Cedar-Ford in Menominee County, and three locations in the Menominee Watershed in Menominee and Dickinson Counties. The four survey locations on the Days River and Tacoosh River, reported a false negative in 15 out of the 20 tests. The Days River Sites consists of a greater than average percentage of Low and Medium Intensity Developed land cover. The Tacoosh River sites have a Woody Wetland percentage >65% at both the local and landscape scale. The two sites on Fortyseven Mile Creek reported a false negative for 8 out of the 10 tests. The two sites on Porterfield Creek reported a false negative for 6 of the 9 tests.

The second area of high concentration of false negatives is in the northeast Lower Peninsula, in Cheboygan, Montmorency, Alcona and Ogemaw counties. These occur within the Cheboygan, Black, Thunder Bay, and Au Sable watersheds. The waterbodies where the inaccurate predictions occur include: Black River, Mud Creek, Cheboygan River, Thunder Bay River, Grass Creek, Weber Creek, Buff Creek, Gimlet Creek, and Harper Creek. Interestingly, Grass Creek, in the Thunder Bay Watershed, within Montmorency County, contains a pair of survey locations that are 100% accurate and 0% accurate, respectively.

The third main concentration is found in central Michigan in Osceola, Isabella, Midland, Gratiot, Saginaw, Shiawassee, Ingham, Genesee, and Lapeer Counties. This includes the

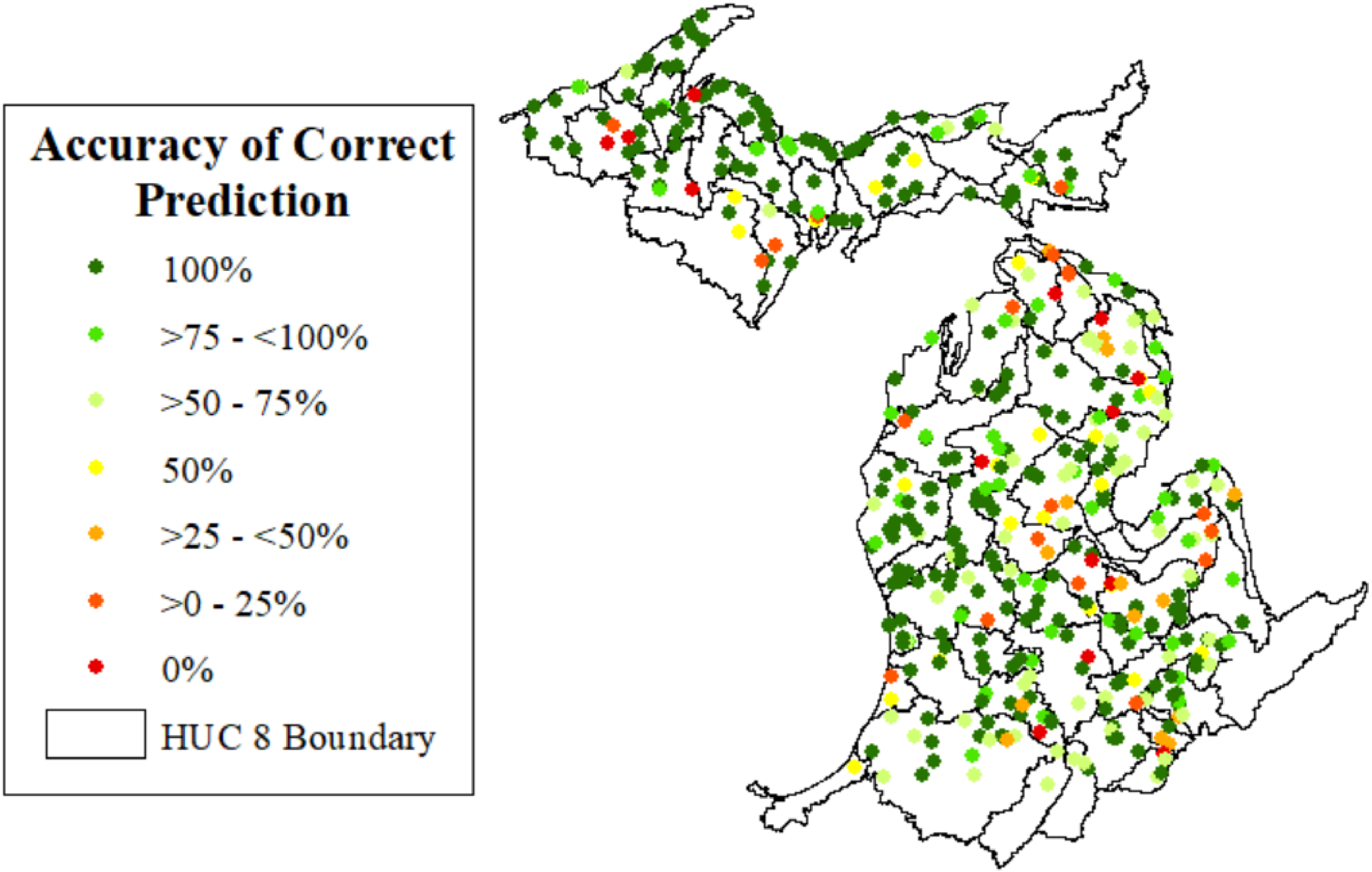

Figure 4.8 – Frequency of Correct *Faxonius rusticus* Presence Findings and Predictions of Possible Invasion Sites

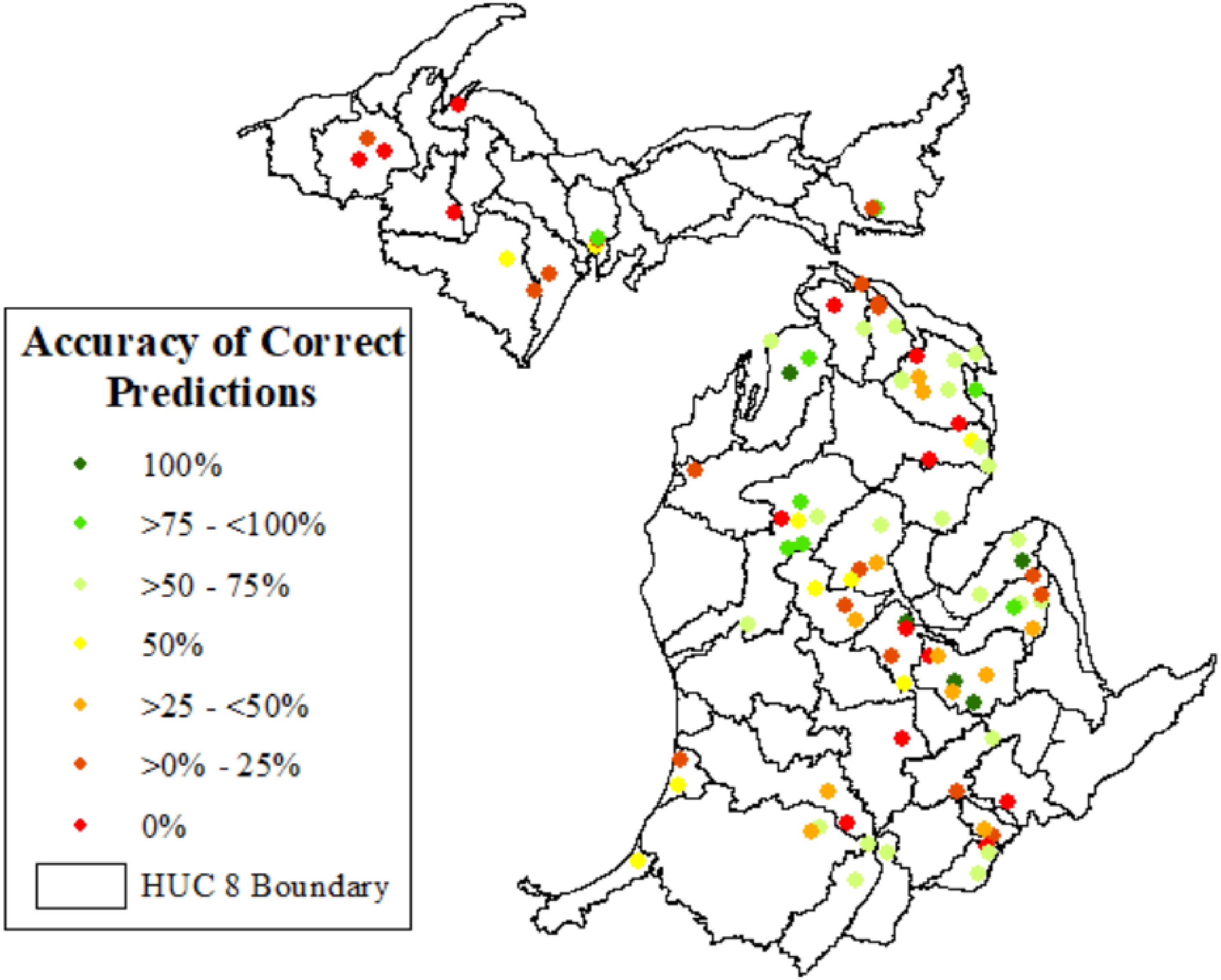

Figure 4.9 – Frequency of Correct *Faxonius rusticus* Presence Findings at Sites where F. rusticus is Present

following watersheds: Muskegon, Tittabawassee, Pine, Saginaw, Upper Grand, and Flint. The waterbodies include: Dishwash Creek, Beebe Creek, Tittabawassee River, South Branch Salt River, Coldwater River, Salt Creek, Pine River, Lamb Creek, Marsh Creek, Brent Run, Misteguay Creek, Shiawassee River, Swartz Creek, Hasler Creek, and Dietz Creek.

Only 14 sites that reported false negatives did not also report a correct positive prediction in at least one of the tests. Conversely, only 10 sites that had correct predictions for positive *F. rusticus* presence had no false negatives.

4.4: Habitat, Physical, and Chemical Limitations of *Faxonius rusticus*

Using the Olden function of the NeuralNetTools package (Beck, 2018), each variable in the model had its relative importance measured. The ten variables with the highest average importance (using absolute values) can be found in Table 4.4. Table 4.4 also contains the average importances with, and without, using the absolute values of these importances. Using standardized values allows us to see the direction of the importance, while the absolute values allow us to see the magnitude of impact for each of the variables.

Table 4.5 provides a further look at the 10 most important values of the study: $CaCO_3$ (high value) concentrations, Barren LULC at the local scale, Open Water LULC at the local scale, the Barren LLC on the landscape scale, the Developed – High Intensity LLC at both scales, the Shrub LLC at the landscape scale, extracid (alkalinity), the Shrub LLC at the local scale, and percent volume of cobble. Five descriptive statistics are provided for each variable: the mean values of the variable when the crayfish being examined is absent; the mean value of the variable when the examined crayfish is present, the mean value of that variable across all sample

Table 4.4 – Ten Most Important Variables to Determining the Presence of *Faxonius rusticus*

Variable	Average Absolute Value of Importance	Average Importance
Norm_caco3_h	1498.6900	-1498.6920
Norm_Barren	1269.6130	-684.8365
Norm_Water	1176.0107	1176.0107
Norm_Barren1k	1152.7069	-351.8329
Norm_DevH	951.1481	462.2786
Norm_DevH1k	938.0564	-847.9884
Norm_Shrub1k	907.3958	-501.9819
Norm_extracid_r	833.2228	-828.8770
Norm_Shrub	793.6285	375.4324
Norm_Cob_All	634.8137	83.9172

locations regardless of crayfish presence, the mean value of the variable at sites deemed vulnerable to invasion (FP); and the mean value of the variable at sites that have been incorrectly declared to be free of the crayfish species. The standard deviation and the standard error of the mean are also provided. Additionally, the means for the 10 most important variables for *F. propinquus* and *F. virilis* are also reported vis-a-vis to *F. rusticus* for comparison (the text is in grey for these variables). A full statistical summary for all variables can be found in Appendix C.

The Developed – High Intensity LULC at the local scale and the landscape scale are the only human made LULC variables among the 10 most important. The variables with a positive contribution with F. rusticus presence were (from greatest magnitude to least) the Open Water LULC at the local scale, the Developed – High Intensity LULC at the Local Scale, the Shrub LULC at the local scale, and the total cobble. The variables with a negative association with F. rusticus were $CaCO_3$ quantity, the Developed – High Intensity LULC at the landscape scale, alkalinity, the Barren LULC at the local scale, the Shrub LULC at the landscape scale, and the Barren LULC at the Landscape Scale.

Additionally, of those 10 variables, only alkalinity (extracid_r), $CaCO_3$ quantity, the Barren LULC at the local and landscape scales, and the Shrub LULC at the local and landscape scales have mean values for Predictions and False Negatives that correspond with the means when *F. rusticus* is present or absent, respectively. The means of Developed High Intensity LULC at the landscape scale are the complete opposite of what one would expect (the prediction mean is closer to the absent mean (.0069 and .0085 respectively) and the false negative mean is closer to the present mean (.0092 and .0091 respectively), as is cobble. The predictions and false positive averages for the Open Water LULC at the local scale both exceeded the average mean when *F. rusticus* was present, while the predictions and false negatives for Developed High Intensity LULC at the local scale both exceeded the average when *F. rusticus* is absent.

Table 4.5 – Variable Averages for Each Crayfish Species' 10 Most Important Variables

Variable Averages for the Each Crayfish Species 10 Most Important Variables													
		Faxonius rusticus				***Faxonius propinquus***				***Faxonius virilis***			
Variables[a]		N[b]	Mean[c]	Std Dev[d]	SEM[e]	N	Mean	Std Dev	SEM	N	Mean	Std Dev	SEM
cobble _all_h	0.0[f]	671	2.6066	5.8463	0.2257	489	3.0429	8.1046	0.3665	668	2.9746	7.5906	0.2937
	1.0[g]	159	3.3836	10.8557	0.8609	341	2.3431	5.2706	0.2854	162	1.8519	4.3184	0.3393
	Ave[h]	830	2.7554	7.0836	0.2459	830	2.7554	7.0836	0.2459	830	2.7554	7.0836	0.2459
	FP[i]	340	1.8647	4.3763	0.2373	701	2.2553	6.2701	1.2309	393	1.9415	4.0398	0.2038
	FN[j]	415	3.1398	9.9002	0.4860	651	2.6160	5.0436	0.1977	428	1.7734	4.1960	0.2028
caco3_h	0.0	671	5.2548	17.5225	0.6764	489	4.2188	13.9657	0.6315	668	4.7799	16.3728	0.6335
	1.0	159	3.4843	13.0446	1.0345	341	5.9150	20.0914	1.0880	162	5.4753	18.3378	1.4408
	Ave	830	4.9157	16.7650	0.5819	830	4.9157	16.7650	0.5819	830	4.9157	16.7650	0.5819
	FP	340	3.2059	12.6221	0.6845	701	4.1141	16.5439	0.3868	393	4.6260	16.6802	0.8414
	FN	415	4.4048	15.3717	0.7546	651	5.3702	18.1330	0.7107	428	5.0070	17.2370	0.8332
extracid _r	0.0	671	4.1352	10.7709	0.4158	489	4.7744	12.5866	0.5692	668	4.1352	11.3684	0.4399
	1.0	159	2.8981	10.6767	0.8467	341	2.6416	7.2177	0.3909	162	2.9210	7.6980	0.6048
	Ave	830	3.8982	10.7575	0.3734	830	3.8982	10.7575	0.3734	830	3.8982	10.7575	0.3734
	FP	340	2.0738	8.9115	0.4833	701	3.8177	10.2403	0.0277	393	4.1623	9.6272	0.4856
	FN	415	3.0520	10.3394	0.5075	651	3.0158	7.5375	0.2954	428	3.9304	9.9956	0.4832
LULC Water	0.0	671	0.0053	0.0233	0.0009	489	0.0066	0.0282	0.0013	668	0.0061	0.0255	0.0010
	1.0	159	0.0135	0.0406	0.0032	341	0.0073	0.0268	0.0014	162	0.0104	0.0349	0.0027
	Ave	830	0.0069	0.0276	0.0010	830	0.0069	0.0276	0.0010	830	0.0069	0.0276	0.0010
	FP	340	0.0135	0.0414	0.0022	701	0.0065	0.0288	0.0011	393	0.0097	0.0297	0.0015
	FN	415	0.0139	0.0402	0.0020	651	0.0080	0.0263	0.0010	428	0.0133	0.0382	0.0018

Table 4.5 – Continued

Variables		*Faxonius rusticus*				*Faxonius propinquus*				*Faxonius virilis*			
		N	Mean	Std Dev	SEM	N	Mean	Std Dev	SEM	N	Mean	Std Dev	SEM
LULC EmWet	0.0	671	0.0137	0.0362	0.0014	489	0.0153	0.0400	0.0018	668	0.0119	0.0318	0.0012
	1.0	159	0.0116	0.0215	0.0017	341	0.0103	0.0223	0.0012	162	0.0191	0.0411	0.0032
	Ave	830	0.0133	0.0339	0.0012	830	0.0133	0.0339	0.0012	830	0.0133	0.0339	0.0012
	FP	340	0.0139	0.0332	0.0018	701	0.0137	0.0325	0.0012	393	0.0160	0.0409	0.0021
	FN	415	0.0118	0.0230	0.0011	651	0.0138	0.0278	0.0011	428	0.0166	0.0359	0.0017
LULC DevM	0.0	671	0.0217	0.0607	0.0023	489	0.0240	0.0652	0.0029	668	0.0184	0.0560	0.0022
	1.0	159	0.0262	0.0632	0.0050	341	0.0205	0.0549	0.0030	162	0.0398	0.0770	0.0060
	Ave	830	0.0225	0.0612	0.0021	830	0.0225	0.0612	0.0021	830	0.0225	0.0612	0.0021
	FP	340	0.0283	0.0670	0.0036	701	0.0221	0.0657	0.0025	393	0.0331	0.0762	0.0038
	FN	415	0.0259	0.0577	0.0028	651	0.0205	0.0571	0.0022	428	0.0265	0.0619	0.0030
LULC DevH	0.0	671	0.0099	0.0469	0.0018	489	0.0091	0.0449	0.0020	668	0.0078	0.0414	0.0016
	1.0	159	0.0085	0.0362	0.0029	341	0.0103	0.0453	0.0025	162	0.0169	0.0572	0.0045
	Ave	830	0.0096	0.0450	0.0016	830	0.0096	0.0450	0.0016	830	0.0096	0.0450	0.0016
	FP	340	0.0129	0.0434	0.0024	701	0.0134	0.0596	0.0023	393	0.0167	0.0626	0.0032
	FN	415	0.0114	0.0497	0.0024	651	0.0064	0.0237	0.0009	428	0.0121	0.0483	0.0023
LULC Grass	0.0	671	0.0228	0.0583	0.0022	489	0.0234	0.0616	0.0028	668	0.0220	0.0562	0.0022
	1.0	159	0.0159	0.0323	0.0026	341	0.0186	0.0416	0.0023	162	0.0193	0.0458	0.0036
	Ave	830	0.0215	0.0543	0.0543	830	0.0215	0.0543	0.0019	830	0.0215	0.0543	0.0019
	FP	340	0.0196	0.0196	0.0458	701	0.0222	0.0595	0.0022	393	0.0175	0.0572	0.0029
	FN	415	0.0160	0.0160	0.0323	651	0.0222	0.0518	0.0020	428	0.0168	0.0472	0.0023

Table 4.5 – Continued

	Faxonius rusticus					***Faxonius propinquus***				***Faxonius virilis***			
Variables		N	Mean	Std Dev	SEM	N	Mean	Std Dev	SEM	N	Mean	Std Dev	SEM
LULC Shrub	0.0	671	0.0057	0.0234	0.0009	489	4.7744	12.5866	0.5692	668	0.0058	0.0233	0.0009
	1.0	159	0.0028	0.0087	0.0007	341	2.6416	7.2177	0.3909	162	0.0024	0.0098	0.0008
	Ave	830	0.0052	0.0214	0.0007	830	3.8982	10.7575	0.3734	830	0.0052	0.0214	0.0007
	FP	340	0.0023	0.0073	0.0004	701	0.0066	0.0261	0.0010	393	0.0033	0.0134	0.0007
	FN	415	0.0040	0.0101	0.0005	651	0.0041	0.0154	0.0006	428	0.0032	0.0123	0.0006
LULC Barren	0.0	671	0.0055	0.0246	0.0010	489	0.0039	0.0201	0.0009	668	0.0037	0.0188	0.0007
	1.0	159	0.0022	0.0065	0.0005	341	0.0062	0.0252	0.0014	162	0.0097	0.0329	0.0026
	Ave	830	0.0049	0.0224	0.0008	830	0.0049	0.0224	0.0008	830	0.0049	0.0224	0.0008
	FP	340	0.0042	0.0179	0.0010	701	0.0042	0.0197	0.0007	393	0.0069	0.0230	0.0012
	FN	415	0.0037	0.0084	0.0004	651	0.0047	0.0220	0.0009	428	0.0078	0.0291	0.0014
LULC DevM1k	0.0	671	0.0199	0.0589	0.0023	489	0.0231	0.0676	0.0031	668	0.0153	0.0500	0.0019
	1.0	159	0.0224	0.0573	0.0045	341	0.0165	0.0422	0.0023	162	0.0414	0.0822	0.0065
	Ave	830	0.0204	0.0586	0.0020	830	0.0204	0.0586	0.0020	830	0.0204	0.0586	0.0020
	FP	340	0.0201	0.0665	0.0036	701	0.0178	0.0545	0.0021	393	0.0288	0.0687	0.0035
	FN	415	0.0192	0.0497	0.0024	651	0.0158	0.0420	0.0016	428	0.0266	0.0657	0.0032
LULC DevH1k	0.0	671	0.0085	0.0359	0.0014	489	0.0082	0.0340	0.0015	668	0.0065	0.0331	0.0013
	1.0	159	0.0091	0.0329	0.0026	341	0.0091	0.0372	0.0020	162	0.0171	0.0425	0.0033
	Ave	830	0.0086	0.0354	0.0354	830	0.0086	0.0354	0.0012	830	0.0086	0.0354	0.0012
	FP	340	0.0069	0.0069	0.0244	701	0.0099	0.0445	0.0017	393	0.0137	0.0491	0.0025
	FN	415	0.0092	0.0092	0.0368	651	0.0061	0.0229	0.0009	428	0.0094	0.0344	0.0017

Table 4.5 – Continued

		Faxonius rusticus				***Faxonius propinquus***				***Faxonius virilis***			
Variables		N	Mean	Std Dev	SEM	N	Mean	Std Dev	SEM	N	Mean	Std Dev	SEM
LULC Shrub1k	0.0	671	0.0096	0.0291	0.0011	489	0.0106	0.0323	0.0015	668	0.0093	0.0279	0.0011
	1.0	159	0.0052	0.0127	0.0010	341	0.0061	0.0153	0.0008	162	0.0065	0.0214	0.0017
	Ave	830	0.0088	0.0268	0.0009	830	0.0088	0.0268	0.0009	830	0.0088	0.0268	0.0009
	FP	340	0.0038	0.0106	0.0006	701	0.0105	0.0255	0.0010	393	0.0072	0.0214	0.0011
	FN	415	0.0076	0.0162	0.0008	651	0.0083	0.0197	0.0008	428	0.0074	0.0238	0.0011
LULC Barren1k	0.0	671	0.0039	0.0177	0.0007	489	0.0032	0.0144	0.0007	668	0.0024	0.0089	0.0003
	1.0	159	0.0025	0.0068	0.0005	341	0.0042	0.0185	0.0010	162	0.0085	0.0316	0.0025
	Ave	830	0.0036	0.0162	0.0006	830	0.0036	0.0162	0.0006	830	0.0036	0.0162	0.0006
	FP	340	0.0040	0.0211	0.0011	701	0.0034	0.0198	0.0007	393	0.0049	0.0119	0.0006
	FN	415	0.0030	0.0068	0.0003	651	0.0037	0.0201	0.0008	428	0.0044	0.0140	0.0007

[a] Names and detailed description of all variables can be found in Table 3.1
[b] Number of survey locations
[c] The mean value of the variable across the selected number of survey locations
[d] Standard deviation of the mean
[e] Standard error of the mean
[f] Crayfish species is not present
[g] Crayfish species is present
[h] Average across all survey locations regardless of crayfish species presence
[i] Crayfish species was predicted to occur at a site it was absent from (false positive)
[j] Crayfish species was predicted to not be present at a location they were present at (false negative)

4.5: *Faxonius propinquus* Accuracy, Presence and Predictions

The ANN model developed to predict presence/absence of *F. rusticus*, was similarly used to predict presence/absence of *F. propinquus*. The sample location Row 812 (FID 853) was never selected and thus was not included in predictions. For *F. propinquus*, the number of times each survey location was sampled varied from 1 to 10 for all sample locations (aside from Row 812).

As with *F. rusticus*, the ANN for *F. propinquus* was ran 10 times with a training set consisting a random selection of half the number of survey locations where *F. propinquus* was present (170) and half the number of survey locations where there were *F. propinquus* was absent (245). The outputs were then converted back to shapesfiles (R to SPSS to Excel to ArcMap). The confusion matrices of the 10 selections and their specificities can be found in Appendix B. The accuracy of the model runs ranged from 0.6602 (Tests 1 and 2) to 0.7060 (Test 7) with an average of 0.6742 (see Table 4.6). Unlike the situation for *F. rusticus*, the predictions indicate a population expansion of a native species, not an invasive one. The specificities ranged from 0.5660 (Test 9) to 0.6322 (Test 3) with an average of 0.6013 (see Table 4.6).

The accuracy of predictions for the presence *F. propinquus* tends to be inaccurate in several locations (see Figure 4.10). The first area of issue is the central-east part of the Lower Peninsula, in the Au Gres River-Rifle, Tittabawassee, Kawkawlin-Pine, and Muskegon Watersheds (Roscommon, Ogemaw, Iosco, Arenac, Gladwin, Clare, and Midland Counties). The waterbodies in the region include: Rifle River, South Branch Eddy Creek (0% accuracy vs. 100% accurate paired survey location), Tittabawassee River (including East Branch, which has a pair

Table 4.6 – Accuracy and Specificity for Each ANN Test Performed on *Faxonius propinquus*

Test	Accuracy	Specificity
1	0.6602	0.5852
2	0.6602	0.5843
3	0.6988	0.6322
4	0.6651	0.5899
5	0.6916	0.6257
6	0.6771	0.6121
7	0.7060	0.6256
8	0.6699	0.5944
9	0.6747	0.5978
10	0.6386	0.5660

of coordinates with ~17% accuracy and 100% accuracy), Knappen Creek (20% accuracy vs. 100% accurate pair), Van Drain, Hale Creek, North Branch Tobacco River, North Branch Pine River, Vaughn Creek (40% vs. 100% accuracy), and Johnson Creek (40% vs. 100% accuracy).

The second major cluster of model inaccuracy is found in the Southeast Lower Peninsula, specifically Oakland, Macomb, St. Clair, and Lapeer Counties. These include the Detroit, Clinton, St. Clair, and Flint Watersheds. The affected waterbodies are: River Rouge (Main and Upper Branch) (20% vs. 100% accuracy for both sets of pairs), Sashabaw Creek, Paint Creek, Clinton River, Rattle Run, Kintz Creek (20% vs. 100% accuracy), Hunter's Creek, Hasler Creek (~17% vs. 100% accuracy), and Cedar Creek (~17% vs. 100% accuracy). This region contains a high percentage of Pasture/Hay and "Developed" lands.

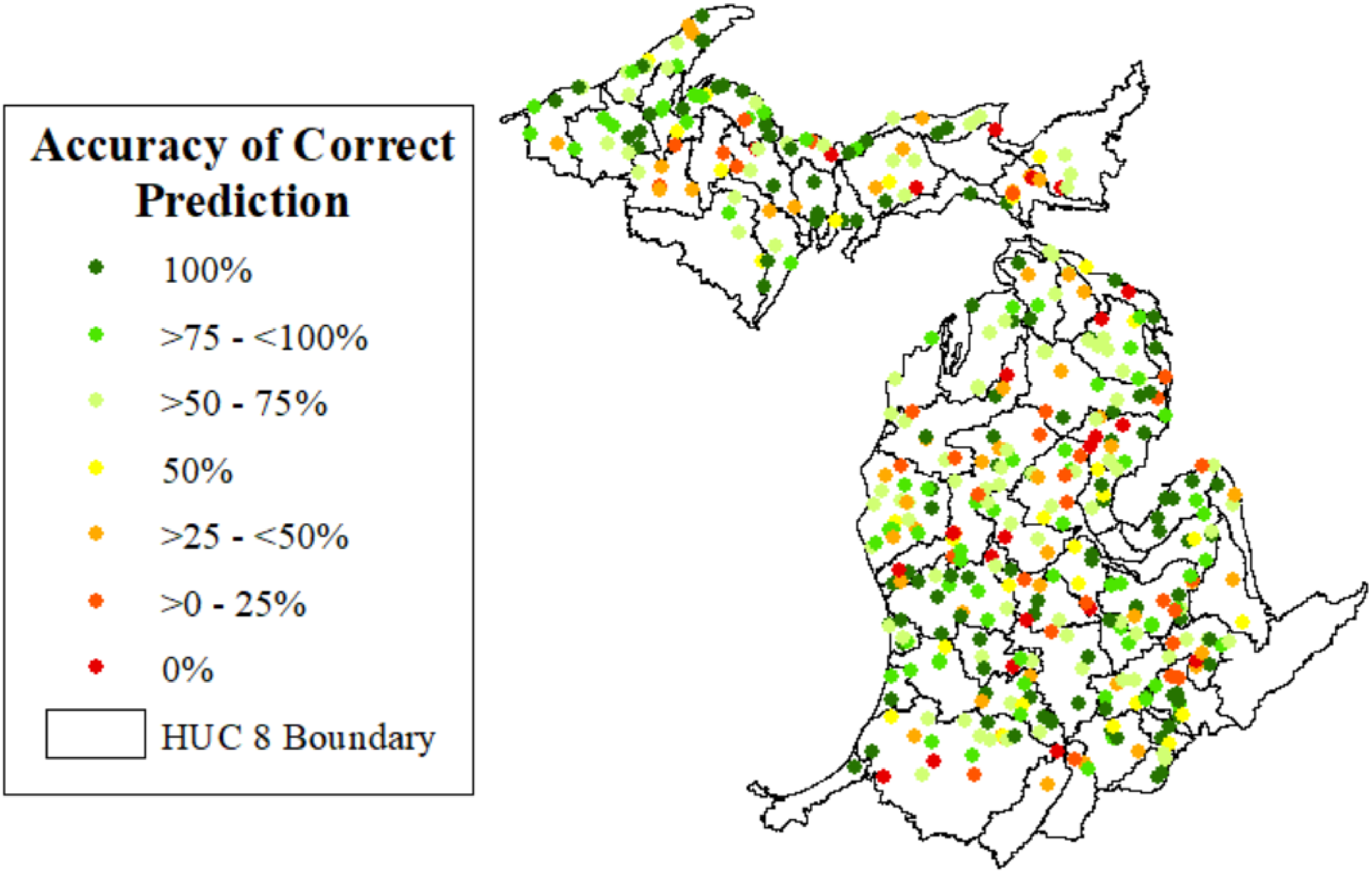

Figure 4.10 – Frequency of Correct *Faxonius propinquus* Presence Findings and Predictions of Possible Invasion Site

The third major cluster of inaccurate model predictions is found in eastern Upper Peninsula, specifically Chippewa and Mackinac Counties. These sites occur in the Tahquamenon, Carp-Pine, and St. Marys Watersheds. The affected waterbodies include: Cheney Creek (0% vs. 100% accuracy for a pair of coordinates), McMahen Creek, Biscuit Creek (20% vs. 100% accuracy), Trout Brook (0% vs. 100% accuracy), Pine River, Taylor Creek, and Carp River (~14% vs. 100% accuracy). The McMahen Creek, Biscuit Creek, Trout Brook, and Pine River sample locations all occur in some of the highest concentrations of Pasture/Hay LLC in Michigan.

Predictions of potential future locations where *F. propinquus* could be present are fairly evenly distributed throughout the entire state (see Figure 4.11).

As with *F. rusticus*, the Olden function of the NeuralNetTools package in R (Beck, 2018) were used to measure the importance of each variable in the equation. The ten variables with the highest average importance (using absolute values) can be found in Table 4.7. Table 4.7 also contains the average importances without using the absolute values. The same five descriptive statistics recorded for *F. rusticus* in Table 4.5 were also recorded for *F. propinquus* and also listed in Table 4.5. A full statistical summary for all variables can be found in Appendix C.

The Developed – High Intensity LULC at the local scale was the only variable to have a positive contribution to *F. propinquus* prediction, with the other 9 variable all having negative contributions.

Of these 10 variables, only the Developed – High Intensity LULC at both local and landscape scales had averages for predictions and false negatives that mirrored the averages

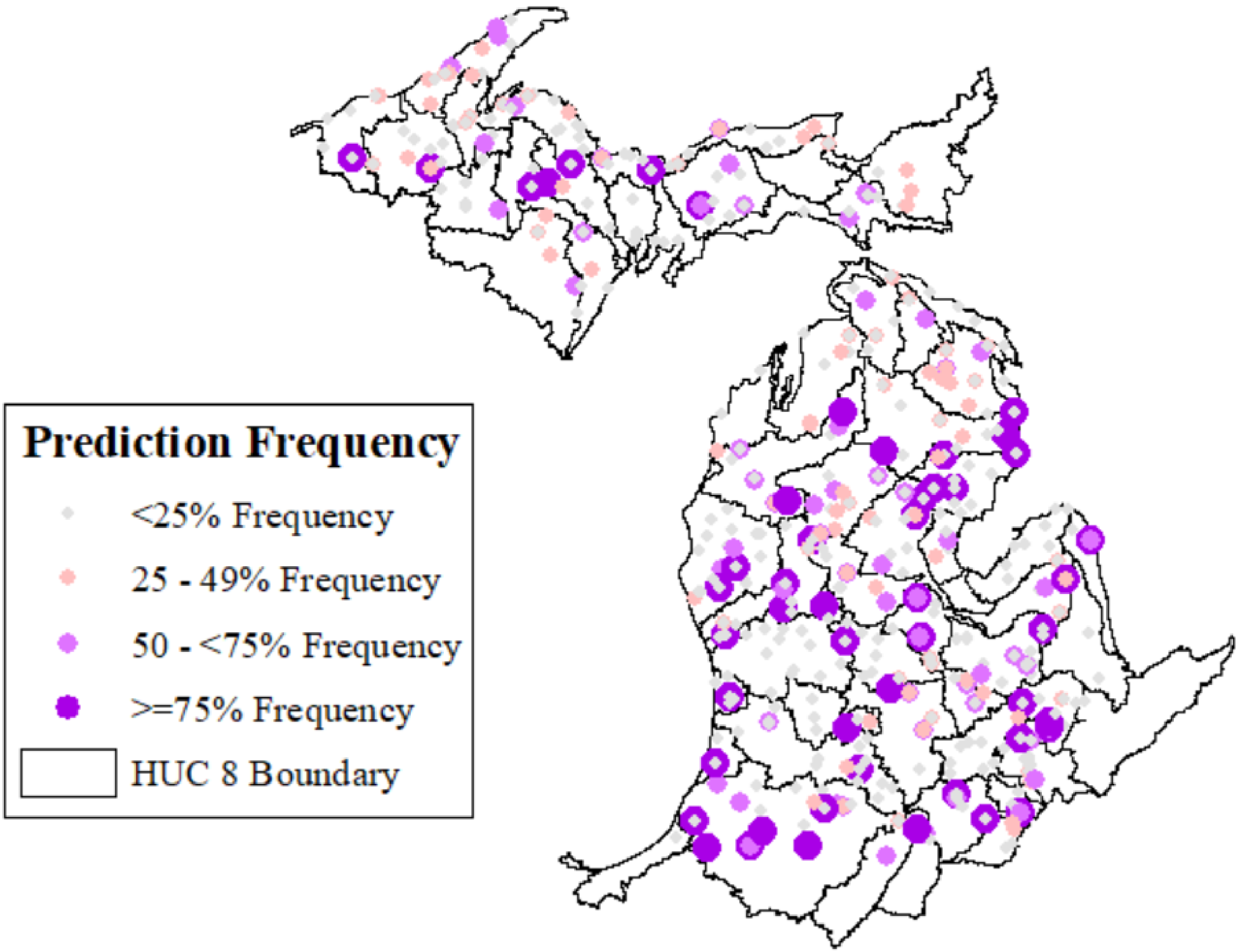

Figure 4.11 – Frequency of *Faxonius propinquus* Prediction at Sites Where They are Currently Absent

Table 4.7 – Ten Most Important Variables to Determining the presence of *F*axonius *propinquus*

Variable	Average Absolute Value of Importance	Average Importance
Norm_DevH	1790.9417	1790.9417
Norm_DevM1k	1475.8996	-1475.8996
Norm_Shrub1k	1364.9095	-1276.6724
Norm_Barren	1349.2817	1246.4763
Norm_Water	1295.7487	-1295.7487
Norm_caco3_h	1238.0520	-64.2185
Norm_EmWet	1046.5363	-799.1689
Norm_Barren1k	1025.4446	-1025.4459
Norm_DevH1k	949.8987	-817.9723
Norm_DevM	900.0215	-568.7712

where *F. propinquus* was present and absent, respectively. $CaCO_3$ quantity, the Open Water at the local scale, Shrub land at the landscape scale, and Barren Land at the landscape scale all had means for the predictions and false positives that were closer to the absent and present means, respectively.

4.6: *Faxonius virilis* Accuracy, Presence and Predictions

The ANN model developed to predict presence/absence of *F. rusticus*, was similarly used to predict presence/absence of *F. virilis*. All 830 survey locations were selected across the 10 tests. For *F. virilis*, the number of times each survey location was sampled varied from 1 to 10 for all sample locations.

After running the ANN, using sites with 81 sample locations with *F. virilis* and 334 without, with 10 iterations, the results were converted back to shapefiles (R to SPSS to Excel to ArcMap). The confusion matrices of the 10 selections and their specificities can be found in Appendix B. The accuracy of these confusion matrices ranged from 0.7807 (Test 2) to 0.8265

(Test 8) with an average of 0.8022 (see Table 4.8). The specificities ranged from 0.4419 (Test 2) to 0.5672 (Test 8) with an average of 0.4949 (see Table 4.8).

Table 4.8 – Accuracy and Specificity for each ANN Test Performed on *Faxonius virilis*

Test	Accuracy	Specificity
1	0.8048	0.5000
2	0.7807	0.4419
3	0.7855	0.4459
4	0.8000	0.4906
5	0.7880	0.4533
6	0.8072	0.5062
7	0.8096	0.5147
8	0.8265	0.5672
9	0.8241	0.5526
10	0.7952	0.4767

The accuracy of predictions in regards to *F. virilis* presence tends to be inaccurate in the following locations: Southwest Lower Peninsula (in the Little Calumet-Galien, (West) St. Joseph, Kalamazoo, Lower Grand, and Thornapple Watersheds) (East Branch Paw Paw River, Le Feure Lake Outlet, Sheldon Creek, Battle Creek, Little Thornapple Creek, Butler Creek, Grand River, and Black Creek), the southeastern Lower Peninsula (in the Huron and Detroit Watersheds) (Huron River, Mill Creek, Fleming Creek, Ecorse River, Middle Branch River Rouge), the Thumb (in the Flint, St. Clair, Birch-Willow, and Pigeon-Wiscoggin Watersheds) (Bryan Drain, Elm Creek, Cedar Creek, Black Creek, Rock Falls Creek, New River, and Wiscoggin Drain), and the Middle of the Upper Peninsula (in the Cedar-Ford, Escanaba, Betsy-Chocolay, and Fishdam-Sturgeon Watersheds) (West Branch Ford River, Squaw Creek, AuTrain River, and Sturgeon River) (see Figure 4.12). The locations within the Lower Peninsula with inaccurate predictions all contain a high percentage of Cultivated Croplands.

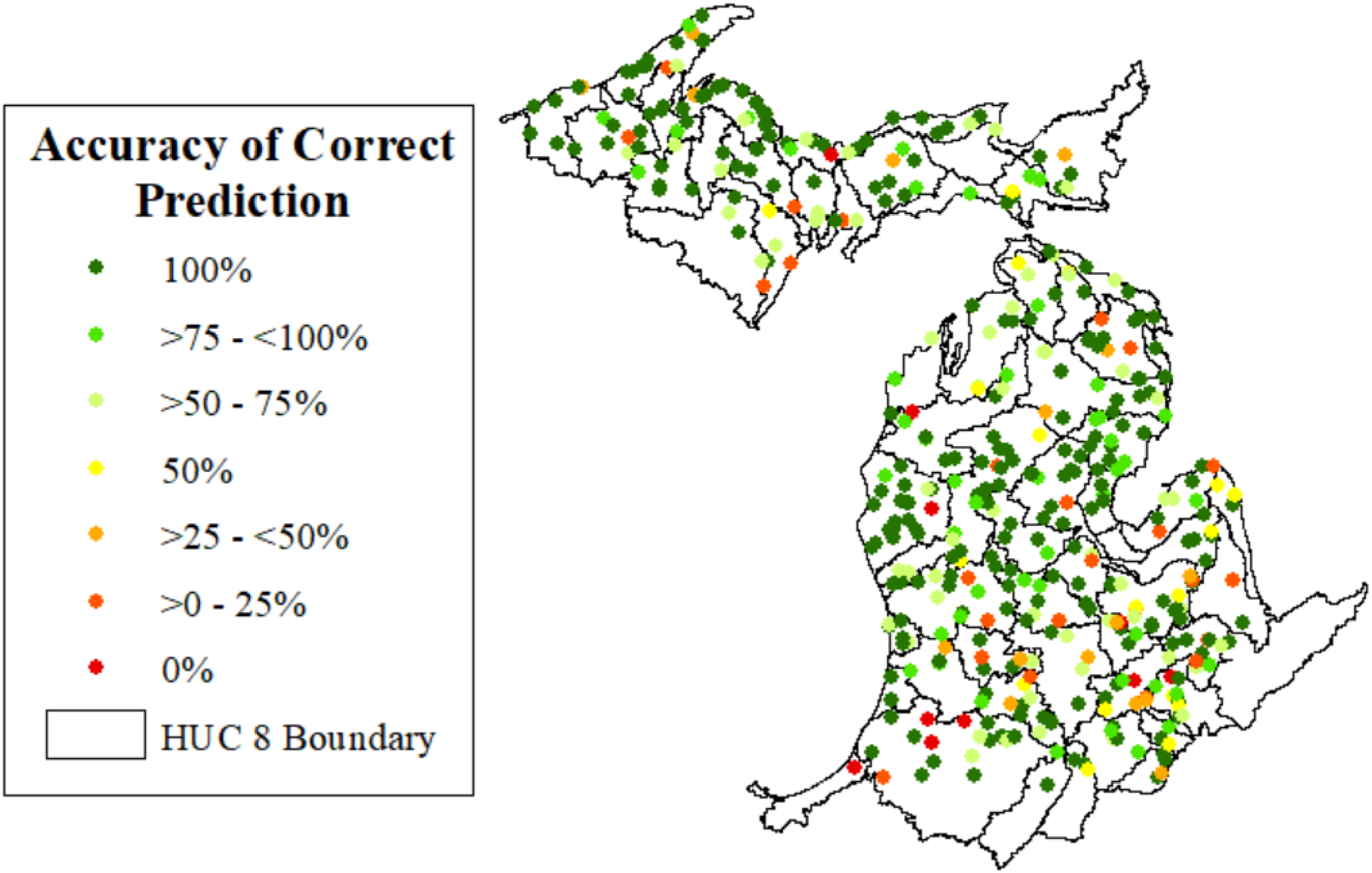

Figure 4.12 – Frequency of Correct *Faxonius virilis* Presence Findings and Predictions of Possible Invasion Sites

Predictions of potential future locations where *F. virilis* are concentrated in the southern region of the Lower Peninsula, forms a pattern resembling an uppercase italic N (*N*) (see Figures 4.13 and 4.14). These predictions of presence predominantly occur in the (West) St. Joseph, Kalamazoo, Thornapple, Maple, Shiawassee, Flint, Huron, Clinton, Detroit, Ottawa-Stony, St. Clair, Cass, Pigeon-Wiscoggin, and Birch-Willow Watersheds. This region primarily consists of Cultivated Cropland, and has the state's highest concentration of "Developed" land. Notable waterbodies where these predictions occurred are in include: the River Rouge (Upper and Middle Branches), Clinton River, Red Run, Davis Creek, Porter Drain, Atherton Drain, Spencer Drain, Gibson Drain, Sandy Creek, Little Lake Creek, Black Creek, Pigeon River (West Branch Extension), East Branch Willow Creek, New River, Williams Creek, Marsh Creek, West Branch Flint River, Bryan Drain, Elm Creek Holden Drain, Stony Creek, Butler Creek, Little Thornapple River, Green Lake Creek, Big Creek, North Branch Rice Creek, Bear Creek, and Church Drain.

As with *F. rusticus* and *F. propinquus*, the Olden function of NeuralNetTools package (Beck, 2018) measured the importance of each variable in the equation. The ten most important variables based on the highest average importance (using absolute values) can be found in Table 4.9. Table 4.9 also provides the average importances with, and without, using the absolute values of these importances. Each of these variables had the same statistical summaries recorded in Table 4.5 as those of *F. rusticus* and *F. propinquus* (a full statistical summary for all variables can be found in Appendix C).

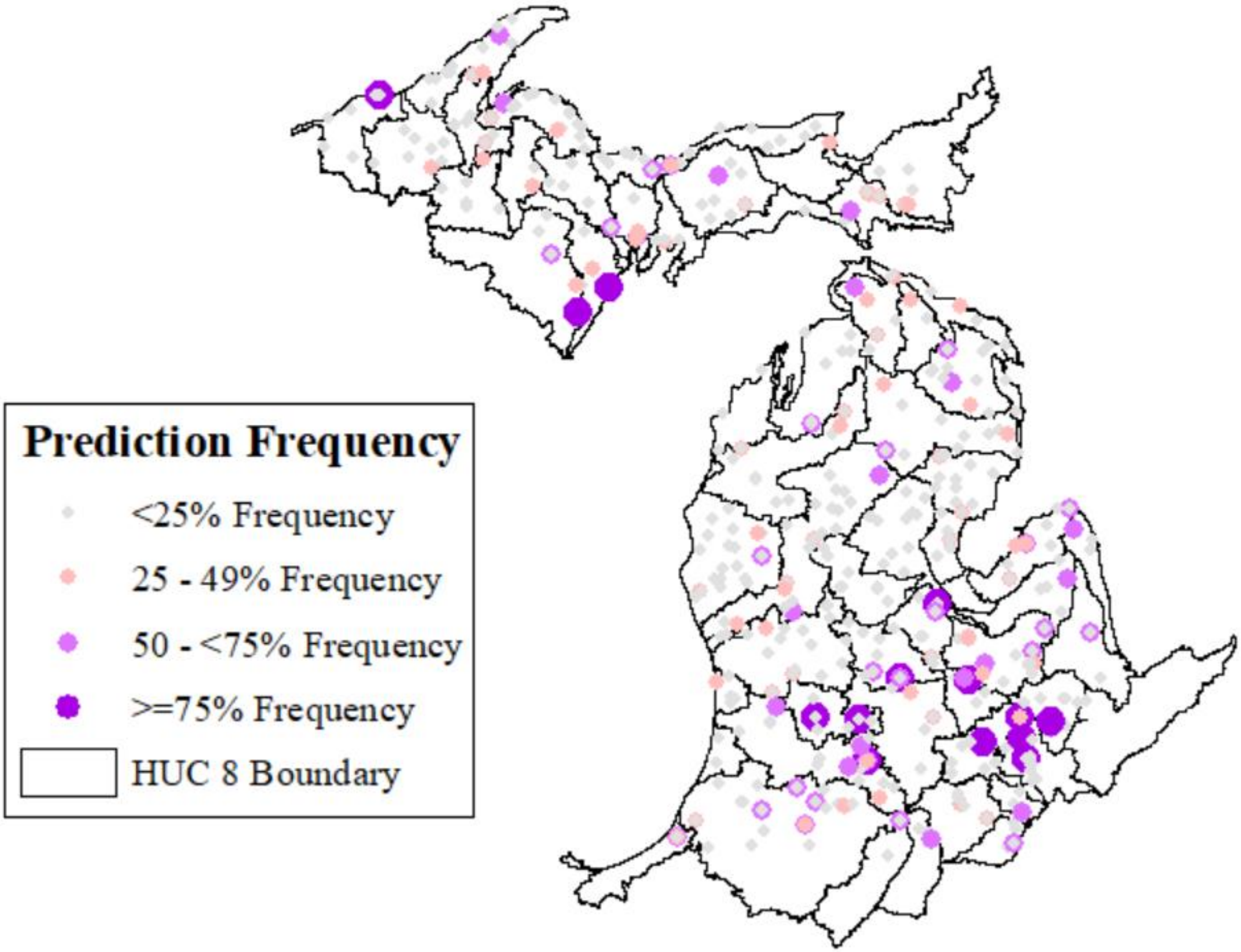

Figure 4.13 – Frequency of *Faxonius virilis* Prediction at Sites Where They are Currently Absent

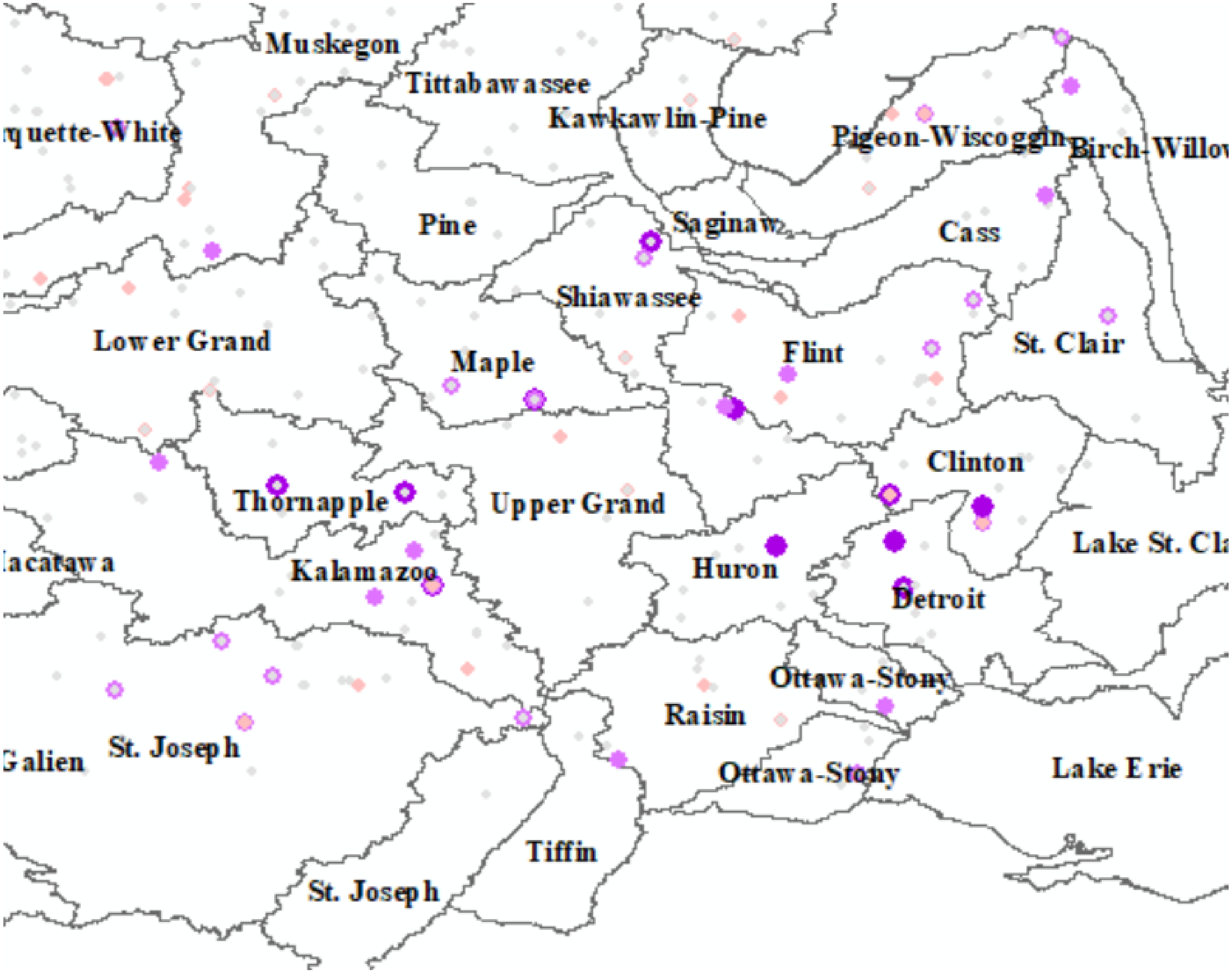

Figure 4.14 – Predictions of *Faxonius virilis* in the Southern Lower Peninsula

Table 4.9 – Ten Most Important Variables to Determining the presence of *Faxonius virilis*

Variable	Average Absolute Value of Importance	Average Importance
Norm_DevH1k	2165.4366	2165.4366
Norm_DevH	1400.8876	-1400.8876
Norm_Barren1k	1351.4926	1351.4926
Norm_Shrub	1134.8926	-1025.5514
Norm_caco3_h	1130.2998	-1087.9820
Norm_EmWet	1000.9253	705.4610
Norm_Grass	965.0900	-570.1171
Norm_DevM	893.8099	-629.8456
Norm_Water	890.6050	-478.8423
Norm_extracid_r	850.1770	-114.9232

Of the 10 variables, the Developed – High Intensity LULC at the landscape scale, the Barren LULC at the landscape scale, and the Emergeant Woody Wetland LULC at the local scale were the only variables to have positive contributions to *F. virilis* predictions. The Developed – High Intensity LULC at the local scale, $CaCO_3$ quantity, Shrub LULC at the local scale, Developed – Medium Intensity LULC at the local scale, Grassland LULC at the local scale, Open Water LULC at the local scale, and alkalinity had negative contributions to *F. virilis* predictions.

Additionally, only High and Medium Intensity Developed land at the local scale and High Intensity Developed Land at the landscape scale had averages for predictions and false negatives that mirrored the averages where *F. virilis* was present and absent, respectively. $CaCO_3$ quantity and Open Water at the local scale had means for the predictions and false positives that were closer to the absent and present means, respectively. At the local scale, Grasslands and Shrubland means for predictions and false negatives were both far below the mean when *F. virilis* was present (by a significant degree in Shrubland (0.0032 and 0.0033 versus 0.0193)).

CHAPTER V

DISCUSSION

The primary purpose of this research was to create a model to determine the waterbodies most vulnerable to invasion by *F. rusticus*, to identify those variables that most relate to that vulnerability and thus could facilitate/impede their invasion pathways, while comparing the presence/absence of the native *F. propinquus* and *F. virilis* using the same variables. This chapter is presented in 5 sections. Section 5.1 highlights the sites vulnerable to *F. rusticus* invasion and the eminence of its threat. Section 5.2 elaborates the assessment of the accuracy of the model. Section 5.3 discusses the variables with the highest importance based on metrics in the model, how they contribute to the predictions, how they compare to the average results of the presence/absence of the each of the three species, and why these variables may have high importance. Section 5.4 explains the limitations of this research. Finally, Section 5.5 suggests future research opportunities based on these findings.

5.1: Predictions of *Faxonius rusticus* Expansion and Threat to Native Species

Table 4.1 and Figure 4.1 showed that *F. rusticus* only co-occurred with *F. propinquus* and/or *F. virilis* at 41 of the 159 total locations where it currently exists (Smith et al., 2018). It was the sole crayfish species occupying the other 118 survey locations (Smith et al., 2018). Those co-occurrences are fewer than the number of co-occurrences between *F. propinquus* and *F. virilis* (57, or 63 if you count the locations where all three species were found together) (Smith et al., 2018). The fewer occurrences of native crayfish species at locations where *F. rusticus* is found is supported by the existing literature. After sampling 99 rivers and streams in southern Ontario, Reid & Nocera (2015) found that *F. propinquus* (the most common crayfish species in the region) made up a mean value of 72.5% of the total crayfish catch in sites where *F. rusticus* was absent, but only made up 12.2% of the mean crayfish catch when *F. rusticus* was present. At Lake of the Woods in Ontario, Canada, Jansen et al. (2009) determined that *F. rusticus* contributed to >80% of the total crayfish catch at 47 of the 78 sites they were the most prevalent species in.

These studies, as well as others (Olden et al., 2006; Olden et al., 2011) highlight the vulnerability of native crayfish to the presence of *F. rusticus*. As stated in the results, *F. rusticus* was predicted to occur at uninvaded waterbodies 340 times at 215 different locations. After removing the sites where the frequency of presence prediction was less than 50%, 45 locations were predicted to be vulnerable to *F. rusticus* invasion as they are similar to other sites where *F. rusticus* occurs (see Figure 4.3). Fourteen of these predicted survey locations contain *F. propinquus*; eight contain *F. virilis*; and four contain both *F. propinquus* and *F. virilis* (see Figure 4.4). The native crayfish at all 26 of these sites are in danger of extirpation. Four of these survey locations are in extreme danger as *F. rusticus* are present in streams within close proximity to them. Specifically, the specific sites are located at Trout Creek (FID 760), North Branch Black River (FID 126), Bluff Creek (FID 355) and Wilkes Creek (FID 631). Additionally, *F. rusticus* were predicted to be within close proximity of the following sampling sites where no species of crayfish were found: Red Run (FID 685), Galien River (FID 674), Harper Creek (FID 538), Marsh Creek (FID 307), Thunder Bay River (FID 585), Sandy Creek (FID 192), and South Branch Cass River (FID 263). *F. rusticus* could already be present in these streams given their proximity to pre-existing *F. rusticus* locations.

Given the possibility that these 11 sites may have already been invaded, the greatest amount of preventative measures should be issued for the other 34 locations (see Figures 5.1 &

5.2). The vulnerable sites containing solely *F. propinquus* are in the North Branch Kalamazoo River (FID 26), Davis Creek (FID 149), Kintz Creek (FID 226), Black Creek (FID 385), Clinton River (FID 417), Pere Marquette River (FIDs 450 & 451), Rifle River (FID 532), Maple River (FID 609), Grand River (FIDs 695), and Fox River (FID 875). The vulnerable sites containing solely *F. virilis* are in the Pierce Drain (FID 91), Healy Drain (FID 248), Rock Falls Creek (FIDs 282 & 283), Hay Marsh Creek (FID 569), Little Iron River (FID 738), and East Branch Sturgeon River (FID 804). The vulnerable sites containing both *F. propinquus* and *F. virilis* are in the Black Creek (FID 324), Little Black River (FIDs 612 & 613), and Grand River (FID 694). The Little Black River in particular may be vulnerable as *F. rusticus* as it is, of these 34 locations, the waterbody closest to a population of *F. rusticus*. The vulnerable sites that currently do not contain any crayfish species are on Blakely Drain (FID 189), Black Creek (216), Elm Creek (FID 258), Red Run (FID 268), Saganing River (FIDs 344 & 345), Flower Creek (FID 403), Knappen Creek (FID 513), Moyer Creek (FID 558), Trout Brook (FID 833), Cheney Creek (FID 851), and an unnamed drain on Indian River (FID 883). These waterbodies are the most vulnerable to invasion, but if action is taken proper monitoring and preventative actions may be able to protect these locations.

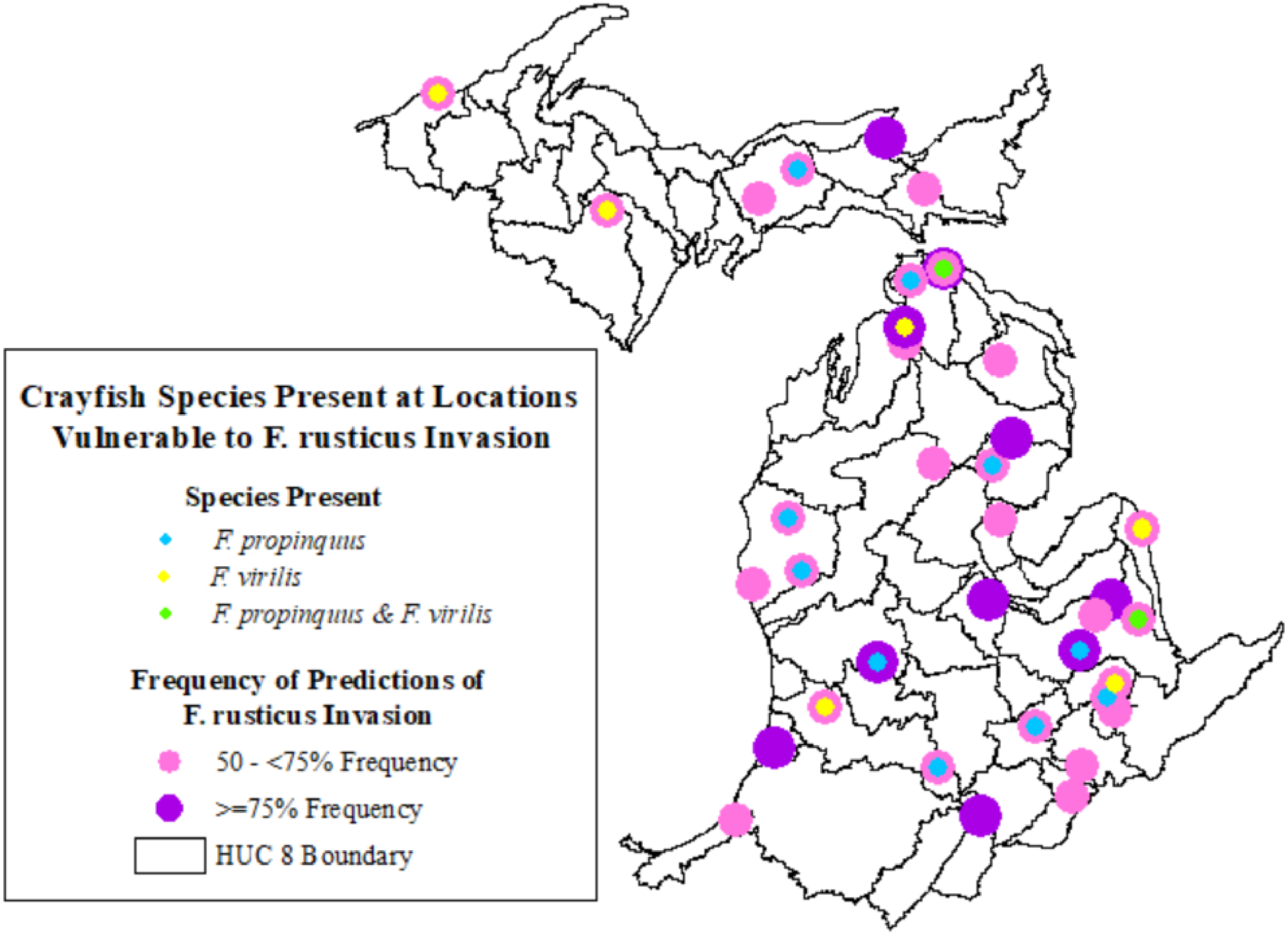

Figure 5.1 – Sites Most Vulnerable to *F. rusticus* Invasion

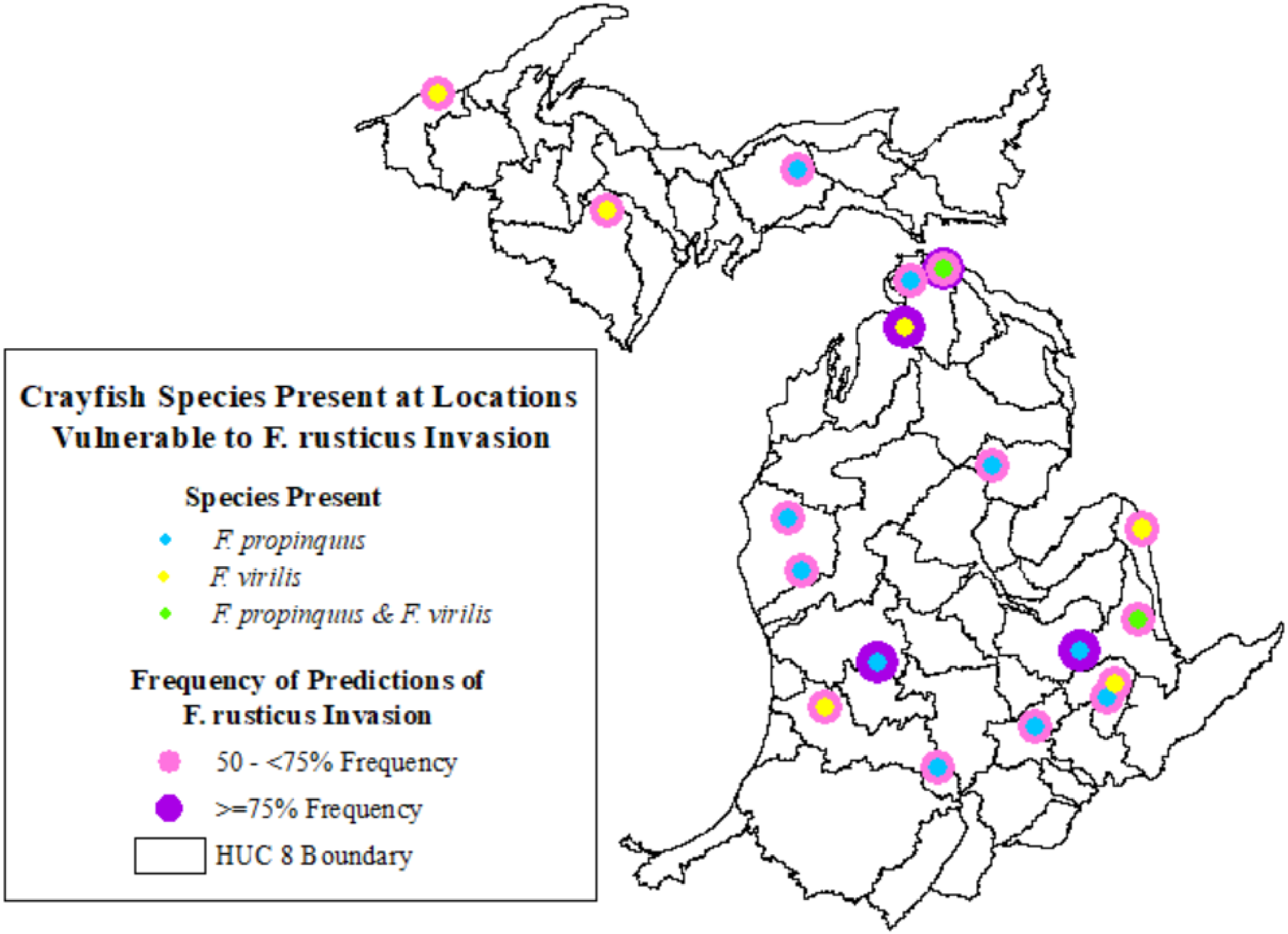

Table 5.2 – Sites Containing Native Crayfish Vulnerable to *F. rusticus* Invasion

5.2: Accuracy Assessment

The model for determining the presence/absence of *F. rusticus* was most effective at correctly predicting locations where *F. rusticus* were absent. This is to be expected as there were many more sites that *F. rusticus* are not present (671) than where they were present (159). Overall, the model had the best accuracy when predicting *F. rusticus* (~90.07%) and *F. virilis* (~89.71%) versus *F. propinquus* (~84.31%). The *F. rusticus* results stand to reason as this model was built around this species, but the similar accuracy for *F. virilis* and the lower accuracy for *F. propinquus* is interesting and raises new issues. The reason for these differences could be related to the number of survey locations selected for each species. Half of the total number of survey locations where each species was present was chosen, as were the survey locations they were absent. 415 sites were sampled for each species, but the ratio of absent survey locations selected to present survey locations selected varied for the three species: 336/79 for *F. rusticus*, 245/170 for *F. propinquus*, and 334/81 for *F. virilis*. *F. rusticus* and *F. virilis* had very similar ratios whereas *F. propinquus*, the most prevalent species in the state, had a higher number of survey locations where they were present and a lower number of survey locations where they were absent. This indicates that the model's accuracy increases as the number of survey locations a species is absent from increases and the number of survey locations a species is present in decreases. This phenomenon was observed during the modelling process.

Olden et al. (2011) used the Receiver Operating Characteristic (ROC) and Cohen's kappa statistic to determine whether the prediction of the model differed from what would be expected from random chance. These statistical measures have not been applied to the model proposed in this research, but will be implemented at a future date.

5.3: Relative Significance of Specific Variables

It was decided that for this analysis, no variables would be removed from the NLCD Land Cover Classifications (either at the local or landscape scale). This decision was reached so as to enable the ANN to evaluate the relationships among all LULC types. Until the underlying factors influencing *F. rusticus* populations are better understood, removing assessment variables could potentially remove the opportunity to identify key interactions between the variables, thus lowering the model's predictive power. Additionally, variables from both the SSURGO and NLCD datasets were among the most important for the three species, necessitating the inclusion of both datasets.

This study supports previous evidence supporting the work of Anderson et al. (2018) that landscape scale LLC is important in determining the context of the local scale LULC, and both contribute to the overall effectiveness of the model. Examining the landscape scale can identify the larger influences *F. rusticus* has on native streams and vice versa while the local scale provides a better snapshot the effects the habitat has on *F. rusticus*. Woody Wetland, Emergent Herbaceous Wetland, the "Developed" Classes, Shrub, and Barren land were all present in greater percentages at the local as compared to the landscape scale. The Open Water, Cultivated Crops, Deciduous Forest, Evergreen Forest, Mixed Forest, Pasture/Hay, and Grasslands classes make up a greater percentage of LULC at the landscape scale than the local scale. It is important to look at both the local and landscape scales, because there is not a consistent pattern between the LULC's within the state. Both scales are needed as LULC makeup at the local scale does not necessarily reflect the LULC makeup at the landscape scale. For example, LULC's that can cover large swaths of land, like the "Developed" or "Wetland" may make up 90% of the local scale's LULC, but only 50% of the LULC at the landscape scale. Thus, context is important to determining the correct scale to examine. The results also have shown an LULC may have different importance levels and directions between the local and landscape scales. This will be highlighted in the following paragraphs.

Of the 10 variables with the greatest average absolute values of importance for *F. rusticus*, only 2 were human associated (Developed – High Intensity LULC at both the local and landscape scales). These two variables along with alkalinity and $CaCO_3$ quantity were the only variables used in the model of Olden et al. (2011) that showed up among the 10 most important variables for F. rusticus. Olden et al. (2011) measured the importance of variable as a percentage of its importance versus all the variables. This research cannot follow that precedent as it the importance of each variable relative to another is too uncertain to be presented as an overall percentage.

In terms of the relative importance of the variables, 5 variables were among the ten most important for all of the three crayfish species: (1) the amount of $CaCO_3$ in the soil, (2) the Open Water area at the local scale, (3) Barren land at the landscape scale, and (4) the High Intensity Developed land at both the local scale and (5) landscape scale.

Two variables were likely of high importance due to their scarcity. $CaCO_3$ is a relatively uncommon variable, boasting a mean value of 4.9157 (quantities range from 0-100) for the sampled locations. The low percentage of sites where it is present, especially when *F. rusticus* is present, likely caused the model to assume that *F. rusticus* were not found at sites that had a percentage of $CaCO_3$ greater than 0. The Barren LLC was also a very uncommon LULC, especially at the landscape scale, where it made up an average of 0.37% of the LULCs of the sample sites within the 1000-acre range, with 524 of the survey locations not containing it. Only 11 survey locations containing *F. rusticus* have >1% of the total landscape LULC composed of

Barren Land. As with $CaCO_3$ it is likely that the model associated the presence of this LULC with the absence of *F. rusticus* due to the little correlation provided with it. Interestingly, the Barren LULC at the local scale has a larger range of means while *F. rusticus* is absent vs present than landscape scale ((0.55% vs 0.22% local). (0.39% vs 0.25% landscape)).

According to the signs of the average importances of $CaCO_3$ quantity and the Barren LULC at the landscape scale these variables had a negative contribution with *F. rusticus* prediction (-1498.6920 and -351.8329, respectively) and *F. propinquus* prediction (-64.2185 and -1025.4459, respectively). *F. virilis* presence had a negative relationship with $CaCO_3$ quantity (-1087.9820), but it had a positive relationship with the Barren LULC at the landscape level (1351.4926).

Interestingly, the Open Water class is less common at the local scale (~0.69% on average) versus the landscape scale (~1.32% on average). This could be due to the increased range of the landscape scale including nearby lakes, as streams are thin bodies that take up relatively little area. The high importance of Open Water LULC at the local scale for all three of the crayfish species makes sense. Crayfish require permanent bodies of water to reside in, and thus would need to reside in the Open Water LULC (though the Woody Wetland and Emergent Herbaceous Wetland LULCs may also suffice). What is unusual is that the Open Water LULC at the local scale only has a positive relationship with *F. rusticus* (1176.0107), with *F. propinquus* and *F. virilis* both having negative relationships with this LULC (-1295.7487 and -478.8423, respectively).

There are a few potential explanations for this result. It is possible that the negative relationship between the two native species and the Open Water LULC is due to an interaction the Open Water LULC has with another variable. As ANN's operate as a black box, it is difficult to clearly identify the interactions the variables have between each other. It is also possible that the negative association is due to the results being influenced by outliers, in which case it may also be necessary to perform run the ANN additional times for each of the three species as 10 runs might not have been enough to smooth out the influence of outliers. One last possibility is that *F. rusticus* itself is responsible for the negative associations. Native species with a positive association with cobble substrate (specifically *F. propinquus* and *F. virilis*) have been shown to gain a negative association with cobble once *F. rusticus* start occupying the same habitat (Smith et al., 2018). This is likely the result of the aggressive *F. rusticus*, which also has a positive association with cobble, forcing the *F. propinquus* and *F. virilis* out of its preferred substrate. It is possible that the *F. rusticus* may similarly be forcing *F. propinquus* (and to a lesser extent) *F. virilis* out of the open water LULC at the local scale. Additional research will be needed to further address this issue.

The interaction of Developed - High Intensity LULC, across spatial scales, is interesting. It is a relatively uncommon LULC, but unlike the previously discussed Barren Land LULC at the landscape scale, this class has areas of extremely high concentrations, particularly for Detroit and its surrounding suburbs. With the exception of *F. rusticus* at the local scale, all three crayfish species occur in areas with higher LULC percentage of Developed - High Intensity LULC at both local and landscape scales. For *F. propinquus* and *F. virilis*, the mean values of predicted locations and false negatives correlate with the average means of sites where the species is present and absent (respectively). This could be expected as one would expect an LULC percentage for a predicted location of a crayfish to have a mean value similar to that of the location when the species is present (the same goes for false negatives and absence). *F. rusticus* does not follow that pattern at all, instead the sampling sections where it is predicted to occur have an LULC percentage mean for the Developed - High Intensity Class at the landscape scale to the mean in locations where *F. rusticus* is absent and an average percentage for false negatives similar to that of sites where *F. rusticus* is present. This disparity indicates that there are interactions among the input variables occurring within ANN modeling process and the interrelationships, especially among developed land covers and other factors, are important for predicting the presence of crayfish. Further research is needed to understand the relationship among these variable interactions, especially since these variables were among the most important identified for every species.

The only other variable that completely inverted the expected means correlations for *F. rusticus* presence and absence was the cobble quantity. The cobble quantity was one of three variables from the SSURGO data (along with $CaCO_3$ quantity and extracid) to be one of the 10 most important variables for *F. rusticus* presence prediction and the only one of the three to have a positive contribution to *F. rusticus* prediction. In fact, *F. rusticus* were more likely to be present at sites with a higher cobble quantity. This supports the findings by Taylor and Redmer (1996) findings that *F. rusticus* has a positive association with cobble habitats. Interestingly, *F. propinquus* and *F. virilis* were more likely to be found at locations with a lower average cobble quantity. This may have less to do with the two native species preferences and more to do with *F. rusticus* actually forcing them out of that habitat. Smith et al. (2018) found that *F. rusticus* had a positive association with cobble and *F. virilis* and *F. propinquus* started preferring non-cobble substrate types when *F. rusticus* was present. This, combined with the aggressive nature of *F. rusticus*, seems to indicate that *F. rusticus* may in fact be forcing the other crayfish out of their preferred cobble habitat.

Other variables that correlate, as would be expected, with *F. rusicus*'s averages means for presence/absence are $CaCO_3$ quantity (which also correlates with *F. virilis*), alkalinity, and the Shrub class at both local and landscape scales. Extracid (the number of soil exchangeable hydrogen atoms) is another relatively uncommon variable, though more prevalent than $CaCO_3$ (it has a value greater that 0 in 332 of the survey locations). It is generally lower when the three crayfish species are present, (though to a far lesser extent with *F. rusticus*) and has a negative relationship with *F. rusticus* (-828.8770) and *F. virilis* (-114.9232). Shrubland, at both local and landscape scales, is an uncommon LULC primarily concentrated in the Upper Peninsula and northern Lower Peninsula. As with Barren Land, the Shrub LLC at the local scale is barely present when *F. rusticus* occurs (29 survey locations have a value >0% and only 14 were >1%). As with $CaCO_3$ and the Barren LLC (at the landscape scale), it is likely that the model associated the presence of this LULC with the absence of *F. rusticus* due to the limited correlation provided with it. Shrubland was more prevalent at the landscape scale for *F. rusticus* (86 survey locations had >0%) and had a positive contribution with predicting *F. rusticus*. Perhaps the larger sampling size provided a greater chance to pick up some of the concentrated clusters of Shrubland.

The only other variables among the 10 most important are the Developed - Medium Intensity LULC, at both local and landscape scales, and the Emergent Herbaceous Wetland LULC at the local scale. Developed - Medium Intensity LULC is one of the top 10 most important variables for *F. propinquus* and *F. virilis* at the local scale, and the second most important variable for the presence *F. propinquus* at the landscape scale. Similar to its High Intensity counterpart, the Developed - Medium Intensity LULC is a relatively uncommon LULC, with areas of extremely high concentrations, particularly in Detroit and its surround suburbs. *F.*

propinquus is much less prevalent in areas with a with a higher concentration of the Developed, Medium Intensity LULC at both local and landscape scales. *F. rusticus* and *F. virilis* have higher average concentrations of these LULCs when present. This does support the findings where *F. rusticus* has an association with human activity (Johnson et al., 2008; Olden et al., 2011; Havel et al., 2015). *F. propinquus* has a negative relationship with the Developed - Medium Intensity LULC at both local (-568.7712) and landscape (-1475.8996) scales and *F. virilis* has a negative relationship with it at the local scale (-629.8456).

Similarly, the Emergent Herbaceous Wetland LULC at the local scale was among the most important variables for predicting the presence of *F. propinquus* and *F. virilis*, but not *F. rusticus*. This variable had a positive correlation with the presence of *F. virilis*, but a negative correlation with *F. propinquus* (also *F. rusticus* presence). As wetlands are a relatively undisturbed LULC, as far as human activity goes, it makes sense that *F. rusticus* would have a weaker association with these LULC types than the native species, although why *F. propinquus* has a negative association with this LULC is not apparent.

5.4: Limitations

Though the results of this study are satisfactory, there are still several limitations present in this research that need to be addressed. Several of these limitations are due to the limitations of the SSURGO database information. Soil information collected by SSURGO was not fully complete for the entire State of Michigan. Many soil sections that occurred at areas that were predominantly underwater had no chemical/compositional information reported for them (these measurements were recorded as 0 in the SSURGO database). If zeroes are meant to represent an unmeasured value and not a value of 0, they cannot be integrated into means tests, Independent Samples T-Tests, and/or ANNs as these null values logically throw off the results. Due to this limitation, any survey location which had an average pH of 0 (which is impossible) was removed, causing 136 of the total sample locations to be omitted from this research due to this lack of data.

This raises an additional limitation: it is impossible to determine whether the 0 values for survey locations that do not have an average pH of 0 are actually recorded values of 0 or unmeasured. It is possible to have recorded values of 0 for fields such as the total volume of substrates (cobble, pebble, silt, etc.) or the percentage of $CaCO_3$, gypsum, clay-like particulates, etc. present at the site; however there are a high number of 0 values for most of the calcium related variables. Less than 50 sites had data for the percentages of gypsum and/or carbonates greater than 0. It is possible that all of these values are accurate, but the fact that 0 is a default leaves that accuracy in question. Ideally this model should be run again using the presence of calcium related values ($CaCO_3$, Gypsum, Carbonates, etc.) once more robust measurements of their presence can be collected. It is possible that the NRCS State Soil Geographic Database (STATSGO) may contain the missing data. STATSGO is a much coarser dataset than SSURGO and parsing it would be beyond the scale of this thesis, but it may be worth looking into in the future. Calcium is too intriguing a variable to be left this unexamined.

One last potential limitation comes from the selection of the survey locations. One of the major criteria that contributed to the exact geographical of the survey locations Smith et al. (2018) selected within the examined streams were how easily the site could be accessed. These sites needed to be sites where the surveyor was legally allowed to access and could safely traverse. They also needed to be sites where the surveyors would minimize the likelihood of the surveyor being disturbed/harassed by local residents. These necessities may have caused contributed to certain results.

5.5: Future Research

There is still much work to be done for analyzing the potential spread of *F. rusticus*, both in Michigan and beyond. The first area of future work is to determine the prevalence of *F. rusticus* in Michigan's lakes. Smith et al.'s (2018) research and this report specifically only examined the streams of Michigan. *F. rusticus* is already present in Lake Huron, Lake Michigan, and Lake Erie (where they may possibly be a native species) and are perfectly capable of expanding their range at a rate that fare exceeds that of native crayfish (Peters et al., 2014). Given the potential danger they pose to lake ecosystems and native species (Wilson et al., 2004), performing a similar survey and modelling procedure on Michigan's lakes warrants investigation.

The scope of this model could also be expanded. It is currently focused only on the State of Michigan, but it is possible that this research could be expanded to the rest of the country as well. The NLCD and SSURGO data sites provide LULC coverage and soil surveys for all of the continental United States. This type of modelling could easily be modified to accommodate any other state interested in exploring this predictive model and its potential results.

Additionally, this model could also be used to examine other invasive crayfish species. For example, Michigan is also being invaded by the red swamp crayfish (*Procambarus clarkii*). though that invasion is far more recent (2013 according to the USGS (2020)). Compared to *F. rusticus*, not as much information is known about its invasion or behavior, given that its recent arrival. Research on this particular Michigan invasion is much less plentiful than *F. rusticus*, but this model could prove applicable when more locational data is compiled.

Finally, *Cambarus robustus* presence could also be determined and compared to *F. rustics* presence/absence in a similar way to that of *F. propinquus* and *F. virilis*. Despite occupying the same habitat type as *F. rusticus*, the relationship between these two species is relatively unexplored, especially as compared to *F. propinquus*, *F. virilis*, and *F. immunis* (Jansen et al., 2009; Roth et al., 2016; Smith et al., 2018). Learning more about the presence of *Cambarus robustus* as compared to *F. rusticus* could provide more information on the effects *F. rusticus* has on native species. There simply were not enough survey locations containing *Cambarus robustus* for this research to make any statistically significant findings on its potential expansion range or habitat limitations. If a survey containing a more expansive range of *Cambarus robustus* locations is conducted, that data could be put in this model for comparison.

REFERENCES

Adams, S. B. (2013). Effects of small impoundments on downstream crayfish assemblages. *Freshwater Science*, 32(4), 1318-1332.

Anderson, M. G., Clark, M. M., Cornett, M. W., Hall, K. R., Olivero Sheldon, A., & Prince, J. (2018). Resilient sites for terrestrial conservation in the Great Lakes and tallgrass prairie. The Nature Conservancy, Eastern Conservation Science and North America Region.

Beck, M. W. (2018). NeuralNetTools: Visualization and analysis tools for neural networks. *Journal of Statistical Software*, 85(11), 1-20.

Berrill, M., Hollett, L., Margosian, A., & Hudson, J. (1985). Variation in tolerance to low environmental pH by the crayfish *Orconectes rusticus*, *O. propinquus*, and *Cambarus robustus*. *Canadian Journal of Zoology*, 63(11), 2586-2589.

Bobeldyk, A. M., & Lamberti, G. A. (2008). A decade after invasion: Evaluating the continuing effects of rusty crayfish on a Michigan river. *Journal of Great Lakes Research*, 34(2), 265-275.

Cairns, A., & Yan, N. D. (2009). A review of the influence of low ambient calcium concentrations on freshwater daphniids, gammarids, and crayfish. *Environmental Reviews*, 17, 67-79.

Capelli, G. M., & Magnuson, J. J. (1983). Morphoedaphic and biogeographic analysis of crayfish distribution in northern Wisconsin. *Journal of Crustacean Biology*, 3(4), 548-564.

Charlebois, P. M., & Lamberti, G. A. (1996). Invading crayfish in a Michigan stream: Direct and indirect effects on periphyton and macroinvertebrates. *Journal of the North American Benthological Society*, 15(4), 551-563.

Cox, D. (1995). Water quality: pH and alkalinity. Retrieved from https://ag.umass.edu/greenhouse-floriculture/fact-sheets/water-quality-ph-alkalinity.

Crandall, K. A., & De Grave, S. (2017). An updated classification of the freshwater crayfishes (Decapoda: Astacidea) of the world, with a complete species list. *Journal of Crustacean Biology*, 37(5), 615-653.

Department of Natural Resources (2019). *Stream river assessment units*. Retrieved from http://gis-midnr.opendata.arcgis.com/datasets/c0f0eef5603e43d5a9864f8fdf6ddfe2_5.

Dougherty, M. M., Larson, E. R., Renshaw, M. A., Gantz, C. A., Egan, S. P., Erickson, D. M., & Lodge, D. M. (2016). Environmental DNA (eDNA) detects the intrusive rusty crayfish *Orconectes rusticus* at low abundances. *Journal of Applied Ecology*, 53, 722-732.

Edwards, B. A., Jackson, D. A., & Somers, K. M. (2009). Multispecies crayfish declines in lakes: Implications for species distributions and richness. *Journal of the North American Benthological Society*, 28(3), 719-732.

Edwards, B. A., Jackson, D. A., & Somers, K. M. (2015). Evaluating the effect of lake calcium concentration on the acquisition of carapace calcium by freshwater crayfish. *Hydrobiologia*, 744, 91-100.

Edwards, B. A., Lewis, V. R. E., Rodd, F. H., & Jackson, D. A. (2013). Interactive effects of calcium decline and predation risk on the potential for a continuing northward range expansion of the rusty crayfish (*Orconectes rusticus*). *Canadian Journal of Zoology*, 91, 328-337.

Environmental Systems Research Institute (ESRI). (2019). ArcMap 10.17.1 (Version 10.17.1) [Computer software]. Available from https://www.esri.com/en-us/home.

Garvey, J. E., Stein, R. A., & Thomas, H. M. (1994). Assessing how fish predation and interspecific prey competition influence a crayfish assemblage. *Ecology*, 75(2), 532-547.

GIS Open Data. (2019a). *Counties (v17a)*. Retrieved from http://gis-michigan.opendata.arcgis.com/datasets/67a8ff23b5f54f15b7133b8c30981441_0.

GIS Open Data. (2019b). *Watershed boundary – 8 digit*. Retrieved from https://gis-michigan.opendata.arcgis.com/datasets/watershed-boundary-8-digit.

Havel, J. E., Kovalenko, K. E., Thomaz, S. M., Amalfitano, S., & Katz, Lee B. (2015). Aquatic invasive species: Challenges for the future. *Hydrobiologia*, 750, 147-170.

Hein, C. L., Roth, B. M., Ives, A. R., & Vander Zanden, M. J. (2006). Fish predation and trapping for rusty crayfish (*Orconectes rusticus*) control: a whole-lake experiment. *Canadian Journal of Fisheries and Aquatic Sciences*, 63, 383-393.

Jansen, W., Geard, N., Mosindy, T., Olson, G., & Turner, M. (2009). Relative abundance and habitat association of three crayfish (*Orconectes virilis, O. rusticus*, and *O. immunis*) near an invasion front of *O. rusticus*, and long-term changes in their distribution in Lake of the Woods, Canada. *Aquatic Invasions*, 4(4), 627-649.

Johnson, P. T. J., Olden, J. D., & Vander Zanden, M. J. (2008). Dam invaders: Impoundments facilitate biological invasions into freshwaters. *Frontiers in Ecology and the Environment*, 6(7), 357-363.

Keller, T. A., & Hazlett B. A. (2010). Thermal preferences and distribution of northern Michigan crayfishes. *Northeastern Naturalist*, 17(4), 615-628.

Kuhlmann, M. L. (2016). Invasion-related change in crayfish density affects a stream macroinvertebrate community. *Northeastern Naturalist*, 23(4), 434-453.

Larson, E. R., Renshaw, M. A., Gantz, C. A., Umek, J., Chandra, S., Lodge, D. M., & Egan, S. P. (2017). Environmental DNA (eDNA) detects the invasive crayfishes *Orconectes rusticus* and *Pacifastacus leniusculus* in large lakes of North America. *Hydrobiologia*, 800, 173-185.

Layne, J. R., Jr., Claussen, D. L., & Manis, M. L. (1987). Effects of acclimation temperature, season, and time of day on the critical thermal maxima and minima of the crayfish *Orconectes rusticus*. *Journal of Thermal Biology*, 12(3), 183-187.

Light, T., & Marchetti, M. P. (2007). Distinguishing between invasions and habitat changes as drivers of diversity loss among California's freshwater fishes. *Conservation Biology*, 21(2), 434-446.

Lippson, R. (1975). The distribution of the crayfishes of Michigan with aspects of their life cycle and physiology (Dissertation). Michigan State University, East Lansing, United States of America.

McCarthy, J. M., Hein, C. L., Olden, J. O., & Vander Zanden, M. J. (2006). Coupling long-term studies with meta-analysis to investigate impacts of non-native crayfish on zoobenthic communities. *Freshwater Biology*, 51(2), 224-235.

Minnesota Sea Grant. (2016). *Rusty crayfish: A nasty invader*. Retrieved from http://www.seagrant.umn.edu/ais/rustycrayfish_invader.

Missouri Department of Conservation. (n.d.). *Calico crayfish (Papershell crayfish)*. Retrieved from https://nature.mdc.mo.gov/discover-nature/field-guide/calico-crayfish-papershell-crayfish.

Multi-Resolution Land Characteristics Consortium. (n.d.a). *Multi-Resolution Land Characteristics (MRLC) Consortium*. Retrieved from https://www.mrlc.gov/.

Multi-Resolution Land Characteristics Consortium. (n.d.b). *National Land Cover Database 2016 (NLCD2016) legend*. Retrieved from https://www.mrlc.gov/data/legends/national-land-cover-database-2016-nlcd2016-legend.

Multi-Resolution Land Characteristics Consortium. (n.d.c). *NLCD 2016 land cover (CONUS)*. Retrieved from https://www.mrlc.gov/data/nlcd-2016-land-cover-conus.

Mundahl, N. D. (1989). Seasonal and diel changes in thermal tolerance of the crayfish *Orconectes rusticus*, with evidence for behavioral thermoregulation. *Journal of the North American Benthological Society*, 8(2), 173-179.

Mundahl, N. D., & Benton, M. J. (1990). Aspects of the thermal ecology of the rusty crayfish *Orconectes rusticus* (Girard). *Oceologia*, 82(2), 210-216.

Nagy, R., A. Fusaro, W. Conard, and C. Morningstar, 2020, *Procambarus clarkii* (Girard, 1852): U.S. Geological Survey, Nonindigenous Aquatic Species Database, Gainesville, FL, https://nas.er.usgs.gov/queries/factsheet.aspx?SpeciesID=217, Revision Date: 1/15/2020, Access Date: 2/15/2020

Olden, J. D., Joy, M. K., & Death, R. G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modeling*, 178, 389-397.

Olden, J. D., McCarthy, J. M., Maxted, J. T., Fetzer, W. W., & Vander Zanden, M. J. (2006). The rapid spread of rusty crayfish (*Orconectes rusticus*) with observations on native crayfish declines in Wisconsin (U.S.A.) over the past 130 years. *Biological Invasions*, 8, 1621-1628.

Olden, J. D., Vander Zanden, M. J., & Johnson, P. T. J. (2011). Assessing ecosystem vulnerability to invasive rusty crayfish (*Orconectes rusticus*). *Ecological Applications*, 21(7), 2587-2599.

Perry, W. L., Jacks, A. M., Fiorenza, D., Young, M., Kuhnke, R, & Jacquemin, S. J. (2013). Effects of water velocity on the size and shape of rusty crayfish, *Orconectes rusticus*. *Freshwater Science*, 32(4), 1398-1409.

Peters, J. A., Cooper, M. J., Creque, S. M., Kornis, M. S., Maxted, J. T., Perry, W. L., Schueler, F. W., Simon, T. P., Taylor, C. A., Thoma, R. F., Uzarski, D. G., & Lodge, D. M. (2014). Historical changes and current status of crayfish diversity and distribution in the Laurentian Great Lakes. *Journal of Great Lakes Research*, 40(1), 35-46.

Price, J. E., & Welch, S. M. (2009). Semi-quantitative methods for crayfish sampling: Sex, size, and habitat bias. *Journal of Crustacean Biology*, 29(2), 208-216.

Reid, S. M., & Nocera, J. J. (2015). Composition of native crayfish assemblages in southern Ontario rivers affected by rusty crayfish (*Orconectes rusticus* Girard, 1852) invasions – implications for endangered queensnake recovery. *Aquatic Invasions*, 10(2), 189-198.

Roth, B. M., Hein, C. L., & Vander Zanden, M. J. (2006). Using bioenergetics and stable isotopes to assess the trophic role of rusty crayfish (*Orconectes rusticus*) in lake littoral zones. *Canadian Journal of Fisheries and Aquatic Sciences*, 63(2), 335-344.

Roth, B. M., Jones, M. L., Hayes, D. B., & Smith, K. (2016). *Distribution of invasive crayfish in Michigan and an assessment of risks they pose to ecosystem health*. Project Completion Report. Michigan Department of Natural Resources. No report number, 56 pages.

Roth, B. M., Tetzlaff, J. C., Alexander, M. L., & Kitchell, J. F. (2007). Reciprocal relationships between exotic rusty crayfish, macrophytes, and *Lepomis* species in northern Wisconsin lakes. *Ecosystems*, 10(1), 75-86.

RStudio Team. (2016). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

Sargent, L. W., & Lodge, D. M. (2014). Evolution of invasive traits in nonindigenous species: Increased survival and faster growth in invasive populations of rusty crayfish (*Orconectes rusticus*). *Evolutionary Applications*, 7(8), 949-961.

Smith, K. R., Roth, B. M., Jones, M. L., Hayes, D. B., Herbst, S. J., & Popoff, N. (2018). Changes in the distribution of Michigan crayfishes and the influence of invasive rusty crayfish (*Faxonius rusticus*) on native crayfish substrate associations. *Biological Invasions*. https://doi.org/10.1007/s10530-018-1852-7.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. Available online at https://websoilsurvey.nrcs.usda.gov/. Accessed January 31, 2019.

Stewart, J. S., Covert, S. A., Estes, N. J., Westenbroek, S. M., Krueger, D., Wieferich, D. J., Slattery, M. T., Lyons, J. D., McKenna, Jr., J. E., Infante, D. M., & Bruce, J. L. (2016). *FishVis, A regional decision support tool for identifying vulnerabilities of riverine habitat and fishes to climate change in the Great Lakes Region* (Scientific Investigations Report 2016–5124). Reston, VA: United States Geological Survey.

Taylor, C. A., & Redmer, M. (1996). Dispersals of the crayfish *Orconectes rusticus* in Illinois, with notes on species displacement and habitat preference. *Journal of Crustacean Biology*, 16(3), 547-551.

Tennessee Wildlife Resources Agency [Photographer]. (n.d.). [Photograph of *Faxonius rusticus*]. Retrieved from https://nas.er.usgs.gov/queries/FactSheet.aspx?speciesID=214.

United States Department of Agriculture – Natural Resources Conservation Service. (2014). *SSURGO metadata – Table column descriptive report*. Retrieved from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/geo/?cid=nrcs142p2_053631.

United States Department of Agriculture – Natural Resources Conservation Service. (2019a). *Web Soil Survey*. Retrieved from https://websoilsurvey.nrcs.usda.gov/.

United States Department of Agriculture – Natural Resources Conservation Service. (2019b). *Web soil survey*. Retrieved from https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx.

United States Department of Agriculture – Natural Resources Conservation Service – Soils. (n.d.). *Description of SSURGO database*. Retrieved from https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_053627.

United States Geological Survey [Map showing the HUC watersheds where *Faxonius rusticus* have been found] (2019). *Faxonius rusticus (Girard, 1852)*. Retrieved January 16, 2020, from https://nas.er.usgs.gov/queries/FactSheet.aspx?SpeciesID=214.

United States Geological Survey – NAS - Nonindigenous Aquatic Species. (2019). *Faxonius rusticus (Girard, 1852)*. Retrieved on January 20, 2020, from https://nas.er.usgs.gov/queries/FactSheet.aspx?speciesID=214.

Westhoff, J. T., & Rosenberger, A. E. (2016). A global review of freshwater crayfish temperature tolerance, preference, and optimal growth. *Reviews in Fish Biology and Fisheries*, 26, 329-349.

Wheatley, M. G., & Gannon, A. T. (1995). Ion regulation in crayfish: Freshwater adaptations and the problem of molting. *American Zoologist*, 35, 49-59.

Wilson, K. A., Magnuson, J. J., Lodge, D. M., Hill, A. M., Kratz, T. K., Perry, W. L., & Willis, T. V. (2004). A long-term rusty crayfish (*Orconectes rusticus*) invasion: Dispersal patterns and community change in a north temperate lake. *Canadian Journal of Fisheries and Aquatic Sciences*, 61, 2255-2266.

Appendix A

Artificial Neural Network (ANN) Code in R

Section A.1: R Code Used to Create Neural Net and Run Tests for *Faxonius rusticus*

```
setwd("D:/Thesis/HUC8_DBF/SimpleThesis")


library(car)
library(carData)
library(dplyr)
library(foreign)
library(haven)
library(lsr)
library(neuralnet)
library(NeuralNetTools)


LLNormF <- read_sav("SimplifiedLatLongSoil4.sav")
#View(LLNormF)


RusAbsF <- read_sav("RusAbsF.sav")
#View(RusAbsF)


RusPresF <- read_sav("RusPresF.sav")
#View(RusPresF)


data1RusFx <- LLNormF
dataRusFx <- data1RusFx


data1RusFAx <- RusAbsF
dataRusFAx <- data1RusFAx


data1RusFPx <- RusPresF
dataRusFPx <- data1RusFPx


indRusFAx <- sample(1:nrow(dataRusFAx), 336)
indRusFPx <- sample(672:nrow(dataRusFPx), 79)


indRusFx <- c(indRusFAx,indRusFPx)


train_dataRusFx <- dataRusFx[indRusFx,]
test_dataRusFx <- dataRusFx[-indRusFx,]


#library(neuralnet)
```

```
npRusFx <- neuralnet(RustyPres~Norm_Cob_All + Norm_Peb_All + Norm_sand_r +
Norm_silt_r + Norm_clay_r + Norm_om_r + Norm_caco3_h + Norm_extracid_r +
Norm_pH_Ave_h + Norm_Water + Norm_WW + Norm_EmWet + Norm_Crops +
Norm_DevOS + Norm_DevL + Norm_DevM + Norm_DevH + Norm_DecFor + Norm_ConFor
+ Norm_MixFor + Norm_Past + Norm_Grass + Norm_Shrub + Norm_Barren + Norm_Water1k
+ Norm_WW1k + Norm_EmWet1k + Norm_Crops1k + Norm_DevOS1k + Norm_DevL1k +
Norm_DevM1k + Norm_DevH1k + Norm_DecFor1k + Norm_ConFor1k + Norm_MixFor1k +
Norm_Past1k + Norm_Grass1k + Norm_Shrub1k + Norm_Barren1k + Norm_BF + Norm_Lat +
Norm_Long, data = train_dataRusFx, hidden = 10, stepmax = 1e+05, linear.output = FALSE)


olden(npRusFx, bar_plot = FALSE)


plot(npRusFx)
outputRusFx <- predict(npRusFx, test_dataRusFx[,-146])
actualRusFx <- dataRusFx[-indRusFx,146]


MSERusFx <- sum((outputRusFx-actualRusFx)^2)/nrow(test_dataRusFx)
acRusFx <- pull(actualRusFx, RustyPres)
rpRusFx <- round(outputRusFx)
rpRusFx[rpRusFx==-1] = 0


confusionMatrix(as.factor(acRusFx),as.factor(rpRusFx))


write.dbf(rpRusFx, "SimRustyFPredx.dbf")
write.dbf(as.data.frame(test_dataRusFx), "SimRustyTestFActualx.dbf")
```

Section A.2: R Code Used to Create Neural Net and Run Tests for *Faxonius propinquus*

```
setwd("D:/Thesis/HUC8_DBF/SimpleThesis")


library(car)
library(carData)
library(dplyr)
library(foreign)
library(haven)
library(lsr)
library(neuralnet)
library(NeuralNetTools)


LLNormN <- read_sav("SimpNCC.sav")
#View(LLNormN)


NCCAbsF <- read_sav("NCCAbsF.sav")
#View(NCCAbsF)


NCCPresF <- read_sav("NCCPresF.sav")
#View(NCCPresF)


data1NCCFx <- LLNormN
dataNCCFx <- data1NCCFx


data1NCCAFx <- NCCAbsF
dataNCCAFx <- data1NCCAFx


data1NCCPFx <- NCCPresF
dataNCCPFx <- data1NCCPFx


indNCCAFx <- sample(1:nrow(dataNCCAFx), 245)
indNCCPFx <- sample(490:nrow(dataNCCPFx), 170)


indNCCFx <- c(indNCCAFx,indNCCPFx)
#View(indNCCFx)


train_dataNCCFx <- dataNCCFx[indNCCFx,]
test_dataNCCFx <- dataNCCFx[-indNCCFx,]
```

Section A.2 - Continued.

```
#library(neuralnet)


npNCCFx <- neuralnet(NCCPres~Norm_Cob_All + Norm_Peb_All + Norm_sand_r +
Norm_silt_r + Norm_clay_r + Norm_om_r + Norm_caco3_h + Norm_extracid_r +
Norm_pH_Ave_h + Norm_Water + Norm_WW + Norm_EmWet + Norm_Crops +
Norm_DevOS + Norm_DevL + Norm_DevM + Norm_DevH + Norm_DecFor + Norm_ConFor
+ Norm_MixFor + Norm_Past + Norm_Grass + Norm_Shrub + Norm_Barren + Norm_Water1k
+ Norm_WW1k + Norm_EmWet1k + Norm_Crops1k + Norm_DevOS1k + Norm_DevL1k +
Norm_DevM1k + Norm_DevH1k + Norm_DecFor1k + Norm_ConFor1k + Norm_MixFor1k +
Norm_Past1k + Norm_Grass1k + Norm_Shrub1k + Norm_Barren1k + Norm_BF + Norm_Lat +
Norm_Long, data = train_dataNCCFx, hidden = 10, stepmax = 1e+05, linear.output = FALSE)


olden(npNCCFx, bar_plot = FALSE)


#plot(npNCCFx)
outputNCCFx <- predict(npNCCFx, test_dataNCCFx[,-147])
actualNCCFx <- dataNCCFx[-indNCCFx,147]


MSENCCFx <- sum((outputNCCFx-actualNCCFx)^2)/nrow(test_dataNCCFx)
acNCCFx <- pull(actualNCCFx, NCCPres)
rpNCCFx <- round(outputNCCFx)
rpNCCFx[rpNCCFx==-1] = 0


confusionMatrix(as.factor(acNCCFx),as.factor(rpNCCFx))


write.dbf(rpNCCFx, "SimNCCFPredFx.dbf")
write.dbf(as.data.frame(test_dataNCCFx), "SimNCCTestFActualFx.dbf")
```

Section A.3: R Code Used to Create Neural Net and Run Tests for *Faxonius virilis*

```
setwd("D:/Thesis/HUC8_DBF/SimpleThesis")


library(car)
library(carData)
library(dplyr)
library(foreign)
library(haven)
library(lsr)
library(neuralnet)
library(NeuralNetTools)


LLNormV <- read_sav("SimpVirile.sav")
#View(LLNormV)

VirAbsFx <- read_sav("VirAbsFx.sav")
#View(VirAbsFx)

VirPresFx <- read_sav("VirPresFx.sav")
#View(VirPresFx)


data1VirFx <- LLNormV
dataVirFx <- data1VirFx


data1VirAFx <- VirAbsF
dataVirAFx <- data1VirAFx


data1VirPFx <- VirPresF
dataVirPFx <- data1VirPFx


indVirAFx <- sample(1:nrow(dataVirAFx), 334)
indVirPFx <- sample(669:nrow(dataVirPFx), 81)


indVirFx <- c(indVirAFx,indVirPFx)


train_dataVirFx <- dataVirFx[indVirFx,]
test_dataVirFx <- dataVirFx[-indVirFx,]


#library(neuralnet)
```

Section A.3 - Continued

```
npVirFx <- neuralnet(VirilePres~Norm_Cob_All + Norm_Peb_All + Norm_sand_r +
Norm_silt_r + Norm_clay_r + Norm_om_r + Norm_caco3_h + Norm_extracid_r +
Norm_pH_Ave_h + Norm_Water + Norm_WW + Norm_EmWet + Norm_Crops +
Norm_DevOS + Norm_DevL + Norm_DevM + Norm_DevH + Norm_DecFor + Norm_ConFor
+ Norm_MixFor + Norm_Past + Norm_Grass + Norm_Shrub + Norm_Barren + Norm_Water1k
+ Norm_WW1k + Norm_EmWet1k + Norm_Crops1k + Norm_DevOS1k + Norm_DevL1k +
Norm_DevM1k + Norm_DevH1k + Norm_DecFor1k + Norm_ConFor1k + Norm_MixFor1k +
Norm_Past1k + Norm_Grass1k + Norm_Shrub1k + Norm_Barren1k + Norm_BF + Norm_Lat +
Norm_Long, data = train_dataVirFx, hidden = 10, stepmax = 1e+05, linear.output = FALSE)


olden(npVirFx, bar_plot = FALSE)


plot(npVirFx)
outputVirFx <- predict(npVirFx, test_dataVirFx[,-148])
actualVirFx <- dataVirFx [-indVirFx,148]


MSEVirFx <- sum((outputVirFx-actualVirFx)^2)/nrow(test_dataVirFx)
acVirFx <- pull(actualVirFx, VirilePres)
rpVirFx <- round(outputVirFx)
rpVirFx[rpVirFx==-1] = 0


confusionMatrix(as.factor(acVirFx),as.factor(rpVirFx))


write.dbf(rpVirFx, "SimVirFPredFx.dbf")


write.dbf(as.data.frame(test_dataVirFx), "SimVirTestFActualFx.dbf")
```

Appendix B

Confusion Matrices, Specificity and Other Statistical Information for All Tests on All Crayfish Species

Section B.1: Confusion Matrices for and Statistical Information for Each Test for *Faxonius rusticus*

Confusion Matrix (*Faxonius rusticus*) (Test 1)			
		Prediction	
		0	1
Actual	0	301	34
	1	46	34

Neural Network Output for *Faxonius rusticus* (Test 1)	
Accuracy	0.8072
95% CI	(0.7659, 0.8441)
No Information Rate	0.8361
P-Value [Acc > NIR]	0.9489
Kappa	0.3431
Mcnemar's Test P-Value	0.2188
Sensitivity	0.8674
Specificity	0.5
Pos Pred Value	0.8985
Neg Pred Value	0.425
Prevalence	0.8361
Detection Rate	0.7253
Detection Prevalence	0.8072
Balanced Accuracy	0.6837
'Positive' Class	0

Section B.1 – Continued

Confusion Matrix (*Faxonius rusticus*) (Test 2)			
		Prediction	
		0	1
Actual	0	310	25
	1	43	37

Neural Network Output for *Faxonius rusticus* (Test 2)	
Accuracy	0.8337
95% CI	(0.7944, 0.8683)
No Information Rate	0.853
P-Value [Acc > NIR]	0.87957
Kappa	0.4127
Mcnemar's Test P-Value	0.03024
Sensitivity	0.8757
Specificity	0.5902
Pos Pred Value	0.9254
Neg Pred Value	0.45
Prevalence	0.853
Detection Rate	0.747
Detection Prevalence	0.8072
Balanced Accuracy	0.7329
'Positive' Class	0

Confusion Matrix (*Faxonius rusticus*) (Test 3)			
		Prediction	
		0	1
Actual	0	301	34
	1	41	39

Neural Network Output for *Faxonius rusticus* (Test 3)	
Accuracy	0.8193
95% CI	(0.7788, 0.8551)
No Information Rate	0.8241
P-Value [Acc > NIR]	0.6311
Kappa	0.3993
Mcnemar's Test P-Value	0.4884
Sensitivity	0.8801
Specificity	0.5342
Pos Pred Value	0.8985
Neg Pred Value	0.4875
Prevalence	0.8241
Detection Rate	0.7253
Detection Prevalence	0.8072
Balanced Accuracy	0.7072
'Positive' Class	0

Section B.1 – Continued

Confusion Matrix (*Faxonius rusticus*) (Test 4)			
		Prediction	
		0	1
Actual	0	305	30
	1	44	36

Neural Network Output for *Faxonius rusticus* (Test 4)	
Accuracy	0.8217
95% CI	(0.7814, 0.8573)
No Information Rate	0.841
P-Value [Acc > NIR]	0.8722
Kappa	0.3862
Mcnemar's Test P-Value	0.1307
Sensitivity	0.8739
Specificity	0.5455
Pos Pred Value	0.9104
Neg Pred Value	0.45
Prevalence	0.841
Detection Rate	0.7349
Detection Prevalence	0.8072
Balanced Accuracy	0.7097
'Positive' Class	0

Section B.1 – Continued

Confusion Matrix (*Faxonius rusticus*) (Test 5)			
		Prediction	
		0	1
Actual	0	311	24
	1	43	37

Neural Network Output for *Faxonius rusticus* (Test 5)	
Accuracy	0.8386
95% CI	(0.7996, 0.8726)
No Information Rate	0.853
P-Value [Acc > NIR]	0.81711
Kappa	0.4297
Mcnemar's Test P-Value	0.02787
Sensitivity	0.8785
Specificity	0.6066
Pos Pred Value	0.9284
Neg Pred Value	0.4625
Prevalence	0.853
Detection Rate	0.7494
Detection Prevalence	0.8072
Balanced Accuracy	0.7425
'Positive' Class	0

Section B.1 – Continued

Confusion Matrix (*Faxonius rusticus*) (Test 6)			
		Prediction	
		0	1
Actual	0	297	38
	1	48	32

Neural Network Output for *Faxonius rusticus* (Test 6)	
Accuracy	0.7928
95% CI	(0.7505, 0.8308)
No Information Rate	0.8313
P-Value [Acc > NIR]	0.9828
Kappa	0.3009
Mcnemar's Test P-Value	0.3318
Sensitivity	0.8609
Specificity	0.4571
Pos Pred Value	0.8866
Neg Pred Value	0.4
Prevalence	0.8313
Detection Rate	0.7157
Detection Prevalence	0.8072
Balanced Accuracy	0.659
'Positive' Class	0

Section B.1 – Continued

Confusion Matrix (*Faxonius rusticus*) (Test 7)			
		Prediction	
		0	1
Actual	0	306	29
	1	40	40

Neural Network Output for *Faxonius rusticus* (Test 7)	
Accuracy	0.8337
95% CI	(0.7944, 0.8683)
No Information Rate	0.8337
P-Value [Acc > NIR]	0.5321
Kappa	0.4363
Mcnemar's Test P-Value	0.2286
Sensitivity	0.8844
Specificity	0.5797
Pos Pred Value	0.9134
Neg Pred Value	0.5
Prevalence	0.8337
Detection Rate	0.7373
Detection Prevalence	0.8072
Balanced Accuracy	0.7321
'Positive' Class	0

Section B.1 – Continued

Confusion Matrix (*Faxonius rusticus*) (Test 8)			
		Prediction	
		0	1
Actual	0	300	35
	1	34	46

Neural Network Output for *Faxonius rusticus* (Test 8)	
Accuracy	0.8313
95% CI	(0.7918, 0.8661)
No Information Rate	0.8072
P-Value [Acc > NIR]	0.1174
Kappa	0.458
Mcnemar's Test P-Value	1
Sensitivity	0.8955
Specificity	0.5625
Pos Pred Value	0.8955
Neg Pred Value	0.5625
Prevalence	0.8072
Detection Rate	0.7229
Detection Prevalence	0.8072
Balanced Accuracy	0.729

Section B.1 – Continued

Confusion Matrix (*Faxonius rusticus*) (Test 9)			
		Prediction	
		0	1
Actual	0	283	52
	1	36	44

Neural Network Output for *Faxonius rusticus* (Test 9)	
Accuracy	0.7880
'Positive' Class	0
95% CI	(0.7454, 0.8263)
No Information Rate	0.7687
P-Value [Acc > NIR]	0.192
Kappa	0.3669
Mcnemar's Test P-Value	0.1098
Sensitivity	0.8871
Specificity	0.4583
Pos Pred Value	0.8448
Neg Pred Value	0.55
Prevalence	0.7687
Detection Rate	0.6819
Detection Prevalence	0.8072
Balanced Accuracy	0.6727
'Positive' Class	0

Confusion Matrix (*Faxonius rusticus*) (Test 10)			
		Prediction	
		0	1
Actual	0	296	39
	1	37	43

Neural Network Output for *Faxonius rusticus* (Test 10)	
Accuracy	0.8145
95% CI	(0.7737, 0.8507)
No Information Rate	0.8048
P-Value [Acc > NIR]	0.3361
Kappa	0.4066
Mcnemar's Test P-Value	1
Sensitivity	0.8862
Specificity	0.5185
Pos Pred Value	0.8836
Neg Pred Value	0.525
Prevalence	0.8048
Detection Rate	0.7133
Detection Prevalence	0.8072
Balanced Accuracy	0.7024
'Positive' Class	0

Section B.2: Confusion Matrices for and Statistical Information for Each Test for *Faxonius propinquus*

Confusion Matrix (*Faxonius propinquus*) (Test 1)			
		Prediction	
		0	1
Actual	0	171	73
	1	68	103

Neural Network Output for *Faxonius propinquus* (Test 1)	
Accuracy	0.6602
95% CI	(0.6124, 0.7057)
No Information Rate	0.5759
P-Value [Acc > NIR]	0.0002678
Kappa	0.3018
Mcnemar's Test P-Value	0.7362219
Sensitivity	0.7155
Specificity	0.5852
Pos Pred Value	0.7008
Neg Pred Value	0.6023
Prevalence	0.5759
Detection Rate	0.412
Detection Prevalence	0.588
Balanced Accuracy	0.6504
'Positive' Class	0

Confusion Matrix (*Faxonius propinquus*) (Test 2)			
		Prediction	
		0	1
Actual	0	170	74
	1	67	104

Neural Network Output for *Faxonius propinquus* (Test 2)	
Accuracy	0.6602
95% CI	(0.6124, 0.7057)
No Information Rate	0.5711
P-Value [Acc > NIR]	0.0001267
Kappa	0.3031
Mcnemar's Test P-Value	0.6133544
Sensitivity	0.7173
Specificity	0.5843
Pos Pred Value	0.6967
Neg Pred Value	0.6082
Prevalence	0.5711
Detection Rate	0.4096
Detection Prevalence	0.588
Balanced Accuracy	0.6508
'Positive' Class	0

Section B.2 – Continued

Confusion Matrix (*Faxonius propinquus*) (Test 3)			
		Prediction	
		0	1
Actual	0	180	64
	1	61	110

Neural Network Output for *Faxonius propinquus* (Test 3)	
Accuracy	0.6988
95% CI	(0.6521, 0.7426)
No Information Rate	0.5807
P-Value [Acc > NIR]	4.54E-07
Kappa	0.38
Mcnemar's Test P-Value	0.858
Sensitivity	0.7469
Specificity	0.6322
Pos Pred Value	0.7377
Neg Pred Value	0.6433
Prevalence	0.5807
Detection Rate	0.4337
Detection Prevalence	0.588
Balanced Accuracy	0.6895
'Positive' Class	0

Confusion Matrix (*Faxonius propinquus*) (Test 4)			
		Prediction	
		0	1
Actual	0	171	73
	1	66	105

Neural Network Output for *Faxonius propinquus* (Test 4)	
Accuracy	0.6651
95% CI	(0.6174, 0.7104)
No Information Rate	0.5711
P-Value [Acc > NIR]	5.61E-05
Kappa	0.3129
Mcnemar's Test P-Value	0.6108
Sensitivity	0.7215
Specificity	0.5899
Pos Pred Value	0.7008
Neg Pred Value	0.614
Prevalence	0.5711
Detection Rate	0.412
Detection Prevalence	0.588
Balanced Accuracy	0.6557
'Positive' Class	0

Section B.2 – Continued

Confusion Matrix (*Faxonius propinquus*) (Test 5)			
		Prediction	
		0	1
Actual	0	180	64
	1	64	107

Neural Network Output for *Faxonius propinquus* (Test 5)	
Accuracy	0.6916
95% CI	(0.6447, 0.7357)
No Information Rate	0.588
P-Value [Acc > NIR]	0.00000832
Kappa	0.3634
Mcnemar's Test P-Value	1
Sensitivity	0.7377
Specificity	0.6257
Pos Pred Value	0.7377
Neg Pred Value	0.6257
Prevalence	0.588
Detection Rate	0.4337
Detection Prevalence	0.588
Balanced Accuracy	0.6817

Section B.2 - Continued

Confusion Matrix (*Faxonius propinquus*) (Test 6)			
		Prediction	
		0	1
Actual	0	180	64
	1	70	101

Neural Network Output for *Faxonius propinquus* (Test 6)	
'Positive' Class	0
Accuracy	0.6771
95% CI	(0.6298, 0.7219)
No Information Rate	0.6024
P-Value [Acc > NIR]	0.0009866
Kappa	0.3301
Mcnemar's Test P-Value	0.6657892
Sensitivity	0.72
Specificity	0.6121
Pos Pred Value	0.7377
Neg Pred Value	0.5906
Prevalence	0.6024
Detection Rate	0.4337
Detection Prevalence	0.588
Balanced Accuracy	0.6661
'Positive' Class	0

Section B.2 - Continued

Confusion Matrix (*Faxonius propinquus*) (Test 7)			
		Prediction	
		0	1
Actual	0	171	73
	1	49	122

Neural Network Output for *Faxonius propinquus* (Test 7)	
Accuracy	0.706
95% CI	(0.6596, 0.7495)
No Information Rate	0.5301
P-Value [Acc > NIR]	1.95E-13
Kappa	0.4058
Mcnemar's Test P-Value	0.03731
Sensitivity	0.7773
Specificity	0.6256
Pos Pred Value	0.7008
Neg Pred Value	0.7135
Prevalence	0.5301
Detection Rate	0.412
Detection Prevalence	0.588
Balanced Accuracy	0.7015
'Positive' Class	0

Section B.2 - Continued

Confusion Matrix (*Faxonius propinquus*) (Test 8)			
		Prediction	
		0	1
Actual	0	171	73
	1	64	107

Neural Network Output for *Faxonius propinquus* (Test 8)	
Accuracy	0.6699
95% CI	(0.6223, 0.715)
No Information Rate	0.5663
P-Value [Acc > NIR]	1.01E-05
Kappa	0.324
Mcnemar's Test P-Value	0.4943
Sensitivity	0.7277
Specificity	0.5944
Pos Pred Value	0.7008
Neg Pred Value	0.6257
Prevalence	0.5663
Detection Rate	0.412
Detection Prevalence	0.588
Balanced Accuracy	0.6611
'Positive' Class	0

Section B.2 - Continued

Confusion Matrix (*Faxonius propinquus*) (Test 9)			
		Prediction	
		0	1
Actual	0	170	74
	1	61	110

Neural Network Output for *Faxonius propinquus* (Test 9)	
Accuracy	0.6747
95% CI	(0.6273, 0.7196)
No Information Rate	0.5566
P-Value [Acc > NIR]	5.95E-07
Kappa	0.3362
Mcnemar's Test P-Value	0.3017
Sensitivity	0.7359
Specificity	0.5978
Pos Pred Value	0.6967
Neg Pred Value	0.6433
Prevalence	0.5566
Detection Rate	0.4096
Detection Prevalence	0.588
Balanced Accuracy	0.6669
'Positive' Class	0

Section B.2 - Continued

Confusion Matrix (*Faxonius propinquus*) (Test 10)

		Prediction	
		0	1
Actual	0	175	69
	1	81	90

Neural Network Output for *Faxonius propinquus* (Test 10)

Accuracy	0.6386
95% CI	(0.5903, 0.6848)
No Information Rate	0.6169
P-Value [Acc > NIR]	0.1957
Kappa	0.2461
Mcnemar's Test P-Value	0.3691
Sensitivity	0.6836
Specificity	0.566
Pos Pred Value	0.7172
Neg Pred Value	0.5263
Prevalence	0.6169
Detection Rate	0.4217
Detection Prevalence	0.588
Balanced Accuracy	0.6248
'Positive' Class	0

Section B.3: Confusion Matrices and Statistical Information for Each Test for *Faxonius virilis*

Confusion Matrix (*Faxonius virilis*) (Test 1)			
		Prediction	
		0	1
Actual	0	306	28
	1	53	28

Neural Network Output for *Faxonius virilis* (Test 1)	
Accuracy	0.8048
95% CI	(0.7634, 0.8419)
No Information Rate	0.8651
P-Value [Acc > NIR]	0.999752
Kappa	0.2965
Mcnemar's Test P-Value	0.007661
Sensitivity	0.8524
Specificity	0.5
Pos Pred Value	0.9162
Neg Pred Value	0.3457
Prevalence	0.8651
Detection Rate	0.7373
Detection Prevalence	0.8048
Balanced Accuracy	0.6762
'Positive' Class	0

Section B.3 - Continued

Confusion Matrix (*Faxonius virilis*) (Test 2)

		Prediction	
		0	1
Actual	0	287	47
	1	43	38

Neural Network Output for *Faxonius virilis* (Test 2)

Accuracy	0.7807
95% CI	(0.7378, 0.8196)
No Information Rate	0.7928
P-Value [Acc > NIR]	0.7494
Kappa	0.318
Mcnemar's Test P-Value	0.675
Sensitivity	0.8693
Specificity	0.4419
Pos Pred Value	0.8563
Neg Pred Value	0.4691
Prevalence	0.7928
Detection Rate	0.6892
Detection Prevalence	0.8048
Balanced Accuracy	0.6556
'Positive' Class	0

Confusion Matrix (*Faxonius virilis*) (Test 3)			
		Prediction	
		0	1
Actual	0	293	41
	1	48	33

Neural Network Output for *Faxonius virilis* (Test 3)	
Accuracy	0.7855
95% CI	(0.7429, 0.8241)
No Information Rate	0.8217
P-Value [Acc > NIR]	0.9745
Kappa	0.2943
Mcnemar's Test P-Value	0.5248
Sensitivity	0.8592
Specificity	0.4459
Pos Pred Value	0.8772
Neg Pred Value	0.4074
Prevalence	0.8217
Detection Rate	0.706
Detection Prevalence	0.8048
Balanced Accuracy	0.6526
'Positive' Class	0

Section B.3 - Continued

Confusion Matrix (*Faxonius virilis*) (Test 4)			
		Prediction	
		0	1
Actual	0	280	54
	1	29	52

Neural Network Output for *Faxonius virilis* (Test 4)	
Accuracy	0.8
95% CI	(0.7582, 0.8374)
No Information Rate	0.7446
P-Value [Acc > NIR]	0.004812
Kappa	0.43
Mcnemar's Test P-Value	0.00843
Sensitivity	0.9061
Specificity	0.4906
Pos Pred Value	0.8383
Neg Pred Value	0.642
Prevalence	0.7446
Detection Rate	0.6747
Detection Prevalence	0.8048
Balanced Accuracy	0.6984
'Positive' Class	0

Section B.3 - Continued

Confusion Matrix (*Faxonius virilis*) (Test 5)			
		Prediction	
		0	1
Actual	0	293	41
	1	47	34

Neural Network Output for *Faxonius virilis* (Test 5)	
Accuracy	0.788
95% CI	(0.7454, 0.8263)
No Information Rate	0.8193
P-Value [Acc > NIR]	0.9552
Kappa	0.3056
Mcnemar's Test P-Value	0.594
Sensitivity	0.8618
Specificity	0.4533
Pos Pred Value	0.8772
Neg Pred Value	0.4198
Prevalence	0.8193
Detection Rate	0.706
Detection Prevalence	0.8048
Balanced Accuracy	0.6575
'Positive' Class	0

Confusion Matrix (*Faxonius virilis*) (Test 6)			
		Prediction	
		0	1
Actual	0	289	45
	1	40	41

Neural Network Output for *Faxonius virilis* (Test 6)	
Accuracy	0.8072
95% CI	(0.7659, 0.8441)
No Information Rate	0.8048
P-Value [Acc > NIR]	0.4803
Kappa	0.3864
Mcnemar's Test P-Value	1
Sensitivity	0.8802
Specificity	0.5062
Pos Pred Value	0.8802
Neg Pred Value	0.5062
Prevalence	0.8048
Detection Rate	0.7084
Detection Prevalence	0.8048
Balanced Accuracy	0.6932
'Positive' Class	0

Section B.3 - Continued

Confusion Matrix (*Faxonius virilis*) (Test 7)			
		Prediction	
		0	1
Actual	0	301	34
	1	46	35

Neural Network Output for *Faxonius virilis* (Test 7)	
Accuracy	0.8096
95% CI	(0.7685, 0.8463)
No Information Rate	0.8361
P-Value [Acc > NIR]	0.9341
Kappa	0.3549
Mcnemar's Test P-Value	0.177
Sensitivity	0.8674
Specificity	0.5147
Pos Pred Value	0.9012
Neg Pred Value	0.4321
Prevalence	0.8361
Detection Rate	0.7253
Detection Prevalence	0.8048
Balanced Accuracy	0.6911
'Positive' Class	0

Section B.3 - Continued

Confusion Matrix (*Faxonius virilis*) (Test 8)			
		Prediction	
		0	1
Actual	0	305	29
	1	43	38

Neural Network Output for *Faxonius virilis* (Test 8)	
Accuracy	0.8265
95% CI	(0.7866, 0.8617)
No Information Rate	0.8386
P-Value [Acc > NIR]	0.7706
Kappa	0.4091
Mcnemar's Test P-Value	0.1255
Sensitivity	0.8764
Specificity	0.5672
Pos Pred Value	0.9132
Neg Pred Value	0.4691
Prevalence	0.8386
Detection Rate	0.7349
Detection Prevalence	0.8048
Balanced Accuracy	0.7218
'Positive' Class	0

Section B.3 - Continued

Confusion Matrix (*Faxonius virilis*) (Test 9)			
		Prediction	
		0	1
Actual	0	300	34
	1	39	42

Neural Network Output for *Faxonius virilis* (Test 9)	
Accuracy	0.8241
95% CI	(0.784, 0.8595)
No Information Rate	0.8169
P-Value [Acc > NIR]	0.3801
Kappa	0.4267
Mcnemar's Test P-Value	0.6397
Sensitivity	0.885
Specificity	0.5526
Pos Pred Value	0.8982
Neg Pred Value	0.5185
Prevalence	0.8169
Detection Rate	0.7229
Detection Prevalence	0.8048
Balanced Accuracy	0.7188
'Positive' Class	0

Section B.3 - Continued

Confusion Matrix (*Faxonius virilis*) (Test 10)			
		Prediction	
		0	1
Actual	0	289	45
	1	40	41

Neural Network Output for *Faxonius virilis* (Test 10)	
Accuracy	0.7952
95% CI	(0.7531, 0.833)
No Information Rate	0.7928
P-Value [Acc > NIR]	0.4806
Kappa	0.363
Mcnemar's Test P-Value	0.6644
Sensitivity	0.8784
Specificity	0.4767
Pos Pred Value	0.8653
Neg Pred Value	0.5062
Prevalence	0.7928
Detection Rate	0.6964
Detection Prevalence	0.8048
Balanced Accuracy	0.6776
'Positive' Class	0

Appendix C

Statistical Summaries for *Faxonius rusticus*, *Faxonius propinquus*, and *Faxonius virilis* for All Variables

Section C.1: Statistical Summaries for *Faxonius rusticus*

Group Statistics *Faxonius rusticus*

Variable[a]		N[b]	Mean[c]	Std Dev[d]	SEM[e]
pebble_all_h	0.0[f]	671	3.6602	6.6802	0.2579
	1.0[g]	159	3.1950	5.6136	0.4452
	Ave[h]	830	3.5711	6.4889	0.2252
	FP[i]	340	3.6735	5.8175	0.3155
	FN[j]	415	3.2699	5.7519	0.2823
cobble_all_h	.00	671	2.6066	5.8463	0.2257
	1.00	159	3.3836	10.8557	0.8609
	Ave	830	2.7554	7.0836	0.2459
	FP	340	1.8647	4.3763	0.2373
	FN	415	3.1398	9.9002	0.4860
sand_r	.00	671	57.4444	31.9463	1.2333
	1.00	159	55.6082	31.8249	2.5239
	Ave	830	57.0927	31.9121	1.1077
	FP	340	62.5450	30.8463	1.6729
	FN	415	52.9308	31.7925	1.5606
silt_r	.00	671	25.6063	22.8176	0.8809
	1.00	159	24.8516	20.6338	1.6364
	Ave	830	25.4617	22.4057	0.7777
	FP	340	22.0012	20.3503	1.1037
	FN	415	27.6361	22.0282	1.0813
clay_r	.00	671	11.7288	10.7062	0.4133
	1.00	159	15.7667	12.9686	1.0285
	Ave	830	12.5023	11.2792	0.3915
	FP	340	13.6832	11.8001	0.6399
	FN	415	15.8186	13.4676	0.6611
om_r	.00	671	13.6827	26.6121	1.0273
	1.00	159	14.4452	26.7684	2.1229
	Ave	830	13.8288	26.6276	0.9243
	FP	340	11.6694	25.1411	1.3635
	FN	415	14.7661	27.6258	1.3561
caco3_h	.00	671	5.2548	17.5225	0.6764
	1.00	159	3.4843	13.0446	1.0345
	Ave	830	4.9157	16.7650	0.5819
	FP	340	3.2059	12.6221	0.6845
	FN	415	4.4048	15.3717	0.7546

Group Statistics *Faxonius rusticus*					
Variable		N	Mean	Std Dev	SEM
extracid_r	.00	671	4.1352	10.7709	0.4158
	1.00	159	2.8981	10.6767	0.8467
	Ave	830	3.8982	10.7575	0.3734
	FP	340	2.0738	8.9115	0.4833
	FN	415	3.0520	10.3394	0.5075
pH_Ave_h	.00	671	7.4450	0.7885	0.0304
	1.00	159	7.6585	0.5594	0.0444
	Ave	830	7.4859	0.7545	0.0262
	FP	340	7.5803	0.7053	0.0382
	FN	415	7.6236	0.6202	0.0304
LULCWater	.00	671	0.0053	0.0233	0.0009
	1.00	159	0.0135	0.0406	0.0032
	Ave	830	0.0069	0.0276	0.0010
	FP	340	0.0135	0.0414	0.0022
	FN	415	0.0139	0.0402	0.0020
LULCWW	.00	671	0.2818	0.2290	0.0088
	1.00	159	0.2893	0.2384	0.0189
	Ave	830	0.2833	0.2307	0.0080
	FP	340	0.2938	0.2207	0.0120
	FN	415	0.2959	0.2327	0.0114
LULCEmWet	.00	671	0.0137	0.0362	0.0014
	1.00	159	0.0116	0.0215	0.0017
	Ave	830	0.0133	0.0339	0.0012
	FP	340	0.0139	0.0332	0.0018
	FN	415	0.0118	0.0230	0.0011
LULCCrops	.00	671	0.1767	0.2755	0.0106
	1.00	159	0.2442	0.2916	0.0231
	Ave	830	0.1896	0.2797	0.0097
	FP	340	0.2084	0.2836	0.0154
	FN	415	0.2181	0.2743	0.0135
LULCDevOS	.00	671	0.0764	0.0703	0.0027
	1.00	159	0.0729	0.0781	0.0062
	Ave	830	0.0757	0.0719	0.0025
	FP	340	0.0851	0.0885	0.0048
	FN	415	0.0674	0.0721	0.0035

Section C.1 – Continued

Group Statistics *Faxonius rusticus*

Variable		N	Mean	Std Dev	SEM
LULCDevL	.00	671	0.0600	0.0700	0.0027
	1.00	159	0.0750	0.0759	0.0060
	Ave	830	0.0629	0.0714	0.0025
	FP	340	0.0740	0.0685	0.0037
	FN	415	0.0764	0.0685	0.0034
LULCDevM	.00	671	0.0217	0.0607	0.0023
	1.00	159	0.0262	0.0632	0.0050
	Ave	830	0.0225	0.0612	0.0021
	FP	340	0.0283	0.0670	0.0036
	FN	415	0.0259	0.0577	0.0028
LULCDevH	.00	671	0.0099	0.0469	0.0018
	1.00	159	0.0085	0.0362	0.0029
	Ave	830	0.0096	0.0450	0.0016
	FP	340	0.0129	0.0434	0.0024
	FN	415	0.0114	0.0497	0.0024
LULCDecFor	.00	671	0.1460	0.1647	0.0064
	1.00	159	0.1070	0.1298	0.0103
	Ave	830	0.1385	0.1593	0.0055
	FP	340	0.1059	0.1350	0.0073
	FN	415	0.1168	0.1331	0.0065
LULCConFor	.00	671	0.0483	0.0863	0.0033
	1.00	159	0.0368	0.0820	0.0065
	Ave	830	0.0461	0.0856	0.0030
	FP	340	0.0340	0.0674	0.0037
	FN	415	0.0346	0.0785	0.0039
LULCMixFor	.00	671	0.0916	0.1426	0.0055
	1.00	159	0.0496	0.0895	0.0071
	Ave	830	0.0836	0.1350	0.0047
	FP	340	0.0634	0.1099	0.0060
	FN	415	0.0617	0.1025	0.0050
LULCPast	.00	671	0.0346	0.0847	0.0033
	1.00	159	0.0446	0.1172	0.0093
	Ave	830	0.0365	0.0918	0.0032
	FP	340	0.0405	0.1112	0.0060
	FN	415	0.0423	0.1230	0.0060

Section C.1 - Continued

Group Statistics *Faxonius rusticus*					
Variable		N	Mean	Std Dev	SEM
LULCGrass	.00	671	0.0228	0.0583	0.0022
	1.00	159	0.0159	0.0323	0.0026
	Ave	830	0.0215	0.0543	0.0019
	FP	340	0.0196	0.0458	0.0025
	FN	415	0.0160	0.0323	0.0016
LULCShrub	.00	671	0.0057	0.0234	0.0009
	1.00	159	0.0028	0.0087	0.0007
	Ave	830	0.0052	0.0214	0.0007
	FP	340	0.0023	0.0073	0.0004
	FN	415	0.0040	0.0101	0.0005
LULCBarren	.00	671	0.0055	0.0246	0.0010
	1.00	159	0.0022	0.0065	0.0005
	Ave	830	0.0049	0.0224	0.0008
	FP	340	0.0042	0.0179	0.0010
	FN	415	0.0037	0.0084	0.0004
LULCWater1k	.00	671	0.0125	0.0370	0.0014
	1.00	159	0.0159	0.0438	0.0035
	Ave	830	0.0132	0.0384	0.0013
	FP	340	0.0154	0.0460	0.0025
	FN	415	0.0200	0.0560	0.0028
LULCWW1k	.00	671	0.2279	0.1875	0.0072
	1.00	159	0.2601	0.1938	0.0154
	Ave	830	0.2341	0.1890	0.0066
	FP	340	0.2483	0.1935	0.0105
	FN	415	0.2648	0.1912	0.0094
LULCEmWet1k	.00	671	0.0093	0.0162	0.0006
	1.00	159	0.0108	0.0157	0.0012
	Ave	830	0.0096	0.0161	0.0006
	FP	340	0.0121	0.0210	0.0011
	FN	415	0.0116	0.0172	0.0008
LULCCrops1k	.00	671	0.2112	0.2764	0.0107
	1.00	159	0.2764	0.2858	0.0227
	Ave	830	0.2237	0.2793	0.0097
	FP	340	0.2387	0.2767	0.0150
	FN	415	0.2558	0.2754	0.0135

Section C.1 - Continued

Group Statistics *Faxonius rusticus*					
Variable		N	Mean	Std Dev	SEM
LULCDevOS1k	.00	671	0.0562	0.0585	0.0023
	1.00	159	0.0563	0.0639	0.0051
	Ave	830	0.0562	0.0596	0.0021
	FP	340	0.0640	0.0694	0.0038
	FN	415	0.0532	0.0530	0.0026
LULCDevL1k	.00	671	0.0441	0.0665	0.0026
	1.00	159	0.0531	0.0703	0.0056
	Ave	830	0.0459	0.0673	0.0023
	FP	340	0.0562	0.0698	0.0038
	FN	415	0.0500	0.0595	0.0029
LULCDevM1k	.00	671	0.0199	0.0589	0.0023
	1.00	159	0.0224	0.0573	0.0045
	Ave	830	0.0204	0.0586	0.0020
	FP	340	0.0250	0.0665	0.0036
	FN	415	0.0189	0.0497	0.0024
LULCDevH1k	.00	671	0.0085	0.0359	0.0014
	1.00	159	0.0091	0.0329	0.0026
	Ave	830	0.0086	0.0354	0.0012
	FP	340	0.0069	0.0244	0.0013
	FN	415	0.0092	0.0368	0.0018
LULCDecFor1k	.00	671	0.1936	0.1659	0.0064
	1.00	159	0.1328	0.1242	0.0099
	Ave	830	0.1820	0.1605	0.0056
	FP	340	0.1567	0.1314	0.0071
	FN	415	0.1367	0.1235	0.0061
LULCConFor1k	.00	671	0.0509	0.0856	0.0033
	1.00	159	0.0309	0.0478	0.0038
	Ave	830	0.0470	0.0801	0.0028
	FP	340	0.0326	0.0543	0.0029
	FN	415	0.0318	0.0472	0.0023

Section C.1 - Continued

Group Statistics *Faxonius rusticus*					
Variable		N	Mean	Std Dev	SEM
LULCMixFor1k	.00	671	0.0936	0.1316	0.0051
	1.00	159	0.0523	0.0834	0.0066
	Ave	830	0.0857	0.1249	0.0043
	FP	340	0.0625	0.0968	0.0052
	FN	415	0.0646	0.0998	0.0049
LULCPast1k	.00	671	0.0353	0.0768	0.0030
	1.00	159	0.0469	0.1000	0.0079
	Ave	830	0.0375	0.0818	0.0028
	FP	340	0.0445	0.1007	0.0055
	FN	415	0.0497	0.1096	0.0054
LULCGrass1k	.00	671	0.0228	0.0450	0.0017
	1.00	159	0.0247	0.0437	0.0035
	Ave	830	0.0231	0.0447	0.0016
	FP	340	0.0279	0.0479	0.0026
	FN	415	0.0227	0.0423	0.0021
LULCShrub1k	.00	671	0.0096	0.0291	0.0011
	1.00	159	0.0052	0.0127	0.0010
	Ave	830	0.0088	0.0268	0.0009
	FP	340	0.0038	0.0106	0.0006
	FN	415	0.0076	0.0162	0.0008
LULCBarren1k	.00	671	0.0039	0.0177	0.0007
	1.00	159	0.0025	0.0068	0.0005
	Ave	830	0.0036	0.0162	0.0006
	FP	340	0.0040	0.0211	0.0011
	FN	415	0.0030	0.0068	0.0003
Proj_baseflow_cfs	.00	671	36.4821	186.8055	7.2115
	1.00	159	72.2648	228.9681	18.1584
	Ave	830	43.3369	195.9434	6.8013
	FP	340	102.0367	404.6798	21.9468
	FN	415	77.0688	223.5989	10.9760

Section C.1 – Continued

Group Statistics *Faxonius rusticus*					
Variable		N	Mean	Std Dev	SEM
Lat	.00	671	44.2252	1.6462	0.0636
	1.00	159	43.9097	1.3772	0.1092
	Ave	830	44.1647	1.6022	0.0556
	FP	340	44.0003	1.4787	0.0802
	FN	415	44.0227	1.4705	0.0722
Long	.00	671	-85.4848	1.7076	0.0659
	1.00	159	-84.6629	1.4701	0.1166
	Ave	830	-85.3274	1.6951	0.0588
	FP	340	-84.8628	1.4417	0.0782
	FN	415	-84.8892	1.6628	0.0816

[a] Names and detailed description of all variables can be found in Table 3.1
[b] Number of survey locations
[c] The mean value of the variable across the selected number of survey locations
[d] Standard deviation of the mean
[e] Standard error of the mean
[f] *Faxonius rusticus* is not present
[g] *Faxonius rusticus* is present
[h] Average across all survey locations regardless of *Faxonius rusticus* presence
[i] *Faxonius rusticus* was predicted to occur at a site it was absent from (false positive)
[j] *Faxonius rusticus* was predicted to not be present at a location they were present at (false negative)

Section C.2: Statistical Summaries for *Faxonius propinquus*

Group Statistics *Faxonius propinquus*

Variable[a]		N[b]	Mean[c]	Std Dev[d]	SEM[e]
pebble_all_h	.00[f]	489	3.2045	5.9756	0.2702
	1.00[g]	341	4.0968	7.1373	0.3865
	Ave[h]	830	3.5711	6.4889	0.2252
	FP[i]	701	3.6462	6.5176	0.2462
	FN[j]	651	3.8080	7.0272	0.1977
cobble_all_h	.00	489	3.0429	8.1046	0.3665
	1.00	341	2.3431	5.2706	0.2854
	Ave	830	2.7554	7.0836	0.2459
	FP	701	2.2553	6.2701	1.2039
	FN	651	2.6160	5.0436	0.1977
sand_r	.00	489	59.6374	32.5711	1.4729
	1.00	341	53.4434	30.6220	1.6583
	Ave	830	57.0927	31.9121	1.1077
	FP	701	57.8238	31.8739	0.8688
	FN	651	53.8773	31.7363	1.2438
silt_r	.00	489	24.4575	23.2261	1.0503
	1.00	341	26.9018	21.1243	1.1439
	Ave	830	25.4617	22.4057	0.7777
	FP	701	25.6680	23.0035	0.4216
	FN	651	28.1647	22.8725	0.8964
clay_r	.00	489	11.6065	11.3269	0.5122
	1.00	341	13.7868	11.1013	0.6012
	Ave	830	12.5023	11.2792	0.3915
	FP	701	12.7991	11.1629	0.9682
	FN	651	13.6524	11.8721	0.4653
om_r	.00	489	13.1170	25.6481	1.1598
	1.00	341	14.8495	27.9793	1.5152
	Ave	830	13.8288	26.6276	0.9243
	FP	701	13.0000	25.6337	0.6249
	FN	651	15.0273	28.2596	1.1076
caco3_h	.00	489	4.2188	13.9657	0.6315
	1.00	341	5.9150	20.0914	1.0880
	Ave	830	4.9157	16.7650	0.5819
	FP	701	4.1141	16.5439	0.3868
	FN	651	5.3702	18.1330	0.7107

Section C.2 – Continued

Group Statistics *Faxonius propinquus*

Variable		N	Mean	Std Dev	SEM
extracid_r	.00	489	4.7744	12.5866	0.5692
	1.00	341	2.6416	7.2177	0.3909
	Ave	830	3.8982	10.7575	0.3734
	FP	701	3.8177	10.2403	0.0277
	FN	651	3.0158	7.5375	0.2954
pH_Ave_h	.00	489	7.4673	0.7576	0.0343
	1.00	341	7.5126	0.7503	0.0406
	Ave	830	7.4859	0.7545	0.0262
	FP	701	7.4242	0.7342	0.0277
	FN	651	7.5300	0.7645	0.0300
LULCWater	.00	489	0.0066	0.0282	0.0013
	1.00	341	0.0073	0.0268	0.0014
	Ave	830	0.0069	0.0276	0.0010
	FP	701	0.0065	0.0288	0.0011
	FN	651	0.0080	0.0263	0.0010
LULCWW	.00	489	0.2631	0.2406	0.0109
	1.00	341	0.3122	0.2128	0.0115
	Ave	830	0.2833	0.2307	0.0080
	FP	701	0.2963	0.2356	0.0089
	FN	651	0.3112	0.2170	0.0085
LULCEmWet	.00	489	0.0153	0.0400	0.0018
	1.00	341	0.0103	0.0223	0.0012
	Ave	830	0.0133	0.0339	0.0012
	FP	701	0.0137	0.0325	0.0012
	FN	651	0.0138	0.0278	0.0011
LULCCrops	.00	489	0.1966	0.2976	0.0135
	1.00	341	0.1796	0.2520	0.0136
	Ave	830	0.1896	0.2797	0.0097
	FP	701	0.1678	0.2599	0.0098
	FN	651	0.1726	0.2620	0.0103
LULCDevOS	.00	489	0.0776	0.0818	0.0037
	1.00	341	0.0729	0.0544	0.0029
	Ave	830	0.0757	0.0719	0.0025
	FP	701	0.0802	0.0819	0.0031
	FN	651	0.0739	0.0616	0.0024

Section C.2 – Continued

Group Statistics *Faxonius propinquus*					
Variable		N	Mean	Std Dev	SEM
LULCDevL	.00	489	0.0636	0.0721	0.0033
	1.00	341	0.0619	0.0703	0.0038
	Ave	830	0.0629	0.0714	0.0025
	FP	701	0.0589	0.0696	0.0026
	FN	651	0.0613	0.0713	0.0028
LULCDevM	.00	489	0.0240	0.0652	0.0029
	1.00	341	0.0205	0.0549	0.0030
	Ave	830	0.0225	0.0612	0.0021
	FP	701	0.0221	0.0657	0.0025
	FN	651	0.0205	0.0571	0.0022
LULCDevH	.00	489	0.0091	0.0449	0.0020
	1.00	341	0.0103	0.0453	0.0025
	Ave	830	0.0096	0.0450	0.0016
	FP	701	0.0134	0.0596	0.0023
	FN	651	0.0064	0.0237	0.0009
LULCDecFor	.00	489	0.1291	0.1642	0.0074
	1.00	341	0.1521	0.1512	0.0082
	Ave	830	0.1385	0.1593	0.0055
	FP	701	0.1454	0.1695	0.0064
	FN	651	0.1612	0.1532	0.0060
LULCConFor	.00	489	0.0513	0.0914	0.0041
	1.00	341	0.0387	0.0760	0.0041
	Ave	830	0.0461	0.0856	0.0030
	FP	701	0.0441	0.0865	0.0033
	FN	651	0.0458	0.0785	0.0031
LULCMixFor	.00	489	0.0973	0.1565	0.0071
	1.00	341	0.0639	0.0929	0.0050
	Ave	830	0.0836	0.1350	0.0047
	FP	701	0.0805	0.1278	0.0048
	FN	651	0.0558	0.1041	0.0041
LULCPast	.00	489	0.0328	0.0842	0.0038
	1.00	341	0.0418	0.1016	0.0055
	Ave	830	0.0365	0.0918	0.0032
	FP	701	0.0381	0.0887	0.0033
	FN	651	0.0432	0.1139	0.0045

Group Statistics *Faxonius propinquus*

Variable		N	Mean	Std Dev	SEM
LULCGrass	.00	489	0.0234	0.0616	0.0028
	1.00	341	0.0186	0.0416	0.0023
	Ave	830	0.0215	0.0543	0.0019
	FP	701	0.0222	0.0595	0.0022
	FN	651	0.0222	0.0518	0.0020
LULCShrub	.00	489	4.7744	12.5866	0.5692
	1.00	341	2.6416	7.2177	0.3909
	Ave	830	3.8982	10.7575	0.3734
	FP	701	0.0066	0.0261	0.0010
	FN	651	0.0041	0.0154	0.0006
LULCBarren	.00	489	0.0039	0.0201	0.0009
	1.00	341	0.0062	0.0252	0.0014
	Ave	830	0.0049	0.0224	0.0008
	FP	701	0.0042	0.0197	0.0007
	FN	651	0.0047	0.0220	0.0009
LULCWater1k	.00	489	0.0123	0.0370	0.0017
	1.00	341	0.0145	0.0403	0.0022
	Ave	830	0.0132	0.0384	0.0013
	FP	701	0.0164	0.0470	0.0018
	FN	651	0.0177	0.0471	0.0018
LULCWW1k	.00	489	0.2226	0.1932	0.0087
	1.00	341	0.2505	0.1818	0.0098
	Ave	830	0.2341	0.1890	0.0066
	FP	701	0.2368	0.1971	0.0074
	FN	651	0.2596	0.1809	0.0071
LULCEmWet1k	.00	489	0.0095	0.0164	0.0007
	1.00	341	0.0097	0.0158	0.0009
	Ave	830	0.0096	0.0161	0.0006
	FP	701	0.0093	0.0157	0.0006
	FN	651	0.0120	0.0192	0.0008
LULCCrops1k	.00	489	0.2184	0.2906	0.0131
	1.00	341	0.2313	0.2623	0.0142
	Ave	830	0.2237	0.2793	0.0097
	FP	701	0.2199	0.2776	0.0105
	FN	651	0.2055	0.2607	0.0102

Section C.2 – Continued

Group Statistics *Faxonius propinquus*					
Variable		N	Mean	Std Dev	SEM
LULCDevOS1k	.00	489	0.0549	0.0608	0.0027
	1.00	341	0.0581	0.0577	0.0031
	Ave	830	0.0562	0.0596	0.0021
	FP	701	0.0583	0.0651	0.0025
	FN	651	0.0565	0.0589	0.0023
LULCDevL1k	.00	489	0.0462	0.0708	0.0032
	1.00	341	0.0453	0.0620	0.0034
	Ave	830	0.0459	0.0673	0.0023
	FP	701	0.0432	0.0699	0.00264
	FN	651	0.0449	0.0656	0.00257
LULCDevM1k	.00	489	0.0231	0.0676	0.0031
	1.00	341	0.0165	0.0422	0.0023
	Ave	830	0.0204	0.0586	0.0020
	FP	701	0.0178	0.0545	0.0021
	FN	651	0.0158	0.0420	0.0016
LULCDevH1k	.00	489	0.0082	0.0340	0.0015
	1.00	341	0.0091	0.0372	0.0020
	Ave	830	0.0086	0.0354	0.0012
	FP	701	0.0099	0.0445	0.0017
	FN	651	0.0061	0.0229	0.0009
LULCDecFor1k	.00	489	0.1786	0.1692	0.0077
	1.00	341	0.1868	0.1471	0.0080
	Ave	830	0.1820	0.1605	0.0056
	FP	701	0.1841	0.1656	0.0063
	FN	651	0.1791	0.1523	0.0060
LULCConFor1k	.00	489	0.0512	0.0876	0.0040
	1.00	341	0.0411	0.0677	0.0037
	Ave	830	0.0470	0.0801	0.0028
	FP	701	0.0459	0.0764	0.0029
	FN	651	0.0457	0.0684	0.0027

Section C.2 – Continued

Group Statistics *Faxonius propinquus*

Variable		N	Mean	Std Dev	SEM
LULCMixFor1k	.00	489	0.1019	0.1460	0.0066
	1.00	341	0.0625	0.0806	0.0044
	Ave	830	0.0857	0.1249	0.0043
	FP	701	0.0826	0.1198	0.0045
	FN	651	0.0767	0.0962	0.0038
LULCPast1k	.00	489	0.0349	0.0766	0.0035
	1.00	341	0.0412	0.0888	0.0048
	Ave	830	0.0375	0.0818	0.0028
	FP	701	0.0379	0.0805	0.0030
	FN	651	0.0440	0.1037	0.0041
LULCGrass1k	.00	489	0.0237	0.0474	0.0021
	1.00	341	0.0224	0.0406	0.0022
	Ave	830	0.0231	0.0447	0.0016
	FP	701	0.0236	0.0463	0.0017
	FN	651	0.0240	0.0414	0.0016
LULCShrub1k	.00	489	0.0106	0.0323	0.0015
	1.00	341	0.0061	0.0153	0.0008
	Ave	830	0.0088	0.0268	0.0009
	FP	701	0.0105	0.0255	0.0010
	FN	651	0.0083	0.0197	0.0008
LULCBarren1k	.00	489	0.0032	0.0144	0.0007
	1.00	341	0.0042	0.0185	0.0010
	Ave	830	0.0036	0.0162	0.0006
	FP	701	0.0034	0.0198	0.0007
	FN	651	0.0037	0.0201	0.0008
Proj_baseflow_cfs	.00	489	39.4656	218.6614	9.8882
	1.00	341	48.8884	157.9091	8.5513
	Ave	830	43.3369	195.9434	6.8013
	FP	701	68.5850	356.7722	13.4751
	FN	651	40.3504	141.0955	5.5300

Section C.2 – Continued

Group Statistics *Faxonius propinquus*					
Variable		N	Mean	Std Dev	SEM
Lat	.00	489	44.3588	1.6868	0.0763
	1.00	341	43.8864	1.4295	0.0774
	Ave	830	44.1647	1.6022	0.0556
	FP	701	44.1509	1.6522	0.0624
	FN	651	44.1644	1.4749	0.0578
Long	.00	489	-85.3523	1.8663	0.0844
	1.00	341	-85.2916	1.4158	0.0767
	Ave	830	-85.3274	1.6951	0.0588
	FP	701	-85.3392	1.7892	0.0676
	FN	651	-85.1516	1.5110	0.0592

[a] Names and detailed description of all variables can be found in Table 3.1

[b] Number of survey locations

[c] The mean value of the variable across the selected number of survey locations

[d] Standard deviation of the mean

[e] Standard error of the mean

[f] *Faxonius propinquus* is not present

[g] *Faxonius propinquus* is present

[h] Average across all survey locations regardless of *Faxonius propinquus* presence

[i] *Faxonius propinquus* was predicted to occur at a site it was absent from (false positive)

[j] *Faxonius propinquus* was predicted to not be present at a location they were present at (false negative)

Section C.3: Statistical Summaries for *Faxonius virilis*

Group Statistics *Faxonius virilis*

Variable[a]		N[b]	Mean[c]	Std Dev[d]	SEM[e]
pebble_all_h	0.0[f]	668	3.5135	6.2928	0.2435
	1.0[g]	162	3.8086	7.2579	0.5702
	Ave[h]	830	3.5711	6.4889	0.2252
	FP[i]	393	2.7939	5.9447	0.2999
	FN[j]	428	3.4159	7.2298	0.3495
cobble_all_h	.00	668	2.9746	7.5906	0.2937
	1.00	162	1.8519	4.3184	0.3393
	Ave	830	2.7554	7.0836	0.2459
	FP	393	1.9415	4.0398	0.2038
	FN	428	1.7734	4.1960	0.2028
sand_r	.00	668	57.0133	32.2475	1.2477
	1.00	162	57.4198	30.5844	2.4029
	Ave	830	57.0927	31.9121	1.1077
	FP	393	55.6059	33.1019	1.6698
	FN	428	55.0453	31.8628	1.5401
silt_r	.00	668	25.3579	22.3216	0.8636
	1.00	162	25.8895	22.8145	1.7925
	Ave	830	25.4617	22.4057	0.7777
	FP	393	26.6784	25.7984	1.3014
	FN	428	27.0084	23.7906	1.1500
clay_r	.00	668	12.5344	11.6539	0.4509
	1.00	162	12.3698	9.6124	0.7552
	Ave	830	12.5023	11.2792	0.3915
	FP	393	11.3545	10.0040	0.5046
	FN	428	13.5070	11.3048	0.5464
om_r	.00	668	14.5424	27.3614	1.0586
	1.00	162	10.8864	23.2055	1.8232
	Ave	830	13.8288	26.6276	0.9243
	FP	393	15.3917	28.5593	1.4406
	FN	428	13.4898	25.7217	1.2433
caco3_h	.00	668	4.7799	16.3728	0.6335
	1.00	162	5.4753	18.3378	1.4408
	Ave	830	4.9157	16.7650	0.5819
	FP	393	4.6260	16.6802	0.8414
	FN	428	5.0070	17.2370	0.8332

Section C.3 – Continued

Group Statistics *Faxonius virilis*					
Variable		N	Mean	Std Dev	SEM
extracid_r	.00	668	4.1352	11.3684	0.4399
	1.00	162	2.9210	7.6980	0.6048
	Ave	830	3.8982	10.7575	0.3734
	FP	393	4.1623	9.6272	0.4856
	FN	428	3.9304	9.9956	0.4832
pH_Ave_h	.00	668	7.4624	0.7789	0.0301
	1.00	162	7.5827	0.6374	0.0501
	Ave	830	7.4859	0.7545	0.0262
	FP	393	7.4651	0.6856	0.0346
	FN	428	7.5217	0.6690	0.0323
LULCWater	.00	668	0.0061	0.0255	0.0010
	1.00	162	0.0104	0.0349	0.0027
	Ave	830	0.0069	0.0276	0.0010
	FP	393	0.0097	0.0297	0.0015
	FN	428	0.0133	0.0382	0.0018
LULCWW	.00	668	0.2920	0.2291	0.0089
	1.00	162	0.2472	0.2344	0.0184
	Ave	830	0.2833	0.2307	0.0080
	FP	393	0.2642	0.2324	0.0117
	FN	428	0.2912	0.2500	0.0121
LULCEmWet	.00	668	0.0119	0.0318	0.0012
	1.00	162	0.0191	0.0411	0.0032
	Ave	830	0.0133	0.0339	0.0012
	FP	393	0.0160	0.0409	0.0021
	FN	428	0.0166	0.0359	0.0017
LULCCrops	.00	668	0.1781	0.2726	0.0105
	1.00	162	0.2370	0.3038	0.0239
	Ave	830	0.1896	0.2797	0.0097
	FP	393	0.2464	0.3116	0.0157
	FN	428	0.2384	0.2987	0.0144
LULCDevOS	.00	668	0.0708	0.0611	0.0024
	1.00	162	0.0959	0.1031	0.0081
	Ave	830	0.0757	0.0719	0.0025
	FP	393	0.0833	0.0947	0.0048
	FN	428	0.0803	0.0823	0.0040

Section C.3 – Continued

Group Statistics *Faxonius virilis*					
Variable		N	Mean	Std Dev	SEM
LULCDevL	.00	668	0.0559	0.0646	0.0025
	1.00	162	0.0916	0.0890	0.0070
	Ave	830	0.0629	0.0714	0.0025
	FP	393	0.0783	0.0905	0.0046
	FN	428	0.0782	0.0835	0.0040
LULCDevM	.00	668	0.0184	0.0560	0.0022
	1.00	162	0.0398	0.0770	0.0060
	Ave	830	0.0225	0.0612	0.0021
	FP	393	0.0331	0.0762	0.0038
	FN	428	0.0265	0.0619	0.0030
LULCDevH	.00	668	0.0078	0.0414	0.0016
	1.00	162	0.0169	0.0572	0.0045
	Ave	830	0.0096	0.0450	0.0016
	FP	393	0.0167	0.0626	0.0032
	FN	428	0.0121	0.0483	0.0023
LULCDecFor	.00	668	0.1546	0.1687	0.0065
	1.00	162	0.0723	0.0853	0.0067
	Ave	830	0.1385	0.1593	0.0055
	FP	393	0.0848	0.1082	0.0055
	FN	428	0.0802	0.0901	0.0044
LULCConFor	.00	668	0.0500	0.0889	0.0034
	1.00	162	0.0302	0.0680	0.0053
	Ave	830	0.0461	0.0856	0.0030
	FP	393	0.0377	0.0856	0.0043
	FN	428	0.0303	0.0734	0.0035
LULCMixFor	.00	668	0.0878	0.1384	0.0054
	1.00	162	0.0660	0.1190	0.0094
	Ave	830	0.0836	0.1350	0.0047
	FP	393	0.0645	0.1164	0.0059
	FN	428	0.0617	0.1055	0.0051
LULCPast	.00	668	0.0351	0.0907	0.0035
	1.00	162	0.0421	0.0964	0.0076
	Ave	830	0.0365	0.0918	0.0032
	FP	393	0.0377	0.0948	0.0048
	FN	428	0.0435	0.0979	0.0047

Section C.3 – Continued

Group Statistics *Faxonius virilis*					
Variable		N	Mean	Std Dev	SEM
LULCGrass	.00	668	0.0220	0.0562	0.0022
	1.00	162	0.0193	0.0458	0.0036
	Ave	830	0.0215	0.0543	0.0019
	FP	393	0.0175	0.0572	0.0029
	FN	428	0.0168	0.0472	0.0023
LULCShrub	.00	668	0.0058	0.0233	0.0009
	1.00	162	0.0024	0.0098	0.0008
	Ave	830	0.0052	0.0214	0.0007
	FP	393	0.0033	0.0134	0.0007
	FN	428	0.0032	0.0123	0.0006
LULCBarren	.00	668	0.0037	0.0188	0.0007
	1.00	162	0.0097	0.0329	0.0026
	Ave	830	0.0049	0.0224	0.0008
	FP	393	0.0069	0.0230	0.0012
	FN	428	0.0078	0.0291	0.0014
LULCWater1k	.00	668	0.0117	0.0356	0.0014
	1.00	162	0.0194	0.0477	0.0038
	Ave	830	0.0132	0.0384	0.0013
	FP	393	0.0251	0.0572	0.0029
	FN	428	0.0182	0.0460	0.0022
LULCWW1k	.00	668	0.2410	0.1910	0.0074
	1.00	162	0.2055	0.1783	0.0140
	Ave	830	0.2341	0.1890	0.0066
	FP	393	0.2159	0.1856	0.0094
	FN	428	0.2355	0.1845	0.0089
LULCEmWet1k	.00	668	0.0089	0.0151	0.0006
	1.00	162	0.0123	0.0196	0.0015
	Ave	830	0.0096	0.0161	0.0006
	FP	393	0.0119	0.0178	0.0009
	FN	428	0.0101	0.0159	0.0008
LULCCrops1k	.00	668	0.2176	0.2763	0.0107
	1.00	162	0.2485	0.2907	0.0228
	Ave	830	0.2237	0.2793	0.0097
	FP	393	0.2655	0.3036	0.0153
	FN	428	0.2523	0.2906	0.0140

Section C.3 – Continued

Group Statistics *Faxonius virilis*					
Variable		N	Mean	Std Dev	SEM
LULCDevOS1k	.00	668	0.0511	0.0523	0.0020
	1.00	162	0.0774	0.0798	0.0063
	Ave	830	0.0562	0.0596	0.0021
	FP	393	0.0646	0.0731	0.0037
	FN	428	0.0636	0.0661	0.0032
LULCDevL1k	.00	668	0.0389	0.0579	0.0022
	1.00	162	0.0746	0.0916	0.0072
	Ave	830	0.0459	0.0673	0.0023
	FP	393	0.0571	0.0800	0.0040
	FN	428	0.0568	0.0757	0.0037
LULCDevM1k	.00	668	0.0153	0.0500	0.0019
	1.00	162	0.0414	0.0822	0.0065
	Ave	830	0.0204	0.0586	0.0020
	FP	393	0.0288	0.0687	0.0035
	FN	428	0.0266	0.0657	0.0032
LULCDevH1k	.00	668	0.0065	0.0331	0.0013
	1.00	162	0.0171	0.0425	0.0033
	Ave	830	0.0086	0.0354	0.0012
	FP	393	0.0137	0.0491	0.0025
	FN	428	0.0094	0.0344	0.0017
LULCDecFor1k	.00	668	0.1963	0.1674	0.0065
	1.00	162	0.1228	0.1103	0.0087
	Ave	830	0.1820	0.1605	0.0056
	FP	393	0.1345	0.1310	0.0066
	FN	428	0.1418	0.1250	0.0060
LULCConFor1k	.00	668	0.0499	0.0810	0.0031
	1.00	162	0.0353	0.0755	0.0059
	Ave	830	0.0470	0.0801	0.0028
	FP	393	0.0386	0.0792	0.0040
	FN	428	0.0386	0.0741	0.0036

Section C.3 – Continued

Group Statistics *Faxonius virilis*

Variable		N	Mean	Std Dev	SEM
LULCMixFor1k	.00	668	0.0920	0.1307	0.0051
	1.00	162	0.0597	0.0930	0.0073
	Ave	830	0.0857	0.1249	0.0043
	FP	393	0.0645	0.1041	0.0052
	FN	428	0.0623	0.0919	0.0044
LULCPast1k	.00	668	0.0347	0.0780	0.0030
	1.00	162	0.0494	0.0955	0.0075
	Ave	830	0.0375	0.0818	0.0028
	FP	393	0.0473	0.0885	0.0045
	FN	428	0.0520	0.0979	0.0047
LULCGrass1k	.00	668	0.0240	0.0466	0.0018
	1.00	162	0.0196	0.0362	0.0028
	Ave	830	0.0231	0.0447	0.0016
	FP	393	0.0196	0.0466	0.0024
	FN	428	0.0192	0.0389	0.0019
LULCShrub1k	.00	668	0.0093	0.0279	0.0011
	1.00	162	0.0065	0.0214	0.0017
	Ave	830	0.0088	0.0268	0.0009
	FP	393	0.0072	0.0214	0.0011
	FN	428	0.0074	0.0238	0.0011
LULCBarren1k	.00	668	0.0024	0.0089	0.0003
	1.00	162	0.0085	0.0316	0.0025
	Ave	830	0.0036	0.0162	0.0006
	FP	393	0.0049	0.0119	0.0006
	FN	428	0.0044	0.0140	0.0007
Proj_baseflow_cfs	.00	668	38.4131	138.6414	5.3642
	1.00	162	63.6401	342.8489	26.9368
	Ave	830	43.3369	195.9434	6.8013
	FP	393	39.5895	150.7821	7.6060
	FN	428	97.7761	432.1851	20.8905

Section C.3 – Continued

Group Statistics *Faxonius virilis*

Variable		N	Mean	Std Dev	SEM
Lat	.00	668	44.2394	1.5984	0.0618
	1.00	162	43.8569	1.5860	0.1246
	Ave	830	44.1647	1.6022	0.0556
	FP	393	43.9496	1.6055	0.0810
	FN	428	43.9839	1.5713	0.0759
Long	.00	668	-85.4140	1.6984	0.0657
	1.00	162	-84.9702	1.6386	0.1287
	Ave	830	-85.3274	1.6951	0.0588
	FP	393	-85.1454	1.6131	0.0814
	FN	428	-84.9213	1.7118	0.0827

[a] Names and detailed description of all variables can be found in Table 3.1
[b] Number of survey locations
[c] The mean value of the variable across the selected number of survey locations
[d] Standard deviation of the mean
[e] Standard error of the mean
[f] *Faxonius virilis* is not present
[g] *Faxonius virilis* is present
[h] Average across all survey locations regardless of *Faxonius virilis* presence
[i] *Faxonius virilis* was predicted to occur at a site it was absent from (false positive)
[j] *Faxonius virilis* was predicted to not be present at a location they were present at (false negative)