SPATIAL ANALYSIS OF TWO AQUATIC INVADERS IN ADIRONDACK LAKES:

A MODELLING APPROACH FOR ENVIRONMENTAL MANAGEMENT

by

Valerie Elaine Bowler

Bachelor of Science, Guelph University, 1986

Bachelor of Education, York University, 2002

Master of Education, State University of New York, 2007

A Thesis presented to Ryerson University

in partial fulfillment of the requirements for the

degree of

Master of Applied Science

in the program of

Environmental Applied Science and Management

Toronto, Ontario, Canada, 2017

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Spatial Analysis of Two Aquatic Invaders in Adirondack Lakes: A Modelling Approach for
Environmental Management

Master of Applied Science, 2017

Valerie Elaine Bowler

Environmental Applied Science and Management, Ryerson University

Abstract

The global expansion of humans has stressed the natural world, removed boundaries between
continents and habitats and exposed natural areas to invasive species. These cause billions of
dollars of damage yet there are limited funds given for their management. Predictive tools can be
used to develop pro-active strategies for managing invasive species and this study developed such
a tool. Publicly available data were used to build predictive models for the presence of two invasive
species, curly-leaf pondweed (*Potamogeton crispus*) and Eurasian watermilfoil (*Myriophyllum
spicatum*) within the Adirondack Park (New York State). Predictors were identified through:
bivariate analysis to test the variables; ordinary least squares regression to build predictive models
and logistic regression to validate those models; geographically weighted logistic regression to
evaluate local impacts. Models were ranked by Aikake information criterion minimization and
evaluated with McFadden's rho-squared, standard coefficients and variance inflation factors. The
top five models for each invasive species established seven predictors for curly-leaf pondweed and
nine predictors for Eurasian watermilfoil. Geographically weighted regression, a local analysis,
was found to be a definite improvement over the global analysis for watermilfoil but not for
pondweed. Two predictors (lake elevation and distance to Interstate-87) were significant in all the
top models for both species. The identified predictors provided a group of characteristics that could
be used to identify vulnerable lakes and prioritize management strategies. Even though these
findings were specific to the Adirondack Park, this approach could be applied to other invasive
species or other areas to help in the decision-making process for management.

Acknowledgements

There are a few people I would like to thank that have been supportive and behind me through the journey of this thesis.

I would like to thank my supervisor and mentor, Dr Richard Shaker who was extremely patient and positive in every circumstance. You have taught me so much about statistics and how it can be practically applied to projects and real-life situations. Also, your data set (that you so diligently gathered and maintained) was generously given to me to use and analyse in order to find out more about invasive species and to work towards a practical application. This shows your dedication to not only the ecological issue of invasive species but also your concern for the Adirondack Park in particular. You were always available and helpful with all my questions and ideas throughout the process of this thesis. I could not have had a better coach.

I would also like to thank my friends who have been tremendously patient with me during this time and my parents who have supported me and encouraged me constantly. I could not have done this without them.

Thank you all for your support!

Table of Contents

List of Tables

List of Figures

List of equations

List of Appendices

List of Abbreviations

Abbreviation	Explanation
AIC	Aikake Information Criterion
AICc	Aikake Information Criterion (corrected)
AIS	Aquatic invasive species
AISR	Aquatic invasive species richness
APA	Adirondack Park Authority
APIPP	Adirondack Park Invasive Plant Program
CBD	Convention on Biological Diversity
CLP	Curly-leaf pondweed
EWI	Eurasian watermilfoil
GAM	Generalized additive models
GWR	Geographically Weighted Regression
GWLR	Geographically Weighted Logistic Regression
I-87	Interstate 87
IS	Invasive species
MLR	Multiple Logistic Regression
OLS	Ordinary Least Squares
Park	Adirondack Park
SAM	Spatial Analysis in Macroecology
SDM	Species Distribution Models
St. Coef	Standard Coefficient
VIF	Variance Inflation Factor

Chapter 1: Introduction

Early colonization is recognized as a primary vehicle for spreading species around the globe (Richardson and Pyšek, 2007). As trade and commerce increasingly crisscrossed the globe, there has been a corresponding increase in the number of 'non-native' species introduced into new and previously unavailable lands. The expansion of the global economy meant that boundaries dissolved and ecosystems that were once separated and contained became interspersed with new alien species (Keller and Perrings, 2011). These non-native species may have survived in the new environment and if they were able to get established can become invasive species (IS). Invasive species are "those taxa that were introduced recently and exert a substantial negative impact on native biota, economic values or human health" (Hellmann et al., 2008). Introductions of these species can be accomplished by a variety of different activities, both intentional and nonintentional. Ten percent of the newly introduced species tend to survive, and of those, only ten percent successfully adapt and are therefore able to spread (Keller and Perrings, 2011). In the new environment, these species can negatively affect established organisms as well as the biological diversity of the community (Hejda et al., 2009). The spread of invasive species is recognized by ecologists and biologists as a serious global burden and a strong driver for the loss of global biodiversity (Bright, 1999; Keller and Perrings, 2011). There are numerous aspects to the decline in global biodiversity of which five of significant importance were summarized by the acronym HIPPO: habitat destruction, invasive species, pollution, population growth, and overharvesting (Wilson, 2002). Scientists argue that IS are second only to habitat loss for causing global decreases in biodiversity (CBD, 2011; Hulme and Firn, 2015). It has been estimated that "approximately 400 of the 958 species listed as threatened [in North America] under the Endangered Species Act are considered to be at risk primarily through competition or predation by non-native species" (Pimentel et al., 2005). The Earth Summit in 1992 put the spotlight on IS by including them in the mandate to protect endangered species because of their potential impact on these species (CBD, 2011). The Aichi Targets were set up at the tenth meeting of the Convention of Biological Diversity (CBD) in 2010 in Aichi, Japan as part of the strategic plan for the conservation of biodiversity. Target 9 specifically involves IS and states the need to learn more and work towards their eradication (CBD, 2011; Brundu, 2015).

The number of IS has continued to rise, affecting every country on earth. The IS issue remains problematic and countries have been allocating funds for attempted eradication or basic

management of these organisms (Pimental et al., 2005). The success of invasive species is thought in part to be determined by propagule pressure (a combination of the number of invasive individuals and the number of times they are introduced) which gives opportunities for such species to arrive in suitable areas and become established (Sheppard et al., 2006; Uden et al., 2015). There is also a distinct correlation between the number of non-native species that successfully became invasive and the amount of human disturbance in that area (Richardson and Pyšek, 2007; Keller and Perrings, 2011). It has been estimated that between one third and one-half of the earth's land cover has been changed by humans (Vitousek et al., 1997). Human activities degrade the environment by changing the physical characteristics, thereby causing stress to the native species, and allowing strong, aggressive IS to prosper (Keller and Perrings, 2011). Early efforts to manage IS focused on identifying potential characteristics that helped them succeed, but that was not enough. The next steps were to determine the mechanics of the spread which required extensive knowledge of the connection between IS and humans. These factors became the focus of many studies and research is ongoing because of the complexity of the problem. Not only did it involve a combination of countries for every IS, but it also required specific regional work because an IS can behave differently in different environments. Thus, there was no simple strategy for dealing with and controlling IS and it was generally accepted that for effective management, IS must be investigated and treated at a regional level (Dullinger et al., 2009). Since the 1980s, research has been done to understand how some species become invasive, but very little research was done on managing the IS spread (Sheppard et al., 2006). Identifying the pattern of invasions allowed preventative strategies to be implemented but finding a single effective strategy was difficult. There were a variety of approaches taken, from looking at available ecological niches, to IS impact on the environment to looking for abiotic conditions that were suitable for IS (Kolar and Lodge, 2001; Peterson et al., 2003; Kearney and Porter, 2009; Pyšek et al., 2012). Predictive studies were more promising as a management strategy, but no one approach has been found. Even though studies looked at IS from different perspectives, all of them have added to the knowledge base needed to develop control measures and policies for IS. Controlling the spread of invasive species is paramount to stopping IS and is referred to as the "weakest link scenario" which suggests that if the spread were controlled, it would make all the other issues (i.e. displacement of native species, the impact on the environment) manageable (Keller and Perrings, 2011). Conversely, if the spread of invasive species is not controlled, the battle against them and the damage they cause, would be

lost. A significant problem with stopping the spread is that IS are able to use a variety of dispersal mechanisms and seem able to adapt to difficulties. This leads to problems for resource managers trying to set up policies to protect natural areas against IS.

Humans have been trying to understand the natural world for centuries. To achieve this, various measuring tools were used and developed over time. Statistical analysis, however, has only been used for environmental investigations (as a way of representing observations quantitatively) since the early 1900s (Hastings et al., 2005). Fisher (1937) was one of the first scientists to use statistics for biological studies and this has become a widely used method to investigate interactions between organisms and environmental conditions. However, research strategies used across the biological fields have advanced and changed in tandem with mathematics and technology (Fortin and Dale, 2005). Ecological research often involves a variety of spatial processes and the environments being studied show spatial heterogeneity which should be considered for effective analysis (Fortin and Dale, 2005). As a result, spatial analysis has developed a variety of techniques and strategies and has now become standard practice for considering spatial patterns quantitatively (Lennon, 2000; Liebhold and Gurevitch, 2002; Wagner and Fortin, 2005; Miller and Hanham, 2011; Páez et al., 2011). According to Liebhold and Gurevitch (2002), spatial analysis is especially important to ecologists because they like to quantify spatial patterns and they "recognize that there can be underlying spatial pattern that can have a significant impact" on the interactions being studied.

An advantage of spatial analysis is that it can be used in various ways and applied to so many different research topics (Liebhold and Gurevitch, 2002; Fortin and Dale, 2005). However, this flexibility may lead to poor results if it is used carelessly and if the assumptions embedded in the traditional statistics are ignored (Legendre, 1993; Lennon, 2000; Liebhold and Gurevitch, 2002). Analysts have begun investigating this problem and Lennon (2000), in his important work on the use of spatial analysis, questioned the effectiveness of even simple statistical analysis if spatial autocorrelation were not adjusted for or acknowledged in the model and he wondered at the validity of past research in light of this. Lennon (2000) found that if spatial autocorrelation was ignored, the statistical outcome could show incorrect significance levels thereby resulting in a bias towards certain ecological patterns over the long term. This questioning of the proper use of spatial analysis has resulted in more rigorous statistical testing and a search for other statistical processes that might deal with autocorrelation (Liebhold and Gurevitch, 2002). Geographically weighted

regression (GWR) has gained attention as an alternative with many benefits. Geographically weighted regression was being used initially in geographic circles, but more recently it has been adopted by biogeographic studies and ecological investigations (Wagner and Fortin, 2005; Tu and Xia, 2008; Windle et al., 2010). Studies found that GWR decreased problems common in spatial analysis used with traditional statistical analysis (Tu and Xia, 2008; Miller and Hanham, 2011). While GWR was originally developed for data with Gaussian distributions it has advanced to the point where it could be used with Poisson distributions as well as logistic distributions. This was advantageous to ecological studies and was employed in this thesis by using geographically weighted logistic regression (GWLR) to explore its suitability for this type of spatial analysis. Previous work completed within Adirondack Park was used as a foundation for this spatial analysis. The overarching guiding question for this research was: can publicly available data be used for meaningful and effective research that could be used by natural resource managers to control or slow the spread of invasive species? In order to answer this question, a data set gathered by Dr. Richard Shaker from public sources was used to develop a transferable strategy or technique that used statistical analysis to build predictive models on which decisions could be based. This thesis used numerous lake and landscape variables to determine the best predictors for the presence of two aquatic invasive species (AIS), curly-leaf pondweed (*Potamogeton crispus*) and Eurasian watermilfoil (*Myriophyllum spicatum*) across 126 lakes in the Adirondack Park of New York State. This provided inferential information on how the lake and landscape characteristics related to both of the aquatic invaders and helped answer this study's guiding research question. The statistical relationships found between the two different aquatic invaders will be compared and contrasted. This was accomplished by testing and formatting the lake and landscape variables so they could be examined using various statistical processes that were used to build predictive models for each aquatic invader and to evaluate the predictors in each model. Geographically weighted logistic regression was also employed for each model to investigate the impact of local variation on the findings. Overall, the goal was to produce predictive models that could be used as a tool to easily identify lakes in the Adirondack Park susceptible to future invasions by these AIS by matching the predictors from the models with the lake and landscape characteristics of the lakes. This would supply a tool for resource managers in their efforts to slow the spread of IS.

Chapter 2: Background

Charles Elton brought IS to the forefront with *The Ecology of Invasions by Animals and Plants* in 1958. He emphasized the connections between the spread of IS and human colonization (Elton, 1958; Richardson and Pyšek, 2007). Several theories on IS and invasion resistance were put forward by Elton and are still believed valid today (Elton, 1958). However, many studies have been conducted to test Elton's ideas such as the "diversity-stability hypothesis" (ecosystems with higher diversity will be more resistant to invasion) (Elton, 1958; McNaughton and Wolf, 1970; Levine and D'Antonio, 1999; Borrvall, 2000; Tilman et al., 2002) and the "simplification of nature" idea that refers to how humans change the environment in a way that helps the (Elton, 1958; Richardson and Pyšek, 2007; Davis, 2011; Besek and McGee, 2014). Elton did not put great emphasis on propagule pressure beyond the idea that invasions were associated with global colonization and land use. More precisely, it was the changes to the environment, brought on by human land use that disrupts the native populations and increases niche availability for invaders (Devine, 1998; Hellmann et al., 2008; Shandas and Alberti, 2009). Land use in proximity to water systems affects the aquatic biota in those water systems, changing the aquatic conditions of that ecosystem (Roth et al., 1996; Shandas and Alberti, 2009; Shaker and Ehlinger, 2014). Kennedy et al. (2002) stated that human activity "breached the isolation of communities" and destroyed a level of resistance within that community thus aiding the import of IS. Human-mediated dispersions have been found to be one of the greatest methods of IS spread, indicating that human activity, rather than other factors such as a species adaptability to new conditions or their generalist tendencies helped IS establish in a new habitat (Trebitz, 2007; Francis and Pyšek, 2012; Hulme and Firn, 2015). Other studies found that the influence of propagule pressure "swamped the effects of other postulated influences on [a species] establishment" because it was the variations in propagule pressure (number of releases in an area vs. number of actual introductions) that determined the success of an IS (Lockwood et al., 2005).

2.1 Understanding aquatic invasions

Fresh water ecosystems usually have high levels of biodiversity but they easily become degraded by human disturbance and development making conditions favourable for IS (Francis and Chadwick, 2012). This is partly due to the high connectivity between freshwater systems as well as the fact that water is such an effective medium for movement and transport (Francis and Chadwick, 2012). As well, aquatic invasions are "particularly prevalent and damaging in aquatic

systems" and freshwater communities are highly vulnerable to biotic exchange (Havel et al., 2015). This is because the strong trophic connections are disturbed by the introduction of new species and thus extensive IS management is needed to protect the integrity of these systems (Kulhanek et al., 2011; Gallardo et al., 2015). Invasive weeds and pests (terrestrial and aquatic) impact economies through direct losses, biological degradation and their effect on human health and wealth (Bright, 1999; Pyseck et al., 2012). Invasive species cross international borders which pose logistic problems for developing policies. Historically, international policies have been difficult to implement and were not very successful due in part to the fact that they often "ran up against far more powerful policies and interests that, in one way or another, encourage invasions…[with] the increasing integration of the world's economies … making a bad situation worse" (Bright, 1999). As a result, it falls to regional authorities to set up their own policies in efforts to protect the natural communities under their domain.

There are three stages generally accepted as the "stages of invasion" by which non-native species become invaders which must be understood for effectively managing biological invasions (Elton, 1958; Mills et al., 1993, Leung et al., 2002). These stages (Figure 1) are referred to as the colonization stage; the lag stage or establishment stage; the expansion stage (Leung et al., 2002; Shea and Chesson, 2002; Hellmann et al., 2008). In the lag phase, the non-native species adjust to the new environment and become established (Elton, 1958; Mills et al., 1993; Sheppard et al., 2006; Hellmann et al., 2008). This time lag between the introduction of the species to a new environment and the point at which its presence is recognized could be extensive (Mills et al., 1993; Johnson and Padilla, 1996; Bright, 1999; Ricciardi, 2006; Pagnucco et al., 2014). Each stage can be managed, but the cost increases and the effectiveness decreases as the stages progress. An aquatic example worth noting is the zebra mussel (*Dreissena polymorphy*) which was believed to have been introduced into Lake St. Clair in the ballast water of Eurasian freighters two years before they were actually documented as an IS in 1988 (Mills et al., 1994; Briski et al., 2014). This lag phase of two years made it difficult for resource managers and invasion biologists, and by the time they were noticed, they were established and may have already been in the expansion stage making control even more difficult (Mills et al., 1993; Buchan and Padilla, 2000). Thus, knowing where an IS might spread or colonize, is paramount to successful management yet advancements in invasive species ecology are often slowed by the difficulties of providing "robust predictions" (Besek and McGee, 2014). One method for management is to identify the vulnerability of a lake

6

to future invasions (Vander Zanden and Olden, 2008). Ideally, this would mean actions could be taken to prevent the IS from entering a habitat which is suitable for them (Mills et al., 1993; Leung et al., 2002; Espeland, 2013). Efforts to identify the invasibility of a lake look at a mix of several factors such as, a) if the needs of the IS can be met in the lake being considered, b) the type of propagule pressure that might occur in that lake and, c) the possible interactions that might occur with the native species – competition or predation (Vander Zanden and Olden, 2008).

Figure 1: Three stages of invasion identified for nonnative species in a new environment, with suggestions for what management strategies might be possible at each stage. (adopted from Leung et al., 2002; Hellmann et al., 2008)

Knowing the process of how non-native species become invasive and why they are successful would also be helpful in preventing their spread (Elton, 1958; Leung et al., 2002). Although Elton (1958) did not consider dispersal and phenotypic plasticity as primary reasons for invasion success, these are both now considered to play a major role in the ability of an IS to become a successful invader (Buchan and Padilla, 1999; MacIsaac et al., 2004; Lockwood et al., 2005; Capers et al., 2009; Espeland, 2013). The zebra mussel, for example, is easily transported to new environments because it can survive in several life cycle stages (larval, juvenile and adult) making it difficult to control them (Johnson and Padilla, 1996). Aquatic plants such as Eurasian watermilfoil (*Myriophyllum spicatum*), have excellent dispersal capabilities and spread by fragments which have the ability to develop roots asexually that grow quickly into an adult plant (Mills et al., 1993; Martin and Valentine, 2014). These fragments are able to survive out of water for some time and are also easily transported via water flow through creeks, on/in boats moving

between lakes or even on fishing equipment (Mills et al., 1993; Buchan and Padilla, 2000; Martin and Valentine, 2014; Bruckerhoff et al., 2015). Species that change growth patterns to adapt to changing conditions, such as to shallow water from the preferred deep water, are well on their way to being able to grow and proliferate in foreign and "unsuitable" environments (Elton, 1958; Gallardo et al., 2015). It was found that aquatic invasive plants, in particular, changed the environment in a way that helped other invasive plants thereby encouraging more invasions (Vitousek et al., 1997; Leung et al., 2002; Pyšek et al., 2012; Brundu, 2015). For instance, the water hyacinth (*Eichhornia crassipes*) can grow in any water depth and forms a floating mat that provides a shallower environment for other aquatic invasive plants that could not grow in deep water (Brundu, 2015). Many invasive plant species are also highly competitive and able to form dense monocultures that compromise the native species (Keane and Crawley, 2002). When entering a novel environment, IS often altered the new area and changed the function of the ecosystem (Mills et al., 1993; Devine, 1998; Besek and McGee, 2014). Thum and Lennon (2010) found that aquatic plant invaders had "a high probability of changing the sediment and nutrient levels within the aquatic environment". This occurs when the IS disrupts the balance of the new ecosystems, displacing similar species, consuming or attacking native species, or simply using nutrients more efficiently than the native species (Mills et al., 1993; Pyšek et al., 2012; Havel et al., 2015).

2.2 Humans and invasive species

Human activity or disturbance also changes the water chemistry and water quality, making the abiotic conditions intolerable to the native species but quite acceptable to IS (Roth et al., 1996; Wilcove et al., 1998; Ricciardi, 2006). Humans also play a significant role in moving IS around and introducing them to new areas. According to Mills et al. (1993), of the invasive aquatic plants introduced to the Great Lakes specifically, most were introduced accidentally. Unfortunately, intentional releases are also common. Lakes used for aquaculture, where fish farmers often raise non-native fish species, have had some of these fish escape into the freshwater ecosystem (Mills et al., 1993; Devine, 1998). In order to entice more visitors, it has been common practice for a lake district to stock its lakes with fish for the enjoyment of anglers, thus releasing fish that could be non-native (Mills et al., 1993; Devine, 1998; APIPP, 2015). The stocking of fish is a deliberate release which in turn attracts more anglers for recreational purposes, bringing in more tourists and money. Unfortunately, the increased movement of anglers between lakes also carries invasive

plants or animals as stowaways or hitchhikers on the boats (Mills et al., 1993; Devine, 1998; Shaker and Erlinger, 2014). Knowledge of possible entry points for AIS can be used to develop prevention strategies that could be used to control their movement (Vitousek et al., 1997; Bright, 1999; Hulme and Firn, 2015). In addition to the impact invasive species have on the natural communities, they also have an impact on human health and welfare (Elton, 1958; Mills et al., 1994; Keller and Perrings, 2011; Gallardo et al., 2015) and the overall economy in a variety of ways (Mills et al., 1993; Devine, 1998; Pimentel et al., 2005). Up-to-date costs were difficult to locate, but pooling information for North America, it seems that AIS cost billions of dollars per year due to the fouling of equipment, blocking of intake pipes used for industry, drying up of needed wetlands or lowering of revenue from tourism - and this estimate did not include costs to control the AIS (Mills et al., 1993; Devine, 1998; Pimentel et al., 2005; Gallardo et al., 2015). In order to counter the influx of new species and slow their spread, there is a need for research that focusses on keeping invasive species' populations in check. Some type of tool that could identify vulnerable areas would be a great advantage for prioritizing management activities and expenditures (Devine, 1998; MacIsaac et al., 2004; Hulme and Firn, 2015).

If a species becomes established despite efforts or through the absence of management, it could increase its population to the point of extensive dispersal – the expansions stage (Figure 1) (Keane and Crawley, 2002; Peterson et al., 2003). Because of this, propagule pressure and human movement as well as structures like boat canals and aqueducts, have helped the IS spread. In the past, if an IS was discovered in a new area, large amounts of money were spent to eradicate or at least control them (Mills et al., 1993; Mills et al., 1994; Devine, 1998; Hansen et al., 2016). Established populations act as "stepping stones" for secondary spread of that species into nearby habitats and therefore those pathways must also be controlled (MacIsaac et al., 2004). One strategy used to control IS spread was to prevent already established populations from reaching the expansion stage through actively managing the population (Buchan and Padilla, 2000). Controlling established populations is time-consuming and expensive, but it is still more cost effective than trying to eradicate them once they reach the expansion stage (Pimentel et al., 2005; Dullinger et al., 2009). However, the best scenario for managing invasions (Figure 1), is to block the initial entry point and then set up a monitoring system to ensure its effectiveness. As noted previously, IS are very difficult to locate and identify in the colonization stage since their population is low and early detection in aquatic environments is even more difficult than on land (Brundu, 2015).

Therefore, another strategy must be used to locate them in a timely manner. The ability to predict where a particular invasive organism might spread, means that energy can be channelled to that body of water in anticipation of an invasion (Les and Mehroff, 1999; Peterson et al., 2003; Keller and Perrings, 2011; Havel et al., 2015). Control barriers and monitoring schedules can then be set up early, to protect that lake (Thum and Lennon, 2010).

The management of freshwater systems has broadly included under natural land management, but since aquatic biota have unique issues, they really need to be dealt with separately. Aquatic organisms are particularly vulnerable due to the connectivity between bodies of water and as a result, freshwater ecosystems are believed to be one of the most endangered ecosystems in the world (Abell et al., 2007; Francis and Pyšek, 2012; Brundu, 2015). Many aquatic invasive plants can reproduce both sexually and asexually and can develop from spores, root tips or fragments, all of which are easily transported (Martin and Valentine, 2014; Brundu, 2015). With regards to managing aquatic invasions, funds were typically limited and thus conservation authorities had a difficult time monitoring all the lakes involved – too many lakes, too little money (Thum and Lennon, 2010; APIPP, 2015). Managers are doing their best with limited resources by using volunteers and prioritizing the lakes according to the presence of IS (Leung et al., 2002; Thum and Lennon, 2010). To help the resource managers determine which lakes need attention, predictive strategies are being investigated. Many focus on the physical conditions needed by invasive species (Buchan and Padilla, 2000; Peterson and Vieglais, 2001; Thum and Lennon, 2010). However, this did not take into account the proximity of "suitable" lakes to established populations (proximity and propagule pressure) and did not take into account the importance of human disturbance (access/impact) (Peterson et al., 2003; Lockwood et al., 2005; Dullinger et al., 2009). The presence of human activity has been found to correlate with IS. As a result, variables that represent human activity should be considered when determining whether or not a lake is suitable for IS. Roth et al. (1996) and Shandas and Alberti (2009) definitively showed that land use and land forms affected nearby aquatic biota and that a change in land use or the removal of inland forest from around an aquatic system was found to change the water quality, the water chemistry and even sedimentation within nearby bodies of water. Riparian vegetation "supports aquatic systems by increasing organic input [through] increasing litter and debris" and a change to these areas impacted the nutrients available within the aquatic system (Roth et al., 1996). Thus including data that represented these activities when developing a predictive model would be

crucial (Roth et al., 1996; Johnson et al., 2001). Past researchers used land cover and land use metrics (along with other characteristics) as proxies for environmental conditions in statistical analysis which is now an acceptable practice if relationships have been established by previous research (Dullinger et al., 2009). Invasion biologists know that preventing the colonization stage is the best strategy for fighting IS but if blocking their transport/entry were not successful, preventing the next stages (establishment and further spread) becomes imperative (Les and Mehrhoff, 1999; Dullinger et al., 2009; Thum and Lennon, 2010). Because funding is often limited compared to the extent of the problem, resource managers need to determine how their efforts can have the greatest success.

2.3 Previous invasive species research

Previous studies by Buchan and Padilla (2000) and Peterson et al. (2001) to name a few, focused on the physical conditions needed by IS – the idea of the IS finding a suitable habitat. The theory is that if conditions typical of the species' native environment were present in a new habitat, then that area could potentially be invaded (Peterson et al., 2003; Gallien et al., 2010; Smolik et al., 2010). Buchan and Padilla (2000) used a variety of easily available, and thus not overly expensive, landscape variables that represent the lake environment to predict the presence or absence of Eurasian watermilfoil using logistic regression. A number of metrics were used to determine if the lakes contained suitable habitat, and it was found that "percent forest cover" in the watershed had a greater effect on the presence of Eurasian watermilfoil (EWI) than did human access to lakes (boat ramps, proximity to highways, etc.) or the species of fish in the lake (Buchan and Padilla, 2000). Buchan and Padilla (2000) suggested the connection between forest cover and EWI was because forest in the watershed affects the levels of dissolved inorganic carbon in the water which is known to affect EWI growth (Keast, 1984; Smith et al., 2002). Other studies on IS were not explicitly predictive, but they looked at the "vulnerability" of a lake or the opposite of what Elton termed "ecological resistance" (Elton, 1958; Johnston et al., 2001; Kearney and Porter, 2009). Johnson et al. (2001) looked at human activity and considered boat activity, combined with the number of boat launch sites on a lake, as an indicator of accessibility to a lake and thus the vulnerability of a particular lake. All these studies used traditional statistical analysis such as ordinary least squares (OLS), linear regression/logistic regression, generalized additive models (GAM) and other traditional statistical techniques that were all global measures. A review of spatial analysis used for studying the spread of invasions by Hastings et al. (2005), acknowledged

that "recent theoretical work has shown that IS spread is a much more complex process than the classical models suggested, as long-range dispersal events can have a large influence on the rate of range expansion through time". The authors go on to discuss the limitations of spatial analysis when large heterogeneity and variations exist across the study. A limitation of previous studies was related to autocorrelation - very common in ecological studies (Buchan and Padilla, 2000). These assumptions are often violated when conducting spatial analysis. Lennon (2000) pointed out that many ecological studies used the traditional methods without proper consideration of their short comings. According to researchers that studied the effectiveness of GWR and its use in biogeographic studies, it corrected some of these issues (Fotheringham et al., 2002; Tu and Xia, 2008; Fotheringham and Oshan, 2016). Such studies indicated that GWR might give better, more accurate results than the traditional methods, especially when used with spatial data.

2.4 Using geographically weighted regression in ecological studies

Geographically Weighted Regression is considered a "local model" in that the calculations incorporate the spatial variation around each observation point (Miller and Hanham, 2011), and gives results that compare variations within a local area. Geographically weighted regression was considered "much more flexible in estimating local parameters" and thus was suitable for a local analysis rather than a global analysis using traditional regression. Furthermore, GWR was able to quantitatively explain these variations in relation to environmental factors (Windle et al., 2010). It can be used to build models, similar to traditional species distribution models (SDMs), but GWR also allows examinations at different scales by changing the size of the bandwidth used within the program (Miller and Hanham, 2011). Bandwidth can be explained as "the distance within which other areas will have influence", and it defines the distance-decay parameter within a study, determining search radius and neighbourhood (Miller and Hanham, 2011). In GWR, bandwidth can be set at a fixed distance or size if desired or it can be set as an "adaptive kernel" that can shrink or expand within the analysis (Miller and Hanham, 2011).

As mentioned previously, spatial analysis has certain assumptions built into it that, if violated, could impact the outcome and show errors that can result in false relationships being accepted (Fotheringham and Oshan, 2016). Liebhold and Gurevitch (2002) suggested that ecological studies have so much embedded spatial dependence that analysis of the ecological data often leads to those assumptions being violated. The assumptions referred to are: the assumption of independence between the subjects/process and the assumption of "stationarity" (Fortin and

Dale, 2005). Stationarity refers to the processes (or models of processes) as being independent of both the direction in space and the location (Fortin and Dale, 2005). When using spatial analysis, it is possible to adjust for the presence of spatial dependence and some spatial autocorrelation can be corrected (Liebhold and Gurevitch, 2002). But Lennon (2002) feared that these adjustments are often not done, resulting in inaccurate conclusions. Fotheringham and Oshan (2016) presented GWR as a robust application that helped to minimize these same issues when compared to simple spatial analysis. Nevertheless, analysts must be aware of the issues in order to ensure the interpretation of GWR results is done correctly. Tu and Xia (2008) examined spatially varying relationships between a set of variables by specifically comparing OLS with GWR and concluded that "GWR models showed great improvements of model performance over their corresponding OLS models". They suggested that GWR would serve as a useful tool for research in environmental issues at many levels; from a local, regional scale to a global scale (Tu and Xia, 2008). Geographic studies have used GWR for decades, but it has only been used for ecological studies since early 2000 (Brunsdon et al., 1998; Harris et al., 2010). However, many researchers have been sceptical of its use and have been testing it against the traditional regression they normally used rather than using it for analysis of their data. They concluded that it could be used for investigating non-stationarity but "caution should be exercised" when drawing conclusions on spatial relationships with GWR only (Páez et al., 2011). Geographically weighted regression was originally designed for Gaussian distribution but scientific knowledge and technology has developed so that GWR methods now include Poisson distribution and logistic distribution (presence/absence) so it can now be applied in more situations. Some recent studies have successfully conducted GWR for predictive purposes, not just to test it against global models. For instance, Windle et al. (2010) used GWR for predicting the distribution of cod in the Pacific Ocean, and Tu et al. (2014) applied it to the relationship between land use and water quality.

Chapter 3: Study purpose

The purpose of this study was to use easily accessible data to conduct spatial analysis at a macroecology level that can then be used by natural resource managers to create effective strategies against IS. This was accomplished by using a number of lake and landscape variables that are regularly collected or readily available from open sources, to build predictive models for the presence of two AIS; curly-leaf pondweed (*Potamogeton crispus*) and Eurasian watermilfoil (*Myriophyllum spicatum*). Modelling is used in ecological studies to quantify the "species-environmental relationships by specifying the underlying processes or by predicting the observed patterns of the organisms from the spatial distribution of environmental factors" and can be done with statistical analysis (Wagner and Fortin, 2005). The freely accessible statistical software Spatial Analysis in Macroecology or SAM (Rangel et al., 2010), was used to build models which could be used by resource managers in the decision-making process for developing invasive species strategies. These predictive models could be used to identify what lakes might be susceptible, allowing for the development of pro-active strategies. This thesis was confined to 126 lakes in the Adirondack Park (referred to as the Park from this point) and focused on two IS, curly-leaf pondweed (CLP) and Eurasian watermilfoil (EWI). These two species of AIS were used to determine if this modelling technique could be employed for individual species or if the benefits of individual analysis were limited, suggesting that analysis of a group of AIS could perhaps be more effective overall. Both of these species were recognized as problem IS in the Park and thus any research on their management would be beneficial (Nichols and Shaw, 1986; APIPP, 2014). A difference in results between the species might suggest an advantage to conducting individual analyses for effective AIS management, especially if one was particularly aggressive or had a greater impact on the environment, such as EWI. This study was also put forth to test if significant statistical analysis could be done on IS if only publicly available data were used. The purpose for doing this was to find out if useful analysis would be possible even if, due to limited funding, there was no money available for specific data collection. The lake and landscape variables used in this thesis were from government sources as well as from environmental organizations that gathered regular data that were publicly accessible.

3.1 Objectives

This thesis was set to achieve three objectives: 1) to conduct bivariate analyses to determine the relationship of a variety of lake and landscape variables to the presence of two AIS, curly-leaf

pondweed and Eurasian watermilfoil separately, 2) to use multiple regression to build predictive models from those variables and evaluate the models for each AIS separately, 3) to explore available software (GWR4, 2016) to conduct local multiple logistic regression by employing GWLR using the same predictive models from the second objective, and determine if GWLR was an improvement for this type of environmental study. In an exploration of the new methods being used within GWR, GWLR was employed and the findings contrasted to determine its usefulness in ecological studies. In order to achieve these objectives, a large variety of lake and landscape metrics (including land cover and configuration) representing different environmental characteristics and parameters, made available from previous research in the Adirondack Park, were analysed (Shaker et al., 2013).

3.2 The study area

　　Various organisations in the Adirondack Park (Park) have been monitoring the lakes and collecting data for at least 15 years, and thus the Park presents itself as a useful study area. The Park is located in New York State and comprises about one-fifth of the State. The Park has over 100 towns/villages within its boundary with approximately 132,000 permanent residents (APA, 2015). Being within a day's drive of New York City, it is a hotspot for outdoor recreation and a popular holiday destination. The Park is a mix of private and public lands resulting in varied types of land-use, with various levels of legislation (APA, 2015). There are different types of industry within the Park such as logging, agriculture and urban development, all of which can affect the natural environment. Despite the presence of different levels of industrial activity, there are some areas designated as "wilderness" that are to remain unoccupied and cannot be sold or leased (APIPP, 2015). However, these wilderness areas are still used for recreation and are thus exposed to IS through anthropogenic sources. There are over 10,000 lakes and ponds within the Park (APA, 2015; APIPP, 2015). Many are linked by rivers and streams giving connectivity between the lakes within the Park, thus permitting transfers and exchanges of AIS. All of these characteristics make the Park a unique area that presents many difficulties to resource managers while at the same time making it an important candidate for this type of study.

　　The lakes in the Park are monitored by the Adirondack Park Invasive Plant Program (APIPP) which is associated with the Adirondack Park Authority (APA) and is responsible for coordinating the Aquatic Invasive Species Project and the Terrestrial Invasive Species Project (APA, 2015). The Park is currently monitored for 17 invasive species of which eleven are aquatic,

eight of these are plant IS (APIPP, 2015). This study includes 126 lakes, (all greater than 25 hectares) located within the Park and monitored by APIPP (2014) (Figure 2). Of the 126 lakes
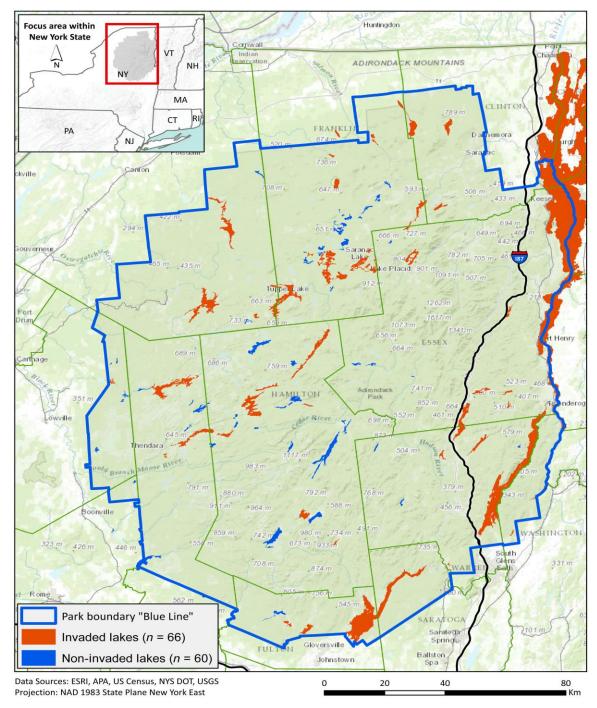

Figure 2: Study area of Adirondack Park delineated by the blue line, with the location of invaded and uninvaded lakes marked. (Shaker et al., 2017)

studied here, 66 are invaded, and 60 are un-invaded according to the data. This area is of great importance because of the collection of watersheds and the fact that still-pristine lakes, ecologically important wetland habitats, as well as remnants of old forest still exist within its border. A portion of the New York Great Lakes Basin, one of the most important freshwater resources in North America, lies within the Park and supplies drinking water and agricultural water to millions of Americans (Martens, 2014). Headwaters from five major drainage basins are located within the Park (APA, 2015). Concerns over the water quality within this region and the impact that IS can have on it has led to the development of the current programs for monitoring and controlling IS within the area (APA, 2015). Over 300 separate waterways are being surveyed by volunteers and Park staff, and as of 2015, 28 percent of those waterways were already invaded (APIPP, 2015; Shaker et al., 2017)

3.3 Species Studied

Two aquatic invasive plant species were studied from the eight found within the Park boundaries (Table 1). Curly-leaf pondweed (*Potamogeton crispus*) and Eurasian watermilfoil (*Myriophyllum spicatum*) were considered significant problems within the Park and lakes are constantly monitored for their presence. Both of these are aquatic macrophytes that produce large submerged beds which can compromise the native species (Mills 1993). Within the Park CLP has been given a threat ranking of 79.79 which was "high" and EWI has been given a threat ranking of 100 which was "very high" and this was the highest of all the aquatic invasive plants in the Park (APIPP, 2015). Four of the other invasive plants in the Park were also listed as "very high" but none were given a ranking of 100. Eurasian watermilfoil has been considered an aggressive invader and was often one of the first plant invaders found in lakes in North America (Les and Mehrhoff, 1999; Martin and Valentine, 2014). Conversely, CLP was usually found in lakes already invaded by other macrophytes (Engelhardt and Ritchie, 2002). Curly-leaf pondweed will grow early in the Spring and become established before the native species start to grow. In this way, they become a serious threat to native macrophytes by using nutrients before other species (Mills, 1993; Engelhardt and Ritchie, 2002). These two AIS have been selected for this in part due to these characteristics as well as their longevity in the Park. Eurasian watermilfoil was first recorded in the Park in 2002 and CLP was first recorded in 2005 (APIPP, 2015).

Table 1: List of aquatic invasive plants monitored in Adirondack Park (APIPP, 2015)

Currently identified aquatic invasive plants	
Common Name	Scientific Name
Eurasian watermilfoil	***Myriophyllum spicatum***
variable-leaf watermilfoil	*Myriophyllum heterophyllum*
water chestnut	*Trapa natans*
curly-leaf pondweed	***Potamogeton crispus***
fanwort	*Cabomba caroliniana*
European frog-bit	*Hydrocharis morsus-ranae*
yellow floating heart	*Nymphoides peltata*
brittle naiad	*Najas minor*
Potential aquatic invasive plants – not yet recorded in Park	
Common Name	Scientific Name
starry stonewort	*Nitellopsis obtuse*
parrotfeather	*Myriophyllum aquaticum*
hydrilla	*Hydrilla verticillata*
Brazilian elodea	*Egeria densa*

Note: Aquatic invasive plants in bold are the two species analysed in this thesis

Chapter 4: Materials and methods

This study looks at two of the eleven AIS currently being monitored and managed by APIPP (APA, 2015). Bivariate analysis plus multiple regression were used to determine bivariate then multivariate relationships to the presence of the two AIS of interest, CLP and EWI, individually. Previous studies on IS most commonly involved either one species being studied or IS as a group.

4.1 Software used

The main software used for this thesis was Spatial Analysis for Macroecology (SAM) (Rangel et all, 2010) and ArcGIS 10.3 (ESRI, 2004) along with the newly updated GWR4 (GWR4 2016). SAM is a freeware application offering a variety of spatial statistical methods that are common to macroecology studies (Rangel et al., 2010). SAM is easy to use and can process large data sets. ArcGIS 10.3 is a "geographic information system that allows management, analysis, and display of geographic information" and was used here for mapping and demonstrating the distribution of each AIS and offered a comparison of invasion patterns across the Park (ESRI, 2004). GWR4 is a powerful software that can carry out simple but common statistical analyses as well as more complicated regression statistics while incorporating regional variations (Rangel et al., 2010; GWR4, 2016). Traditional multiple regression is a global measure, but in order to look at local influences, GWR4 was used. It is a freeware application, updated in 2016, that conducts GWR, (or more specific to this study, GWLR) which can then be compared to traditional regression models to see whether including the local variation in the analysis changes the outcome and would therefore be valuable to the research. The level of any improvement would be determined by comparing the Akaike Information Criterion (AIC) values of the global (traditional statistics) to the local (GWLR) findings. An improvement when using local analysis, would suggest that for ecologically damaging IS it would be worthwhile running this type of statistical analysis so that specific influences were considered and would give more detailed information for any decision making process.

4.2 Data used

The data set used here was compiled by Dr Richard Shaker in "Predicting aquatic invasion in Adirondack lakes: a spatial analysis of lake and landscape characteristics" (Shaker et al., 2017). The data has been collected since 2012 and is continually updated; the current thesis will build on

the findings of the 2016 study (Shaker et al., 2017). Dr Shaker et al. (2017) looked at aquatic invasive species richness (AISR) of eleven species across the Adirondacks, using the same data set. By using the same data set, greater in-depth analysis of the same study area using the same sampling units and metrics can prove advantageous to resource managers. The data set includes numerous lake and landscape metrics that are found within the Adirondack Park and can be used to better understand the characteristics of the Park as well as the lakes within it. This study differs from the study by Dr Shaker et al. (2017) in several ways. This study deals specifically with two of the eleven AIS used by Dr. Shaker and basically computed separate studies for each of the AIS of interest here, CLP and EWI. Dr Shaker et al. (2017) used the Getis-Ord Gi statistic to determine spatial hot and cold spots for AIS richness and by so doing identified 20 lakes that were at risk to future invasions.

Many different sources were used to collect the original data, and specific data from the Adirondack Park Invasive Plant Program (APIPP) and the Adirondack Park Authority (APA) were included along with readily available data from government organisations such as the Department of Transportation (Table 2). ArcGIS was used to gather metrics on distances and other measurable variables (Shaker et al., 2013) and landscapes were assessed using a 300-m riparian zone for each

Table 2: Sources used for research data (Shaker et al., 2017)

Park Associated Sources:	
APIPP	Adirondack Park Invasive Plant Program
APA	Adirondack Park Agency
PRISM	Partners with Regional Invasive Species Management
General Public Sources:	
DOT	Department of Transportation
NYDEC	New York State Department of Environmental Conservation
USGS	United States Geology Survey
SCFG	Sportsman's Connection Fishing Guide
ESRI	Environmental Systems Research Institute

lake (Shaker et al., 2017). *In situ* measurements including elevation, mean lake depth plus area and perimeter measurements were collected and consolidated as well (Shaker et al., 2013). Information on the collection methods as well as any calculations carried out can be found in Shaker et al. (2013) and Shaker et al. (2017). In the 2013 study, Shaker et al. examined landscape and lake variables in order to investigate the significance of public and private boat launch sites on aquatic invasions in twenty-six lakes within the Park. Since the 2013 study, Dr Shaker has

continued to collect data in the Park and these data have been made available for this study (Shaker et al., 2013). The data set included 126 lakes, and this thesis selected a total of 53 possible variables and metrics from that data set (Appendix A).

4.2.1 Preparing the data

This study started with a list of lake and landscape metrics that was reduced to 53 possible variables after searching the literature and considering current IS research. Appendix A lists all variables found by previous research to have some correlation with the spread of IS which were also available within the Park. These might aid in transporting IS, lead to degradation of the habitats, change water quality, affect species diversity, or the number of invasive organisms being introduced, to name a few. Those variables found to have some connection to IS, whether they affected water quality or other environmental conditions that were believed to support IS, were retained for further analysis (Keast, 1984; Lodge, 1993; Buchan and Padilla, 2000; Peterson and Vieglais, 2001; Olson et al., 2012; Kimbel, 2016). The variables were then examined, and any duplicates or redundant variables were removed (i.e. perimeter in kilometres and perimeter in miles), and only one representative variable was retained. Also, diversity measurements were given in both Simpson's indices and Shannon's indices. These are very similar, and to avoid redundancy being carried into the calculations, Shannon's indices were used (SHEI – Shannon's evenness index; SHDI – Shannon's diversity index), and the Simpson's indices were discarded. The remaining variables were put through several steps to prepare them for analysis. It is important to note that parametric statistics have two inherent assumptions: data are normally distributed (normality) and, the data are independent of each other (random). Variables often do not show a normal distribution and can be skewed (Tu and Xia, 2008).

In order to meet these assumptions, the data were first checked for normality. The distribution of the variable should show a "normal" curve or the typical bell curve in order for the statistical tests to be accurate. Using SAM, the variables were formatted, examined, and their distributions graphed (Rangel et al., 2010). Those variables that were not normally distributed were transformed in order to achieve a normal distribution (Osbourne, 2002). In SAM, the variables not normally distributed were transformed using appropriate methods; either log10(x+1) for continuous data or Arcsine – square root for proportional data, such as percentages (Osbourne, 2002; Shaker et al., 2013). Secondly, once the data were accepted as being normally distributed, the variables were checked for autocorrelation (non-randomness) and independence. As previously

mentioned, when spatial analysis is used for ecology, autocorrelation often occurs, and this can bias the results towards the autocorrelated variables, with an increasing strength in the correlation coefficient. This results in high significance levels being given to the correlations more frequently than is acceptable (Lennon 2000). Randomness can be difficult to ensure since it can be affected by the method of sampling. The very nature of environmental parameters is such that they are interconnected with each other and show levels of autocorrelation and thus are not random. With respect to environmental studies, it is generally expected that autocorrelation (non-randomness) is present to some degree (Rangel et al., 2010; Olson et al., 2012; Shaker et al., 2013). However, the validity of the conclusions drawn from these traditional analyses links to the assumption of randomness, and thus it must be at least acknowledged and considered when preparing the data as well as the interpretation of the results (Lennon, 2000). There is a reciprocal relationship between environmental structure and the underlying processes that must also be acknowledged (and incorporated into the interpretation of the statistical analysis) if the ecosystem is to be fully understood (Siesa et al., 2011). This relationship results in a correlation between the variables studied and as such, points to a lack of randomness. Autocorrelation is one measure used to determine randomness, but it is important to note that non-correlation does not necessarily or automatically indicate randomness (Wagner and Fortin, 2005). Spatial autocorrelation is the "lack of independence between pairs of observations at a given distance in time and space" (Dormann et al., 2007). Data points close together can be influenced by the same processes resulting in autocorrelation because these biologically important processes occur within geographically close locations. In environmental studies, it occurs because of the presence of biologically important processes between geographic locations. It should be noted that some ecologists ignore autocorrelation as unimportant while others consider the very presence of autocorrelation as vital to the understanding of a particular environmental issue or problem and thus must be considered when interpreting the findings (Lennon, 2000; Rangel et al., 2010). This thesis did not ignore autocorrelation and the use of GWLR for analysis identified some of these correlations.

4.3 Analysis for the study

SAM was used initially to determine the relationship between the individual variables and the presence of each AIS through bivariate analysis (Figure 3). This bivariate analysis produced a number of values: standardised coefficient of determination, McFadden's rho-squared (ρ^2), Chi-square and p-values of significance. These values were examined to determine

the strength of the relationship of each variable to the presence of each AIS. McFadden's ρ^2 and p-values were used as primary indicators of a relationship. A significance threshold of 90% was applied, and only those variables that fell within this range (at or above 90%) were retained for further analysis. Each AIS showed different variables of significance. This left a relevant subset of variables that already showed a relationship to each AIS as indicated by the p-values. The set of variables for each AIS was then ready for multiple regression.
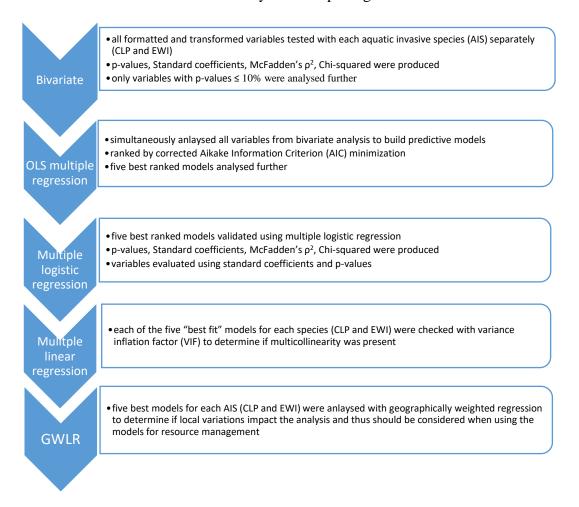
Bivariate
- all formatted and transformed variables tested with each aquatic invasive species (AIS) separately (CLP and EWI)
- p-values, Standard coefficients, McFadden's ρ^2, Chi-squared were produced
- only variables with p-values ≤ 10% were analysed further

OLS multiple regression
- simultaneously anlaysed all variables from bivariate analysis to build predictive models
- ranked by corrected Aikake Information Criterion (AIC) minimization
- five best ranked models analysed further

Multiple logistic regression
- five best ranked models validated using multiple logistic regression
- p-values, Standard coefficients, McFadden's ρ^2, Chi-squared were produced
- variables evaluated using standard coefficients and p-values

Mulitple linear regression
- each of the five "best fit" models for each species (CLP and EWI) were checked with variance inflation factor (VIF) to determine if multicollinearity was present

GWLR
- five best models for each AIS (CLP and EWI) were anlaysed with geographically weighted regression to determine if local variations impact the analysis and thus should be considered when using the models for resource management

Figure 3: Steps taken for the analysis, in order to meet objectives of this thesis

Because environmental variables are not found in isolation in nature, all the variables identified as significant in the bivariate analysis were simultaneously analysed using multiple regression. Ecosystems work synergetically with their components, changing the relationships because of the interactions between them, and thus it was important to run multiple regression in order to test the significance of the variables when not considered in isolation. Usually, when

variables are found to be insignificant they are discarded. In this thesis, those variables selected by the "best fit" models using OLS multiple regression were retained because once a variable was established as being part of a "best fit" model in the initial OLS test, they were kept for further analysis. However, this is not usual. Since there is no exact modelling technique that exists for creating multiple regression for spatial analysis, a multi-model framework was used as a heuristic approach to accomplish the objectives of this study. This multi-model framework used one type of multiple regression (OLS) to build the models by determining which combined variables had the highest level of correlation to each AIS. Secondly, logistic multiple regression was done to analyse and validate the variables in each model. Then as a way to check for collinearity (one of the issues with spatial analysis), the variance inflation factors (VIFs) were produced for each model. This was not a perfect technique and may be the reason why the second analysis (multiple logistic regression) resulted in a lack of significance for some of the retested variables.

Multiple regression was applied to all the variables found to be significantly correlated in the bivariate analysis to build predictive models. To establish the best predictive models, the "Model selection and multi-modal inference" tool was selected in SAM. SAM selected OLS for this initial multiple regression analysis which produced thousands of models by combining all the variables in every possible combination for each AIS (65,535 possible models for CLP and 524,287 for EWI) (Appendix B). The OLS regression calculated corrected Akaike Information Criterion (AIC_c) scores for each model. Aikake information criterion scores generally represent a "goodness of fit" statistic for each model and are accepted as a suitable measure for comparing models (Windle et al., 2010). The AIC scores also indicate the predictive ability of each model. In this initial OLS regression, all of the models were ranked according to their AICc scores with the lowest AICc being ranked at the top. The lower the AIC value, the closer the model is to reality (Fotheringham et al., 2002; Windle et al., 2010; Shaker et al., 2013).

For each of the AIS in this study, the first five models selected with AIC minimization in the OLS regression were tested further. The top models underwent multiple logistic regression analysis in SAM to evaluate the relationship between the variables established in each model and validate their findings. This second multiple logistic regression (MLR) produced a number of statistical values of which St. Coeff and p-values were primarily used to investigate the nature of the relationships between the variables and each AIS. The St. Coeff values were important because they indicate the strength and direction of the relationships. This second analysis (MLR) was also

done as a way to validate the findings and test for collinearity. Variance inflation factors were produced as a basic method to determine the presence of multicollinearity and give the errors associated with multicollinearity between the covariates (Shaker et al., 2017). Variance inflation factors will indicate if there is a correlation between the variable estimates, but they do not include the intercept term in the calculations and do not show where the collinearity occurs only that it is present (Fotheringham and Oshan, 2016). Because of these limitations, it is argued that it is not as useful as some researchers have proposed. Nevertheless, it does give an indication of multicollinearity, which if present, would require further investigation. Generally, a VIF > 10 indicates a problem with multicollinearity (Shandas, 2009; Fotheringham and Oshan, 2016) and a VIF between 2.5 and 10 would indicate possible issues but VIF values < 2.5 indicated that there are no issues with multicollinearity within the models (Shaker and Ehlinger, 2014).

Standard (traditional) global regression methods assume randomness and independence (previously discussed) which means all the data could be treated the same across space. However, environmental issues that deal with regional-scale data tend not to have independence because they are impacted by local conditions (Wagner and Fortin, 2005; Shaker et al., 2017). Spatial analysis is used to assess those processes in an effort to understand what might be responsible for generating the typically non-random patterns seen (Wagner and Fortin, 2005). As a result, applying global regression to these variables that might not meet the assumptions, can produce type I errors – a "false positive" (when a true hypothesis is incorrectly rejected) (Lennon, 2000; Fortin and Dale, 2005). Geographically weighted regression, which is basically a modern global regression that includes geographic coordinates to give a local estimate, incorporates spatial data through applying a weighted distance-decay to the observations and in this way is able to address spatial non-stationarity (Windle et al., 2010). Any observations around the sample point are given more weight, the assumption being that those points at a smaller distance have a greater impact on the local parameter estimates for that location (Tu and Xia, 2008; Shaker and Ehlinger, 2014).

The final objective of this thesis, to determine the effectiveness of using GWLR for this type of study, was accomplished by using the GWR4 software to test the first five models for each AIS. This was done to determine if GWLR improved the results, thus indicating that there was a local impact affecting the presence of the AIS. In order to include geographic influences, location points had to be incorporated into the regression, which in this study was done by including the longitudinal and latitudinal coordinates (Rangel et al., 2010).

25

The equation used for GWLR was:

$$y_i \sim \text{Bernoulli}\left[p_i\right]$$
$$\text{logit}(p_i) = \sum_k \beta_k(u_i, v_i) x_{k,i}$$

(1)

For this, the dependent variable must be 0 or 1 and p_i is the modelled probability that the dependent variable becomes one (Nakaya et al., 2016).

Within the GWR4 software, there were a number of choices available that allowed the researcher to control the analysis. The main difference with using GWR for analysis was the geographic aspects of the data and distance between points. This was referred to as distance-decay and was introduced by using a specific radius for analysing the local model and was set by applying a particular bandwidth (Nakaya et al., 2014). Analysis within the GWR4 software allowed a "moving window regression for each observation point" (Windle et al., 2010). In this study, the bandwidth selection was set for the "Golden section search" which allowed the program to search for the best bandwidth size for that particular data set (Nakaya et al., 2014). The kernel type which controls the distance-decay function was set to "adaptive" which used the number of samples rather than a specific distance (Appendix D) (Windle et al., 2010). The criteria used to determine bandwidth could be set to use different values, and for this study, AICc was used. The model type could also be selected within GWR4, to determine which tests were run. In this study, the logistic options used were the "standardisation of independent variables" and "geographical variability test". The results of the GWLR summarised these settings and then established the values associated with these parameters (Appendix D).

When using GWR4, local multiple logistic regression was conducted with the selected kernel size and thus incorporated local variation (Nakaya et al., 2016). At the same time within the GWR4 program, traditional multiple logistic regression was also computed, using the same data set. This resulted in AIC scores for both model types which was easily compared between the two types of models. In this way, using the same criteria for computation, the AIC scores were produced and compared to easily determine whether using local statistics (GWLR) did indeed improve on the global results. Since AIC values showed the "goodness of fit", this evaluation was achieved by comparing the AIC scores calculated by both multiple regression analyses (global and local). In this study, the GWLR (local) AIC scores were compared to the MLR (global) AIC scores that had been validated. When comparing AIC scores between two models, a difference of more than 3.0 between the AIC scores being compared indicated a significant improvement for the lower

value over the higher value (Lu et al., 2014). ∆AIC scores were calculated between the global and local analysis for each model using:

$$\Delta AIC = \text{Classic AIC}_G \text{ (Global/MLR)} - \text{Classic AIC}_L \text{ (Local/GWLR)} \tag{2}$$

∆AIC indicated the level of differences between the two statistical methods numerically. The ∆AIC was calculated for each of the five models to allow direct comparisons and determine if there was an improvement of one type over the other. Because low AIC scores indicate a better model, positive values for ∆AIC indicated the local model would be better than the global.

Chapter 5: Results

Using regression analysis and a large dataset of 126 lakes and more than 50 lake and landscape explanatory characteristics, the results revealed variables that showed statistically significant correlations to the two chosen AIS, curly-leaf pondweed and Eurasian watermilfoil in the Adirondack Park. The initial test for significance among the variables used bivariate regression and those variables found to be significant were then analysed using multiple regression in several ways to predict the presence of specific AIS. Bivariate analysis of each variable against each dependent variable resulted in a number of significant variables which were different for each AIS which were then used for further analysis with multiple regression.

5.1 Bivariate analysis

5.1.1 Curly-leaf pondweed

A total of sixteen variables that showed significant bivariate correlations for curly-leaf pondweed (Table 3). Of the lake and landscape variables, there were eight found to be significant for CLP, (game fish, state launch sites, lake elevation, maximum depth, minimum depth, distance to Interstate 87 (I-87), lake area and distance to census collection point). There were four land-cover composition metrics found to be significant for CLP (developed at high intensity, developed open space, percent deciduous forest, percent shrub/scrub). There were four landscape diversity/land cover class configuration variables that were significant for CLP, (SHEI, RPR, AI_DO and PRD). Of all the predictor variables found significantly related to CLP with bivariate analaysis, the strongest positive relationship was with the presence of game fish ($\rho^2 = 0.1805$, p-values < 0.0001, and St. Coeff = 4.198).

5.1.2 Eurasian watermilfoil

There were nineteen variables showing significant bivariate correlations to Eurasian watermilfoil (Table 3). Nine of these were lake and landscape variables (game fish, boat access type, boat launch sites, state launch sites, lake elevation, min depth, distance to I-87, nearest invaded lake and distance to census collection point). There were six land-cover composition variables significantly correlated to EWI (developed at low intensity, developed open space, percent deciduous forest, percent mixed forest, percent shrub/scrub, percent open water). There were four landscape diversity/land cover class configuration variables that were significant for EWI (SHEI, RPR, AI_DO and ENNAM_DO). Of these nineteen predictor variables, the strongest

positive correlation to EWI was percent developed open space (ρ^2 = 0.11, p-value = 0.0002, and St. Coef = 1.77).
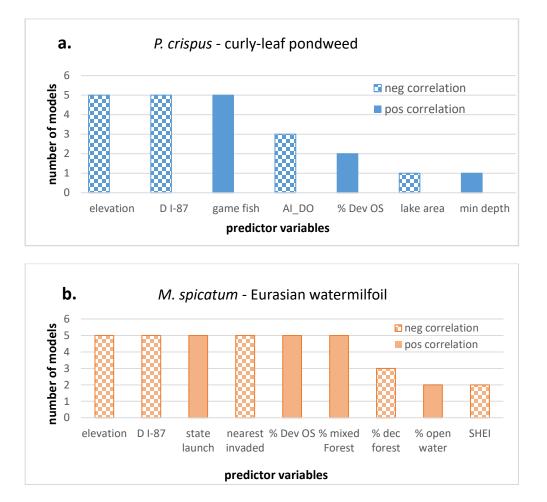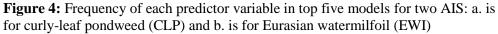
Table 3: Significant variables established with the bivariate analysis in SAM for each invasive species, curly-leaf pondweed, (*Potamogeton crispus)* and Eurasian watermilfoil (*Myriophyllum spicatum*).

Variables		Curly-leaf pondweed				Eurasian watermilfoil			
Variable	Transformed	McFadden's Rho²	Chi-square	St. Coeff.	p-value	McFadden's Rho²	Chi-square	St. Coeff.	p-value
Game Fish presence	---	0.1805	15.866	4.198	**<0.0001**	0.0577	9.329	1.261	0.0023
Access Type	---	----	----	----	----	0.0311	5.025	0.924	0.0250
Boat Launch	---	----	----	----	----	0.0199	3.218	0.764	0.0728
State Launch	---	0.0756	6.645	2.52	0.0099	0.0303	4.896	0.884	0.0269
Dev High Intensity	ArcSin-SqRt	0.0648	5.692	1.619	0.0170	----	----	----	----
Dev Low Intensity	ArcSin-SqRt	----	----	----	----	0.0888	14.371	1.558	**0.0002**
Dev Open Space	ArcSin-SqRt	0.1314	11.547	2.948	**0.0007**	0.1064	17.213	1.767	**<0.0002**
Deciduous Forest	ArcSin-SqRt	0.1422	12.501	- 3.416	**0.0004**	0.1586	25.656	- 2.329	**<0.0003**
Mixed Forest	ArcSin-SqRt	----	----	----	----	0.0370	5.988	0.983	<0.0144
Shrub-Scrub	ArcSin-SqRt	0.0632	5.554	2.8	0.0184	0.0565	9.134	- 1.422	0.0025
Open Water	---	----	----	----	----	0.0381	6.158	0.996	0.0131
Elevation Ft	---	*0.3266*	*28.709*	*- 4.326*	**<0.0003**	0.1285	20.788	- 2.046	**<0.001**
Max Depth Ft	Log(x+1)	0.0318	2.796	1.465	0.0945	----	----	----	----
Min Depth Ft	Log(x+1)	0.0908	7.986	3.185	**0.0047**	0.0268	4.340	0.858	0.0372
Distance to i-87 exit	---	*0.404*	*35.515*	*- 7.656*	**<0.0001**	*0.1945*	*31.461*	*- 2.794*	**<0.0001**
Area of lake (Ha)	Log(x+1)	0.0763	6.708	2.077	0.0096	0.0282	4.558	0.85	0.0328
Nearest invaded lake	Log	----	----	----	----	0.0541	8.729	- 1.19	0.0031
Distance to Census	Log(x+1)	0.1019	8.954	- 2.44	0.0028	----	----	----	----
SHEI	---	0.0445	3.912	- 1.792	0.0479	0.0197	3.187	- 0.719	0.0073
PRD	Log(x+1)	0.0343	3.016	- 1.614	0.0824	----	----	----	----
RPR	Log(x+1)	0.1391	12.229	3.843	**0.0005**	0.0483	7.816	1.184	0.0052
AI_DO	---	0.0351	3.084	1.97	0.0791	0.0783	12.661	1.724	**0.0004**
ENNAM_DO	Log(x+1)	----	----	----	----	0.0288	4.657	0.922	0.0309

Values with <0.001 p-values are in **bold**. Data with McFadden's rho-square values ≥0.2 are in ***bold and italics***. SHEI = Shannon's Evenness index; RPR = Relative Patch Richness; AI = Aggregation Index; ENNAM_DO = Euclidean Nearest Neighbour distance.

5.2 Multiple regression

Bivariate analysis showed the basic and direct correlation between two variables, but when other independent variables are added and tested simultaneously, the combination of variables showed different relationships. Of the thousands of models built with the initial model building analysis (ordinary least squares multiple regression) only the top five for each AIS, (ranked by AIC_c minimization), were investigated further. Once the models were built with OLS regression, the top five models were examined again with MLR for a closer examination and classic AIC values were recorded from that analysis. The top five models for

Figure 4: Frequency of each predictor variable in top five models for two AIS: a. is for curly-leaf pondweed (CLP) and b. is for Eurasian watermilfoil (EWI)

each species produced a group of variables that showed strongly correlations. There were significant differences between the two AIS predictive models if one looks at the frequency of the predictor variables in the models as well as the degree of significance for those variables (Figure

4). The best predictive models for CLP contained fewer variables compared to EWI, seven and nine respectively, and also had only three variables common across the models (Figure 4a). The best models for CLP included seven variables (game fish, lake elevation, distance to I-87, min depth, lake area, percent developed open space and AI_DO) and the best EWI models together included nine variables (state launch, lake elevation, distance to I-87, distance to nearest invaded lake, percent open space, percent deciduous forest, percent mixed forest, percent open water and SHEI). When comparing the two species with the first five models, CLP had fewer predictor variables than did EWI. There were two important overlaps among the variables found in all of the top models for both species: lake elevation and distance to I-87 (Table 4). All the models for both AIS included these two variables with good Standard Coefficients which were all significant for distance to I-87 at 5% or less and all but two (models 2 & 4 for CLP) were significant for lake elevation at 5%.

5.2.1 Curly-leaf pondweed

After analysing the top models separately again with MLR, the classic AIC values changed for each model and ranged from 46.793 to 50.743 (Table 4). According to this global regression, model three (AIC = 46.793, ρ^2 = 0.5815) had the lowest AIC, followed by model one (AIC = 47.093, ρ^2 = 0.5553). Some predictor variables (predictors) were constant across all five "best fit" models for CLP: lake elevation, distance to I-87 and game fish. Two of these (game fish and distance to I-87) were significant at 5% or less (Table 4). These had the highest correlation to the presence of CLP with three models (2, 4 & 5) that included AI-DO, two models (2 & 5) included % developed open space and one model (3) included lake area (Table 4). According to AIC scores from the MLR, the model with the lowest AIC (AIC = 47.093) was model three and it included the initial repeating predictors plus lake area. The second "best fit" model (AIC = 46.793) was model one and had the fewest predictors (lake elevation, distance to I-87 and game fish). The model ranked third by the AIC scores was model four (AIC =49.072) and had the three initial predictors with the aggregation index, AI_DO. The next ranked model, model number two (AIC = 49.130), included the initial predictors plus % developed open space and AI_DO added. The fifth ranked model (AIC = 50.743) was model five and had the most predictors with the initial three plus minimum depth, % developed open space and AI_DO.

Table 4: Values produced with multiple logistic regression between lake and landscape metrics and the two aquatic invasive species (curly-leaf pondweed (CLP) and Eurasian watermilfoil (EWI)) for five models each - with Standard Coefficients, AICs, VIF and McFadden's ρ^2 values

Statistical measures and Predictor Variables	Curly-leaf pondweed (CLP)					Eurasian watermilfoil (EWI)				
	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*	*Model 1*	*Model 2*	*Model 3*	*Model 4*	*Model 5*
AIC	47.093	49.130	46.793	49.072	50.743	99.173	99.805	101.152	98.743	98.303
McFadden's ρ^2	0.5553	0.5776	0.5815	0.5550	0.5820	0.4982	0.4943	0.4736	0.4885	0.4912
VIF max values	1.158	2.041	1.612	1.261	2.107	1.644	1.652	1.637	1.446	1.439
Standardized Constant	0.000	0.000	0.000*	0.000	0.000	0.000***	0.000***	0.000***	0.000***	0.000***
Lake and Landscape										
Lake Elevation	-2.790**	-1.930	-3.777**	-2.828**	-1.842	-1.446**	-1.434 **	-1.475 **	-1.504 **	-1.488 **
Distance to I-87	-5.458***	-6.030**	-5.520**	-5.430***	-6.043**	-2.715**	-2.691**	-2.414**	-3.080**	-3.166***
Game Fish	3.485***	3.678**	5.261***	3.55**	3.479**	–	–	–	–	–
State Launch	–	–	–	–	–	1.070 *	1.096***	1.279**	1.147***	1.082*
Nearest invaded lake	–	–	–	–	–	-2.717***	-2.566***	-2.638***	-2.711***	-2.881***
Min Depth	–	–	–	–	1.101	–	–	–	–	–
Lake Area(Ha)	–	–	-2.764	–	–	–	–	–	–	–
Land-cover composition										
% Dev. Open Space	–	5.675	–	–	3.237	1.820***	1.818***	1.657***	1.909***	1.921***
% Mixed Forest	–	–	–	–	–	1.197	0.995	0.685	1.093	1.269
% Deciduous Forest	–	–	–	–	–	-0.865	-0.836	-1.148	–	–
% Open Water	–	–	–	–	–	–	1.247*	–	1.440**	–
Landscape diversity										
SHEI	–	–	–	–	–	-1.479*	–	–	–	-1.646**
Land-cover class configuration										
ai_do	–	-2.207	–	-0.321	-2.972	–	–	–	–	–

Note: The values recorded are Standard coefficients from logistic regression for each tested variable. Variable categories: Lake and Landscape measurements, Land-cover composition in percentages, Landscape diversity measures (SHEI = Shannon's evenness index) and Land-cover class configuration (aggregation index). Symbols used: * significant at 10 %, ** significant at 5%, *** significant at 1%.

Multiple linear regression used to gather more information about the predictors produced the VIF. All the VIF scores were below 2.107 (range of 1.261 to 2.107) for CLP which indicated that multicollinearity was not an issue for these models. Standardised coefficients gave directional relationships that indicated game fish, minimum depth and % developed open space were positively correlated with the presence of CLP, meaning that as their levels increased so did the likelihood of CLP being present. On the other hand, lake elevation, distance to I-87, lake area, and AI_DO were negatively correlated with the presence of CLP, showing an inverse relationship – meaning that as the variable levels (for instance elevation or distance) increased the likelihood of CLP being present decreased. These relationships, whether positively or negatively correlated, were consistent across all five models. The distance to I-87 and game fish predictors were found significant in all five models. Distance to I-87 was at 5% significance for three models (2, 3 & 5) and was significant at 1% in two models (1 & 4). Game fish was significant at 5% for three models (2, 3 & 5) and at 1% for two models (1 & 3). Lake elevation was significant at 5% in three models (1, 3 & 4) but was <u>not</u> significant in two models (2 & 5). The other predictors that occurred in the five models were not significant for CLP. The strongest positive correlation to the presence of CLP that was also significant was with game fish (St. Coef = 5.261, p-value = 0.01) in model three. The strongest negative correlation that was also significant was with distance to I-87 (St. Coef = -6.043, p-value = 0.05) in model five.

5.2.2 Eurasian watermilfoil

Further examination of the best five models produced new classic AIC scores which ranged from 98.303 to 101.152. According to the global MLR, model five was the best fitting model with an AIC score of 98.303 ($\rho^2 = 0.4912$) and model four as the second best fit with an AIC score of 98.743 ($\rho^2 = 0.4885$). Evaluation of the top five models demonstrated the relationship between the predictor variables (predictors) in the models. The predictors with the strongest correlation to EWI were lake elevation, distance to I-87, state launch, distance to nearest invaded lake, % developed open space, % mixed forest, % deciduous forest, % open water and SHEI (Figure 4). Predictors that were constant or repeating for EWI across all five models were lake elevation, distance to I-87, state launch, distance to nearest invaded lake, % developed open space and % mixed forest, of which all but % mixed forest were significant with p-values at 10% or less. Along with these initial predictors, three models added % deciduous forest (1, 2 & 3), two models included % open water (2 & 4) and two models included SHEI (1 & 5) (Table 4). The best model with the lowest AIC

(AIC = 98.303) was model five and contained the initial six predictors (lake elevation, distance to I-87, state launch, distance to nearest invaded lake, % developed open space and % mixed forest) and SHEI. The second ranked model (AIC = 98.743) was model four and included the initial six predictors plus % open water. The third ranked model (AIC = 99.173) included the initial six variables, plus % deciduous forest and SHEI. The fourth ranked model (AIC = 99.805) was model two and included the initial six predictors plus % deciduous forest and % open water. The last model here (AIC = 101.152), was model three and included the six initial variables plus % deciduous forest.

SAM was used to calculate the VIFs in order to validate the models further. All VIF scores for the EWI models were below 1.652 with a range between 1.439 and 1.652. This indicated that multicollinearity was not an issue for these variables. Standardised coefficients, showing directional relationships, indicated that state launch, % developed open space (as well as % mixed forest, but this was not significant) were positively correlated to the presence of EWI, thus as those predictors increased (in number or % land used) the likelihood of EWI being present also increased. The other predictors, (lake elevation, distance to I-87, distance to nearest invaded lake, and SHEI) were all negatively correlated with the presence of EWI, indicating that as those values increased the likelihood of EWI being present decreased (an inverse relationship). Percent deciduous forest was also negatively correlated, yet it was not a significant relationship according to MLR. These relationships, whether they were positively or negatively correlated, were again consistent across all five models.

The repeating predictors in the top models, (except % mixed forest), were all significant for EWI. Lake elevation was significant at 10% across all models. Distance to I-87 in models one to four was significant at 5%, but in the fifth model was significant at 1%. The state launch predictor had different significance levels, with models one and five being significant at 10%, model three at 5% and models two and four at 1%. Nearest invaded lake was significant at 1% for all predictors as was % developed open space, making these the most significant of all the predictors. Percent deciduous forest (only found in models one, two and three), was not significant in any of them. The % open water in model two was significant at 10% and in model four at 5%. SHEI in model one was significant at 10% and in model five at 5%. Overall, there were more predictors in the top models for EWI, and they showed greater significance than the CLP predictors. The strongest positive correlation to the presence of EWI that was also

34

significant was with % developed open space (St. Coef = 1.921, p-value = 0.01) in model five. The strongest negative correlation that was also significant, was for distance to I-87 (St. Coef = -3.166, p-value = 0.01) in model five as well. Both of these were in the model with the lowest AIC score, model five.

5.3 Using geographically weighted logistic regression

Using geographically weighted logistic regression in predictive models gives a number of values, (ρ^2, % deviance, AIC) which are selected by the user. These values can be used in a variety of ways, but for this study, classic AIC scores were primarily used as a comparison measure.

5.3.1 Curly-leaf pondweed

The presence of CLP estimated and observed across the 126 lakes using GWLR, with AIC as the selection criteria, gave classic AIC scores that ranged from 46.416 to 48.957 (Table 5). The lowest AIC score using GWLR, was produced for model one at 46.416, establishing it as the best fitting model using GWLR with model three as the second best fit at 46.825. This was different from the MLR analysis, where model three was the "best fit". When a comparison of the traditional

Table 5: Comparing the local analysis of Geographic Weighted Logistic Regression with the global analysis of multiple logistic regression, as shown by comparing AIC scores for both aquatic invasive species.

	curly-leaf pondweed						Eurasian watermilfoil				
	Model 1	Model 2	Model 3	Model 4	Model 5		Model 1	Model 2	Model 3	Model 4	Model 5
	Multiple Logistic Regression Models (global)										
Classic AIC_G	47.093	49.130	46.793	49.072	50.743		99.173	99.805	101.152	98.743	98.303
McFadden's ρ^2	0.5550	0.5776	0.5815	0.5555	0.5820		0.4982	0.4943	0.4736	0.4885	0.4912
	Geographically Weighted Logistic Regression Models (local)										
Classic AIC_L	46.416	48.957	46.825	48.941	48.957		94.103	92.962	94.576	91.015	92.208
% deviance	0.6112	0.5965	0.5942	0.5713	0.5965		0.5727	0.5794	0.5554	0.578	0.569
Bandwidth %	47	100	100	100	100		54	54	53	52	53
Δ AIC	0.677	0.173	-0.032	0.131	1.786		5.070	6.843	6.576	7.728	6.095

Notes: all models are significant at p<0.001. Δ AIC = Classic AIC_G (Global) – Classic AIC_L (Local). The Δ AIC for model 3 for CLP is the only model that did not show an improvement with GWLR.

logistic regression models (a global measure) to the GWLR models (a local measure) was made, all of the models showed a slight improvement with the local (GWLR) over global (multiple regression) except model three which showed a negative value of -0.032. This negative ΔAIC indicates that GWLR was not a closer approximation of reality for the grouped variables within that model. Model three also had the lowest AIC score with the global MLR and thus was deemed the 'best fit'. However, these values were in fact very close: an AIC score of 46.793 for the global model and 46.825 for the local model.

The ΔAIC for the other four models ranged from 0.131 to 1.786. According to the AIC scores, the local models had lower AIC scores (making them better models), but none of them showed a clear improvement since none of the ΔAIC scores were ≥ 3 (Figure 5). However, a comparison of McFadden's rho squared (for logistic multiple regression) and % deviance (for GWLR) indicated better values using GWLR. It could be argued that these two values were not comparable yet in this study the global regression completed with the GWR4 program produced

Figure 5: Bar graph showing ΔAIC, where a positive value demonstrated an improvement for geographically weighted logistic regression (local analysis) compared to the traditional logistic regression (global analysis) for curly-leaf pondweed (CLP) and Eurasian watermilfoil (EWI).

Note: The horizontal line is at Δ AIC = 3 (the acceptable level that indicates a significant improvement from one model to another).

% deviance values that were almost identical to the ρ^2 values given in the MLR from SAM (given in Appendix D). For example, the ρ^2 value for model one with MLR was 0.5550, and the % deviance with global models in GWR4, was 0.5552. Nevertheless, AIC scores were more broadly accepted in ecological studies and are therefore the primary comparison tools used here. Another indication of the suitability of using GWLR was evident in the bandwidth selected within the program. For CLP, the bandwidth selected by the program was at 100% for four models with only one model less than 100% at 47%. When bandwidth is very large, it approaches a global analysis. Thus a bandwidth of 100% is equivalent to a global analysis and showed that GWLR was not really appropriate or effective.

5.3.2 Eurasian watermilfoil

The presence of EWI was investigated across the 126 monitored lakes to determine the most important predictor variables, and this was then estimated with GWLR and evaluated using AIC (Table 5). Classic AIC (local) scores for the top models ranged from 91.015 to 94.576 with the lowest score produced by model four at AIC = 91.015. This model was thus ranked as the "best fit" with GWLR, and model five was the next "best fit" at AIC = 92.208. A comparison of the traditional MLR models (a global measure) to the GWLR models (a local measure), was done by their AIC scores. All five models showed significant improvement in the fit of the models as indicated by the ΔAIC scores (ranging from 5.070 to 7.728). These scores were well above the accepted ΔAIC ≥ 3 (Figure 5). This indicated the local EWI models were a definite improvement over the global analysis. The fourth model had the lowest AIC_L score and the greatest improvement (Δ AIC = 7.728), and the next best improvement was the second model (Δ AIC = 6.843)

When comparing the EWI models using McFadden's ρ^2 (for global regression) and % deviance (for local GWLR), all models showed improvements as well. The traditional global McFadden's ρ^2 ranged from 0.4735 to 0.4982 and % deviance from GWLR ranged from 0.5554 to 0.5794, which were all higher (Table 5). The GWLR analysis was also more precise for EWI because the bandwidth used was smaller overall. The bandwidth for EWI ranged from 52% to 54% across all of the models, with none of them utilizing 100% bandwidth as occurred with CLP.

Chapter 6: Discussion

The guiding question for this thesis was to determine if it would be possible to use publicly available data to produce meaningful and useful research that could be used to help control or slow the spread of IS. Even though this study used only publicly available data, the findings were strong and highlighted differences between the two species of interest. Publicly available data, especially in North America, were numerous and covered a variety of types, which allowed extensive research to be conducted. This study was able to look at factors that related to propagule pressure, human activity, land cover, lake characteristics, and topography all using publicly available data. After robust analyses, strong statistical findings were produced for both of the AIS of interest. This involved applying layers of statistical analyses, both bivariate and multiple regression, and building predictive models that indicated which lake and landscape variables correlated to the presence of either of the aquatic invaders. Those predictor variables that showed strong correlations, as evidenced by their presence in the first five models, to either CLP or EWI could then be used to identify yet uninvaded lakes that might be vulnerable to invasions. Identifying variables that were closely related to IS allowed a greater understanding of what environmental characteristics or human activities aid IS and thus would need to be managed.

6.1 The importance of predictor variables

Patterns of invasive species spread can be understood by finding variables that relate to each IS and use them as predictors for that species. Many researchers (Buchan and Padilla, 2000; Capers et al., 2009; Dullinger et al., 2009; Thum and Lennon, 2010) have worked to identify these variables but because organisms can adapt and change according to various conditions, they do not react predictably. Therefore, any broad assumptions about a relationship between environmental characteristics and IS may not apply across many ecosystems. Instead, researchers have looked at the role that the environmental characteristics played in the success of IS (e.g. available nutrients, abiotic conditions) and analysed those in order to develop strategies to manage IS spread. Invasive species researchers have identified certain factors as strongly correlated to IS success and spread, many of which were selected for this thesis. Examples of these would be propagule pressure, large nutrient loading, monotypic habitat with little resilience, high human activity or degraded habitat.

One aspect of the findings for CLP that raised some questions in this study, was the predictors that lacked significance in the multiple regression models, even though they were significant in the bivariate models. Whether or not to retain them for further analysis was debated, but as previously discussed, using a heuristic approach for this study meant that the predictor variables might not show the expected statistical significance even if they were relevant. Only variables found significant in the bivariate analysis were used for the model building, and all those predictor variables found in the first five predictive models were retained. The first five models selected in the initial OLS regression were established as the "best fit" and this is why all the variables selected by those models were retained through all further analyses, regardless of their statistical significance. Besides the heuristic approach, there was another possible explanation for why the covariates were not found to be significant for CLP and that was sample size. The total sample size for the study was 126 lakes with 66 of them being invaded, which would be an acceptable sample size for analysis. However, by focusing on only two IS and the lakes that they had invaded as reference points for determining predictor variables, the working sample size became smaller. More specifically, the lakes invaded with CLP totalled only 14 and those invaded with EWI totalled 43 (Figure 6). Therefore, the analysis for CLP had 14/126 lakes (only 11%) providing data and the analysis for EWI had 43/126 lakes (34%) contributing data. When smaller sample sizes are used with regression, the accuracy of the results is lower which might explain the lack of significance in the multivariate models (Bingham and Fry, 2010). It was also important to note that all the lakes invaded by CLP also contained EWI (Figure 6c). This was to be expected according to research because CLP has been found to be less aggressive than many other AIS and is rarely found as a solitary IS (Engelhardt, 2002). As a result CLP was commonly found in already invaded lakes (Engelhardt, 2002).

6.1.1 Predictors for propagule pressure

Any research on IS spread will involve propagule pressure, and Lockwood et al. (2005) felt that it was the key to truly understanding IS success. Invasive species in particular use human-mediated dispersal as a significant part of their propagation. Thus any management strategies must consider (and even target) human activity that leads to greater propagule pressure (Keller and Perrings, 2011; Hulme and Firn, 2015). Recreational boating has been tagged as the greatest propagator of AIS and as a result, predictor variables (such as launch sites, favourite fish species,

lake size, access to lakes and human activity in and around lakes) that relate to or encourage this activity were heavily represented in this study (Johnson et al., 2001; Rothlisberger et al., 2010; Shaker et al., 2017). The top predictors found in this study (game fish, distance to I-87 and elevation) were related to propagule pressure and often correlated to IS in general because of what
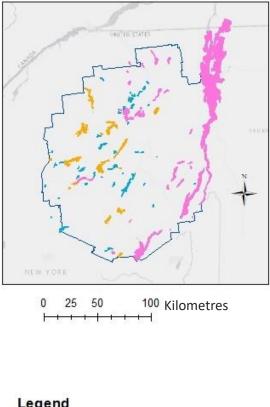
a. Lakes with curly-leaf pondweed (CLP)

b. Lakes with Eurasian watermilfoil (EWI)

c. Lakes with both CLP and EWI

Figure 6: Maps of Adirondack Park showing invaded lakes for each aquatic invasive species in the study. (a) all the lakes with curly-leaf pondweed (14); (b) all the lakes with Eurasian watermilfoil (43); (c) the lakes that have both CLP and EWI together (14). (produced using ArcGIS 10.3.1 from ESRI, 2004)

they meant to the spread of IS. This was supported by other studies and, according to Thum and Lennon (2010), propagule pressure was one of the most important influences on invasion success (Lockwood et al., 2005; Sheppard et al., 2006). Propagule pressure includes both the number being introduced (how many organisms) and the number of times these introductions occur (Gertzen et al., 2011). Variables connected to accessing a lake (game fish, access, boat launch, state launch as well as distance to I-87) were found to be significant for both species in the bivariate analyses and related to the use of the lakes. Fishing or recreational boating was considered a measure of propagation potential for AIS, and the significance of such variables in this study supported propagule pressure as being important in determining the presence of CLP and EWI within the Adirondack lakes (Devine, 1998; Bruckerhoff et al., 2015).

Distance to Interstate 87 was the primary predictor across all five multiple regression models (for both AIS) and elevation (which was considered a natural influence that impacts dispersal through water flow) was also strongly related. These were both part of human access to lakes because lakes at higher elevations would presumably be more difficult to access due to rough terrain and steep topography. The distance to I-87 exits, associated with recreational use of the lakes, had a more direct association because as (Euclidean) distance increased between a lake and available exits from the highway, there were fewer of either AIS in the lakes. The distance to I-87 was statistically significant for both AIS, indicating that the closer a lake was to an I-87 exit, the greater the probability of either AIS being present. Buchan and Padilla (2000) found that EWI was most likely to spread to lakes that had a high occurrence of boat traffic. This variable (distance to I-87) had a significant ($p < 5\%$) negative correlation to both AIS that was consistent through all tests, showing that as the distance between I-87 exits and a lake increased, the likelihood of either AIS being found would decrease. In theory, the closer a lake is to the main thoroughfare the greater the boat traffic to that lake. It was evident on the map (Figure 7) that there were two strips of invaded lakes, one along the I-87 (blue strip) and one going from the middle of the north section, across the Park in a southwestern direction (pink strip). The blue strip followed the direction of I-87 and the pink strip was more associated with lower elevations.

Elevation was expected to be strongly correlated to both AIS, especially in a mountainous region like the Park. The significant negative (or inverse) correlation to elevation as a predictor for both species suggested that as elevation increased, the likelihood of either species being present would decrease. Headwaters are usually in higher elevations, and because water flows to lower

elevations, it was to be expected that AIS would spread from high to low elevations with the water, increasing in numbers as water flowed through lakes and onward. It would be possible for IS to spread from low to high elevation (against the water current) but this would require outside help such as through human activity and movement (Johnson et al., 2001).

This movement related to those predictors (such as fishing and recreational boating) that drive human activities (Buchan and Padilla, 2000; Dullinger et al., 2009; Bruckerhoff et al., 2015). The second strip (pink) on the map (Figure 7) follows the lower land to the west of the mountain range, (high elevation points were shaded in brown between the two strips of invaded lakes), rather than the highway. This area might be more affected by the elevation variable. A more specific study would be needed to distinguish between these two predictors and their mapped locations. Regardless, both of these strips of invaded lakes appeared to have a significant influence on the presence of AIS. Another feature of elevation was that because water flows to lower elevations, everything in that water (chemicals, pollutants, nutrients, sediments) also moves to lower elevations (Francis and Chadwick, 2012). Thus lakes at lower elevations would then be more polluted, have more nutrients and sedimentation and other components that might affect water quality and change the natural environmental conditions to favour AIS. Invasive species researchers (Devine, 1998; Trebitz & Taylor, 2007; Shusky et al., 2009; Pysek et al., 2012) found that it was common for greater numbers of, and a greater variety of, IS to be associated with areas that are nutrient rich (agricultural runoff). Water chemistry was not directly monitored in this study, but agricultural activity and % developed areas were intended as representative measures as suggested by several studies (Roth et al., 1996; Peterson and Vieglais, 2001; Shandas and Alberti, 2009). However, these predictors did not show strong correlations to either AIS. In order to clarify this, since these other studies show strong correlations, further studies are needed (Elton, 1958; Capers et al., 2009; Shusky et al., 2009).

Just as distance to the I-87 was important for human-mediated dispersal, the distance from an invaded lake to a new, uninvaded lake was important as a source of spread and influenced propagule pressure. Shorter travel distances between these lakes would result in greater possible propagule pressure. Thus the distance to the nearest invaded lake predictor would be expected to show a strong negative correlation to the presence of AIS. While it was strongly (and negatively) correlated with the presence of EWI, and was highly significant (at 1%), it was not selected by any of the CLP models. This was puzzling since many researchers agree that habitats nearby would be
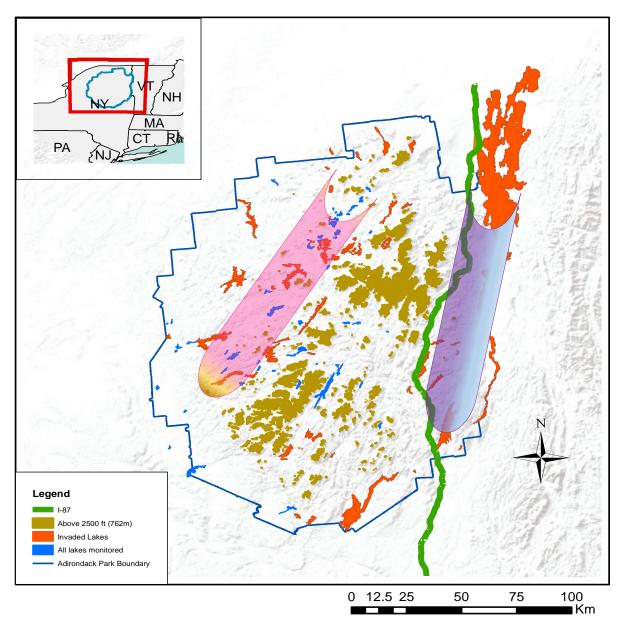
Figure 7: Map showing distribution of invaded lakes and elevation in Adirondack Park from World Maps and ArcGIS https://www.worldmapsonline.com

more likely to be invaded due to their proximity (Kearney and Porter, 2009; Francis and Chadwick, 2012; Espeland, 2013; Havel et al., 2015) and, the first law of geography states that locations closer to each other are more similar than those farther apart and thus should support the same species easily (Espeland, 2013). An explanation for the lack of significance for CLP might be due to the fact that turions are the main propagule and these often use human-mediated dispersal, requiring humans to move between the lakes. However, there may not be easy or direct access between the

lakes being considered, either due to poor roads between them or no direct water connection. It has been suggested that for many IS, sexual reproduction is "energetically expensive" compared to asexual propagation, and as a result, sexual reproduction (which produces seeds) is rarely the main strategy for their spread (Engelhardt and Ritchie, 2002; Martin and Valentine, 2014; Bruckerhoff et al., 2015). CLP does not tend to use sexual reproduction as much as it does fragments or the vegetative buds (turions) which move with water flow (elevation) or human-mediated dispersal (Mills et al., 1993; Havel et al., 2014). This would need more study to distinguish between lakes that have direct and easy access to other lakes and those that were a bit harder to actually reach with trailered boats. Other aspects of the lake such as the populations of fish species present or the abiotic factors, would impact how humans use the lake. This may have more bearing on dispersal success than does the Euclidean distance between lakes that was used for the nearest invaded lake predictor. Nevertheless, travel time between lakes does impact the success of AIS spread. Since nearby lakes would be more similar (First law of Geography), they should support the same AIS. This appeared true for EWI because the nearest invaded lake predictor was highly significant (99%) and was important to its propagation since EWI can use seeds which move without needing roads or prescribed paths. Therefore a variable using Euclidean distance to a nearby lake, rather than actual travel distance, may be more appropriate for IS that use this type of dispersal (Trebitz et al., 2007; Shusky et al., 2009). The significance of the nearest invaded lake metric to EWI in this study supported other research with respect to EWI's ability to spread by natural means and by human activities, although, EWI primarily spreads by simple fragments that get transported in some way (Francis and Chadwick, 2012; Martin and Valentine, 2014; Bruckerhoff et al., 2015). Boating activity has been found as the main vector for transporting and dispersing EWI, especially in a recreational area like the Park (Johnson and Padilla, 1996; Shaker et al., 2013). The movement of boats and aquatic recreational gear between lakes transports fragments and the less time the fragment is out of the water, the better it can survive (Johnson et al., 2001; Bruckerhoff et al., 2015). It would be the nearby lakes that could be easiest to reach while fragments were still viable (Bruckerhoff et al., 2015). Fragments from EWI, whether leaves or stem sections, have survived out of water for 18 hours (longer than CLP is known to survive), but if they were embedded within a mass of plant material, EWI fragments survived up to 48 hours (Bruckerhoff et al., 2015). This allowed fragments more travel time between lakes.

Curly-leaf pondweed was found to depend primarily on turions (specialized vegetative buds that act as propagules) for dispersal rather than seeds or fragments and as a result had a slow dispersal rate (Mills et al., 1993; Caffrey and Monahan, 2006).These turions were not as easily spread by attaching themselves to boats or propellers, but could be found in coils of other aquatic plants stuck on boats (Xie et al., 2014; Bruckerhoff et al., 2015). Bruckerhoff et al. (2015) found that asexual reproduction of CLP by fragments was not very successful because their fragments dry out within 12 hours. Conversely, the turions could survive 28 days before becoming inviable (Bruckerhoff et al., 2015). Once turions were released into the lakes or rivers, they moved with the current (high to low elevation) and stayed dormant until a change in conditions (temperature or light levels) could cause them to germinate (Xia et al., 2014). Once they gained access to a new area and developed to the point of producing turions, CLP can be very hard to control and manage, allowing them to become well-established in that lake. Lake area was found to be significant to both EWI and CLP in the bivariate analysis, yet was only found in one multivariate model for CLP, which was not statistically significant (Table 3). It showed a negative correlation, supporting the fact that CLP was more prominent in small lakes and did seem to like shallower water (Mills et al., 1993). It was surprising that lake area, or any other variables related to the type of lake, was not more prominent in determining the presence of EWI since it can grow in large, deep lakes (although it also grew well in medium depth water) (Nichols and Shaw, 1986). This lack of significance in this study could be simply because lake area was more significant when looking at AISR that included a number of AIS (Trebitz et al., 2007). A common finding in the ecological literature is that species-area relationships exist meaning that a larger area will have more species (Shaker et al., 2017). Thus, larger lakes support more species as well as more IS and many studies that found lake area to be correlated to AIS were studying AISR (Engelhardt and Ritchie, 2002; Capers et al., 2009), rather than only one or two species (Shaker et al., 2017). EWI, is common in both shallow and deep water and is found in large or small lakes, although it becomes dominant, and thus a greater problem, in shallower water and smaller lakes (Trebitz et al., 2007). Ultimately, EWI can live in deep water and grow 10 metres to reach light if needed, and it can grow submerged, or it can float. Because of this adaptability, neither predictor (lake area or lake depth) was found to be significant to the presence of EWI alone.

Nevertheless, these are not the only variables that represented propagation. Other variables that could act as proxies for propagule pressure were game fish and state launch which were both

found to be significantly correlated in the bivariate analyses (Dullinger, 2009). Both of these were shown to be significantly correlated with CLP in the bivariate analysis, but just game fish was significant in all five multiple regression models. EWI, however, did not show a multivariate relationship to game fish (Table 3). The game fish variable measured the presence of three fish (yellow perch, small mouth bass, rainbow trout) that were favoured by the anglers who bring in boats and fishing gear, thereby aiding in dispersal (Dullinger, 2007; Bruckerhoff et al., 2015). The presence of these fish would be the incentive for anglers to enter a particular lake (Buchan and Padilla, 2000; Johnson et al., 2001; Shaker et al., 2017). The state launch predictor represented public launch sites that were accessible to tourists and residents and acted as an access point for dispersal. Although game fish did not appear in the top models for EWI, state launch was in all five models. This was understood to show a connection to recreational fishing because anglers might enter the lakes via launch sites. So even if the game fish predictor did not show strongly for EWI, this does not mean anglers were not represented by the predictors that were – such as state launch sites. State launch sites would also include water activities other than fishing, (water skiing, speed boating and other water sports) and thus indicated a connection to broader human activities.

6.1.2 Predictors related to land use and land cover type

The presence of mixed forest around the lakes did correlate to EWI indicating that forest appeared to impact in some way the environment that EWI preferred. It was a positive correlation showing that higher percent mixed forest supported higher EWI. Shandas and Alberti (2009) found that the presence of riparian vegetation strongly affected instream biological conditions and Roth et al. (1996) suggested that it supported aquatic systems by increasing organic debris. Buchan and Padilla (2000), whose study was similar to this thesis, found that percent forest cover was more important than human access or development for EWI because it related to nutrients in the water. Inorganic carbon, a known limiting factor for EWI, can be supplied by the presence of riparian vegetation, and thus the presence of forest becomes an important negative predictor (Keast, 1984; Buchan and Padilla, 2000; Smith et al., 2002; Thum and Lennon, 2010). The percent deciduous forest was negatively correlated to only two (EWI) models in this study yet the correlation to percent mixed forest was positive which might be related to different carbon levels from each. However, another explanation for this correlation was found in research by Shaker and Rapp (2013) that investigated the lake and landscape predictor of lake associations. The findings

suggested a connection between deciduous forest and land ownership. They were looking at lake associations and their connection to the predictors, but the thought processes might also be applied to this study (Shaker and Rapp, 2013). They proposed that the presence of large areas of deciduous forest around a lake was an indication of state ownership of large tracts of forest and thus that particular lake would have less privately owned properties along its shore. With respect to the negative correlation between percent deciduous forest and the presence of EWI, it could indeed be related to the fact that the land around the lake was state owned forest, resulting in fewer access points for boats. This might reduce boat entry to that lake which in turn, would present fewer opportunities for EWI to be introduced. This connection was speculative and requires more specific studies to determine if the state owned land around lakes does indeed limit the number and type of access points/launch sites for boats which in turn would mean there were fewer entry points for human-mediated dispersal of AIS. Future studies could also investigate a more direct relationship to dissolved inorganic carbon and if this would be different between the two types of forest (mixed and deciduous).

It has long been accepted that IS are connected, and helped, by human activities (Elton, 1958). Therefore, it would be expected that there was some level of correlation between those factors and IS in Adirondack Park as well (Elton, 1958; Mills et al., 1993; Devine, 1998; Bruckerhoff et al., 2015). Not surprising, these AIS showed a level of correlation to the development of nearby land and the type of land cover yet it was smaller than suggested by other studies (Elton, 1958; Buchan and Padilla, 2000; Francis and Chadwick, 2012). Despite a stronger bivariate relationship to levels of human development and human land cover use, the type of human activity in the multiple regression analysis that related to CLP centred around those activities (fishing) that aided dispersal and not those that might affect aquatic conditions (agriculture or industry). This demonstrated that characteristics commonly accepted as significantly correlated to IS in general, may not be significant for a single species. This would need to be studied with more specific details on the location of the sample along with evidence of the nutrient source and a direct measure of water chemistry. Eurasian watermilfoil was found by numerous researchers to be connected to human activity and disturbance and this study supported that to some degree (Mills et al., 1993; Buchan and Padilla, 2000; Trebitz, 2007; Shusky et al., 2009). As a result, it was expected that correlations would occur between many variables, such as percent cultivated areas or areas with a high level of development. However, these relationships

were not strongly represented in the best fit models. The lack of significance of these variables could simply be because their occurrence across a study area that had such a variety of land use types, was low. The raw data showed that these variables were not well represented across all the invaded lakes. As a result, their representation in the Park, and particularly in proximity to the sampled lakes, was low. In an area with high levels of industry and development, a similar analysis would likely show a strong correlation between those variables and these AIS.

6.2 Modelling to management

The synergy of ecology should be incorporated into ecological analysis if true understanding is to be gained and if it is to be applied to the real world. In this thesis, multiple regression was conducted which gave a more realistic picture of the relationships between variables and the aquatic invasive species. The initial OLS analysis compared covariates simultaneously and only reported the overall statistical fit but did not show the individual significance of the variables. Therefore, multiple logistic regression was performed on each model separately to evaluate the variables within the model. The predictive multivariate models indicated the level of the relationship between the predictors and the AIS (the dependent variable). The predictors found in each model were then considered as a unit that acted together and gave the most effective predictions. This gave more useful information than just the bivariate models. By evaluating the models, none of them had issues with collinearity as indicated by VIFs below 2.5, the accepted level, and as such the results were accepted with confidence knowing they were not violating assumptions.

Adirondack Park has over 10,000 bodies of water which spread over a large area (almost 25,000 km^2), with various levels of legislation governing it. There are over 30 organizations working together to manage the Park with upwards of 700 volunteers (APIPP, 2015). Even with all these groups working together to manage IS it is difficult to cover the whole area with complete efficiency. If nothing else, IS are hard to manage because they act differently in different environments and are often very adaptable to new conditions (Pimental et al., 2005; Havel et al., 2015). Because of this, as well as limited resources (both funding and manpower), predictive models have been gaining attention as a management tool for environmental planners. Using multiple regression to build these models has proven to be a practical strategy that incorporated many predictor variables and took into account how they might interact with each other to influence the presence of AIS. These multivariate models gave specific information that was more

suited to decision-making than were straight bivariate models. Within the Park, environmental planners focus on prevention and early detection, which then leads to rapid response actions (APIPP, 2015). Limited funds that must be spread across many departments and numerous lakes mean that actions must be prioritized. Models such as those built in this thesis could be used to determine which lakes were the most likely to be invaded and thus require extra attention. Each model should be considered as a group of predictors with strong correlations to the AIS tested. The first model produced for CLP using the multi-model framework had the fewest number of variables (game fish, elevation, distance to I-87), and those same variables were found together in the other top models. These variables were also the only ones that had statistically significant coefficients within the CLP models. Such consistency was not present for the EWI models. The top "best fit" model for EWI (model five) had four predictors (distance to I-87 exits, lake elevation, state launch and nearest invaded lake) that acted as proxies for propagule pressure and three other predictors (% developed open space, % mixed forest and SHEI) (Table 4). Those predictors most related to propagule pressure indicated that lakes close to an I-87 exit had a shorter distance for transporting boats and the state launch predictor showed which lakes had public piers or launch sites suitable for large boats. Both of these, like the common variables for CLP, encouraged human activity and thus would aid in the dispersal of EWI. This human activity created a connection between those lakes with high boat activity and the number of entry points for EWI, thus ensuring a successful introduction. Both distance to nearest invaded lake and distance to I-87 were negatively correlated which could be interpreted as: lakes that were easy to get to from I-87 and that were also close to already invaded lakes were more likely to be vulnerable to EWI, even better if they had several state launches. Resource managers could then use these predictors to identify vulnerable lakes.

The top "best fit" model for CLP (model three), demonstrated that the presence of anglers and their boats was the main source of AIS dispersal (Buchan and Padilla, 1999; Johnson et al., 2001; Bruckerhoff et al., 2015). Knowing these predictors (lake elevation, distance to I-87, game fish and lake area) and their significant relationship to the presence of CLP would allow resource managers to locate lakes matching these characteristics and to actively protect the lake by targeting any points of entry and initiating boat inspections and cleaning stations. Smaller lakes at low elevations, (with any of the three popular fish species and easy access from the I-87), could harbour a certain vulnerability for CLP. Lakes having all of the repeating predictor variables to some

degree, should be prioritized for constant monitoring in order to catch invasions as early as possible. The other predictors (AI_DO, % dev open space and lake area) that were strongly correlated with CLP should also be considered when looking for matching lakes (whether they are positively or negatively correlated is also important to consider). However, just these predictors (ai_do, percent developed open space and lake area) without the other repeating predictors would not be useful for predicting the possibility of CLP spreading to that lake. The value of using multivariate predictive models was that this gave a combination of predictors that were most significant when found together. For instance, for EWI, the presence of state launches on a lake put those lakes on the map for recreational boaters and anglers. Public launch sites were easier to monitor, so the fact that state launches were found to be highly correlated to the presence of EWI would be good news from a monitoring point of view. (However, the presence of this predictor in the best models might also be due to the fact that there were more data available for this variable, because it was public, than for the private sites. Overall, considering the top predictors for EWI (distance to I-87, nearest invaded lake and state launch as well as lake elevation) are categorized as propagule pressure, it would be a starting point to help resource managers identify vulnerable lakes. It would require searching for all of these predictors, as a group, present for one lake and label it as vulnerable to that AIS. Resource managers could then set up monitoring for those identified lakes, to proactively prevent the spread by creating a protocol to protect them. The Park, specifically APIPP, already has several programs (i.e., boat wash stations, boat launch stewards, awareness programs, public signs and warnings) set up to manage AIS. The use of predictive models would help determine particularly sensitive areas that would benefit from these programs (APIPP, 2014; Shaker et al., 2017). For managing EWI, according to the repeating predictors in the top models, a low elevation lake with an abundance of mixed forest and developed open space in its riparian zone that is also close to an Interstate 87 exit with invaded lakes nearby, might be considered vulnerable to future EWI invasion. The other predictors that might increase the likelihood of EWI invasions were the presence of open water or lakes with low SHEI values. Small amounts of deciduous forest around the lakes might also support EWI. Again, just like with CLP, the presence of these variables individually would not be as useful for predicting the presence of EWI as would using the models as a whole by grouping the predictors together. This illustrated how such an analysis can be used for managing IS and guiding decisions for future management plans within the Park.

Conducting a study like this that focussed on only two known AIS, showed differences between the two species that supported the need for running individual analyses. As it pertained to the use of these models for managing AIS, the ability to analyse the species separately may only be important if a particularly damaging species were found in an area and its spread required serious attention. Otherwise, the distinctions between the known invaders may be of little or no help to resource managers, and in fact, they may even prefer a broader scope that would cover the majority of AIS with which they were concerned rather than spending resources (e.g. money) to fight individual species. The significance here was that EWI was known as one of the most aggressive and damaging AIS and as such, finding a difference between it and another less damaging AIS (such as CLP) would allow more stringent actions against EWI (Buchan and Padilla, 2000; Martin and Valentine, 2014). Also, by analysing these species separately, it was evident that predictor variables could certainly be narrowed down if a particular species required more attention due to the damage it caused.

The final objective for this study was to use GWLR to determine if this statistical method was useful for dealing with environmental issues. Because spatial analysis was needed for ecological studies and standard statistical analysis had assumptions that were not always valid for spatially explicit data, traditional analysis might be less than ideal for environmental issues. Using GWLR for analysis of spatially varying data incorporated the regional variations into the multiple regression via the distance-decay and bandwidth size and thus allowed a more specific application. Using the lower AIC score as the better model with a difference of three between the two model types being considered a significant difference, there was enough information to declare the higher ΔAIC score as a definite improvement. The comparison (for CLP) between traditional MLR and GWLR did not show much of an improvement which again could be because of the small sample size. GWLR analysis has been considered as more accurate as sample size increased (Fotheringham and Oshan, 2016). In fact, Fotheringham and Oshan (2016), felt that sample sizes for GWR in general should not be lower than 100 for accuracy. If samples were small, analysts felt the bandwidth would approach global and then using GWR would be meaningless. This was seen here with CLP. The bandwidth size approached 100% for four of the models (Table 5). The total number of lakes was over 100 (126 in total), but as the number of invaded lakes used to determine the relationships between the covariates was lower than the total of 126, the statistical accuracy might be affected. Conversely, EWI was present in more lakes and thus provided more

data on which to base the calculations. EWI showed significant improvements for local models with ΔAIC scores for all five models greater than 5.07 which indicated that all of them were better models by incorporating regional variations with model four having the greatest improvement (ΔAIC = 7.728). Thus, using GWLR analysis for predictive models with regional variation should show a more accurate association between the dependent and independent variables, providing suitable data and sample sizes were available. In summary, there should be more confidence in analysing predictive models using GWLR providing there was an appropriate sample size.

Chapter 7: Conclusion

7.1 Limitations

This thesis focused on only two IS which resulted in smaller sample sizes for analysis than if all the AIS within the Park were used. This must be considered when using the findings for decision-making. Looking at one species that was present in only a small percentage of monitored lakes, may not provide enough data for analysis, specifically when looking at correlations between the characteristics of the lake and the presence of that IS. The number of lakes invaded by CLP was only 14 out of the 126 monitored lakes, thus providing limited data points. This may explain why the results for CLP showed less significance than those for EWI. However, even though the statistical findings for EWI, which had more invaded lakes than CLP, showed greater significant values, there were still significant predictors produced for CLP.

It is important to note that this thesis was specific to Adirondack Park in New York State. For management purposes this can be a limitation since the specific findings should only be applied to the lakes of that Park. For instance, the significance of the lake elevation and the distance to I-87 to the presence of both IS can be applied to the Adirondack Park area. The same predictors would probably not play such a significant role in another natural resource area that had little variation in elevation and thus could not be assumed to relate to the presence of CLP or EWI in all areas. Never the less, the technique used and the process of using the predictors from that technique, could be applied to other areas if suitable data were available. With this technique, it was also important to use predictors from several of the top models (the top five) not just the single top model, since this would show which predictors were consistently present plus give other predictors for consideration.

In this thesis, several conclusions were explained using the available data (i.e. the strong impact of lake elevation or the lack of significance for human development within this Park). Other scientific studies (Trebitz, 2007; Capers et al., 2009; Francis and Chadwick, 2012) have found the presence of human development (agriculture, industry) as a key element in the success of IS in general and thus those characteristics were expected to provide greater correlations in this study. These variables were proxies for other lake and landscape characteristics (such as % agriculture represented nutrient loading). One important aspect of this thesis and its applicability to environmental management was to use data that were publicly available and thus easily accessed. Because only certain types of data were available, many metrics were proxies for other

areas of interest. This may have introduced a lower level of clarity in the statistical analysis but using proxies is an accepted practice as long as the relationship has previously been proven (Klein, 1979; Wagner and Fortin, 2005; Dullinger et al., 2009; Shusky et al., 2009). In order to ascertain the connections, the variables used in this thesis were researched to ensure there were distinct connections that demonstrated these proxies were acceptable, and thus their use was justified (Roth et al., 1996; Dullinger et al., 2009). Other measures that might not be specifically found in publicly available data, such as information on native species or direct measures of water chemistry, might add layers of understanding to modelling for managing IS. However, the findings of this study were strong and definite correlations were produced while using the publicly available lake and landscape variables.

7.2 Future Direction

This thesis was limited in the number of IS analysed and was specific to one Park, but it demonstrated the feasibility of using publicly available data in a methodology that could realistically be used for developing management strategies. This thesis also demonstrated that predictors could be found for specific IS that would allow those species to be targeted independently if they were particularly damaging. By taking all the predictors found in the best models and searching for their combined presence in a lake (or near a lake), vulnerable lakes could be given higher priorities. While this research supplied a group of predictors for both CLP and EWI, the next step would be to locate those lakes in the Park using the predictors. This would allow a practical application in fighting IS. Also, applying this methodology to the other AIS in the Park, the top predictors for all of them could be compared to find if the same overlapping predictors from this study (lake elevation and distance to I-87) showed strong correlations to the other AIS. If they did, then it would suggest that those predictors should definitely be given top priority when setting up management strategies. While the findings of this study were valuable there is still much work to be done, within the Adirondack Park to find predictors for the other AIS, as well as in other managed natural resources with data specific to those areas. This same model building could be expanded if other more specific data were available such as which native species were present, measures of water chemistry (especially related to pollution) or even evidence of climate change (water temperature) in efforts to fine tune the predictive ability. Even without adding other variables, the significance of the predictors established while using the publicly available data, was strong for those related to propagule pressure and human movement

within the Park. This gave valuable information that could be applied to management strategies right away to make protecting the vulnerable lakes more cost effective for the Park authorities.

In conclusion, conducting multiple regression on variables found to be significant with bivariate analysis was used to build predictive models for managing invasive species spread. This thesis met the objectives set forth with several significant findings that could be practically applied to establish pro-active policies for protecting lakes against IS to some degree. Using only those data that were publicly available from government organizations or environmental agencies provided clear predictors and established statistical relationships between the independent variables and the two different aquatic invaders studied. Using ordinary least squares regression to build the models produced thousands of models of which the five "best fit" models for each AIS of interest, CLP and EWI were analysed further. By building multivariate predictive models, a group of predictor variables were identified that when found together became strong predictors for the presence of the AIS studied. These models gave clear indications of what characteristics might impact the ability of an AIS to access a lake as well as the lake conditions/characteristics that were strongly connected to either CLP or EWI. The predictors with the strongest correlation to CLP or EWI were those related in some way to propagule pressure. These were the distance from the I-87, lake access (elevation for water flow, launch sites for boat access) and the presence of nearby lakes that were previously invaded. Because these related to propagule pressure and thus dispersal, the vectors (boats, fishing equipment, and human use of lakes) that were part of these predictors could be targeted by resource managers in efforts to limit the spread of these two AIS. There were still other predictors, such as developed open space for both AIS, percent mixed forest for EWI and the aggregation index for CLP that could also be used to identify potentially vulnerable lakes. The value of multivariate models was that lakes with all of the selected predictors could be specifically detected rather than a slew of lakes with a few individual variables identified by bivariate analysis. As the sample size increased, using GWLR produced a more accurate or better fit model compared to the global analysis. When determining the significance of the predictors, GWLR incorporated local variations for a more meaningful estimate. This also allowed for better spatial analysis overall since GWLR considered the heterogeneity present in environmental research that could be missed with traditional statistical approaches.

The data were used in this thesis to build predictive models with significant findings (especially for Eurasian watermilfoil), but it must be stressed that this type of analysis is best when

used within a specific area. Therefore, the predictors found significant in this study should not be used to identify vulnerable lakes in another conservation area, for example Algonquin Park in Ontario, Canada. The same technique could be applied to local public data collected within Algonquin, and predictive models could then be built for that particular park. For instance, an area without human disturbance or recreational activity (no boats being transported between lakes) might have completely different predictors than those established by this modelling approach. This thesis did not find large correlations between the two AIS being studied (CLP and EWI) and environmental conditions or human disturbance as other studies found. While this is important to acknowledge here and examine, it also emphasized that IS research using predictive models must be done on a regional level. There were several findings in this thesis that might be improved by using more specific data (water chemistry, agricultural management practices, etc.) in order to determine if other factors might be acting on the lakes that were not represented by the variables employed in this analysis. One key point from the guiding research question was to use data that were publicly available in order to demonstrate that simple but useful analysis could be undertaken using available data to identify predictors that would help prioritize lakes. This was accomplished as demonstrated by the findings. Once the lakes needing urgent attention were identified they could be protected against the aquatic invaders. The last objective was to investigate GWLR to determine its applicability to environmental studies and in particular to invasive species management. The evaluation was done by comparing the local GWLR values (AIC scores) which incorporated local variations, to the traditional global models that used multiple logistic regression. The findings from this study showed that if the sample size was suitable, GWLR models for EWI were definitely an improvement over the global models as established by a difference in the AIC scores that were greater than three.

The initial research question was answered by successfully establishing predictors that could be used by resource managers and it also developed a workable technique to determine the likelihood of either aquatic invader spreading to new lakes, but it should not stop there. While these findings were valuable, there is still much work to be done and more research required. With more modelling conducted, a broader list of predictors for AIS in general within the Park would be possible which would help with decision-making by natural resource managers working to slow the spread of AIS across the Park. Ideally, more studies like this could be used in other areas to manage invasive species despite limited funding.

Appendix A: Initial list of variables from Dr Shaker's data set. (These were formatted and tested at the beginning of the study).

Lake and Landscape variables	Composition Metrics	Landscape Diversity Metrics
Boat Launch	**% Deciduous Forest**	**PR**
State Launch	**% Mixed Forest**	**PRD**
Campground	**% Evergreen Forest**	**RPR**
Game Fish	**% Shrub Scrub**	SHDI
Access type	**% Herbaceous**	SIDI
Lake Elevation	**% Hay/Pasture**	SHEI
Distance to I-87	**% Cultivated Crop**	SIEI
Perimeter (Km)	**% Emergent Herb Wetland**	MSIEI
Perimeter (Mi)	**% Woody Wetland**	AI_DO
Max Depth	% Open Water	AI_EF
Min Depth	% Barren Land	PLADJ_DO
Lake Area (Ha)	**% Developed High Intensity**	PLADJ_EF
Area (Km)	**% Developed Med Intensity**	PLADJ_EF
Area (Acres)	**% Developed Low Intensity**	**ENNAM_DO**
Nearest invaded lake	**% Developed Open Space**	ENNAM_EF
Distance to Census point		AreaM-DO
Distance to populated area		AreaM-EF
Population 2000		
PA-ratio		
Shape Length		
Shape Area		

Note: The transformed variables are in bold.

Note: AI, aggregation index; PLADJ, percentage of like adjacency; AREA_AM, area-weighted mean patch area; ENN_AM, area-weighted mean Euclidean nearest neighbour distance; RPR, relative patch richness; SHDI, Shannon's diversity index; SHEI, Shannon's evenness index; SIDI, Simpson's diversity index; SIEI, Simpson's Evenness index; Game Fish indicated the number of game fish together.

Appendix B: Building models in SAM using the "model selection and multi-modal inference" tool

B.1: The models built for curly-leaf pondweed (CLP) using all retained variables

```
Total of 65535 OLS models run, in 2.94 minutes

Response variable is clp.
Explanatory variables are:
    Variable #1: GameFish
    Variable #2: StateLnc
    Variable #3: DevHInMd
    Variable #4: DevOSpMd
    Variable #5: DecForMd
    Variable #6: ShrubMd
    Variable #7: ElevatFt
    Variable #8: MaxDpMd
    Variable #9: MinDpMd
    Variable #10: d_i87
    Variable #11: AreaHaMd
    Variable #12: D_CensMd
    Variable #13: shei
    Variable #14: prdMd
    Variable #15: rprMd
    Variable #16: ai_do
```

Results for the OLS Model Selection procedure, sorted by Akaike Information Criterion (AICc).

Model	Variables	nVars	r²	Cond.Num.	AICc	Delta AICc	L(gi|x)	AICc wi
Mod #32131	1, 7, 10	3	0.44	1.491	3.35	0	1	0.008
Mod #28099	1, 4, 7, 10, 16	5	0.458	2.735	3.631	0.28	0.869	0.007
Mod #32132	1, 7, 10, 11	4	0.444	2.232	4.558	1.208	0.547	0.005
Mod #32194	1, 7, 10, 16	4	0.444	1.724	4.62	1.27	0.53	0.004
Mod #27972	1, 4, 7, 9, 10, 16	6	0.463	2.865	4.665	1.315	0.518	0.004
Mod #32188	1, 7, 10, 14	4	0.444	1.94	4.691	1.341	0.512	0.004
Mod #28036	1, 4, 7, 10	4	0.443	1.726	4.956	1.606	0.448	0.004
Mod #32004	1, 7, 9, 10	4	0.442	1.773	5	1.649	0.438	0.004
Mod #15748	1, 2, 7, 10	4	0.442	1.648	5.042	1.692	0.429	0.004
Mod #31876	1, 7, 8, 10	4	0.442	1.787	5.099	1.749	0.417	0.003
Mod #23940	1, 3, 7, 10	4	0.442	1.636	5.155	1.805	0.406	0.003
Mod #32180	1, 7, 10, 13	4	0.441	1.88	5.307	1.957	0.376	0.003
Mod #11716	1, 2, 4, 7, 10, 16	6	0.461	2.78	5.313	1.963	0.375	0.003
Mod #30084	1, 5, 7, 10	4	0.44	1.988	5.464	2.114	0.348	0.003
Mod #31108	1, 6, 7, 10	4	0.44	1.974	5.497	2.147	0.342	0.003
Mod #32005	1, 7, 9, 10, 11	5	0.45	2.458	5.537	2.186	0.335	0.003
Mod #32192	1, 7, 10, 15	4	0.44	2.025	5.546	2.196	0.334	0.003
Mod #32164	1, 7, 10, 12	4	0.44	1.8	5.552	2.202	0.332	0.003
Mod #28092	1, 4, 7, 10, 13, 16	6	0.459	3.479	5.637	2.287	0.319	0.003
Mod #31749	1, 7, 8, 9, 10	5	0.449	2.399	5.672	2.321	0.313	0.003
Mod #28098	1, 4, 7, 10, 15, 16	6	0.459	3.248	5.712	2.362	0.307	0.003
Mod #32061	1, 7, 9, 10, 14	5	0.449	2.149	5.719	2.369	0.306	0.003
Mod #19908	1, 3, 4, 7, 10, 16	6	0.459	2.891	5.757	2.407	0.3	0.003
Mod #27844	1, 4, 7, 8, 10, 16	6	0.458	2.901	5.83	2.48	0.289	0.002
Mod #26052	1, 4, 5, 7, 10, 16	6	0.458	2.937	5.859	2.509	0.285	0.002
Mod #28084	1, 4, 7, 10, 12, 16	6	0.458	2.871	5.876	2.526	0.283	0.002
Mod #27076	1, 4, 6, 7, 10, 16	6	0.458	2.826	5.903	2.552	0.279	0.002
Mod #28068	1, 4, 7, 10, 11, 16	6	0.458	3.732	5.912	2.561	0.278	0.002
Mod #28096	1, 4, 7, 10, 14, 16	6	0.458	3.473	5.912	2.562	0.278	0.002
Mod #27717	1, 4, 7, 8, 9, 10, 16	7	0.468	3.028	5.998	2.648	0.266	0.002
Mod #15749	1, 2, 7, 10, 11	5	0.447	2.368	6.077	2.727	0.256	0.002

58

B.2: The first few models built for Eurasian watermilfoil (EWI) using all retained variables

```
Total of 524287 OLS models run, in 30.09 minutes

Response variable is ewi.
Explanatory variables are:
   Variable #1: GameFish
   Variable #2: Access
   Variable #3: Boat_Lnc
   Variable #4: StateLnc
   Variable #5: DevLInMd
   Variable #6: DevOSpMd
   Variable #7: DecForMd
   Variable #8: MixForMd
   Variable #9: ShrubMd
   Variable #10: OpenWatr
   Variable #11: ElevatFt
   Variable #12: MinDpMd
   Variable #13: d_i87
   Variable #14: AreaHaMd
   Variable #15: NearDMd
   Variable #16: shei
   Variable #17: rprMd
   Variable #18: ai_do
   Variable #19: ennmDOMd


Results for the OLS Model Selection procedure, sorted by Akaike Information Criterion (AICc).
```

Model	Variables	nVars	r²	Cond.Num.	AICc	Delta AICc	L(gi|x)	AICc wi
Mod #476840	4, 6, 7, 8, 11, 13, 15, 16	8	0.505	2.418	102.902	0	1	0.004
Mod #476328	4, 6, 7, 8, 10, 11, 13, 15	8	0.501	2.407	103.884	0.982	0.612	0.002
Mod #476839	4, 6, 7, 8, 11, 13, 15	7	0.491	2.265	103.976	1.074	0.584	0.002
Mod #480423	4, 6, 8, 10, 11, 13, 15	7	0.49	2.199	104.12	1.218	0.544	0.002
Mod #480935	4, 6, 8, 11, 13, 15, 16	7	0.49	2.249	104.133	1.231	0.54	0.002
Mod #444072	3, 6, 7, 8, 11, 13, 15, 16	8	0.499	2.491	104.395	1.493	0.474	0.002
Mod #411305	3, 4, 6, 7, 8, 11, 13, 15, 16	9	0.508	2.613	104.417	1.515	0.469	0.002
Mod #476809	4, 6, 7, 8, 11, 13, 14, 15, 16	9	0.507	3.536	104.718	1.816	0.403	0.001
Mod #411304	3, 4, 6, 7, 8, 11, 13, 15	8	0.497	2.481	104.793	1.891	0.388	0.001
Mod #448167	3, 6, 8, 11, 13, 15, 16	7	0.487	2.27	104.934	2.032	0.362	0.001
Mod #476847	4, 6, 7, 8, 11, 13, 15, 16, 19	9	0.506	3.04	105.07	2.168	0.338	0.001

Appendix C: Analysis of first five models using Multiple Logistic Regression Analysis

C.1: Results of multiple logistic regression for curly-leaf pondweed (CLP)

Model 1 (3 Variables)

```
Regression Settings | Analytical Results | Graphical Results | Spatial Structure | Maps

 Logistic Regression Analysis
  Results for clp as a response variable, and 3 predictor variable(s).

     n: 126    m: 3     McFadden's Rho-Squared: 0.5553    Chi Square: 48.813    P: <.0001
    LL(0): -43.9528     LL(N): -19.5464    2*[LL(N)-LL(0)]: 48.8128    AIC: 47.0929


     Variable     Coefficient    Std Coeff.    Std Error      t       P Value
     Constant         1.767          0           1.758      1.005     0.315
     GameFish         1.224        3.485         0.518      2.36      0.018
     ElevatFt        -0.003       -2.79          0.001     -1.962     0.05
        d_i87        <.001        -5.458        <.001      -2.431     0.015


 Odds Ratios and 95% Confidence Intervals:
     Variable     Odd Ratio          Low            High
     Constant        5.856          0.187        183.584
     GameFish        3.4            1.231          9.393
     ElevatFt        0.997          0.995          1
        d_i87        1              1              1


 Classification Table and Measures of Accuracy:

 Classification Table and Measures of Accuracy:

               Predicted
            ---------------
 Observed    1    0    Total                    Overall Accuracy: 0.937
            ---------------                         Sensitivity: 0.5
    1     |  7 |  7  |  14                          Specificity: 0.991
    0     |  1 | 111 | 112                       Kappa Statistic: 0.604
            ---------------              True Skill Statistic (TSS): 0.491
 Total    |  8 | 118 | 126
            ---------------
```

Model 2 (5 Variables)

```
Regression Settings | Analytical Results | Graphical Results | Spatial Structure | Maps

 Logistic Regression Analysis
  Results for clp as a response variable, and 5 predictor variable(s).

     n: 126    m: 5     McFadden's Rho-Squared: 0.5776    Chi Square: 50.776    P: <.0001
    LL(0): -43.9528     LL(N): -18.5649    2*[LL(N)-LL(0)]: 50.7759    AIC: 49.1297


     Variable     Coefficient    Std Coeff.    Std Error      t       P Value
     Constant         0.888          0           2.535      0.35      0.726
     GameFish         1.291        3.678         0.583      2.214     0.027
     ElevatFt        -0.002       -1.93          0.002     -1.159     0.246
        d_i87        <.001        -6.03         <.001      -2.188     0.029
     DevOSpMd         7.593        2.743         5.675      1.338     0.181
       ai_do        -0.029       -2.207         0.037     -0.784     0.433


 Odds Ratios and 95% Confidence Intervals:
     Variable     Odd Ratio          Low            High
     Constant        2.429          0.017        349.57
     GameFish        3.638          1.16          11.411
     ElevatFt        0.998          0.995          1.001
        d_i87        1              1              1
     DevOSpMd     1984.99           0.029        >1×10^8
       ai_do        0.971          0.903          1.045
```

```
Classification Table and Measures of Accuracy:

            Predicted
          ---------------
Observed   1     0    Total               Overall Accuracy: 0.937
          ---------------                      Sensitivity: 0.5
   1     |  7  |  7  |  14                      Specificity: 0.991
   0     |  1  | 111 | 112                  Kappa Statistic: 0.604
          ---------------            True Skill Statistic (TSS): 0.491
Total    |  8  | 118 | 126
          ---------------
```

Model 3 (4 Variables)

```
Regression Settings | Analytical Results | Graphical Results | Spatial Structure | Maps

Logistic Regression Analysis
 Results for clp as a response variable, and 4 predictor variable(s).

   n: 126    m: 4    McFadden's Rho-Squared: 0.5815    Chi Square: 51.113    P: <.0001
   LL(0): -43.9528    LL(N): -18.3964    2*[LL(N)-LL(0)]: 51.1128    AIC: 46.7929


   Variable      Coefficient    Std Coeff.    Std Error      t        P Value
   Constant         5.214           0           3.19       1.635       0.102
   GameFish         1.847         5.261         0.723       2.557       0.011
   ElevatFt        -0.003        -3.777         0.002      -2.276       0.023
      d_i87        <.001         -5.52         <.001       -2.286       0.022
   AreaHaMd        -1.46         -2.765         0.986      -1.481       0.138


Odds Ratios and 95% Confidence Intervals:
   Variable      Odd Ratio          Low            High
   Constant      183.875           0.354        95417.138
   GameFish        6.344           1.539          26.149
   ElevatFt        0.997           0.994            1
      d_i87        1               1                1
   AreaHaMd        0.232           0.034           1.603
```

```
Classification Table and Measures of Accuracy:

            Predicted
          ---------------
Observed   1     0    Total               Overall Accuracy: 0.937
          ---------------                      Sensitivity: 0.5
   1     |  7  |  7  |  14                      Specificity: 0.991
   0     |  1  | 111 | 112                  Kappa Statistic: 0.604
          ---------------            True Skill Statistic (TSS): 0.491
Total    |  8  | 118 | 126
          ---------------
```

Model 4 (4 Variables)

```
Regression Settings | Analytical Results | Graphical Results | Spatial Structure | Maps

Logistic Regression Analysis
 Results for clp as a response variable, and 4 predictor variable(s).

   n: 126    m: 4    McFadden's Rho-Squared: 0.5555    Chi Square: 48.834    P: <.0001
   LL(0): -43.9528    LL(N): -19.5359    2*[LL(N)-LL(0)]: 48.834    AIC: 49.0717


   Variable      Coefficient    Std Coeff.    Std Error      t        P Value
   Constant         1.978           0           2.266      0.873       0.383
   GameFish         1.247         3.55          0.544       2.293       0.022
   ElevatFt        -0.003        -2.828         0.001      -1.954       0.051
      d_i87        <.001         -5.43         <.001       -2.424       0.015
      ai_do        -0.004        -0.321         0.029      -0.147       0.883


Odds Ratios and 95% Confidence Intervals:
   Variable      Odd Ratio          Low            High
   Constant        7.23            0.085         613.348
   GameFish        3.478           1.199          10.095
   ElevatFt        0.997           0.995            1
      d_i87        1               1                1
      ai_do        0.996           0.941           1.053
```

```
Classification Table and Measures of Accuracy:

          Predicted
        ----------------
Observed   1    0    Total                    Overall Accuracy: 0.929
        ----------------                           Sensitivity: 0.429
   1    |  6 |   8 |  14                           Specificity: 0.991
   0    |  1 | 111 | 112                        Kappa Statistic: 0.537
        ----------------               True Skill Statistic (TSS): 0.42
Total   |  7 | 119 | 126
        ----------------
```

Model 5 (6 Variables)

```
Regression Settings | Analytical Results | Graphical Results | Spatial Structure | Maps

Logistic Regression Analysis
 Results for clp as a response variable, and 6 predictor variable(s).

   n: 126     m: 6     McFadden's Rho-Squared: 0.582    Chi Square: 51.163    P: <.0001
   LL(0): -43.9528     LL(N): -18.3713    2*[LL(N)-LL(0)]: 51.163    AIC: 50.7427


   Variable      Coefficient     Std Coeff.    Std Error      t        P Value
   Constant          0.607           0           2.628      0.231      0.817
   GameFish          1.222         3.479         0.602      2.03       0.042
   ElevatFt         -0.002        -1.842         0.002     -1.02       0.308
      d_i87         <.001         -6.043        <.001      -2.094      0.036
   DevOSpMd          8.964         3.237         6.414      1.397      0.162
   MinDpMd           0.545         1.101         0.886      0.615      0.539
      ai_do         -0.039        -2.972         0.041     -0.951      0.342


Odds Ratios and 95% Confidence Intervals:
   Variable     Odd Ratio          Low             High
   Constant         1.835          0.011          316.581
   GameFish         3.393          1.043           11.034
   ElevatFt         0.998          0.995            1.002
      d_i87         1              1                1
   DevOSpMd      7812.706          0.027          >2×10^9
   MinDpMd          1.724          0.304            9.782
```

```
Classification Table and Measures of Accuracy:

          Predicted
        ----------------
Observed   1    0    Total                    Overall Accuracy: 0.937
        ----------------                           Sensitivity: 0.5
   1    |  7 |   7 |  14                           Specificity: 0.991
   0    |  1 | 111 | 112                        Kappa Statistic: 0.604
        ----------------               True Skill Statistic (TSS): 0.491
Total   |  8 | 118 | 126
        ----------------
```

C2: Results of multiple logistic regression for Eurasian watermilfoil (EWI)

Model 1 (8 Variables)

```
Regression Settings  Analytical Results  | Graphical Results | Spatial Structure | Maps |

Logistic Regression Analysis
 Results for ewi as a response variable, and 8 predictor variable(s).

    n: 126    m: 8     McFadden's Rho-Squared: 0.4982    Chi Square: 80.579    P: <.0001
    LL(0): -80.8761     LL(N): -40.5865    2*[LL(N)-LL(0)]: 80.5792    AIC: 99.1731
```

Variable	Coefficient	Std Coeff.	Std Error	t	P Value
Constant	21.311	0	5.657	3.767	<.001
StateLnc	1.017	1.07	0.628	1.62	0.105
DevOSpMd	7.604	1.82	3.035	2.505	0.012
DecForMd	-2.411	-0.865	2.266	-1.064	0.287
MixForMd	5.715	1.197	3.941	1.45	0.147
ElevatFt	-0.002	-1.446	0.001	-1.975	0.048
d_i87	<.001	-2.715	<.001	-2.076	0.038
NearDMd	-3.705	-2.717	1.07	-3.462	<.001
shei	-6.628	-1.479	3.521	-1.882	0.06

Odds Ratios and 95% Confidence Intervals:

Variable	Odd Ratio	Low	High
Constant	>1×10^9	27551.133	>1×10^14
StateLnc	2.765	0.808	9.467
DevOSpMd	2006.971	5.233	769760.222
DecForMd	0.09	0.001	7.613
MixForMd	303.476	0.134	686810.56
ElevatFt	0.998	0.996	1
d_i87	1	1	1
NearDMd	0.025	0.003	0.2
shei	0.001	<.001	1.314

Classification Table and Measures of Accuracy:

```
              Predicted
         ---------------
 Observed   1    0    Total                Overall Accuracy: 0.841
         ---------------                        Sensitivity: 0.581
     1  |  25 |  18 |  43                       Specificity: 0.976
     0  |   2 |  81 |  83                    Kappa Statistic: 0.612
         ---------------            True Skill Statistic (TSS): 0.557
  Total  |  27 |  99 | 126
         ---------------
```

Model 2 (8 Variables)

```
Regression Settings  Analytical Results  | Graphical Results | Spatial Structure | Maps |

Logistic Regression Analysis
 Results for ewi as a response variable, and 8 predictor variable(s).

    n: 126    m: 8     McFadden's Rho-Squared: 0.4943    Chi Square: 79.947    P: <.0001
    LL(0): -80.8761     LL(N): -40.9027    2*[LL(N)-LL(0)]: 79.947    AIC: 99.8053
```

Variable	Coefficient	Std Coeff.	Std Error	t	P Value
Constant	14.253	0	5.144	2.771	0.006
StateLnc	1.041	1.096	0.629	1.657	0.098
DevOSpMd	7.595	1.818	3.005	2.528	0.011
DecForMd	-2.33	-0.836	2.393	-0.974	0.33
MixForMd	4.754	0.995	3.636	1.308	0.191
ElevatFt	-0.002	-1.434	0.001	-1.933	0.053
d_i87	<.001	-2.691	<.001	-2.069	0.039
NearDMd	-3.5	-2.566	1.056	-3.314	<.001
OpenWatr	4.95	1.247	2.8	1.768	0.077

```
Odds Ratios and 95% Confidence Intervals:
   Variable       Odd Ratio            Low              High
   Constant        >1×10^6           64.78           >3×10^10
   StateLnc         2.833            0.826              9.711
   DevOSpMd      1989.168            5.506         718657.849
   DecForMd         0.097           <.001             10.582
   MixForMd       116.057            0.093         144439.064
   ElevatFt         0.998            0.996                  1
     d_i87              1                1                  1
   NearDMd           0.03            0.004              0.239
   OpenWatr       141.108            0.583          34126.695


Classification Table and Measures of Accuracy:

             Predicted
         ---------------
 Observed   1     0    Total                    Overall Accuracy: 0.841
         ---------------                            Sensitivity: 0.581
     1   | 25  | 18  |  43                          Specificity: 0.976
     0   |  2  | 81  |  83                       Kappa Statistic: 0.612
         ---------------                True Skill Statistic (TSS): 0.557
 Total   | 27  | 99  | 126
         ---------------
```

Model 3 (7 Variables)

```
Logistic Regression Analysis
 Results for ewi as a response variable, and 7 predictor variable(s).

   n: 126     m: 7     McFadden's Rho-Squared: 0.4736     Chi Square: 76.6        P: <.0001
   LL(0): -80.8761     LL(N): -42.5762     2*[LL(N)-LL(0)]: 76.5999     AIC: 101.1524


   Variable     Coefficient    Std Coeff.    Std Error      t        P Value
   Constant        17.328          0           4.74       3.655      <.001
   StateLnc         1.215        1.279         0.601       2.024      0.043
   DevOSpMd         6.923        1.657         2.902       2.386      0.017
   DecForMd        -3.199       -1.148         2.128      -1.503      0.133
   MixForMd         3.27         0.685         3.255       1.005      0.315
   ElevatFt        -0.002       -1.475        <.001       -2.111      0.035
     d_i87         <.001        -2.414        <.001       -2.053      0.04
   NearDMd         -3.597       -2.638         1.013      -3.549      <.001


Odds Ratios and 95% Confidence Intervals:
   Variable       Odd Ratio            Low              High |
   Constant        >3×10^7          3093.61          >3×10^11
   StateLnc         3.371            1.039             10.937
   DevOSpMd      1015.764            3.442          299760.42
   DecForMd         0.041           <.001              2.645
   MixForMd        26.306           0.045          15505.751
   ElevatFt         0.998            0.996                  1
     d_i87              1                1                  1
   NearDMd          0.027            0.004                0.2


Classification Table and Measures of Accuracy:

             Predicted
         ---------------
 Observed   1     0    Total                    Overall Accuracy: 0.833
         ---------------                            Sensitivity: 0.535
     1   | 23  | 20  |  43                          Specificity: 0.988
     0   |  1  | 82  |  83                       Kappa Statistic: 0.585
         ---------------                True Skill Statistic (TSS): 0.523
 Total   | 24  | 102 | 126
         ---------------
```

64

Model 4 (7 Variables)

```
Logistic Regression Analysis
 Results for ewi as a response variable, and 7 predictor variable(s).

   n: 126    m: 7    McFadden's Rho-Squared: 0.4885    Chi Square: 79.01    P: <.0001
   LL(0): -80.8761    LL(N): -41.3713    2*[LL(N)-LL(0)]: 79.0096    AIC: 98.7427
```

Variable	Coefficient	Std Coeff.	Std Error	t	P Value
Constant	13.883	0	5.171	2.685	0.007
StateLnc	1.09	1.147	0.633	1.722	0.085
DevOSpMd	7.975	1.909	2.941	2.712	0.007
MixForMd	5.22	1.093	3.762	1.388	0.165
ElevatFt	-0.002	-1.504	0.001	-2.014	0.044
d_i87	<.001	-3.08	<.001	-2.401	0.016
NearDMd	-3.696	-2.711	1.051	-3.515	<.001
OpenWatr	5.719	1.44	2.738	2.089	0.037

```
Odds Ratios and 95% Confidence Intervals:
```

Variable	Odd Ratio	Low	High
Constant	>1×10^6	42.431	>2×10^10
StateLnc	2.973	0.86	10.272
DevOSpMd	2907.678	9.125	926569.581
MixForMd	184.954	0.116	294460.598
ElevatFt	0.998	0.996	1
d_i87	1	1	1
NearDMd	0.025	0.003	0.195
OpenWatr	304.73	1.423	65244.796

```
Classification Table and Measures of Accuracy:

          Predicted
        ---------------
Observed   1    0    Total              Overall Accuracy: 0.825
        ---------------                      Sensitivity: 0.558
   1   | 24 | 19 |   43                       Specificity: 0.964
   0   |  3 | 80 |   83                   Kappa Statistic: 0.573
        ---------------          True Skill Statistic (TSS): 0.522
 Total | 27 | 99 | 126
        ---------------
```

Model 5 (7 Variables)

```
Logistic Regression Analysis
 Results for ewi as a response variable, and 7 predictor variable(s).

   n: 126    m: 7    McFadden's Rho-Squared: 0.4912    Chi Square: 79.449    P: <.0001
   LL(0): -80.8761    LL(N): -41.1517    2*[LL(N)-LL(0)]: 79.4488    AIC: 98.3035
```

Variable	Coefficient	Std Coeff.	Std Error	t	P Value
Constant	21.854	0	5.693	3.839	<.001
StateLnc	1.028	1.082	0.633	1.625	0.104
DevOSpMd	8.023	1.921	2.969	2.702	0.007
MixForMd	6.063	1.269	4.029	1.505	0.132
ElevatFt	-0.002	-1.488	0.001	-2.038	0.042
d_i87	<.001	-3.166	<.001	-2.491	0.013
NearDMd	-3.929	-2.881	1.064	-3.693	<.001
shei	-7.377	-1.646	3.47	-2.126	0.034

```
Odds Ratios and 95% Confidence Intervals:
   Variable      Odd Ratio           Low                High
   Constant       >3×10^9         44173.153           >2×10^14
   StateLnc          2.797            0.809                9.671
   DevOSpMd       3050.132            9.056             >1×10^6
   MixForMd        429.745            0.16              >1×10^6
   ElevatFt          0.998            0.996                1
      d_i87          1                1                    1
    NearDMd          0.02             0.002                0.158
       shei         <.001            <.001                 0.562


Classification Table and Measures of Accuracy:

            Predicted
         ---------------
 Observed   1     0      Total              Overall Accuracy: 0.857
         ---------------                         Sensitivity: 0.605
     1   | 26 | 17 |  43                         Specificity: 0.988
     0   |  1 | 82 |  83                      Kappa Statistic: 0.651
         ---------------            True Skill Statistic (TSS): 0.593
 Total   | 27 | 99 | 126
         ---------------
```

Appendix D: The statistical output using geographically weighted logistic regression in the GWR4 program - for the first five models of each aquatic invasive species

D.1: Statistical output from GWLR for curly-leaf pondweed (CLP)
```
**************************************************************************
*          Semiparametric Geographically Weighted Regression          *
*                    Release 1.0.90 (GWR 4.0.90)                      *
*                         12 May 2015                                 *
*              (Originally coded by T. Nakaya: 1 Nov 2009)            *
*                                      *
*          Tomoki Nakaya(1), Martin Charlton(2), Chris Brunsdon (2)   *
*          Paul Lewis (2), Jing Yao (3), A Stewart Fotheringham (4)   *
*                    (c) GWR4 development team                        *
* (1) Ritsumeikan University, (2) National University of Ireland, Maynooth, *
*      (3) University of Glasgow, (4) Arizona State University        *
**************************************************************************
```

Model 1: Curly-leaf pondweed (CLP)

Program began at 4/16/2017 7:23:31 PM
Model settings--------------------------------
Model type: Logistic
Geographic kernel: adaptive Gaussian
Method for optimal bandwidth search: Golden section search
Criterion for optimal bandwidth: AICc
Number of varying coefficients: 4
Number of fixed coefficients: 0

Modelling options--------------------------------
Standardisation of independent variables: On
Testing geographical variability of local coefficients: On
Local to Global Variable selection: OFF
Global to Local Variable selection: OFF
Prediction at non-regression points: OFF

Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field11: clp
Offset variable is not specified

Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field15: GameFish
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87

```
****************************************************************************
  Global regression result
****************************************************************************
  < Diagnostic information >
Number of parameters:              4
Deviance:                    39.092896
Classic AIC:                 47.092896
AICc:                        47.423475
BIC/MDL:                     58.438024
Percent deviance explained         0.555286


Variable          Estimate   Standard Error   z(Est/SE)     Exp(Est)
------------------ -------------- -------------- -------------- --------------
Intercept          -3.831883     0.750591       -5.105152     0.021669
GameFish            1.093022     0.463071        2.360379     2.983276
ElevatFt           -0.876844     0.446930       -1.961926     0.416094
d_i87              -1.714018     0.704973       -2.431325     0.180141


****************************************************************************
  GWR (Geographically weighted regression) bandwidth selection
****************************************************************************


Bandwidth search <golden section search>
 Limits: 58,  126
 Golden section search begins...
 Initial values
  pL       Bandwidth:   58.000 Criterion:    47.157
  p1       Bandwidth:   59.447 Criterion:    47.209
  p2       Bandwidth:   60.342 Criterion:    47.218
  pU       Bandwidth:   61.790 Criterion:    47.242
 iter   1 (p1) Bandwidth:   59.447 Criterion:    47.209 Diff:    0.895
 iter   2 (p1) Bandwidth:   58.895 Criterion:    47.157 Diff:    0.553
Best bandwidth size 58.000
Minimum AICc     47.157


****************************************************************************
  GWR (Geographically weighted regression) result
****************************************************************************
  Bandwidth and geographic ranges
Bandwidth size:          58.894582
Coordinate        Min          Max          Range
-------------- -------------- -------------- --------------
X-coord        477564.308000  636437.598000  158873.290000
Y-coord       4773262.610000 4963309.550000  190046.940000


  Diagnostic information
Effective number of parameters (model: trace(S)):            6.152986
Effective number of parameters (variance: trace(S'WSW^-1)):  0.176094
```

Degree of freedom (model: n - trace(S)):   119.847014
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):   113.870122
Deviance: 34.110361
Classic AIC: 46.416333
AICc: 47.156987
BIC/MDL: 63.867936
Percent deviance explained   0.611966
****

Program terminated at 4/16/2017 7:23:48 PM

Model 2: Curly-leaf pondweed (CLP)

Program began at 4/16/2017 7:25:41 PM
Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field11: clp
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field15: GameFish
Independent variable with varying (Local) coefficient: field21: DevOSpMd
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field36: ai_do
****
  Global regression result
****
 < Diagnostic information >
Number of parameters: 6
Deviance: 37.136735
Classic AIC: 49.136735
AICc: 49.842617
BIC/MDL: 66.154426
Percent deviance explained   0.577539

Variable	Estimate	Standard Error	z(Est/SE)	Exp(Est)
Intercept	-4.073254	0.904339	-4.504121	0.017022
GameFish	1.152430	0.520637	2.213498	3.165875
DevOSpMd	0.858927	0.642915	1.335989	2.360627
ElevatFt	-0.606886	0.523062	-1.160257	0.545045
d_i87	-1.893362	0.865518	-2.187549	0.150565
ai_do	-0.690319	0.883038	-0.781755	0.501416

```
*******************************************************************************
  GWR (Geographically weighted regression) bandwidth selection
*******************************************************************************
Bandwidth search <golden section search>
 Limits: 62,  126
 Golden section search begins...
 Initial values
  pL        Bandwidth:   62.000 Criterion:    51.032
  p1        Bandwidth:   86.446 Criterion:    50.406
  p2        Bandwidth:  101.554 Criterion:    50.179
  pU        Bandwidth:  126.000 Criterion:    49.840
  iter   1 (p2) Bandwidth:  101.554 Criterion:    50.179 Diff:    15.108
  iter   2 (p2) Bandwidth:  110.892 Criterion:    49.995 Diff:     9.337
  iter   3 (p2) Bandwidth:  116.663 Criterion:    49.938 Diff:     5.771
  iter   4 (p2) Bandwidth:  120.229 Criterion:    49.905 Diff:     3.567
  iter   5 (p2) Bandwidth:  122.433 Criterion:    49.870 Diff:     2.204
The upper limit in your search has been selected as the optimal bandwidth size.
Best bandwidth size 126.000
Minimum AICc     49.840
*******************************************************************************
  GWR (Geographically weighted regression) result
*******************************************************************************
  Bandwidth and geographic ranges
Bandwidth size:            126.000000
Coordinate          Min            Max            Range
---------------     ---------------     ---------------     ---------------
X-coord       477564.308000   636437.598000   158873.290000
Y-coord       4773262.610000 4963309.550000   190046.940000


  Diagnostic information
Effective number of parameters (model: trace(S)):                    6.742477
Effective number of parameters (variance: trace(S'WSW^-1)):          -0.021306
Degree of freedom (model: n - trace(S)):                   119.257523
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):       112.493741
Deviance:                    35.472302
Classic AIC:                 48.957255
AICc:                        49.840133
BIC/MDL:                     68.080820
Percent deviance explained         0.596473
*******************************************************************************
  GWR Analysis of Deviance Table
*******************************************************************************
Source           Deviance          DOF   Deviance/DOF
-----------      ----------------   - ---------- ----------------
Global model    37.137      120.000       0.309
GWR model       35.472      112.494       0.315
Difference       1.664        7.506       0.222
*******************************************************************************
```

70

Model 3: Curly-leaf pondweed (CLP)
Program began at 4/16/2017 7:27:06 PM
Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field11: clp
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field15: GameFish
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field30: AreaHaMd
****

  Global regression result
****

 < Diagnostic information >
Number of parameters: 5
Deviance: 36.787573
Classic AIC: 46.787573
AICc: 47.287573
BIC/MDL: 60.968982
Percent deviance explained   0.581511

Variable	Estimate	Standard Error	z(Est/SE)	Exp(Est)
Intercept	-4.094107	0.845804	-4.840491	0.016671
GameFish	1.650865	0.645602	2.557094	5.211485
ElevatFt	-1.187299	0.521518	-2.276621	0.305044
d_i87	-1.733932	0.758635	-2.285595	0.176589
AreaHaMd	-0.863359	0.582180	-1.482975	0.421743

****
  GWR (Geographically weighted regression) bandwidth selection
****

Bandwidth search <golden section search>
 Limits: 60, 126
 Golden section search begins...
 Initial values
  pL   Bandwidth: 60.000 Criterion: 48.541
  p1   Bandwidth: 85.210 Criterion: 48.022
  p2   Bandwidth: 100.790 Criterion: 47.790
  pU   Bandwidth: 126.000 Criterion: 47.440
 iter   1 (p2) Bandwidth: 100.790 Criterion: 47.790 Diff: 15.580
 iter   2 (p2) Bandwidth: 110.420 Criterion: 47.592 Diff: 9.629

iter   3 (p2) Bandwidth: 116.371 Criterion: 47.565 Diff: 5.951
iter   4 (p2) Bandwidth: 120.049 Criterion: 47.493 Diff: 3.678
iter   5 (p2) Bandwidth: 122.322 Criterion: 47.451 Diff: 2.273
The upper limit in your search has been selected as the optimal bandwidth size.
Best bandwidth size 126.000
Minimum AICc   47.440

****
 GWR (Geographically weighted regression) result
****
 Bandwidth and geographic ranges
Bandwidth size: 126.000000
Coordinate   Min   Max   Range
--------------   --------------   --------------   --------------
X-coord   477564.308000   636437.598000 158873.290000
Y-coord   4773262.610000 4963309.550000 190046.940000

 Diagnostic information
Effective number of parameters (model: trace(S)):   5.576769
Effective number of parameters (variance: trace(S'WSW^-1)):   -0.114910
Degree of freedom (model: n - trace(S)):   120.423231
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):   114.731552
Deviance: 35.672005
Classic AIC: 46.825543
AICc: 47.439781
BIC/MDL: 62.642832
Percent deviance explained   0.594201

****
Program terminated at 4/16/2017 7:27:29 PM

Model 4: Curly-leaf pondweed (CLP)
Program began at 4/16/2017 7:28:04 PM
Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field11: clp
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field15: GameFish
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field36: ai_do

```
****************************************************************************
  Global regression result
****************************************************************************
 < Diagnostic information >
Number of parameters:              5
Deviance:                    39.071736
Classic AIC:                 49.071736
AICc:                        49.571736
BIC/MDL:                     63.253146
Percent deviance explained        0.555527


Variable          Estimate     Standard Error    z(Est/SE)      Exp(Est)
----------------  --------------   --------------   -----------   ---------------

Intercept         -3.814153       0.755003         -5.051838      0.022056
GameFish           1.113367       0.485534          2.293077      3.044592
ElevatFt          -0.888486       0.454797         -1.953589      0.411278
d_i87             -1.705212       0.703434         -2.424123      0.181734
ai_do             -0.100547       0.683028         -0.147208      0.904342


****************************************************************************
  GWR (Geographically weighted regression) bandwidth selection
****************************************************************************
Bandwidth search <golden section search>
 Limits: 60,  126
 Golden section search begins...
 Initial values
  pL        Bandwidth:   60.000 Criterion:    50.291
  p1        Bandwidth:   85.210 Criterion:    50.057
  p2        Bandwidth:  100.790 Criterion:    49.816
  pU        Bandwidth:  126.000 Criterion:    49.566
 iter   1 (p2) Bandwidth:  100.790 Criterion:    49.816 Diff:    15.580
 iter   2 (p2) Bandwidth:  110.420 Criterion:    49.663 Diff:     9.629
 iter   3 (p2) Bandwidth:  116.371 Criterion:    49.628 Diff:     5.951
 iter   4 (p2) Bandwidth:  120.049 Criterion:    49.613 Diff:     3.678
 iter   5 (p2) Bandwidth:  122.322 Criterion:    49.583 Diff:     2.273
The upper limit in your search has been selected as the optimal bandwidth size.
Best bandwidth size 126.000
Minimum AICc     49.566
****************************************************************************
  GWR (Geographically weighted regression) result
****************************************************************************
  Bandwidth and geographic ranges
Bandwidth size:            126.000000
Coordinate        Min          Max          Range
---------------  --------------  --------------  ---------------

X-coord       477564.308000   636437.598000   158873.290000
Y-coord       4773262.610000  4963309.550000   190046.940000
```

Diagnostic information
Effective number of parameters (model: trace(S)): 5.627171
Effective number of parameters (variance: trace(S'WSW^-1)): 0.146752
Degree of freedom (model: n - trace(S)): 120.372829
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)): 114.598905
Deviance: 37.686496
Classic AIC: 48.940839
AICc: 49.565641
BIC/MDL: 64.901083
Percent deviance explained 0.571285
****
Program terminated at 4/16/2017 7:28:26 PM

Model 5: Curly-leaf pondweed (CLP)

Program began at 6/14/2017 10:57:32 PM
Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field11: clp
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field15: GameFish
Independent variable with varying (Local) coefficient: field21: DevOSpMd
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field28: MinDpMd
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field31: NearDMd
Independent variable with varying (Local) coefficient: field36: ai_do
****
 Global regression result
****
 < Diagnostic information >
Number of parameters: 8
Deviance: 29.058635
Classic AIC: 45.058635
AICc: 46.289405
BIC/MDL: 67.748891
Percent deviance explained 0.669434

Variable	Estimate	Standard Error	z(Est/SE)	Exp(Est)
Intercept	-6.036598	1.835230	-3.289288	0.002390
GameFish	1.313375	0.638236	2.057821	3.718703
DevOSpMd	1.325874	0.879623	1.507321	3.765476

ElevatFt	-0.711676	0.774867	-0.918449	0.490821
MinDpMd	0.708141	0.725834	0.975624	2.030213
d_i87	-3.368972	1.497689	-2.249447	0.034425
NearDMd	-1.547528	0.676546	-2.287394	0.212773
ai_do	-1.364563	1.226437	-1.112624	0.255492

****
 GWR (Geographically weighted regression) bandwidth selection
****

Bandwidth search <golden section search>
 Limits: 66, 126
 Golden section search begins...
 Initial values
  pL   Bandwidth: 66.000 Criterion: 46.345
  p1   Bandwidth: 88.918 Criterion: 46.035
  p2   Bandwidth: 103.082 Criterion: 46.060
  pU   Bandwidth: 126.000 Criterion: 46.019
  iter   1 (p1) Bandwidth: 88.918 Criterion: 46.035 Diff: 14.164
The upper limit in your search has been selected as the optimal bandwidth size.
Best bandwidth size 126.000
Minimum AICc   46.019

****
 GWR (Geographically weighted regression) result
****
 Bandwidth and geographic ranges
Bandwidth size: 126.000000
Coordinate   Min   Max   Range
--------------- --------------- --------------- ---------------
X-coord   477564.308000 636437.598000 158873.290000
Y-coord   4773262.610000 4963309.550000 190046.940000

 Diagnostic information
Effective number of parameters (model: trace(S)):   8.814509
Effective number of parameters (variance: trace(S'WSW^-1)):   0.176785
Degree of freedom (model: n - trace(S)):   117.185491
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):   108.547766
Deviance: 26.901236
Classic AIC: 44.530255
AICc: 46.019427
BIC/MDL: 69.530689
Percent deviance explained   0.693976

****
Program terminated at 6/14/2017 10:58:35 PM

D.2: Statistical output from GWLR for Eurasian watermilfoil (EWI)

Model 1: Eurasian watermilfoil (EWI)

```
**************************************************************************
*        Semiparametric Geographically Weighted Regression        *
*               Release 1.0.90 (GWR 4.0.90)               *
*                   12 May 2015                   *
*          (Originally coded by T. Nakaya: 1 Nov 2009)          *
*                                        *
*        Tomoki Nakaya(1), Martin Charlton(2), Chris Brunsdon (2)    *
*        Paul Lewis (2), Jing Yao (3), A Stewart Fotheringham (4)    *
*               (c) GWR4 development team               *
* (1) Ritsumeikan University, (2) National University of Ireland, Maynooth, *
*        (3) University of Glasgow, (4) Arizona State University        *
**************************************************************************
```

Program began at 6/14/2017 10:42:10 PM
Model settings-------------------------------
Model type: Logistic
Geographic kernel: adaptive Gaussian
Method for optimal bandwidth search: Golden section search
Criterion for optimal bandwidth: AICc
Number of varying coefficients: 9
Number of fixed coefficients: 0

Modelling options-------------------------------
Standardisation of independent variables: On
Testing geographical variability of local coefficients: On
Local to Global Variable selection: OFF
Global to Local Variable selection: OFF
Prediction at non-regression points: OFF

Variable settings-------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field10: ewi
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field18: StateLnc
Independent variable with varying (Local) coefficient: field21: DevOSpMd
Independent variable with varying (Local) coefficient: field22: DecForMd
Independent variable with varying (Local) coefficient: field23: MixForMd
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field31: NearDMd
Independent variable with varying (Local) coefficient: field33: shei

```
****************************************************************************
 Global regression result
****************************************************************************
 < Diagnostic information >
Number of parameters:          9
Deviance:              81.145039
Classic AIC:           99.145039
AICc:                 100.696763
BIC/MDL:              124.671576
Percent deviance explained      0.498338


Variable         Estimate     Standard Error    z(Est/SE)     Exp(Est)
--------------------  ---------------   ---------------   ---------------   ---------------
Intercept        -1.291932     0.358256     -3.606169      0.274739
StateLnc          0.506364     0.312331      1.621241      1.659247
DevOSpMd          0.863370     0.344310      2.507535      2.371138
DecForMd         -0.404135     0.381201     -1.060163      0.667554
MixForMd          0.549944     0.377532      1.456682      1.733156
ElevatFt         -0.686071     0.347090     -1.976633      0.503551
d_i87            -1.285287     0.619440     -2.074919      0.276571
NearDMd          -1.265628     0.365716     -3.460686      0.282062
shei             -0.702619     0.372787     -1.884770      0.495287
****************************************************************************
 GWR (Geographically weighted regression) bandwidth selection
****************************************************************************
Bandwidth search <golden section search>
 Limits: 68,  126
 Golden section search begins...
 Initial values
  pL       Bandwidth:   68.000 Criterion:    97.098
  p1       Bandwidth:   69.235 Criterion:    97.204
  p2       Bandwidth:   69.998 Criterion:    97.204
  pU       Bandwidth:   71.232 Criterion:    97.292
The lower limit in your search has been selected as the optimal bandwidth size.
A new sesssion is recommended to try with a smaller lowest limit of the bandwidth search.
Best bandwidth size 68.000
Minimum AICc    97.098
****************************************************************************
 GWR (Geographically weighted regression) result
****************************************************************************
 Bandwidth and geographic ranges
Bandwidth size:          68.000000
Coordinate        Min             Max            Range
---------------     ---------------    ---------------    ---------------
X-coord       477564.308000   636437.598000  158873.290000
Y-coord       4773262.610000  4963309.550000  190046.940000


 Diagnostic information
```

Effective number of parameters (model: trace(S)):   12.490323
Effective number of parameters (variance: trace(S'WSW^-1)):   -0.289648
Degree of freedom (model: n - trace(S)):   113.509677
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):   100.729707
Deviance: 69.122437
Classic AIC: 94.103083
AICc: 97.098354
BIC/MDL: 129.529159
Percent deviance explained   0.572665
****
Program terminated at 6/14/2017 10:43:59 PM

Model 2: Eurasian watermilfoil (EWI)

Program began at 4/17/2017 1:12:25 PM
Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field10: ewi
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field18: StateLnc
Independent variable with varying (Local) coefficient: field21: DevOSpMd
Independent variable with varying (Local) coefficient: field22: DecForMd
Independent variable with varying (Local) coefficient: field23: MixForMd
Independent variable with varying (Local) coefficient: field25: OpenWatr
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field31: NearDMd
****
 Global regression result
****
 < Diagnostic information >
Number of parameters: 9
Deviance: 81.779462
Classic AIC: 99.779462
AICc: 101.331186
BIC/MDL: 125.305999
Percent deviance explained   0.494415

Variable	Estimate	Standard Error	z(Est/SE)	Exp(Est)
Intercept	-1.320941	0.365735	-3.611746	0.266884
StateLnc	0.518497	0.312657	1.658358	1.679502
DevOSpMd	0.862415	0.340884	2.529937	2.368874
DecForMd	-0.390381	0.402677	-0.969464	0.676799

```
MixForMd         0.457637     0.348287     1.313968     1.580336
OpenWatr         0.592004     0.334532     1.769647     1.807607
ElevatFt        -0.680183     0.351591    -1.934587     0.506524
d_i87           -1.274427     0.616169    -2.068309     0.279591
NearDMd         -1.195427     0.360790    -3.313357     0.302575
```

****
 GWR (Geographically weighted regression) bandwidth selection
****

Bandwidth search <golden section search>
 Limits: 68, 126
 Golden section search begins...
 Initial values
  pL   Bandwidth: 68.000 Criterion: 95.945
  p1   Bandwidth: 69.235 Criterion: 96.081
  p2   Bandwidth: 69.998 Criterion: 96.081
  pU   Bandwidth: 71.232 Criterion: 96.229
The lower limit in your search has been selected as the optimal bandwidth size.
A new session is recommended to try with a smaller lowest limit of the bandwidth search.
Best bandwidth size 68.000
Minimum AICc   95.945
****
 GWR (Geographically weighted regression) result
****
  Bandwidth and geographic ranges
Bandwidth size: 68.000000
Coordinate   Min   Max   Range
--------------- --------------- --------------- ---------------
X-coord   477564.308000   636437.598000   158873.290000
Y-coord   4773262.610000 4963309.550000   190046.940000

  Diagnostic information
Effective number of parameters (model: trace(S)):   12.465380
Effective number of parameters (variance: trace(S'WSW^-1)):   0.181334
Degree of freedom (model: n - trace(S)):   113.534620
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):   101.250574
Deviance: 68.030804
Classic AIC: 92.961564
AICc: 95.944665
BIC/MDL: 128.316896
Percent deviance explained   0.579414
****
Program terminated at 4/17/2017 1:13:52 PM
```

## Model 3: Eurasian watermilfoil (EWI)

Program began at 4/17/2017 1:14:29 PM
Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field10: ewi
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field18: StateLnc
Independent variable with varying (Local) coefficient: field21: DevOSpMd
Independent variable with varying (Local) coefficient: field22: DecForMd
Independent variable with varying (Local) coefficient: field23: MixForMd
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field31: NearDMd


```
****
  Global regression result
****
 < Diagnostic information >
```
Number of parameters:              8
Deviance:               85.135088
Classic AIC:            101.135088
AICc:                   102.365857
BIC/MDL:                123.825343
Percent deviance explained        0.473670

| Variable | Estimate | Standard Error | z(Est/SE) | Exp(Est) |
|---|---|---|---|---|
| Intercept | -1.179764 | 0.332505 | -3.548106 | 0.307351 |
| StateLnc | 0.604725 | 0.298668 | 2.024742 | 1.830749 |
| DevOSpMd | 0.785894 | 0.329135 | 2.387759 | 2.194369 |
| DecForMd | -0.537207 | 0.358025 | -1.500475 | 0.584378 |
| MixForMd | 0.314800 | 0.311652 | 1.010099 | 1.369985 |
| ElevatFt | -0.699359 | 0.331162 | -2.111835 | 0.496904 |
| d_i87 | -1.142769 | 0.556764 | -2.052518 | 0.318935 |
| NearDMd | -1.228873 | 0.346265 | -3.548940 | 0.292622 |

```
*******************************************************************************
  GWR (Geographically weighted regression) bandwidth selection
*******************************************************************************
```

Bandwidth search <golden section search>
 Limits: 66, 126

Golden section search begins...
Initial values
pL   Bandwidth: 66.000 Criterion: 97.033
p1   Bandwidth: 67.277 Criterion: 97.101
p2   Bandwidth: 68.067 Criterion: 97.250
pU   Bandwidth: 69.344 Criterion: 97.399
iter   1 (p1) Bandwidth: 67.277 Criterion: 97.101 Diff: 0.789
iter   2 (p1) Bandwidth: 66.789 Criterion: 97.033 Diff: 0.488
Best bandwidth size 66.000
Minimum AICc   97.033
****
 GWR (Geographically weighted regression) result
****
 Bandwidth and geographic ranges
Bandwidth size: 66.789337
Coordinate   Min   Max   Range
--------------   --------------   --------------   --------------
X-coord   477564.308000   636437.598000   158873.290000
Y-coord   4773262.610000   4963309.550000   190046.940000

 Diagnostic information
Effective number of parameters (model: trace(S)):   11.327022
Effective number of parameters (variance: trace(S'WSW^-1)):   0.168796
Degree of freedom (model: n - trace(S)):   114.672978
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):   103.514753
Deviance: 71.921921
Classic AIC: 94.575964
AICc: 97.032633
BIC/MDL: 126.702590
Percent deviance explained   0.555358
****
Program terminated at 4/17/2017 1:15:32 PM

Model 4: Eurasian watermilfoil (EWI)

Program began at 4/17/2017 1:16:15 PM
****
Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field10: ewi
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field18: StateLnc
Independent variable with varying (Local) coefficient: field21: DevOSpMd
Independent variable with varying (Local) coefficient: field23: MixForMd

Independent variable with varying (Local) coefficient: field25: OpenWatr
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field31: NearDMd

****
  Global regression result
****
 < Diagnostic information >
Number of parameters: 8
Deviance: 82.708095
Classic AIC: 98.708095
AICc: 99.938864
BIC/MDL: 121.398350
Percent deviance explained   0.488674

Variable	Estimate	Standard Error	z(Est/SE)	Exp(Est)
Intercept	-1.399836	0.370082	-3.782502	0.246637
StateLnc	0.542482	0.314668	1.723982	1.720272
DevOSpMd	0.905590	0.333591	2.714675	2.473390
MixForMd	0.502526	0.360166	1.395261	1.652891
OpenWatr	0.683735	0.327019	2.090814	1.981264
ElevatFt	-0.713597	0.354157	-2.014919	0.489879
d_i87	-1.457844	0.607647	-2.399162	0.232738
NearDMd	-1.262460	0.359271	-3.513945	0.282957

****
  GWR (Geographically weighted regression) bandwidth selection
****
Bandwidth search <golden section search>
 Limits: 66, 126
 Golden section search begins...
 Initial values
 pL   Bandwidth: 66.000 Criterion: 93.493
 p1   Bandwidth: 67.277 Criterion: 93.520
 p2   Bandwidth: 68.067 Criterion: 93.690
 pU   Bandwidth: 69.344 Criterion: 93.842
 iter   1 (p1) Bandwidth: 67.277 Criterion: 93.520 Diff: 0.789
 iter   2 (p1) Bandwidth: 66.789 Criterion: 93.493 Diff: 0.488
Best bandwidth size 66.000
Minimum AICc   93.493

```
****************************************************************************
  GWR (Geographically weighted regression) result
****************************************************************************
  Bandwidth and geographic ranges
Bandwidth size:          66.789337
Coordinate        Min              Max              Range
-------------- -      -------------    -------------- -    -------------
X-coord      477564.308000   636437.598000   158873.290000
Y-coord      4773262.610000  4963309.550000  190046.940000

  Diagnostic information
Effective number of parameters (model: trace(S)):              11.373791
Effective number of parameters (variance: trace(S'WSW^-1)):       -0.193185
Degree of freedom (model: n - trace(S)):                      114.626209
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)):     103.059233
Deviance:             68.267819
Classic AIC:          91.015401
AICc:                 93.492592
BIC/MDL:             123.274678
Percent deviance explained        0.577948


****************************************************************************
Program terminated at 4/17/2017 1:17:17 PM
```

Model 5: Eurasian watermilfoil (EWI)
Program began at 4/17/2017 1:17:51 PM

```
****************************************************************************

Variable settings--------------------------------
Areal key is not specified
Easting (x-coord): field13 : labx
Northing (y-coord): field14: laby
Cartesian coordinates: Euclidean distance
Dependent variable: field10: ewi
Offset variable is not specified
Intercept: varying (Local) intercept
Independent variable with varying (Local) coefficient: field18: StateLnc
Independent variable with varying (Local) coefficient: field21: DevOSpMd
Independent variable with varying (Local) coefficient: field23: MixForMd
Independent variable with varying (Local) coefficient: field26: ElevatFt
Independent variable with varying (Local) coefficient: field29: d_i87
Independent variable with varying (Local) coefficient: field31: NearDMd
Independent variable with varying (Local) coefficient: field33: shei
```

```
*****************************************************************************
  Global regression result
*****************************************************************************
 < Diagnostic information >
Number of parameters:              8
Deviance:                  82.266843
Classic AIC:               98.266843
AICc:                      99.497612
BIC/MDL:                  120.957098
Percent deviance explained      0.491402


Variable          Estimate      Standard Error    z(Est/SE)      Exp(Est)
-------------------   ---------------   ---------------    ---------------   ---------------
Intercept         -1.373714      0.359162       -3.824779       0.253165
StateLnc           0.512013      0.314891        1.626002       1.668647
DevOSpMd           0.910937      0.336761        2.704999       2.486651
MixForMd           0.583488      0.385819        1.512335       1.792279
ElevatFt          -0.706161      0.346278       -2.039287       0.493535
d_i87             -1.498644      0.602013       -2.489389       0.223433
NearDMd           -1.342033      0.363572       -3.691241       0.261314
shei              -0.781710      0.367346       -2.127997       0.457623


*****************************************************************************
  GWR (Geographically weighted regression) bandwidth selection
*****************************************************************************


Bandwidth search <golden section search>
 Limits: 66,  126
 Golden section search begins...
 Initial values
 pL        Bandwidth:   66.000 Criterion:    94.629
 p1        Bandwidth:   67.277 Criterion:    94.626
 p2        Bandwidth:   68.067 Criterion:    94.771
 pU        Bandwidth:   69.344 Criterion:    94.903
 iter   1 (p1) Bandwidth:   67.277 Criterion:    94.626 Diff:    0.789
Best bandwidth size 67.000
Minimum AICc     94.626


*****************************************************************************
  GWR (Geographically weighted regression) result
*****************************************************************************
  Bandwidth and geographic ranges
Bandwidth size:           67.277174
Coordinate        Min             Max              Range
---------------    ---------------    ---------------     ---------------
X-coord      477564.308000  636437.598000  158873.290000
Y-coord      4773262.610000  4963309.550000  190046.940000
```

Diagnostic information
Effective number of parameters (model: trace(S)): 11.240113
Effective number of parameters (variance: trace(S'WSW^-1)): 0.190190
Degree of freedom (model: n - trace(S)): 114.759887
Degree of freedom (residual: n - 2trace(S) + trace(S'WSW^-1)): 103.709964
Deviance: 69.727344
Classic AIC: 92.207570
AICc: 94.626354
BIC/MDL: 124.087699
Percent deviance explained 0.568925
****

Program terminated at 4/17/2017 1:18:48 PM

Literature cited

Abell, R., Allan, J. D., and Lehner, B. (2007). Unlocking the potential of protected areas for freshwaters. *Biological Conservation, 134*(1), 48-63. doi:10.1016/j.biocon.2006.08.017

APA (Adirondack Park Agency). (2015). Annual Report. Retrieved on March 22, 2016 from http://apa.ny.gov/

APIPP – Adirondack Park Invasive Plant Program. (2014). *Adirondack park invasive plant program annual report*. Retrieved on March 19, 2016 from www.adkinvasives.com

APIPP – Adirondack Park Invasive Plant Program. (2015). The actual and potential economic impacts of invasive species on the Adirondack park: A preliminary assessment. Prepared by: Yellow Wood Associates, Inc. Retrieved on March 19, 2016 from www.adkinvasives.com)

Besek, J. F., and McGee, J. A. (2014). Introducing the ecological explosion. *International Journal of Sociology, 44*(1), 75-93. doi:10.2753/IJS0020-7659440105

Bingham, N. H., and Fry, J. M. (2010). *Regression: Linear models in statistics* Springer London.

Borrvall, C., Ebenman, B., and Tomas Jonsson, T. J. (2000). Biodiversity lessens the risk of cascading extinction in model food webs. *Ecology Letters, 3*(2), 131-136. doi:10.1046/j.1461-0248.2000.00130.x

Bright, C. (1999). Invasive species: Pathogens of globalization. *Foreign Policy, 116*(116), 50-64.

Briski, E., Bailey, S. A., Cristescu, M. E., and Macisaac, H. J. (2010). Efficacy of 'saltwater flushing' in protecting the great lakes from biological invasions by invertebrate eggs in ships' ballast sediment. *Freshwater Biology, 55*(11), 2414. doi:10.1111/j.1365-2427.2010.02449.x

Bruckerhoff, L., Havel, J., and Knight, S. (2015). Survival of invasive aquatic plants after air exposure and implication for dispersal by recreational boats. *Hydrobiologia*, 746: 113-121. Doi:10.1007/s 10750.014.1947.9

Brundu, G. (2015). Plant invaders in European and Mediterranean inland waters: Profiles, distribution, and threats. *Hydrobiologia, 746*(1), 61-79.

Brunsdon, C., Fotheringham, S., and Charlton, M. (1998). Geographically weighted regression-modelling spatial non-stationarity. *Journal of the Royal Statistical Society. Series D (the Statistician), 47*(3), 431-443.

Buchan, Lucy A. J. and Padilla, D. K. (1999). Estimating the probability of long-distance overland dispersal of invading aquatic species. *Ecological Applications, 9*(1), 254-265.

Buchan, Lucy A. J. and Padilla, D. K. (2000). Predicting the likelihood of Eurasian watermilfoil presence in lakes, a macrophyte monitoring tool. *Ecological Applications, 10*(5), 1442-1455.

Capers, R. S., Selsky, R., Bugbee, G. J., and White, J. C. (2009). Species richness of both native and invasive aquatic plants influenced by environmental conditions and human activity. *Botany, 87*(3), 306-314. doi:10.1139/B08-144

Davis, Mark A. (2011). Researching invasive species fifty years after Elton: A cautionary tale. In D. M. Richardson (Ed.), *Fifty years of invasion ecology: the legacy of Charles Elton*. (pp 269-276). West Sussex: Blackwell Publishing Ltd. doi: 10.1002/9781444329988

Devine, B., and National Geographic Society (U.S.). (1998). *Alien invasion: America's battle with non-native plants and animals* National Geographic Society.

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., et al. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: A review. Ecography, 30, 609-628.

Dullinger, S., Kleinbauer, I., Peterseil, J., Smolik, M., and Essl, F. (2009). Niche based distribution modelling of an invasive alien plant: Effects of population status, propagule pressure and invasion history. *Biological Invasions, 11*(10), 2401-2414. doi:10.1007/s10530-009-9424-5

Engelhardt, K., and Ritchie, M. (2002). The Effect of aquatic plant species richness on wetland ecosystem processes. *Ecology*, 83(10), 2911-2924.

Espeland, E. K. (2013). Predicting the dynamics of local adaptation in invasive species. *Journal of Arid Land, 5*(3), 268-274. doi:10.1007/s40333-013-0163-1

ESRI. (2004). What is ArcGIS? ESRI, Redlands, California.

Fisher, R. A. (1937). The wave of advance of advantageous genes. *Annals of Human Genetics, 7*(4), 355-369. doi:10.1111/j.1469-1809.1937.tb02153.x

Fortin, M.-J. and Dale, M.R.T. (2005). Spatial Analysis: A guide for Ecologists. Cambridge University Press, Cambridge, UK.

Fotheringham, A. S., Brundson, C., and Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships* Wiley.

Fotheringham, A. S., and Oshan, T. M. (2016). Geographically weighted regression and multicollinearity: Dispelling the myth. *Journal of Geographical Systems, 18*(4), 303-329. doi:10.1007/s10109-016-0239-5

Francis, R. A. and Chadwick, M.A. (2012). Chapter 1: Invasive alien species in aquatic ecosystems: A brief overview. In *A handbook of global freshwater invasive species. (pp.3-21).* Abingdon, Oxon, UK: Earthscan.

Francis, R. A. and Pyšek, P. (2012). Chapter 35: Management of freshwater invasive alien species. In *A handbook of global freshwater invasive species. (pp.435-446).* Abingdon, Oxon, UK: Earthscan.

GWR 4. (2016). User Manual: Windows Application for Geographically Weighted Regression Modelling. GWR 4 Development Team (Nakaya, Tomoki).

Gallardo, B., Clavero, M., Sánchez, M. I., and Vilà, M. (2015). Global ecological impacts of invasive species in aquatic ecosystems. *Global Change Biology, 22*(1), 151-163. doi:10.1111/gcb.13004

Gallien, L., Munkemuller, T., Albert, C. H., Boulangeat, I., and Thuiller, W. (2010). Predicting potential distributions of invasive species: Where to go from here? *Diversity and Distributions, 16*(3), 331-342. doi:10.1111/j.1472-4642.2010.00652.x

Gertzen, E. L., Leung, B. and Yan, N. D. (2011). Propagule pressure, Allee effects and the probability of establishment of an invasive species (*Bythotrephes longimanus*). *Ecosphere,* 2(3);art 30. 1-17. doi:10.1890/ES10-00170.1

Hansen, M. J., Madenjian, C. P., Slade, J. W., Steeves, T. B., Almeida, P. R., and Quintella, B. R. (2016). Population ecology of the sea lamprey (*Petromyzon marinus*) as an invasive species in the Laurentian Great Lakes and an imperiled species in Europe. *Reviews in Fish Biology and Fisheries, 26*(3), 509-535. doi:10.1007/s11160-016-9440-3

Harris, P., Fotheringham, A. S., Crespo, R., and Charlton, M. (2010). The use of geographically weighted regression for Spatial prediction: An Evaluation of models using simulated data sets. *Mathematical Geosciences, 42*(6), 657-680. doi:10.1007/s11004-010-9284-7

Hastings, A., Cuddington, K., Davies, K. F., Dugaw, C. J., Elmendorf, S., Freestone, A., . . . Thomson, D. (2005). The spatial spread of invasions: New developments in theory and evidence. *Ecology Letters, 8*(1), 91. doi:10.1111/j.1461-0248.2004.00687.x

Havel, J. E., Kovalenko, K. E., Thomaz, S. M., Amalfitano, S., and Kats, L. B. (2015). Aquatic invasive species: Challenges for the future. *Hydrobiologia, 750*(1), 147-170. doi:10.1007/s10750-014-2166-0

Hejda, M., Pyšek, P., and Jarošík, V. (2009). Impact of invasive plants on the species richness, diversity and composition of invaded communities. *Journal of Ecology, 97*(3), 393-403. doi:10.1111/j.1365-2745.2009.01480.x

Hellmann, J. J., Byers, J. E., Bierwagen, B. G., and Dukes, J. S. (2008). Five potential consequences of climate change for invasive species. *Conservation Biology*, *22*(3), 534–543. http://doi.org/10.1111/j.1523-1739.2008.00951.x

Hulme, P. E., and Firn, J. (2015). Invasion pathways at a crossroad: Policy and research challenges for managing alien species introductions. *Journal of Applied Ecology, 52*(6), 1418-1424.

Johnson, L. E., and Padilla, D. K. (1996). Geographic spread of exotic species: Ecological lessons and opportunities from the invasion of the zebra mussel *Dreissena polymorpha*. *Biological Conservation, 78*(1), 23-33. doi:10.1016/0006-3207(96)00015-8

Johnson, L. E., Ricciardi, A., and Carlton, J. T. (2001). Overland dispersal of aquatic invasive species: a risk assessment of transient recreational boating. *Ecological Applications, 11*(6), 1789-1799. doi:10.1890/1051-0761(2001)011[1789:ODOAIS]2.0.CO;2

Keane, R. M., and Crawley, M. J. (2002). Exotic plant invasions and the enemy release hypothesis. *Trends in Ecology and Evolution, 17*(4), 164-170. doi:10.1016/S0169-5347(02)02499-0

Kearney, M., and Porter, W. (2009). Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. *Ecology Letters, 12*(4), 334-350. doi:10.1111/j.1461-0248.2008.01277.x

Keast, Allen. (1984). The introduced aquatic macrophyte Myriophyllum spicatum, as habitat for fish as habitat for fish and their invertebrate prey. *Canadian Journal of Zoology*. 62(7): 1289-1303. doi:10.1139/z84-186

Keller, R. P., and Perrings, C. (2011). International policy options for reducing the environmental impacts of invasive species. *Bioscience, 61*(12), 1005-1012. doi:10.1525/bio.2011.61.12.10

Kennedy, T. A., Naeem, S., Howe, K. M., Knops, J. M. H., and al, e. (2002). Biodiversity as a barrier to ecological invasion. *Nature, 417*(6889), 636-8. doi:10.1038/nature00776

Kimbel, J. C. (2016). Factors influencing potential intra-lake colonization by *Myriophyllum spicatum L. Aquatic Botany*. 14, 295-307.

Kolar, C., and Lodge, D. M. (2001). Progress in invasion ecology: predicting invaders. *TRENDS in Ecology and Evolution*. 16(4), 199-204.

Kulhanek, S. A., Ricciardi, A., and Leung, B. (2011). Is invasion history a useful tool for predicting the impacts of the world's worst aquatic invasive species? *Ecological Applications, 21*(1), 189-202. doi:10.1890/09-1452.1

Legendre, P. (1993). Spatial autocorrelation: Trouble or new paradigm? *Ecology, 74*(6), 1659-1673. doi:10.2307/1939924

Lennon, J. J. (2000). Red-shifts and red herrings in geographical ecology. *Ecography*. 23, 101-113.

Leung, B., Lodge, D. M., Finnoff, D., Shogren, J. F., Lewis, M. A., and Lamberti, G. (2002). An ounce of prevention or a pound of cure: Bioeconomic risk analysis of invasive species. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 269*(1508), 2407-2413. doi:10.1098/rspb.2002.2179

Levine, J. M., and D'Antonio, C. M. (1999). Elton revisited: A review of evidence linking diversity and invasibility. *Oikos, 87*(1), 15-26.

Liebhold, A. M., and Gurevitch, J. (2002). Integrating the statistical analysis of spatial data in ecology. *Ecography, 25*(5), 553-557. doi:10.1034/j.1600-0587.2002.250505.x

Les, D.H. and Mehrhoff, L.J. (1999). Introduction of nonindigenous aquatic vascular plants in southern New England: A historical perspective. Biological Invasions 1:281-300.

Lodge, D. M. (1993). *Biological invasions: Lessons for ecology*. England: Elsevier Ltd. doi:10.1016/0169-5347(93)90025-K

MacIsaac, H. J., Julianna V. M. Borbely, Muirhead, J. R., and Graniero, P. A. (2004). Back-casting and forecasting biological invasions of inland lakes. *Ecological Applications, 14*(3), 773-783. doi:10.1890/02-5377

Martens, Joe. (2014). New York Great Lakes basin interim action agenda. New York State Department of Environmental Conservation. Retrieved from http://www.dec.ny.gov/lands/91881.html

Martin, C. W., and Valentine, J. F. (2014). Sexual and asexual reproductive strategies of invasive Eurasian milfoil (*Myriophyllum spicatum*) in estuarine environments. *Hydrobiologia, 727*(1), 177-184.

McNaughton, S. J., and Wolf, L. L. (1970). Dominance and the niche in ecological systems. *Science, 167*(3915), 131-139. doi:10.1126/science.167.3915.131

Miller, Jennifer A. and Hanham, Robert Q. (2011). Spatial nonstationarity and the scale of species-environment relationships in the Mojave Desert, California, USA. *International Journal of Geographical Information Science*, 25(3),423-438.

Mills, E. L., Leach, J. H., Carlton, J. T., and Secor, C. L. (1993). Exotic species in the Great Lakes: A history of biotic crises and anthropogenic introductions. *Journal of Great Lakes Research, 19*(1), 1-54. doi:10.1016/S0380-1330(93)71197-1

Mills, E. L., Leach, J. H., Carlton, J. T., and Secor, C. L. (1994). Exotic species and the integrity of the Great Lakes: Lessons from the past. *Bioscience, 44*(10), 666.

Nakaya, T., Charlton, M., Lewis,P., Brunsdon,C., Yao, J. and Fotheringham, S. (2016). GWR4 user manual. GWR4: window application for geographically weighted regression modelling. https://geodacenter. asu.edu/drupal_files/gwr/GWR4manual.pdf

Nichols, S. A., and Shaw, B. H. (1986). Ecological life histories of the three aquatic nuisance plants, *Myriophyllum spicatum, Potamogeton crispus* and *Elodea canadensis*. *Hydrobiologia, 131*(1), 3-21. doi:10.1007/BF00008319

Olson, E. R., Ventura, S. J., and Zedler, J. B. (2012). Merging geospatial and field data to predict the distribution and abundance of an exotic macrophyte in a large Wisconsin reservoir. *Aquatic Botany, 96*(1), 31-41. doi:10.1016/j.aquabot.2011.09.007

Osbourne, J. W. (2002). Notes on the use of data transformation. *Practical Assessment, 8*(6), 1-7.

Páez A, Farber S, Wheeler D. A. (2011). Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships. *Environment and Planning A*. 2011;43:2992-3010.

Pagnucco, K. S., Maynard, G. A., Fera, S. A., Yan, N. D., Nalepa, T. F., and Ricciardi, A. (2014). The future of species invasions in the Great Lakes-St. Lawrence river basin. *Journal of Great Lakes Research,* doi:10.1016/j.jglr.2014.11.004

Peterson, A. T., and Vieglais, D. A. (2001). Predicting species invasions using ecological niche modeling: New approaches from bioinformatics attack a pressing problem. *Bioscience, 51*(5), 363-371.

Peterson, A. T., Papes, M., and Kluza, D. A. (2003). Predicting the potential invasive distributions of four alien plant species in North America. *Weed Science, 51*(6), 863-868. doi:10.1614/P2002-081

Pimentel, D., Zuniga, R., and Morrison, D. (2005). Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics, 52*(3), 273-288.

Pyšek, P., Jarošík, V., Hulme, P. E., Pergl, J., Hejda, M., Schaffner, U., and Vilà, M. (2012). A global assessment of invasive plant impacts on resident species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment. *Global Change Biology, 18*(5), 1725-1737. doi:10.1111/j.1365-2486.2011.02636.x

Rangel, T. F., Diniz-Filho, J. A. F., and Bini, L. M. (2010). SAM: A comprehensive application for spatial analysis in macroecology. *Ecography, 33*(1), 46-50. doi:10.1111/j.1600-0587.2009.06299.x

Ricciardi, A. (2006). Patterns of invasion in the Laurentian Great Lakes in relation to changes in vector activity. *Diversity and Distributions, 12*(4), 425-433. doi:10.1111/j.1366-9516.2006.00262.x

Richardson, D. M., and Pyšek, P. (2007). Elton, C.S. 1958: The ecology of invasions by animals and plants. *Progress in Physical Geography, 31*(6), 659-666. doi:10.1177/0309133307087089

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

Rothlisberger, J. D., Chadderton, L., McNulty, J., and Lodge, D. (2010). Aquatic invasive species transport via trailered boats: What is being moved, who is moving it, and what can be done. *Fisheries*. 35(3), 121-132. doi:10.1577/1548-8446-35.3.121

Shaker, R. Richard, Rapp, C.,Yakubov, A. (2013). Examining Patterns of Aquatic Invasions within the Adirondacks: An OLS and GLM Approach. *Middle States Geographer*, 46:1-11.

Shaker, R. Richard, Rapp, C. (2013). Investigation Aquatic Invasive Species Propagation within the Adirondack Region of New York: A Lake and Landscape approach. *Papers in Applied Geography*, 36:200-209.

Shaker, Richard R. and Ehlinger, T. (2014). Exploring non-linear relationships between landscape and aquatic ecological condition Southern Wisconsin: A GWR and ANN Approach. *International Journal of Applied Geospatial Research*, 5(4):1-20.

Shaker, Richard. R., A. D. Yakubov, S. M. Nick, E. Vennie-Vollrath, T. J. Ehlinger, and K. Wayne Forsythe. (2017). Predicting aquatic invasion in Adirondack lakes: a spatial analysis of lake and landscape characteristics. *Ecosphere* 8(3):e01723. 10.1002/ecs2.1723

Shandas, V., and Alberti, M. (2009). Exploring the role of vegetation fragmentation on aquatic conditions: linking upland with riparian areas in Puget Sound lowland streams. *Landscape and Urban Planning, 90*(1), 66-75. doi:10.1016/j.landurbplan.2008.10.016

Shea, K., and Chesson, P. (2002). Community ecology theory as a framework for biological invasions. *Trends in Ecology and Evolution, 17*(4), 170-176. doi:10.1016/S0169-5347(02)02495-3

Sheppard, A., Shaw, R., and Sforza, R. (2006). Top 20 environmental weeds for classical biological control in Europe: A review of opportunities, regulations and other barriers to adoption. *Weed Research, 46*(2), 93-117. doi:10.1111/j.1365-3180.2006.00497.x

Siesa, M. E., Manenti, R., Padoa-Schioppa, E., De Bernardi, F., and Ficetola, G. F. (2011). Spatial autocorrelation and the analysis of invasion processes from distribution data: A study with the crayfish *Procambarus clarkii. Biological Invasions, 13*(9), 2147-2160. doi:10.1007/s10530-011-0032-9

Smolik, M. G., Dullinger, S., Essl, F., Kleinbauer, I., Leitner, M., Peterseil, J.. . Vogl, G. (2010). Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeography, 37*(3), 411-422.

Thum, R. A., and Lennon, J. T. (2010). Comparative ecological niche models predict the invasive spread of variable-leaf milfoil (myriophyllum heterophyllum) and its potential impact on closely related native species. *Biological Invasions, 12*(1), 133-143.

Tilman, D., Kennedy, T. A., Naeem, S., Howe, K. M., Reich, P., and Knops, J. M. H. (2002). Biodiversity as a barrier to ecological invasion. *Nature, 417*(6889), 636-638. doi:10.1038/nature00776

Tu, J. and Xia, Z. (2008). Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluation. *Science of the Total Environment.* 407 (358-378).

Uden, D. R., Allen, C. R., Angeler, D. G., Corral, L., and Fricke, K. A. (2015). Adaptive invasive species distribution models: A framework for modeling incipient invasions. *Biological Invasions, 17*(10), 2831-2850.

Vander Zanden, M. J., and Olden, J. D. (2008). A management framework for preventing the secondary spread of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences, 65*(7), 1512-1522. doi:10.1139/F08-099

Wagner, H.H., and Fortin, M.J. (2005). Spatial analysis of landscapes: concepts and statistics. *Ecology.* 86: 1975-1987.

Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A., and Losos, E. (1998). Quantifying threats to imperiled species in the United States. *Bioscience, 48*(8), 607-615.

Wilson, E. O. (2002). The future of Life. Vintage Books, New York, New York, USA

Windle, M. J. S., Rose, G. A., Devillers, R., and Fortin, M. (2010). Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): An example from the northwest atlantic. *ICES Journal of Marine Science, 67*(1), 145-154. doi:10.1093/icesjms/fsp224

Xie, D., Yu, D., Xia, C., and You, W. (2014). Stay dormant or escape sprouting? turion buoyancy and sprouting abilities of the submerged macrophyte potamogeton crispus L. *Hydrobiologia, 726*(1), 43-51. doi:10.1007/s10750-013-1749-5