Developing Species Distribution Models for Wood Turtle (*Glyptemys insculpta*) in Atlantic Canada

by

Thomas Hendrik Alexander Baker

Submitted in partial fulfillment of the requirements
for the degree of Master of Environmental Studies

at

Dalhousie University
Halifax, Nova Scotia
August 2022

© Copyright by Thomas Hendrik Alexander Baker, 2022

To all those close to me whose crucial support has made this possible

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The wood turtle, *Glyptemys insculpta,* is listed as threatened federally in Canada and provincially in Nova Scotia (NS) and New Brunswick (NB). Historic surveying for *G. insculpta* in these provinces has been arbitrary and geographic knowledge gaps persist. To address these gaps, this research developed species distribution models for NS and NB using Maxent software, *G. insculpta* occurrence data, and environmental data relevant to the species' ecology. Resulting important model variables included 'Elevation', 'Distance to Alder', and 'Watercourse Density'. The model outputs were used to guide field surveys, which recorded *G. insculpta* occurrences at twelve new sites. In NS, the model output's prediction of *G. insculpta* distribution overlapped with 80.4% of identified core habitat, but only 4.6% of this predicted distribution was within protected areas. The findings will contribute to conservation of this at-risk species by predicting its distribution, and in doing so, inform future survey efforts and conservation decisions.

LIST OF ABBREVIATIONS USED

ACCDC	Atlantic Canada Conservation Data Centre
AUC	Area Under the Curve
AIC_c	Akaike's Information Criterion corrected for small sample sizes
CARP	Clean Annapolis River Project
COSEWIC	Committee On the Status of Endangered Wildlife In Canada
DEM	Digital Elevation Model
EROS SRTM	Earth Resources Observation and Science - Shuttle Radar Topography Mission
ESA	Endangered Species Act
ESRI	Environmental Systems Research Institute
G. insculpta	*Glyptemys insculpta*
GBIF	Global Biodiversity Information Facility
GIS	Geographic Information System
km	Kilometre/s
m	Metre/s
Maxent	Maximum entropy
MTRI	Mersey Tobiatic Research Institute
NB	New Brunswick
NB NRED	New Brunswick Department of Natural Resources and Energy Development
NCC	Nature Conservancy of Canada
NGO	Non-Governmental Organization
NL	Newfoundland and Labrador
NS	Nova Scotia
NS DFA	Nova Scotia Department of Fisheries and Aquaculture
NS NRR	Nova Scotia Department of Natural Resources and Renewables
PEI	Prince Edward Island
PI	Permutation Importance

SDM	Species Distribution Model/Modelling
StD	Standard Deviation
USGS	United States Geological Survey
UTM	Universal Transverse Mercator

ACKNOWLEDGMENTS

I first need to thank my co-supervisors, Dr. Karen F. Beazley and Dr. Lisa I. Doucette who have tirelessly supported me through my thesis journey, from the initial conceptualizations of my research to the final product that you now have before you. I also would like to thank my committee member Dr. Alana Westwood, for providing her expertise that was crucial as I stepped into the daunting world of species distribution modelling. I would also like to extend a thank you to Dr. Allison Moody of Parks Canada for agreeing to be the external examiner for my thesis.

For providing crucial direction, support, and data to my research, I would like to recognize the Nova Scotia Department of Natural Resources and Renewables, New Brunswick Department of Natural Resources and Energy Development, Nature Conservancy of Canada, and Nova Scotia Reptile and Amphibian Recovery Team. I would like to thank many individuals from these organizations for all their help including Kevin Connor, Jason Cyr, Adam Hadley, Dr. Tom Herman, Dr. Donna Hurlburt, Frances Mackinnon, Jeffie McNeil, Dr. Stephen Mockford, Paula Noel, Patrick Nussey, Dr. Mark Pulsifer, Sarah Spencer, Dr. Maureen Toner, and Shaylyn Wallace. I would also like to thank Dr. Constance Browne of the University of New Brunswick, James Churchill of the Atlantic Canada Conservation Data Centre, and Katie McLean of Clean Annapolis River Project for their assistance. I would like to extend a special thanks to Mike Parker and Rick Mosher of East Coast Aquatics, for taking a then-rookie out into the field and showing me how to find wood turtles. For helping with my GIS inquiries, I would also like to thank Jennifer Strang and James Boxall of the Dalhousie GIS Centre.

I would like to acknowledge the financial support I received from the Neil Munro Parks and Protected Areas Award, Colin Stewart Scholarship for Studies on Parks and Protected Areas in the Acadian Forest Region, Nova Scotia Habitat Conservation Fund, the Nature Conservancy of Canada (through the Dr. Bill Freedman Conservation Intern position), the New Brunswick Department of Natural Resources and Energy Development, Dalhousie University graduate funding, and research assistant funding from Dr. Karen F.

Beazley’s research grant. Without all this generosity, my research would not have been possible.

Finally, I am grateful that I have had such a great MES cohort to share the ups and downs of the graduate student experience with, and I would like to express my appreciation to them. I also need to especially thank my close friends, parents, brother, and partner Emily. Their endless support and encouragement has helped me through this process and I am incredibly grateful for them.

CHAPTER 1: INTRODUCTION

In recent decades it has become accepted that Earth is experiencing a biodiversity crisis as the persistence of thousands of species is now imperiled by a multitude of anthropogenic stressors, including habitat destruction, overexploitation, and climate change (Steffen et al., 2015; Ceballos et al., 2015; IPBES, 2019). Nineteen percent of reptiles are estimated to be threatened by extinction (Böhm et al., 2013). Reptiles are more susceptible to anthropogenic threats due to their generally narrower niche requirements and lower mobility, yet are often poorly represented in conservation planning compared to birds and mammals (Böhm et al., 2013; Roll et al., 2017). Climate change also poses a grave threat to reptiles due to their ectothermic physiology and reliance on behaviour to avoid thermal extremes (Sinervo et al., 2010). Turtles, tortoises, and terrapins (Order *Testudines*) are disproportionately more affected by anthropogenic stressors than other reptiles with 51% of *Testudines* species being at risk of extinction (Böhm et al., 2013). This is due to their life history traits of long lives and late sexual maturity, as well as their need for temporally-stable nesting habitat (Heppell et al., 1996; Enneson & Litzgus, 2008; Böhm et al., 2013; Spencer et al., 2017; Lovich et al., 2018). The wood turtle, *Glyptemys insculpta*, is one such at-risk turtle found throughout northeastern North America (Gilhen, 1984; International Union for Conservation of Nature, 2011).

Glyptemys insculpta is a medium-sized freshwater turtle, listed federally as threatened in Canada since 2007 as well as provincially in the Atlantic Canadian provinces of Nova Scotia (NS) and New Brunswick (NB) since 2013 (Endangered Species Act, 1998; Species at Risk Act, 2002; Species at Risk Act, 2012; Environment Canada, 2016; COSEWIC, 2018; NS Department of Lands and Forestry, 2020). Therefore, as of the writing of this thesis, the *G. insculpta* subpopulations in Atlantic Canada have been listed as threatened for almost a decade. As climate change threatens the more southerly areas of this species' range, northern areas such as Atlantic Canada may provide vital climate refugia (Mothes et al., 2020). However, in order for these northern areas to provide sufficient refuge against climate change, important *G. insculpta* habitat must be identified and protected, especially as anthropogenic habitat modification can increase the thermal stress experienced by this species (Hughes & Litzgus, 2019). Besides climate change, *G. insculpta* is imperiled by many other threats throughout Canada including road mortalities, agriculture, habitat modification, illegal collection for the pet trade and subsidized predation from problematic native species (Environment Canada, 2016; COSEWIC, 2018). Of these threats, the direct adult mortality from collisions with vehicles on roads and agricultural machinery in fields remains the most prominent reason for the continued decline of *G. insculpta* subpopulations across Canada (Saumure & Bider, 1998; Saumure et al., 2007; Environment Canada, 2016; COSEWIC, 2018; Wallace et al., 2020). *Glyptemys insculpta* and other turtle species strongly exhibit Type III survivorship (i.e., have naturally high juvenile mortality but low adult mortality) and defy traditional categories of r- versus K-selection (i.e., have large clutch sizes with minimal parental care but exhibit late sexual maturity; Spencer et al., 2017; Mullin et al., 2020). As such, even apparently negligible rates of annual adult mortality can drive turtle subpopulations, including those of *G. insculpta*, toward extirpation (Heppell et al., 1996; Daigle & Jutras, 2005; Enneson & Litzgus, 2008; Spencer et al., 2017; Mullin et al., 2020).

The federal recovery strategy for *G. insculpta* speaks to the urgent need for the reduction of adult mortality, injury, and illegal collection, as well as the conservation, management, and restoration of *G. insculpta* habitat in Canada (Environment Canada, 2016). Despite this urgency, surveying for *G. insculpta* in the Atlantic Canadian provinces of NS and NB has historically been arbitrary and limited compared to the work undertaken in Ontario and Quebec, the other Canadian provinces where *G. insculpta* are found (COSEWIC, 2018). As a result, substantial geographic knowledge gaps remain within NS and NB concerning the distribution of this species. The primary way of addressing such geographic knowledge gaps for *G. insculpta* is the use of visual field surveys. However, these surveys can be time-consuming and thus expensive from a staff-hours perspective, and this expense is exacerbated by the difficulty of finding individual *G. insculpta* (Biggar, 2008; Flanagan et al., 2013).

Consultation with researchers, conservation managers, and Reptile and Amphibian Recovery Team members in NS and NB indicated that species distribution models (SDMs) would serve as useful tools for informing *G. insculpta* survey locations and would thus advance the conservation efforts in these provinces. SDMs are predictive models that can be used to spatially predict the current, future, or historic distribution of a species by modelling the relationship between the occurrences of a species or suite of species and a set of environmental variables or covariates (Austin, 2007; Miller, 2010; Franklin, 2013). Thus, my thesis research primarily aimed to produce SDMs for *G. insculpta* in NS and NB that can be used by researchers and conservation managers to expand our understanding of this species' distribution in Atlantic Canada, inform effective targeting of survey locations and ultimately further the protection of this species. This research was undertaken with the assistance and involvement of multiple governmental and non-governmental organizations within Atlantic Canada, including the NS Department of Natural Resources and Renewables (NS NRR), NB Department of Natural Resources and Energy Development (NB NRED), Nature Conservancy of Canada (NCC), Mersey Tobiatic Research Institute (MTRI), and Clean Annapolis River Project (CARP).

1.1 Research goals and objectives

The primary goal of my thesis is to contribute to improving our understanding of *G. insculpta* distribution in Atlantic Canada through the development of SDMs. In doing so, my objectives are to: 1) identify and assess the most robust environmental variables for predicting the occurrence of *G. insculpta* in Atlantic Canada; 2) share the resulting model outputs with NS NRR and NB NRED to guide survey efforts designed to discover new *G. insculpta* subpopulations; 3) for NS, estimate the portion of areas predicted to have a high probability of *G. insculpta* occurrence that are currently identified as important habitat or are formally protected; and 4) undertake this research using a coproduction-based approach to improve the actionability of the results and conclusions. In doing so, I hope to contribute to the conservation of this species and the expansion of SDM applications in Atlantic Canada, as well as more broadly to the modelling of the distribution of *G. insculpta* and other species with similar habitat requirements.

1.2 *Glyptemys insculpta* Habitat Selection

To model the distribution of a species, it is crucial that one understands their ecology and habitat selection, as well as the niche theory that forms the theoretical foundation for understanding the former. Hutchinson (1957) defined the niche of a species as an ‘n-dimensional hypervolume’ in environmental space, within which the species is able to survive. This ‘hypervolume’ is conceptualized as being composed of the fundamental niche of the species (i.e., limited by climate and resource access) and the usually overlapping but smaller realized niche (i.e., limited by dispersal and biotic interactions, including predation and competition; Pulliam, 2000; Guisan & Zimmermann, 2000; Miller, 2010). However, for species with high dispersal potential and that inhabit landscapes with sufficient connectivity, the realized niche may be larger than the fundamental niche, if enough individuals disperse from source habitats where environmental conditions are suitable to sink habitats where the conditions are not (Pulliam, 2000). *Glyptemys insculpta* is a dispersal-limited species, with most individuals staying within 300 m of their home river and rarely moving between watercourses (Arvisais et al., 2002; Bouchard et al., 2019). Therefore, it is likely that for most *G. insculpta* subpopulations, the realized niche is smaller than the fundamental niche.

Glyptemys insculpta inhabit riparian areas throughout their range in northeastern North America (Gilhen, 1984; COSEWIC, 2018). Unlike many other turtle species in this region, *G. insculpta* extensively uses terrestrial habitat during the active season (i.e., late spring to early fall), but unlike the fully terrestrial eastern box turtle (*Terrepene carolina*), it still requires freshwater habitat for overwintering and active season thermoregulation (Gilhen, 1984; Greaves & Litzgus, 2007; Jones & Willey, 2015). *Glyptemys insculpta* experience two distinct periods in their life history annually: an overwintering season during which they shelter on the bottom of water bodies, predominantly watercourses; and an active season during which they forage and lay their eggs on land while occasionally returning to the water for overnight shelter and easier movement throughout the watershed (Gilhen, 1984).

For *G. insculpta,* watercourses with higher sinuosity, and lower gradient and flow rate produce more suitable habitat (Greaves & Litzgus, 2007; Jones & Willey, 2015; COSEWIC, 2018). These types of streams and rivers are ideal as their slow meandering nature produces two important habitat features: deep pools for overwintering, and sand/gravel bars for nesting (Greaves & Litzgus, 2007; M. Parker, pers. comm., May 2020). In addition to general watercourse variables, *G. insculpta* also select for overwintering sites that are sheltered by underwater structures including roots, log jams, single logs, root balls, beaver lodges, and muskrat burrows (Greaves & Litzgus, 2007; White 2013). During their active season, *G. insculpta* use a plethora of terrestrial habitats, including forests, meadows, scrub-shrub habitats, and agricultural fields (Kaufmann, 1992; Compton et al., 2002; Arisais et al., 2004; Wesley, 2006; Tingley et al., 2009). Likewise, they are generalist foragers, feeding on berries, mushrooms, slugs, earthworms, and other invertebrates (Compton et al., 2002). *Glyptemys insculpta* select heterogeneous landscapes composed of both open and forested habitat in close proximity, and this seems to be primarily driven by their need to balance thermoregulation and foraging (Compton et al., 2002; Dubois et al., 2009). As ectothermic animals, *G. insculpta* must use their environment to regulate their body temperature. Areas with open canopy provide ample sites for basking, while forest cover protects them from the sun during the hot summer days when hyperthermia risk is high (Compton et al., 2002; Arvisais et al., 2004; Hughes & Litzgus, 2019). Additionally, many rich food sources, including slugs, berries, and mushrooms, are most readily found in forested areas or along their edges, while others such as earthworms are more plentiful in fields and meadows (Compton et al., 2002; Wallace, 2020). As *G. insculpta* make use of both open and forested habitat, they have been described as an 'edge species' (Kaufmann, 1992). The dependence of *G. insculpta* on both suitable watercourse and terrestrial habitat is a key aspect of their unique ecology and should be reflected in SDMs for this species.

1.3 Species Distribution Modelling

Predicated upon niche theory, SDM uses statistical and machine-learning based techniques to produce spatial predictions of the current, future, or historic distribution of a species or suite of species (Austin, 2007; Miller, 2010; Franklin, 2013). This requires two sets of data: 1) georeferenced species occurrence data, wherein each set of coordinates represents a location where the species has been observed (a presence) or where it has been determined that the species is not present (an absence); and 2) a stack of environmental variables, commonly referred to as covariates, in raster format. The environmental covariates included in the modelling should reflect habitat features important to the study species (Guisan & Zimmermann, 2000; Austin, 2002). Depending on the chosen SDM technique, the modeller will produce some form of algorithm or equation that models the relationship of the chosen environmental covariates to the presence and, depending on data availability, absence or pseudo-absence of the study species. This model represents an approximation of the niche of the species which is then translated from environmental space to geographic space within the extent of the study area (Guisan & Zimmermann, 2000; Austin, 2002). The resulting model output is usually a map of the predicted distribution of the species which can be applied to conservation efforts or further analyzed for research purposes (Austin, 2002; Austin, 2007; Franklin, 2013).

Prior to beginning the modelling efforts detailed in this thesis, I conducted a targeted literature search of the Biological Abstracts database using the search string: ‘Topic = “Glyptemys insculpta” OR “Clemmys insculpta” OR “wood turtle*” AND Topic = “species distribution model*” OR “ecological niche model*” OR “environmental niche model*”’. This literature search revealed one published study (Mothes et al., 2020) that produced SDMs for *G. insculpta*, and also referred to a report in the grey literature in which *G. insculpta* SDMs were developed (Jones & Willey, 2015). The SDM work of Jones & Willey (2015) has since been published (Willey et al., 2022). Both studies (Mothes et al., 2020; Willey et al., 2022) modelled *G. insculpta* occurrence across the northeastern United States, from the state of Maine in the north to West Virginia in the south. However, this is where the similarities between the models ended for the most part. Willey et al. (2022) split the northeastern United States into three sets of divisions; states, watersheds, and ecoregions, and produced a separate model for each division. This approach resulted in models that were regional and applicable to local conservation efforts. In contrast, Mothes et al. (2020) treated the entire northeastern United States as a single study area and produced one model for the entire area, limiting the local applicability of this model and their study’s ability to reflect regional variation in *G. insculpta* occurrence.

The SDMs developed by Willey et al. (2022) were stream-based, meaning that all the streams in a hydrographic layer were split into roughly 1-kilometre (km) long segments and the model generated a relative probability of *G. insculpta* occurrence. These probabilities were modelled using logistic regression. In contrast, Mothes et al. (2020) used the modelling software Maxent, which uses machine learning and the principle of maximum entropy to create a grid or raster across the study area, with each grid cell containing a relative probability of occurrence (Phillips et al., 2004). Applying a model to linear stream segments as Willey et al. (2022) did is an approach usually reserved for fully aquatic fluvial species, while raster-based approaches like that used by Mothes et al. (2020) are more commonly used for terrestrial species (Mynsberge et al., 2009; Hamilton et al., 2022). As *G. insculpta* uses both watercourse and terrestrial habitat, both approaches are valid. However, in producing a stream-based model, Willey et al. (2022) omitted terrestrial habitat features that may be conducive to *G. insculpta* occurrence, instead using a mix of only watercourse variables (e.g., stream gradient, flow accumulation, etc.) and climate data (e.g., minimum January temperature, 30-year normal precipitation, etc.). Although not published, Browne (2020) produced a *G. insculpta* SDM for NB following the methodology of Willey et al. (2022; originally described in Jones & Willey, 2015), but this model also omitted explicitly terrestrial habitat features. For my thesis, I sought to borrow methodological choices from these studies by producing separate raster-based models for NS and NB that account for both the watercourse and terrestrial habitat requirements of *G. insculpta*. While an earlier model was produced for *G. insculpta* in NB by Browne (2020), the NS models detailed in this thesis are the first *G. insculpta* SDMs ever produced for NS.

In addition to studies that modelled the distribution of *G. insculpta*, I also conducted a literature search of the Biological Abstracts database to find other SDM studies that modelled the distribution of semiaquatic riparian species. To this end, I used the search string: 'Topic = riparian OR river* OR watercourse* OR stream* AND Topic = "species distribution model*" OR "ecological niche model*" OR "environmental niche model*"'. *Glyptemys insculpta* exhibits a distinct ecology as a semiaquatic vertebrate species that is principally dependent on watercourses over other freshwater ecosystems (Gilhen, 1984; Compton et al., 2002; Jones & Willey, 2015; COSEWIC, 2018). My targeted search of the literature returned an apparent dearth of modelling studies focused on species with those characteristics, likely because many semiaquatic freshwater vertebrates (e.g., capybaras, plain-bellied watersnakes, yellow anacondas, neotropical river otters, etc.) will also use palustrine and/or lacustrine habitat (Ferraz et al., 2009; Makowsky et al., 2010; Kershaw et al., 2013; Gomez et al., 2014). Gomez et al. (2014) did attempt to include some watercourse covariates ('River Length' and 'River Width') in their modelling of neotropical river otters (*Lontra longicaudis*) in Argentina, but they were not determined to be important. During this literature search, I found only one study that modelled the distribution of semiaquatic species that are solely dependent on watercourses over other freshwater ecosystems. Kurnaz and Sahin (2021) developed SDMs for two Anatolian mountain brook newts, *Neurergus strauchii* and *Neurergus barani* in Eastern Turkey and parts of neighbouring Syria, Iraq, Iran, and Armenia. However, they were principally interested in investigating the climate-based niche portioning between the two species and predicting future impacts of climate change. As such, Kurnaz and Sahin (2021) only incorporated climate variables into their models and omitted any watercourse or terrestrial habitat features that may be important for their focal species. Therefore, my targeted literature search did not reveal any SDM studies focused on semiaquatic obligate-riparian vertebrate species which incorporated both watercourse and terrestrial habitat features.

1.4 Study Area

Atlantic Canada comprises the provinces of Nova Scotia (NS), New Brunswick (NB), Prince Edward Island (PEI), and Newfoundland and Labrador (NL). Of these, *G. insculpta* are known to inhabit NS and NB and there is no evidence available suggesting their current or historical presence in the other two provinces (Figure 1.1; COSEWIC, 2018). Both NS and NB are situated within the traditional and unceded territory of the L'nu (Mi'kmaw) people, known as Mi'kma'ki. Climatically, both provinces fall within the Atlantic Maritime Ecozone, characterized by a moderate, cool, and moist climate with long, mild winters, and cool summers (Ecological Stratification Working Group, 1995). The forest ecosystems range from boreal coniferous-dominated forests in Northern NB and parts of coastal and highland NS to northern temperate ecosystems including the prominent and unique Wabanaki-Acadian Forest and a small area of the Great Lakes - St. Lawrence River region (Ecological Stratification Working Group, 1995; Neily et al. 2003; NBDNR 2007).

The application of species distribution modelling to at-risk species conservation is relatively new in Atlantic Canada but has been gaining momentum in recent years, especially with the creation of the Atlantic Canada Species At Risk Habitat Modelling Community of Practice in 2021. As of the writing of this thesis, SDMs for multiple species of birds (e.g., Canada Warbler, Olive-sided Flycatcher, Rusty Blackbird, etc.), plants (e.g., *Goodyera pubescens*, *Hepatica nobils var. obtusa*, and *Fraxinus nigra*), and lichens (*Erioderma pedicellatum* and *Fuscopannaria leucosticta*) have been developed in Atlantic Canada (Cameron & Neily, 2008; Staicer et al., 2015; Pearson et al., 2018; Haughian et al., 2018; Bale et al., 2020). As mentioned earlier, an SDM for *G. insculpta* has previously been produced for NB by Browne (2020).

1.5 Thesis Structure

For this thesis, I initially produced SDMs for both NS and NB, which were subsequently used to guide survey effort in 2021, and then refined the NS SDM based on the preliminary findings. Chapter 2 covers the development, results, and application of the initial SDMs. After this preliminary round of modelling, I incorporated additional environmental covariates in the production of a refined SDM for NS. Chapter 3 covers the development, results, and application of this refined SDM, including how it overlaps with the areas of NS that are currently identified as important *G. insculpta* habitat or are formally protected. This chapter is written as a standalone research paper intended for submission to a journal and, as such, there is redundancy between it and the contents of other chapters, particularly within its Methods section. Chapter 4 describes the differences in modelling *G. insculpta* occurrence in NS versus NB, as well as the differences in the initial (2021) and refined (2022) NS SDMs that I produced. Finally, Chapter 5 discusses application of the SDMs to informing survey effort, the challenges and limitations faced during this thesis, and recommendations for future *G. insculpta* modelling in Atlantic Canada and potentially in other regions with similar characteristics and contexts.

1.6 Figures

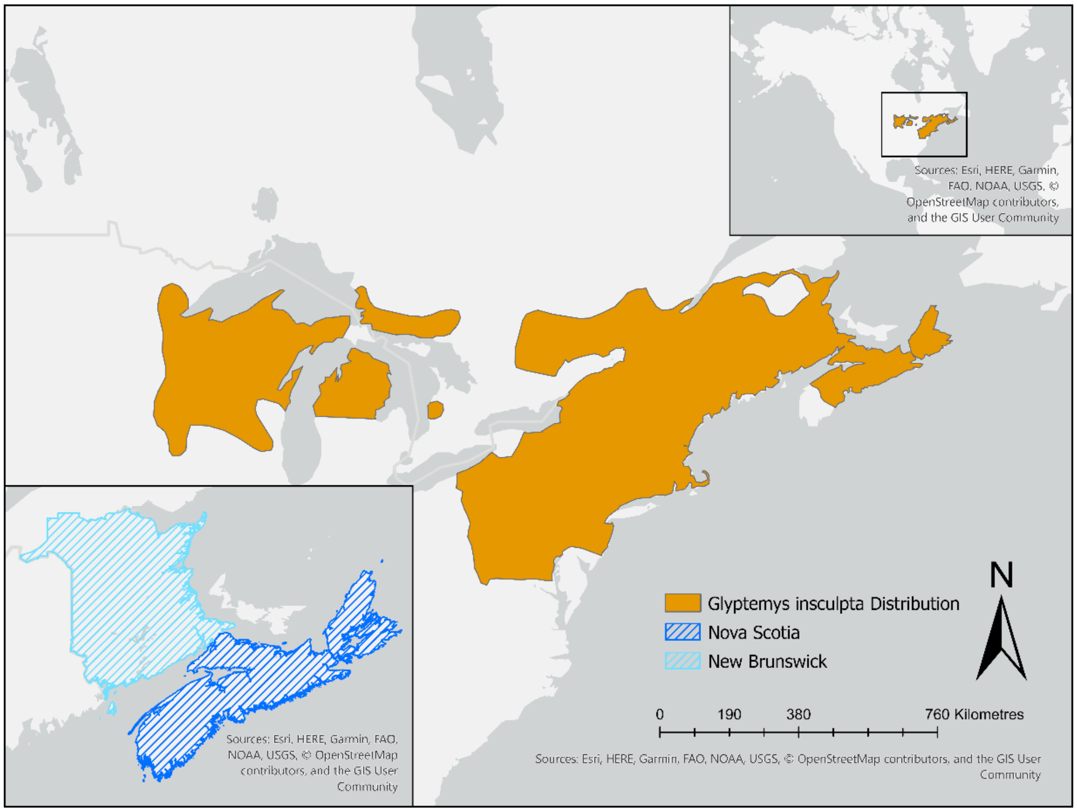


Figure 1.1 Coarse-scale distribution of *Glyptemys insculpta* in orange (provided by S. Giguere of Environment and Climate Change Canada, pers. comm., June 2022) with an inset map highlighting the two study areas, Nova Scotia in dark blue and New Brunswick in light blue.

CHAPTER 2: PRELIMINARY METHODS AND RESULTS

2.1 Research Philosophy

Scientific research is a powerful approach for understanding the natural world. However, to address urgent issues such as the conservation of at-risk species or the threat of climate change, it is not sufficient to solely produce scientific knowledge without taking action in response to it (Meadow et al., 2015; Beier et al., 2017). Knowledge must be effectively coproduced and shared with those individuals and organizations that can effect positive change; including governments, non-governmental organizations (NGOs), and Indigenous communities. A growing body of literature has formed concerning the knowledge-action gap that can exist when knowledge producers (i.e., researchers) make their research decisions in isolation from knowledge users (i.e., managers, executives, community members, etc.), and how this gap often detrimentally affects the actionability of the research (Cvitanovic et al., 2015; Beier et al., 2017; Westwood et al., 2020; Cooke et al., 2021). To address the problem of the knowledge-action gap, coproduction has been proposed as the best strategy to produce actionable scientific knowledge (Beier et al., 2017; Nguyen et al., 2019; Westwood et al., 2020; Cooke et al., 2021).

Coproduction is a collaborative approach to research that involves including end-users as partners throughout the design and execution of a research project, from the initial conceptualizations to the final communications of results and conclusions. (Beier et al., 2017; Westwood et al., 2020; Cooke et al., 2021). In doing so, the goal is that the research produced will be both more relevant to the intended management problem and more respectful of the end-users' values than had it been produced with less end-user participation (Meadow et al., 2015; Nel et al., 2016; Fernández, 2016; Beier et al., 2017). For the end-users, they not only benefit from the production of research that is more applicable to their work, but also gain a deeper understanding of the strengths and limitations of the research which should theoretically improve their management decisions (Cvitanovic et al., 2015; Beier et al., 2017; Reyes et al., 2019). Additionally, engaging in meaningful coproduction with communities, especially Indigenous communities, can result in a research process that is more inclusive and just than more traditional research models (Needham et al., 2020; Westwood et al., 2020; Cooke et al., 2021).

From inception, this thesis applied a coproduction approach to proactively address any potential knowledge-action gaps. It was conceptualized based on crucial needs identified through meetings with two relevant government departments in NS and NB, the Department of Natural Resources and Renewables (NRR) and the Department of Natural Resources and Energy Development (NRED) respectively, as well as the Nature Conservancy of Canada (NCC) and the Reptile and Amphibian Recovery Team in NS. While Beier et al. (2017) recommend holding at least one in-person meeting with all project partners when coproducing research, this was not feasible due to the COVID-19 pandemic and the geographic separation between individuals of the above-mentioned organizations. As a result, virtual meetings were held using video conferencing software. Through these initial meetings, NS NRR and NB NRED emerged as the primary partner organizations for my research. Subsequently, I met with staff in these departments throughout this project to discuss methodological choices as well as applications of the resulting SDMs. By taking a coproduction-based approach to my research, I sought to produce results that would be relevant and applicable to the on-going efforts by NS NRR and NB NRED to promote *G. insculpta* recovery in NS and NB.

2.2 Methodological Approach

Glyptemys insculpta is a cryptic species that is difficult to survey for, which presents a significant limitation when developing SDMs based on presence-absence data, as multiple surveys are needed to confirm its absence at a site (Biggar, 2008; Flanagan et al., 2013). This difficulty of finding *G. insculpta* means that there are no systematically collected absence data for this species in NS or NB with which to build models or guide conservation practices. Although SDM methods that incorporate known absences show generally higher performance than presence-only methods (Zaniewski et al., 2002; Phillips et al., 2004; Elith et al., 2006; Austin, 2007; Miller, 2010), using such a method was not an option for this study. Accordingly, I chose to use the Maxent software package due to its proven robustness when working with presence-only data and its ease of use (Phillips et al., 2004; Phillips et al., 2006; Elith et al., 2006, 2011). Maxent operates using the principle of maximum entropy, whereby the best approximation of an unknown probability distribution is that which models the most uniform distribution possible (i.e., has the maximum entropy), while satisfying the known constraints (Jaynes, 1957; Phillips et al., 2006). In the context of Maxent SDM, the known constraints are determined by averaging the values of the environmental covariates at two sets of georeferenced locations: 1) at species presences; and 2) at randomly selected background points representing a sample of the study area (Phillips et al., 2006; Elith et al., 2011). Maxent modelling necessitates the use of georeferenced presence-only species occurrence data, as described above, along with a stack of environmental variables or covariates in raster-form with the same extent, resolution, and projection.

2.3 Species Occurrence Data

I obtained occurrence data for *G. insculpta* for NB from the Atlantic Canada Conservation Data Centre (ACCDC) and for NS from the ACCDC and NS NRR. The NB ACCDC dataset contained 1935 occurrences while the two NS datasets combined contained 14,509 occurrences. The occurrences in these three datasets include those found during visual field surveys and radio-tracking studies as well as incidental observations. The large number of recorded occurrences for such a threatened and cryptic species in the NS datasets is largely due to the radio-tracking studies conducted on a few subpopulations, in which individual turtles' locations were recorded repeatedly over the course of the studies. No systematically collected presence-absence data are available for *G. insculpta* in NB or NS.

Instances of unauthorized translocations of individual *G. insculpta* in Atlantic Canada have been noted in the past, but no data on translocated individuals are available. To reduce the risk of including occurrences representing translocated or misidentified turtles in the modelling, I defined occurrences as corroborated and thus likely representative of actual *G. insculpta* presence if their nearest neighboring occurrence was within 10 km, as was done by Willey et al. (2022). This exclusionary threshold distance of 10 km accounts for the possibility of unusually large dispersals by individuals as the furthest single-year straight-line distances travelled by *G. insculpta* have been recorded up to 8.3 km (Daigle 1997; Adams 2002; Cameron et al. 2002; Smith 2002; Wesley et al.

2004). I omitted all occurrences which did not meet this requirement. In total, 37 occurrences were omitted from the NS dataset and 29 from the NB dataset due to being uncorroborated, leaving 14,472 NS occurrences and 1906 NB occurrences.

In Atlantic Canada, survey and research effort for *G. insculpta* has been historically uneven, resulting in spatial bias in the datasets. To account for spatial bias, I used the R package 'spThin' to thin the dataset, so that no two occurrences were within 500 m of each other (Aiello-Lammens et al., 2015; R Core Team, 2021; RStudio Team, 2022). The 'spThin' script that I used can be found in Appendix A (Figure A.1). For NS, this reduced the number of occurrences from 14,472 to 607, while for NB, this reduced the occurrences from 1906 to 544. As such, while the original NS dataset was far larger than the NB dataset, after thinning, both provincial datasets contained similar amounts of occurrence data, possibly due to the extensive radio-tracking studies undertaken in some NS watercourses having recorded many occurrences within close proximity (i.e., within the chosen thinning distance of 500 m) of each other. Regarding spatial bias, it is important to recognize that the removal of 'uncorroborated' occurrences as described earlier may have slightly reinforced the spatial bias in the dataset. However, because of the potential risk of translocated or misidentified turtles in the dataset causing unsuitable areas to be considered occupied by *G. insculpta*, I considered this trade-off appropriate.

2.4 Environmental Data

Based on the literature on *G. insculpta* ecology, the previous modelling work for this species (Mothes et al., 2020; Willey et al., 2022), and discussions with regional experts, I identified three habitat features important to *G. insculpta* that could be modelled spatially, either directly or through proxy data. These are the presence of 1) watercourse habitat (Greaves & Litzgus, 2007; Jones & Willey, 2015; COSEWIC, 2018); 2) favourable watercourse characteristics including higher sinuosity, lower gradient and flowrate, and higher flow accumulation (Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018); and 3) terrestrial habitat with either spatially or temporally variable canopy closure (Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019). Together with appropriate climate variables, these habitat features make up the niche of *G. insculpta* that can be modelled spatially. I omitted climate variables from this modelling as, apart from changes in elevation (which I incorporated using an 'Elevation' covariate), temperatures and precipitation are relatively consistent across NS and NB, and known *G. insculpta* subpopulations are located across much of the two provinces (Figure 1.1; Ecological Stratification Working Group, 1995). Many of the covariates that I selected to account for the three habitat features reflect those that were found to be important in the modelling studies by Mothes et al. (2020) and Willey et al. (2022), including, 'Distance to Watercourse', 'Elevation', 'Size of Nearest Watercourse', and 'Gradient of Nearest Watercourse'. I developed these covariates from publicly available data layers and by using ArcGIS Pro (Table 2.1; ESRI Inc., 2020). The initial suite of variables for NS consisted of 10 covariates, while those for NB consisted of 8 covariates (Table 2.2). Detailed descriptions of how I extracted each of these covariates can be found in Appendix B.

For NS, all the covariate layers were projected to NAD 1983 CSRS UTM Zone 20N, transformed into rasters with 250 m cells, and clipped to the extent of NS, excluding Sable Island. For NB, all the covariate layers were projected to NAD 1983 UTM Zone 19N, transformed into rasters with 250 m cells, and clipped to the extent of NB. Before running Maxent, I computed correlation matrices for each suite of initial covariates using the ArcGIS Pro tool 'Band Collection Statistics'. These correlation matrices may be found in Appendix C. Correlation values of 0.7 - 0.75 are often used as a threshold for strong correlation between environmental covariates (e.g., Kershaw et al., 2013; Gomez et al., 2014; Kurnaz & Şahin, 2021). Thus, for covariate pairs with correlation values above 0.7, I planned to remove the covariate that was presumed to be less ecologically relevant. Correlation values of 0.4 to 0.7 prompted consideration but did not immediately warrant removal of either covariate. No covariate pair had a correlation value over 0.7 in either the NS or NB covariate suites (Appendix C). The only covariate pairs that had correlation values over 0.4 were 'Distance to Alder - Distance to Brush' in NS and 'Elevation - Gradient of Nearest Watercourse' and 'Elevation - Tidal Influence of Nearest Watercourse' in NB. As these correlation values were not severe, I elected to retain all the initial covariates for both NS and NB.

2.5 Modelling Procedure

Maxent can model the relationship between an environmental covariate and species presence using one or a combination of different mathematical approaches, referred to as feature types, including a untransformed continuous covariate ("Linear"), the square of a continuous covariate ("Quadratic"), the product of two continuous covariates ("Product"), and thresholds applied to a continuous covariate ("Threshold" and "Hinge"; Phillips et al., 2006). Additionally, to reduce the risk of overfitting to the training data, a regularization multiplier can be applied to Maxent to force the software to limit the number of features used and theoretically produce a more parsimonious distribution (Phillips et al., 2006). In order to select the most optimal set of feature types and the best regularization multiplier for our Maxent modelling, I used the R package 'ENMevaluate' (Muscarella et al., 2014; Kass et al., 2021). The 'ENMevaluate' script that I used can be found in Appendix A (Figures A.2, A.3, and A.4). I chose the combination of feature types and regularization multiplier that produced the lowest corrected Akaike Information Criterion (AICc) value. AICc is a statistical measure of how much information has been lost in the construction of a model, corrected for small sample size to minimise potential model overfitting (Akaike, 1974).

As Maxent is a presence-only modelling approach, it uses pseudo-absences or background points meant to represent the average environmental variables in the study area and stand in for true absence data (Phillips et al., 2004; Phillips et al., 2006). While rigorously-collected absence data are preferable, the performance of pseudo-absences can be improved if their selection is guided by the sampling bias in the presence data (Zaniewski et al., 2002; Phillips et al., 2006). To accomplish this, I used an R script to create spatial-bias files for both NS and NB, a measure of the kernel density of the occurrence data, which the Maxent software can use to weight its placement of pseudo-absences more heavily in areas where surveys have been conducted (as recommended in a Maxent tutorial by J. Banta). The script I used to create the spatial bias files can be found in Appendix A (Figure A.3), and the resulting spatial bias files for both NS and NB are visualized in Appendix D (Figures D.1 and D.2).

For each model run, I chose to have Maxent produce 10 cross-validated model replicates, meaning that the Maxent software randomly split the occurrence data into Training (90%) and Test (10%) data 10 times and ran separate models for each randomly split set. For each replicate, Maxent uses the test data to assess the predictive power of the model by computing the area under the receiver operating characteristic curve (AUC) for both the training and test data. AUC represents the probability that a random positive instance (i.e., a presence) and a random negative instance (i.e., an absence or pseudo-absence) are correctly identified by the model and this value ranges from 0 to 1 (Phillips et al., 2006). For each Maxent run, the model output was the arithmetic average of these 10 replicates, with averaged Test AUC and Training AUC. I used a process of reverse stepwise elimination to narrow down the initial list of covariates to a final suite which was most important for the modelling. This was accomplished by sequentially removing the covariate that had the lowest permutation importance in the last-run model and running a new model without it. The permutation importance of each covariate is derived only from the final model, not the path that Maxent took to produce it, thereby providing a superior measure of covariate importance compared to percent contribution (Phillips, 2017). I evaluated all the models produced and chose a final model based on Test AUC. However, as AUC has been shown to reward overfit models, I also assessed the 10th Percentile Training Presence Test Omission rate and the difference between the Training and Test AUC values to evaluate the risk of overfitting (Lobo et al., 2008; Warren & Seifert, 2011; Radosavljevic & Anderson, 2014). Rates of 10th Percentile Training Presence Test Omission close to 0.1 and small differences between Training and Test AUCs both indicate low levels of overfitting (Warren & Seifert, 2011; Radosavljevic & Anderson, 2014).

2.6 Model Application to Survey Effort

Once I had finalized models for NS and NB, I exported the resulting raster layers into ArcGIS Pro for analysis and application (ESRI Inc., 2020). Each 250 m cell in these rasters contained a predicted relative probability of *G. insculpta* occurrence between 0 and 1. For NS and in communication with experts at NS NRR, I decided to apply an arbitrary threshold of 0.99 to the NS model to identify survey sites for a field season in May – June, 2021, so as to select the most highly predicted sites for surveying. As the principal goal of this field season was to confirm *G. insculpta* presence at sites with no prior recorded occurrences, I created a 2-km buffer around all the corroborated occurrences. I used this buffer and the Erase tool to create a layer containing only those cells with predicted probabilities of occurrence over 0.99 and which were at least 2 km away from a known *G. insculpta* occurrence. The 2-km buffer distance was chosen as it is the distance which NS NRR designates a buffer around *G. insculpta* occurrences for identification of significant habitat and application of special conservation management practices. I then overlayed the resulting raster with the NS hydrographic network and created start and end points for *G. insculpta* surveys on the main watercourses with which each piece of the >0.99 raster overlapped. I supplied these points to NS NRR as potential survey transects. The surveys were conducted by NS NRR staff, using their protocol for *G. insculpta* surveys.

I initially intended to use the same approach for identifying survey sites in NB. However, a legal difference between how the two provinces treat watercourse ownership resulted in a need to take a broader approach for NB. In NS, all watercourses are Crown land (i.e., public land) and cannot be privately owned, while in NB the river bed can be partially or entirely considered part of the adjoining land parcel (La Forest, 1957). The repercussion of this for *G. insculpta* surveying is that in NS, NRR staff can survey in any watercourse in the province without the need to request permission for access, while in NB, NRED staff would need landowner permission to survey in a privately-owned section of a watercourse. To provide NB NRED with the greatest possible options, I applied two thresholds to the NB model, 0.99 as with the NS model and a broad threshold of 0.85. I then removed cells of the model that were within 2 km of known *G. insculpta* occurrences in the same manner as described for NS and provided NB NRED with two rasters, one for 0.99-1 and one for 0.85-0.99. These two rasters were shared with the regional NB NRED offices, and the staff used them to choose their own survey transects.

NS NRR's and NB NRED's survey protocols for *G. insculpta* are informed by the work of Flanagan et al. (2013) and involve searching 1-5-km lengths of watercourse, with observers searching out to 20 m from the watercourse. However, the exact preferred methods differ between provinces. According to the NB NRED protocols, there should be at least two observers who travel together up one bank of the watercourse before surveying the other bank on the way back. One observer surveys from the water's edge out to 10 m, while the second observer surveys from the 10-m mark out to 20 m from the watercourse. NS NRR recommends that their staff survey both sides of the watercourse concurrently if turtles are not being marked, with 1-2 observers on either side searching from the water's edge out to 20 m.

2.7 NS Initial Model Results

The NS model showed high averaged AUC across its 10 replicates for modelling Test data (0.82 +/- 0.03). Based on the recommendations of the R package 'ENMevaluate', all feature types were used in this model including Linear, Quadratic, Hinge, Product, and Threshold features (Kass et al., 2021). For the regularization multiplier, a value of 2.5 was used. The covariates used in this model (Table 2.3) in order of permutation importance were 'Elevation' (48%), 'Distance to Alder' (31.8%), 'Distance to Watercourse' (11.3%), 'Size of Nearest Watercourse' (6.8%), and 'Distance to Brush' (2.1%). The relationship between the predicted probabilities of occurrence and 'Elevation' was negative with higher probabilities predicted at lower elevations (Figure 2.1). The 'Distance to Alder', 'Distance to Watercourse', and 'Distance to Brush' covariates also showed negative relationships, meaning that higher probabilities of occurrence were predicted closer to alder stands, watercourses, and brush. The effect of the 'Size of Nearest Watercourse' covariate was such that rivers (categorized as 'Small Rivers' and 'Medium Rivers') were granted higher predicted probabilities of occurrence than 'Headwaters and Creeks', with 'Small Rivers' having the highest probabilities of occurrence. The difference between the averaged Test AUC and Training AUC values was <0.01 and the 10th Percentile Training Presence Test Omission rate was 0.10, both indicating low levels of overfitting. For each environmental covariate used in a model run, Maxent computes how the model would perform if built using only that covariate, as well as if all covariates except that covariate were used. Of the five environmental covariates retained in the NS model, the covariate that produced the highest AUC when run by itself was 'Distance to Alder', with an AUC of 0.74 (Table 2.5). This was followed by 'Distance to Watercourse' and 'Elevation', both at 0.73. Accordingly, the removal of 'Distance to Alder' resulted in the greatest reduction in AUC compared to the full model (Table 2.5).

2.8 NB Initial Model Results

The NB model showed even higher averaged AUC across its 10 replicates for modelling Test data (0.89 +/- 0.01). Based on the recommendations of the R package ‘ENMevaluate’, only the Hinge feature type was used in this model (Kass et al., 2021). For the regularization multiplier, a value of 0.5 was used. The covariates used in this model (Table 2.4) in order of permutation importance were ‘Elevation’ (65.7%), ‘Distance to Watercourse’ (26.5%), ‘Size of Nearest Watercourse’ (7.8%), and ‘Distance to Alder’ (0.1%). As with the NS model, the relationship between the predicted probability of occurrence and ‘Elevation’ was the strongest relationship and was negative, with the highest probabilities predicted at lower elevations (Figure 2.2). The ‘Distance to Watercourse’ covariate also similarly showed a negative relationship, in that higher probabilities of occurrence were predicted closer to watercourses. Additionally, ‘Small Rivers’ and ‘Medium Rivers’ were once again granted higher predicted probabilities of occurrence than ‘Headwaters and Creeks’, with ‘Small Rivers’ having the highest. However, NB contains a fourth size category, ‘Large Rivers’, in the NCC Stream Classification layer that NS does not, which includes the St. John River, and the tidal portions of the Miramichi and Restigouche Rivers (Atlantic Science, 2019). The NB model scored these ‘Large Rivers’ only marginally better than the ‘Headwaters and Creeks’. Unlike in the NS model, the ‘Distance to Alder’ covariate had a very weak influence on the final NB model. Despite its low permutation importance, I retained ‘Distance to Alder’, as staff at NB NRED thought that an SDM with only three covariates could potentially be too simple. The difference between the averaged Test AUC and Training AUC values was <0.01 and the 10th Percentile Training Presence Test Omission rate was 0.11, both once again indicating low levels of overfitting. Of the four environmental covariates retained in the NB model, the covariate that produced the highest AUC when run by itself was ‘Elevation’, with an AUC of 0.84 (Table 2.6). This was followed by ‘Distance to Watercourse’ at 0.73 and ‘Size of Nearest Watercourse’ at 0.69. Accordingly, the removal of ‘Elevation’ resulted in the greatest reduction in AUC compared to the full model (Table 2.6).

2.9 Field Survey Results

Using my NS model results, I identified 64 potential survey sites in NS that had not been surveyed before or had been last surveyed many years ago (with no previously recorded occurrences), which I shared with NS NRR. Of these, NS NRR staff surveyed 32 sites at least once between the dates of May 18th and June 25th, 2021. These surveys resulted in new *G. insculpta* occurrences being recorded at two sites (L. Doucette, pers. comm., July 2021). On one of the rivers, which we refer to as Site A for data sensitivity reasons, three female *G. insculpta* were found, along with at least 24 turtle nests, many of which showed evidence of predation. The nests could not be definitively determined to be those of *G. insculpta* but the sheer number of nests and the discovery of three individual turtles in one survey suggest that Site A is home to a notable subpopulation of *G. insculpta* and deserves management attention. For NB, I shared the model output with NB NRED and allowed their staff to identify survey sites that they could access and which fit their program. In total, NB NRED staff surveyed 36 sites at least once between the dates of May 13th and June 21st, 2021. These surveys resulted in new *G. insculpta* occurrences being recorded at five of these sites (A. Hadley, pers. comm., July 2021).

Another NS field-survey season was planned and conducted in 2022 using 26 sites chosen by NS NRR in the top 0.99 threshold of a refined NS model (described in Chapter 3), including new watercourses as well as several surveyed in 2021 that warranted further survey effort. Of those 26 sites, 21 were surveyed between May 3rd and June 7th. These surveys resulted in new occurrences being recorded at five sites, meaning that 24% of the sites surveyed in NS in 2022 returned presence data (L. Doucette, pers. comm., June 2022). At one site, which I refer to as Site B, seven living *G. insculpta* were found in a 3-km stretch, including two males, one female, and four juveniles. A deceased *G. insculpta* was also found at Site B, though the cause of mortality was undetermined. These occurrences represent the highest number of individual *G. insculpta* found in any of the model-informed field surveys and thus, as with Site A mentioned earlier, Site B deserves management attention. However, for the many sites that did not return occurrence data in 2021 and 2022, it is important to reiterate that the lack of confirmed occurrences at these sites does not represent absence data, as survey effort was insufficient to confirm absence. Therefore, these sites may warrant further surveying in the future.

2.10 Implications for Further Modelling

Both the NS and NB models showed strong predictive power with high averaged Test AUC values and neither model appeared to be overfit to the Training data. However, they both ended up being ecologically simple models, especially the NB model which appeared to predominantly model the presence of low-elevation rivers. The investigation and incorporation of more covariates that address the habitat features I identified as important for *G. insculpta* could result in SDMs that better model this species' complex habitat requirements. For both the NS and NB models, 'Elevation' was the most important covariate, with the highest probabilities of *G. insculpta* occurrence being predicted at lower elevations. This likely reflects areas where the watercourses of NS and NB are lower in gradient and flowrate, and higher in sinuosity (Hohensinner et al., 2018). The strength of this covariate indicated that it may be beneficial to test the inclusion of more covariates related to the presence of watercourses with those characteristics that are favourable for *G. insculpta*, such as 'Watercourse Density', 'Standard Deviation of Elevation' and 'Standard Deviation of Slope'.

The inclusion of the 'Standard Deviation of Crown Closure' covariate in the NS model was an attempt to incorporate a general measure of terrestrial habitat with variable canopy closure but it was not retained in the final model. Instead, covariates that represent terrestrial habitat types that may provide variable canopy closure, 'Distance to Alder' and 'Distance to Brush', were retained. Perhaps, the measurements of crown closure collected via aerial photograph interpretation in the NS Forest Inventory are too inaccurate and temporally specific to be used for this application. Additionally, it may have been inappropriate to consider measures of crown closure across different forest types (e.g., coniferous softwood forests versus deciduous hardwood forests) as equivalent and directly comparable. These reflections warrant further investigation but, in the meantime, perhaps classifications of habitat types that have variable canopy closure such as 'Alder' and 'Brush' are potentially more specific and temporally stable, and thus more useful. Other habitat types that can be extracted from the NS Forest Inventory, including 'Hardwood' and 'Mixedwood' stands, modelled as 'Distance to Hardwood' and 'Distance to Mixedwood' covariates could also be useful for modelling the distribution of *G. insculpta*. These forest types may be beneficial for the ectothermic *G. insculpta*, as they contain high proportions of deciduous broadleaf trees which are leafless and provide an open canopy for basking in early spring when air temperatures are cooler, but then leaf out in late spring and provide a closed canopy and shade during the hottest times of year (Compton et al., 2002; Dubois et al., 2009).

Based on these reflections, I chose to produce and use the following covariates in a second round of modelling in NS: ‘Watercourse Density’, ‘Standard Deviation of Elevation’, ‘Standard Deviation of Slope’, ‘Distance to Hardwood’, and ‘Distance to Mixedwood’. I also chose to convert the categorical ‘Tidal Influence of Nearest Watercourse’, which did not perform well in either the NS or NB model, into a continuous ‘Distance to Estuary’ covariate. I decided to also produce and include a ‘Distance to Chain Pickerel’ covariate, because of the potential threat to young *G. insculpta* posed by invasive chain pickerel (*Esox niger*), given their known predation on turtle hatchlings in NS (Coastal Action, 2015; Loeza-Quintana et al., 2021). A detailed description of this second round of NS modelling is presented in the following chapter. Additionally, further comparison between the NS and NB models can be found in Chapter 4.

2.11 Tables and Figures

Table 2.1 Spatial data layers used for the extraction of environmental covariates for Maxent modelling of *G. insculpta* in Nova Scotia and New Brunswick, Canada in 2021

Spatial Data Layer	**Description**	**Data Year**	**Resolution (m)**	**Rights**	**Citation**
NS Forest Inventory	Forest inventory of NS, interpreted from aerial photography	1992 - 2018	Vector (Polygon)	NS Department of Natural Resources and Renewables (NS NRR)	(NS NRR, 2021)
NS Hydrographic Network	Hydrographic network of NS, updated and maintained from aerial photography	2020	Vector (Line)	Service Nova Scotia and Internal Services	(Service Nova Scotia and Internal Services, 2020)
NS Enhanced Digital Elevation Model (DEM)	Hydrologically correct digital elevation model for NS	2006	20 m	NS NRR	(NS NRR, 2006, A)
Nature Conservancy of Canada (NCC) Stream Classification v2.0 layer	Hydrographic network of the Northern Appalachian-Acadian Region of Canada and cross-border watersheds of the United States	2019	Vector (Line)	NCC	(Atlantic Science, 2019)
United States Geological Society (USGS) EROS SRTM Digital Elevation Model (DEM)	Near-global digital elevation model, derived from satellite-acquired radar data	2000	1 arc-second (~30 m)	USGS	(USGS, 2000)
NB Landbase Data	Forest, Non-forest, and Wetland data for NB	Unknown	Vector (Polygon)	NB Department of Natural Resources and Energy Development (NB NRED)	(Provided by K. Connor of NB NRED, pers. comm., March 2021)

Table 2.2 Environmental covariates used in Maxent modelling of *G. insculpta* in Nova Scotia and New Brunswick, Canada in 2021. Descriptions of how each covariate was extracted from the source data are available in Appendix *B*

Covariate	**Brief Rationale**	**Literature or Expert Backing**	**Extraction Method**	**NS Source Data**	**NB Source Data**	**Data Type**
Presence of Watercourse Habitat						
Distance to Watercourse	Watercourse Habitat Presence	(Greaves & Litzgus, 2007; Jones & Willey, 2015; COSEWIC, 2018)	Euclidean Distance	Hydrographic Network	NCC Stream Classification	Continuous
Favourable Watercourse Characteristics						
Elevation	Favourable Watercourse Characteristics (Proxy); Climate (Proxy)	(Mothes et al., 2020; L. Doucette, pers. comm., February 2021; Willey et al., 2022)	No Extraction Required	DEM	USGS EROS SRTM DEM	Continuous
Size of Nearest Watercourse	Favourable Watercourse Characteristics	(Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018; Willey et al., 2022)	Euclidean Allocation	NCC Stream Classification	NCC Stream Classification	Categorical
Gradient of Nearest Watercourse	Favourable Watercourse Characteristics	(Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018; Willey et al., 2022)	Euclidean Allocation	NCC Stream Classification	NCC Stream Classification	Categorical
Terrestrial Habitat with Variable Canopy Closure						
Distance to Alder	Terrestrial Habitat with Variable Canopy Closure; Watercourse Habitat Presence (Proxy); Favourable Watercourse Characteristics (Proxy)	(Kaufmann, 1992; Compton et al., 2002; Arvisais et al., 2004; Dubois et al., 2009a; Hughes & Litzgus, 2019; M. Parker, pers. comm., May 2020; M. Pulsifer, pers. comm., 2020)	Euclidean Distance	Forest Inventory	Land-base Data	Continuous
Distance to Brush	Terrestrial Habitat with Variable Canopy Closure	(Kaufmann, 1992; Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019)	Euclidean Distance	Forest Inventory	Not used in NB modelling	Continuous

Covariate	Brief Rationale	Literature or Expert Backing	Extraction Method	NS Source Data	NB Source Data	Data Type
Standard Deviation of Crown Closure	Terrestrial Habitat with Variable Canopy Closure	(Kaufmann, 1992; Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019; Mothes et al., 2020)	Focal Statistics	Forest Inventory	Not used in NB modelling	Continuous
Other Habitat Features						
Temperature of Nearest Watercourse	Climate (Proxy)	(Greaves & Litzgus, 2007; COSEWIC, 2018)	Euclidean Allocation	NCC Stream Classification	NCC Stream Classification	Categorical
Alkalinity of Nearest Watercourse	General Watercourse Characteristic	(COSEWIC, 2018)	Euclidean Allocation	NCC Stream Classification	NCC Stream Classification	Categorical
Tidal Influence of Nearest Watercourse	Hypothesized Avoidance of Brackish Water	(COSEWIC 2018)	Euclidean Allocation	NCC Stream Classification	NCC Stream Classification	Categorical

Table 2.3 Average permutation importance and percent contribution to model prediction, across 10 cross-validated Maxent replicates predicting probability of occurrence of *G. insculpta* in Nova Scotia, Canada in 2021

Covariate	**Average Permutation Importance**	**Average Percent Contribution**
Elevation	48.0	56.9
Distance to Alder	31.8	11.0
Distance to Watercourse	11.3	9.0
Size of Nearest Watercourse	6.8	21.8
Distance to Brush	2.1	1.2

Table 2.4 Average permutation importance and percent contribution to model prediction, across 10 cross-validated Maxent replicates predicting probability of occurrence of *G. insculpta* in New Brunswick, Canada in 2021

Covariate	**Average Permutation Importance**	**Average Percent Contribution**
Elevation	65.7	20.8
Distance to Watercourse	26.5	45.0
Size of Nearest Watercourse	7.8	34.1
Distance to Alder	0.1	0.2

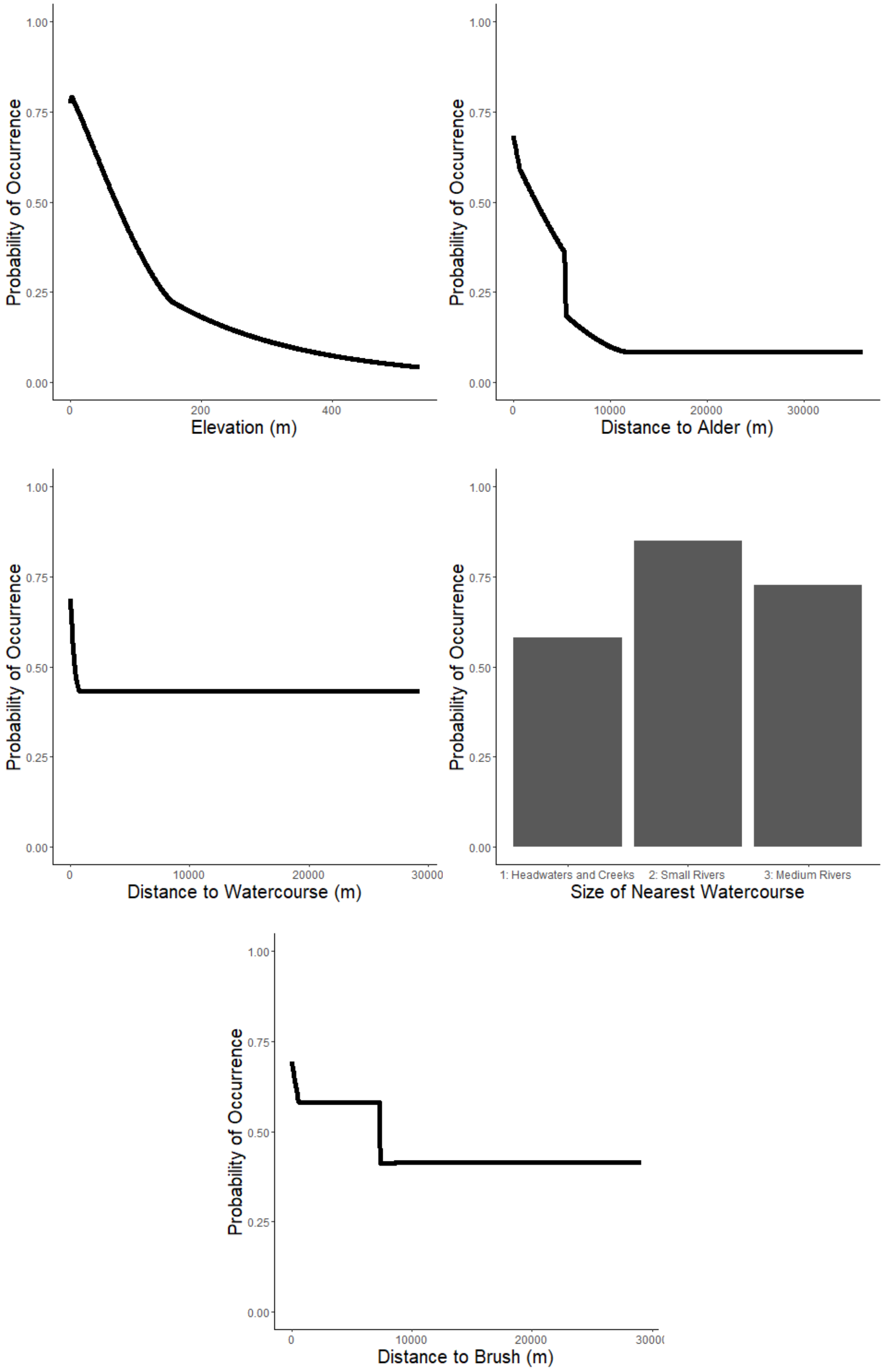


Figure 2.1 Species-response curves indicating the relationship between the predicted probability of occurrence of *G. insculpta* and environmental covariates in the Maxent model for Nova Scotia, Canada in 2021

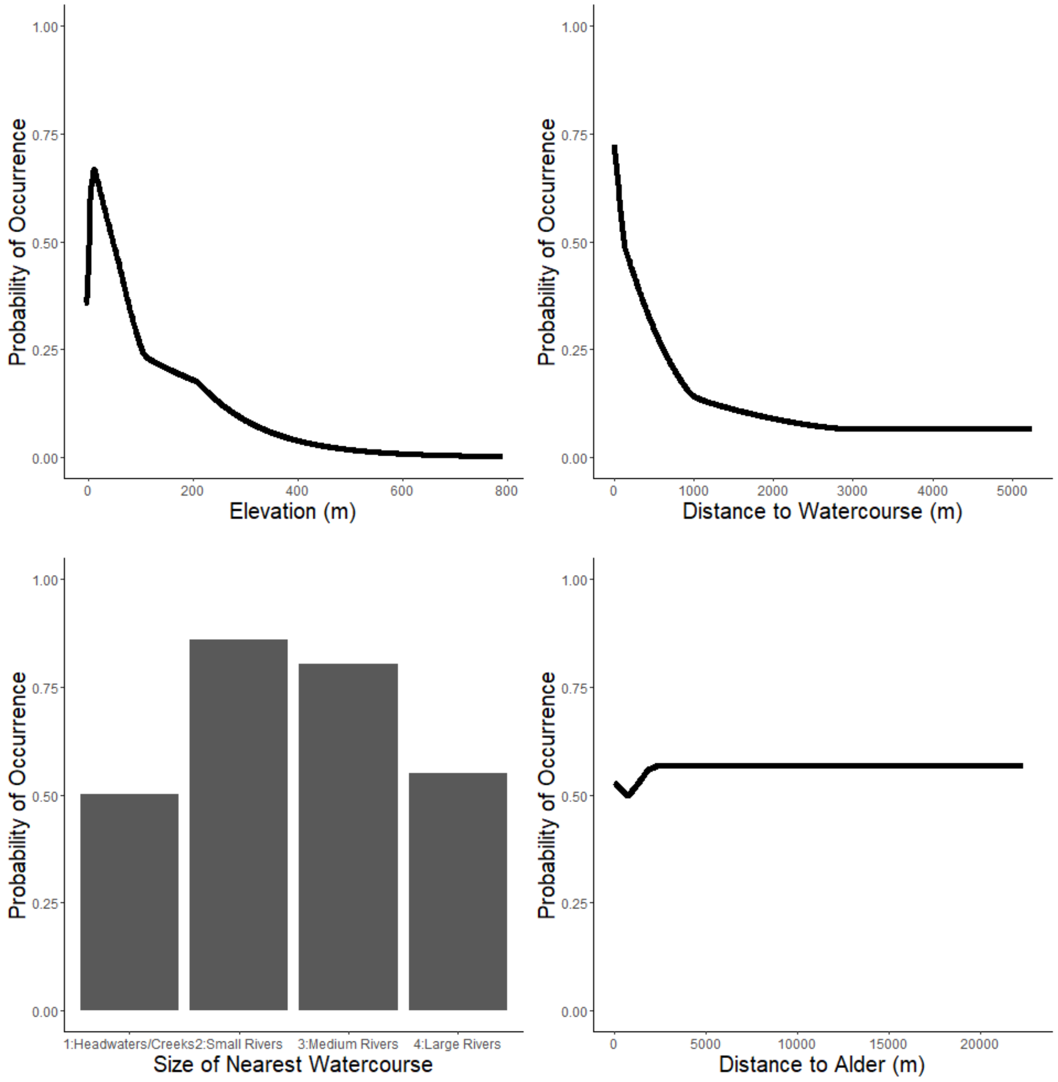


Figure 2.2 Species-response curves indicating the relationship between the predicted probability of occurrence of *G. insculpta* and environmental covariates in the Maxent model for New Brunswick, Canada in 2021

Table 2.5 Area under the receiver operating characteristic curve (AUC) values for the Nova Scotia models that were 1) built with only one covariate; and 2) built using the final suite of covariates minus one covariate.

Covariate	**AUC of model with only this covariate**	**AUC of full model without this covariate**
Distance to Alder	0.74	0.80
Distance to Watercourse	0.73	0.81
Elevation	0.73	0.82
Size of Nearest Watercourse	0.64	0.81
Distance to Brush	0.62	0.82

Table 2.6 Area under the receiver operating characteristic curve (AUC) values for the New Brunswick models that were 1) built with only one covariate; and 2) built using the final suite of covariates minus one covariate.

Covariate	**AUC of model with only this covariate**	**AUC of full model without this covariate**
Elevation	0.84	0.80
Distance to Watercourse	0.73	0.87
Size of Nearest Watercourse	0.69	0.87
Distance to Alder	0.61	0.89

CHAPTER 3: DEVELOPING A SPECIES DISTRIBUTION MODEL FOR THE CONSERVATION OF WOOD TURTLES (GLYPTEMYS INSCULPTA) IN NOVA SCOTIA, CANADA

This chapter is intended as a stand-alone paper to be submitted for potential publication in the target journal, *Conservation Biology*. The co-authors are Thomas H.A. Baker, Lisa I. Doucette, Alana Westwood, and Karen F. Beazley. THAB designed the research with support and input from the other co-authors, acquired and prepared the requisite occurrence and environmental data, developed and analyzed the species distribution model, and led the writing of the paper; LID co-supervised the work, provided expert knowledge on the focal species, organized the field surveys undertaken by the Nova Scotia Department of Natural Resources and Renewables, and contributed feedback to the writing process; AD provided expert knowledge on species distribution modelling which informed the methodological choices made and contributed feedback to the writing process. KFB co-supervised the work, provided expert knowledge on species at risk conservation in Nova Scotia, and contributed feedback to the writing process.

3.1 Introduction

Across the planet, many species are imperilled by a multitude of anthropogenic stressors, constituting a global biodiversity crisis (Steffen et al., 2015; Ceballos et al., 2015; IPBES, 2019). Turtles, tortoises, and terrapins (order *Testudines*) face a disproportionately high extinction risk due to their life history traits of long lives and late sexual maturity, as well as their need for temporally-stable nesting habitat (Heppell et al., 1996; Enneson & Litzgus, 2008; Spencer et al., 2017; Lovich et al., 2018). In fact, 51% of all *Testudines* species are estimated to be at risk of extinction (Böhm et al., 2013). The wood turtle, *Glyptemys insculpta*, is a globally endangered freshwater turtle found throughout northeastern North America (Gilhen, 1984; International Union for Conservation of Nature, 2011). In Canada, this species is officially listed as at-risk at both national (Environment Canada, 2016; COSEWIC, 2018) and provincial levels. The current status of this species warrants the development and application of tools to support its recovery (Environment Canada, 2016).

Glyptemys insculpta are dependent upon watercourses that are sinuous and have low to moderate gradients (Jones & Willey 2015; COSEWIC, 2018). The slow meandering nature of these types of streams and rivers produces two crucial habitat features: deep pools for overwintering, and sufficient sand or gravel bars for nesting (Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018). Relative to other freshwater turtle species in Canada, *G. insculpta* is considerably more terrestrial in nature, foraging on land for much of its active season (late spring to early fall; Gilhen, 1984). During their active season, *G. insculpta* use a plethora of terrestrial habitats, including forests, meadows, alder swales, other scrub-shrub habitats, and agricultural fields (Kaufmann, 1992; Compton et al., 2002; Wesley, 2006; Tingley et al., 2009). Due to their ectothermic physiology and more terrestrial nature relative to other turtle species, *G. insculpta* select for heterogeneous landscapes composed of variable amounts of canopy closure, as areas with open canopy provide ample sites for basking, while nearby closed canopies protect them from the sun during the hottest summer days (Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019). This unique suite of aquatic and terrestrial habitat requirements sets them apart from other turtle species in Canada, in terms of considerations for habitat management and conservation demands.

While listed as endangered globally, *G. insculpta* is listed as threatened across Canada and within the Atlantic Canadian provinces of Nova Scotia (NS) and New Brunswick (NB; *Endangered Species Act*, 1998; *Species at Risk Act*, 2002; International Union for Conservation of Nature, 2011; *Species at Risk Act*, 2012; Environment Canada, 2016; COSEWIC, 2018; NS Department of Lands and Forestry, 2020). Because the more southerly habitat of *G. insculpta* in the United States faces dramatic potential range contraction from climate change, conservation of this species in Canada could be especially important if this region provides climate refugia (Mothes et al., 2020). However, many threats exist for *G. insculpta* in Canada including agriculture, road mortalities, habitat modification, illegal collection, and subsidized predation from problematic native species (Environment Canada, 2016; COSEWIC, 2018). Of these, the direct adult mortality caused by collisions with vehicles on roads and agricultural machinery in fields is considered the most concerning threat (Environment Canada, 2016; COSEWIC, 2018). *Glyptemys insculpta* and other turtle species strongly exhibit Type III survivorship (i.e., have naturally high juvenile mortality but low adult mortality) and defy traditional categories of r- versus K-selection (i.e., have large clutch sizes with minimal parental care but exhibit late sexual maturity; Spencer et al., 2017; Mullin et al., 2020). As such, even seemingly low numbers of yearly adult fatalities, especially of females, can push turtle subpopulations, including those of *G. insculpta*, towards local extinction (Heppell et al., 1996; Daigle & Jutras, 2005; Enneson & Litzgus, 2008; Spencer et al., 2017; Mullin et al., 2020). The provincial status of *G. insculpta* in NS was uplisted from vulnerable to threatened in 2013 in response to the local decline of this species, further emphasizing the need for greater conservation efforts in NS (Environment Canada, 2016; NS Lands and Forestry, 2020).

The federal recovery strategy for *G. insculpta,* which was adopted by the NS government in 2020, speaks to an urgent need for the reduction of adult mortality, injury, and illegal collection, as well as the conservation, management, and restoration of *G. insculpta* habitat in Canada (Environment Canada, 2016; NS Department of Lands and Forestry, 2020). Despite this urgency, surveying for *G. insculpta* in NS has historically been arbitrary and geographic knowledge gaps persist (COSEWIC, 2018). *Glyptemys insculpta* habitat occupancy is the primary reason for designating a site as critical habitat and requires that either at least two distinct *G. insculpta* individuals have been observed at the site or the same individual has been observed over multiple years in the last 40 years (Environment Canada, 2016). Within NS, the provincial equivalent of 'critical habitat' is 'core habitat', which is defined as those areas essential to the survival and recovery of an at-risk species (*Endangered Species Act*, 1998). Currently, core habitat has been identified in NS based on the federal identification of critical habitat but has not been legally designated. Once core habitat is designated, further regulations under the NS *Endangered Species Act* (ESA) may be applied. Additionally, NS NRR has identified significant habitat (a designation broader than both critical and core habitat) for *G. insculpta* based on the observations of at least one individual at a site. Special management practices, which include such measures as timing restrictions for forestry activity and prohibitions on motorized vehicle use near watercourses, amongst others, apply where significant habitat has been identified on Crown land (i.e., public land; NS Department of Natural Resources, 2012). Individual *G. insculpta,* their nests and dwellings, and the areas they habitually occupy are also protected under the NS ESA (1998) and turtle nests specifically are additionally protected under the NS *Wildlife Act* (1989). However, for these protections to be applied, the distribution and habitat use of *G. insculpta* must first be identified, and numbers quantified through field-survey effort. In the context of limited capacity and resources for conservation, reliable methods are needed for targeting survey efforts to sites with high probabilities of occurrence.

Species distribution modelling (SDM) has become an increasingly common tool for conservation research and management. SDM can be used to interpolate species-environment relationships at known species presence sites to unsampled sites within the same region where species occurrence is unknown, and thus map the potential distribution of the species within a region (Austin, 2007; Miller, 2010; Franklin, 2013). This is accomplished by using a broad range of statistical and machine-learning based techniques to model the relationship between the georeferenced occurrences of a species or suite of species, called the response variable, and a set of environmental covariates, called predictor variables (Austin, 2007; Miller, 2010; Franklin, 2013). These modelling approaches apply niche theory to estimate the realized niche of a species or multiple species and then translate this prediction in environmental space to geographic space within the extent of the study area (Guisan & Zimmerman, 2000; Austin, 2002; Phillips et al., 2006; Austin, 2007; Franklin, 2013). The application of SDM to at-risk species within Atlantic Canada remains relatively new but is growing with the recent creation of the Atlantic Canada Species At Risk Habitat Modelling Community of Practice in 2021. To address the knowledge gaps concerning the distribution of *G. insculpta* in Atlantic Canada this study sought to produce the first SDM for this species in the Atlantic Canadian province of NS, with potential utility for prioritizing field-survey sites and for application within other, similar geographic contexts. Comprising a peninsula attached to mainland North America by a narrow isthmus, NS may provide a useful study area for species distribution modelling, as its political boundaries align naturally with its geographic boundaries (Figure 3.1).

As a semiaquatic species with an obligate dependence on watercourse habitat over other freshwater habitats and which also extensively uses terrestrial habitat, *G. insculpta* presents an interesting case study for species distribution modelling. I conducted a literature search of the Biological Abstracts database using the search string: ‘Topic = “Glyptemys insculpta” OR “Clemmys insculpta” OR “wood turtle*” AND Topic = “species distribution model*” OR “ecological niche model*” OR “environmental niche model*”’. This literature search revealed one published study (Mothes et al., 2020) that produced SDMs for *G. insculpta*, and also referred to a report in the grey literature in which *G. insculpta* SDMs were developed (Jones & Willey, 2015). The SDM work of Jones & Willey (2015) has since been published (Willey et al., 2022). Both studies (Mothes et al., 2020; Willey et al., 2022) modelled *G. insculpta* occurrence across the Northeastern United States, from the state of Maine in the north to West Virginia in the south. However, these models differ in substantial ways.

The modelling described in Willey et al. (2022) is applicable to conservation at multiple scales, as separate models were developed for each state, watershed, and ecoregion. Their models were vector-based and predominantly incorporated watercourse variables and climate data. This approach to species distribution modelling in which the model is applied to stream reaches is usually reserved for fully-aquatic fluvial species, while raster-based approaches in which the model is applied continuously across the landscape are generally applied to terrestrial species (Mynsberge et al., 2009; Hamilton et al., 2022). The use of solely watercourse variables (e.g., stream gradient, flow accumulation, etc.) and climate data (e.g., minimum January temperature, 30-year normal precipitation, etc.) in the modelling by Willey et al. (2022) omits the potential role that terrestrial habitat features may have in predicting *G. insculpta* occurrence during its active season. In contrast, Mothes et al. (2020) used a raster-based model that incorporated a relatively diverse suite of environmental variables including some terrestrial features, and which they applied to the entire range of *G. insculpta* in the northeastern United States. Their objective was to predict the change in distribution of *G. insculpta* in the northeastern United States due to projected climate change (Mothes et al., 2020).

For my research, I drew from different aspects of these two studies (Mothes et al., 2020; Willey et al., 2022) by limiting my modelling to the NS region and applying it at a landscape scale, while incorporating a mix of watercourse and terrestrial habitat features in a raster-based format. I also conducted a literature search of the Biological Abstracts database to find other SDM studies that modelled the distribution of semiaquatic riparian species, by using the search string: ‘Topic = riparian OR river* OR watercourse* OR stream* AND Topic = “species distribution model*” OR “ecological niche model*” OR “environmental niche model*”’. This targeted search returned an apparent dearth of modelling studies focused on semiaquatic freshwater vertebrate species dependent upon riparian habitat, likely because many semiaquatic freshwater vertebrates (e.g., capybaras, plain-bellied watersnakes, yellow anacondas, neotropical river otters, etc.) will also use palustrine and/or lacustrine habitat (Ferraz et al., 2009; Makowsky et al., 2010; Kershaw et al., 2013; Gomez et al., 2014). One study that explicitly concerned species dependent on riparian habitat modelled the distribution of two Anatolian mountain brook newts, *Neurergus strauchii* and *Neurergus barani*, but as this study was focused on elucidating climate-based niche partitioning between the two species and predicting future impacts of climate change, only climate variables were included as covariates (Kurnaz & Şahin, 2021). As such, no studies were revealed that modelled the distribution of a riparian-dependent semiaquatic vertebrate species using both watercourse and terrestrial habitat features.

In order to be effective, research concerning at-risk species needs to be coproduced and shared with the individuals and organizations that can apply the resulting knowledge to positive action (Meadow et al., 2015; Beier et al., 2017). Otherwise, the phenomenon of the knowledge-action gap can manifest, in which even well-conducted research may not be found or fully understood by its intended end-users (i.e., managers, executives, community members, etc.) because it was not produced with their involvement (Cvitanovic et al., 2015; Beier et al., 2017; Westwood et al., 2020; Cooke et al., 2021). Coproduction presents a collaborative approach to research that proactively seeks to address any potential knowledge-action gaps by including end-users as partners throughout the design and execution of a research project (Beier et al., 2017; Westwood et al., 2020; Cooke et al., 2021). Through this approach, the goal is to produce more effective and relevant research that is more accessible and useful to end-users (Cvitanovic et al., 2015; Nel et al., 2016; Fernández, 2016; Beier et al., 2017; Reyes et al., 2019). When engaging with communities, especially Indigenous communities, meaningful coproduction can also result in a research process that is more inclusive and just than traditional research models. (Needham et al., 2020; Westwood et al., 2020; Cooke et al., 2021).

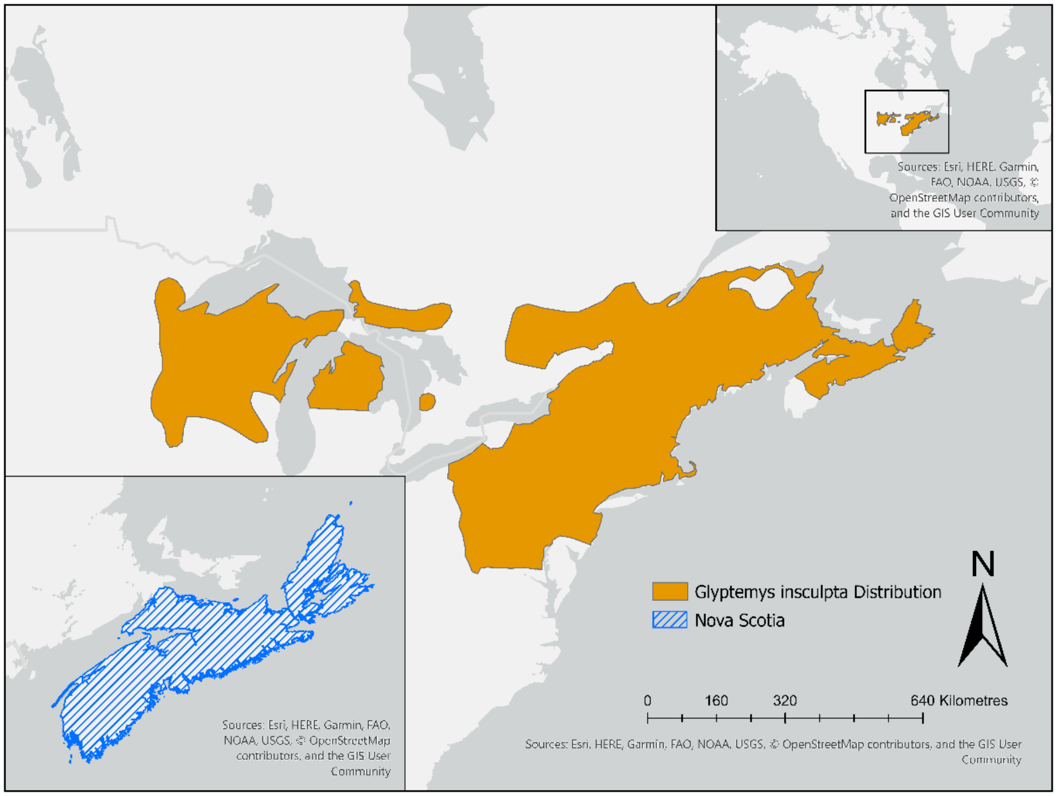
To address these gaps in SDMs, knowledge and action, this study aimed to contribute to the development of SDMs for semiaquatic obligate-riparian species for use in conservation applications while also improving our understanding of *G. insculpta* distribution in Atlantic Canada by coproducing a SDM for *G. insculpta* in NS. The partner organization and intended end-user of this research has been NS NRR, who have been involved in the initial conceptualization methodological design of this study, as well as the direct application of its resulting SDM. My objectives were to: 1) identify and assess the most robust environmental variables for predicting the occurrence of *G. insculpta* in NS; 2) identify areas predicted to have a high probability of occurrence; and 3) estimate the portion of predicted high probability occurrence area that is currently protected or identified as important *G. insculpta* habitat by NS NRR. In doing so, I produced spatial layers and maps of areas where *G. insculpta* may occur that can be used by both government and qualified non-governmental organizations (NGOs) to guide ongoing field-survey efforts for this species aimed at addressing the broader purpose of this study, namely closing the geographic knowledge gaps in Atlantic Canada. The SDM methods, including environmental variables, have potential relevance for application to semiaquatic obligate-riparian species in other geographies with similar contexts.

3.2 Methods

3.2.1 Study Area

Nova Scotia is a coastal peninsula in northeastern North America and one of Canada's easternmost provinces (Figure 3.1). It is also situated within the traditional and unceded territory of the L'nu (Mi'kmaw) people, known as Mi'kma'ki. It is characterized by a cool temperate maritime climate that is milder and more humid than adjacent inland regions (Neily et al., 2003). Nova Scotia is located within the Atlantic Maritime Ecozone (Ecological Stratification Working Group, 1995). Its mainland is primarily dominated by the Wabanaki-Acadian Forest ecosystem type, a unique temperate forest comprised of a mix of deciduous and coniferous tree species, though more boreal ecosystem types are found in the highlands of Cape Breton-Unama'ki and exposed coastal areas of the province

(Neily et al., 2003). The topography of NS primarily consists of low hills, with the highest elevations reaching 550 m (metres) in the Cape Breton-Unama'ki highlands (The Canadian Encyclopedia, 2021).



Sources: Esri, HERE, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community

Sources: Esri, HERE, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community

Sources: Esri, HERE, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community

Figure 3.1 Coarse-scale distribution of *Glyptemys insculpta* in orange (provided by S. Giguere of Environment and Climate Change Canada, pers. comm., June 2022) with an inset map highlighting the study area (Nova Scotia) in blue.

3.2.2 Methodological Approach

Glyptemys insculpta is a cryptic species and difficult to locate, which presents a significant limitation when developing SDMs as multiple surveys are needed to confirm its absence at a site (Biggar, 2008; Flanagan et al., 2013). This difficulty of finding *G. insculpta* means that there are no systematically collected presence-absence data for this species in NS with which to build models. Although SDM methods that incorporate known absences generally show higher performance than presence-only methods (Zaniewski et al., 2002; Phillips et al., 2004; Elith et al., 2006; Austin, 2007; Miller, 2010), using such a method was not an option for this study. Accordingly, I chose instead to use the Maxent software package due to its proven robustness when working with presence-only data and its ease of use (Phillips et al., 2004; Phillips et al., 2006; Elith et al., 2006, 2011). Maxent operates using the principle of maximum entropy, whereby the best approximation of an unknown probability distribution is that which models the most uniform distribution possible (i.e., has the maximum entropy), while satisfying the known constraints (Jaynes, 1957; Phillips et al., 2006). In the context of Maxent SDM, the known constraints are determined by averaging the values of the environmental covariates at two sets of georeferenced locations: 1) species presences, and 2) randomly selected background points representing a sample of the study area (Phillips et al., 2006; Elith et al., 2011). Maxent modelling necessitates the use of georeferenced presence-only species occurrence data, along with a stack of environmental variables or covariates in raster-form at a consistent extent, resolution, and projection.

3.2.3 Species Occurrence Data

Occurrence data for *G. insculpta* were obtained from the Atlantic Canada Conservation Data Centre (ACCDC) and the NS Department of Natural Resources and Renewables (NRR). Together, the two datasets represent the most complete known aggregation of *G. insculpta* occurrences for NS and include 14,509 presences spread across most of the province. These occurrences include those recorded during visual field surveys and radio-tracking studies as well as incidental observations. The large number of recorded occurrences for a threatened and cryptic species is primarily a result of radio-tracking studies conducted on a few subpopulations, in which some individuals' locations were repeatedly recorded over time. As described earlier, no systematically collected presence-absence data are available for *G. insculpta* in NS.

Instances of unauthorized translocations of individual marked *G. insculpta* in NS have been noted in the past, but no data on translocated individuals are available. To reduce the risk of occurrences representing translocated or misidentified turtles being used in the modelling, I defined occurrences as being corroborated and thus likely representative of actual *G. insculpta* presence if their nearest neighboring occurrence was within 10 km, as was done by Willey et al. (2022). This exclusionary distance of 10 km accounts for the possibility of unusually large dispersals by individuals as the furthest single-year straight-line distances travelled by *G. insculpta* have been recorded up to 8.3 km (Daigle 1997; Adams 2002; Cameron et al. 2002; Smith 2002; Wesley et al. 2004). I omitted all occurrences which did not meet this requirement. In total, 37 occurrences were omitted from the dataset due to being uncorroborated, leaving 14,472 occurrences.

Much of the historic survey effort for *G. insculpta* in NS has been focused on a handful of well-known subpopulations, resulting in spatial bias in the datasets. To account for spatial bias, I used the R package 'spThin' to thin the dataset, so that no two occurrences were within 500 m of each other (Aiello-Lammens et al., 2015; R Core Team, 2021; RStudio Team, 2022). The 'spThin' script that I used can be found in Appendix A. This substantially reduced the number of occurrences from 14,472 to 607. The remaining 607 occurrences represent the presences used in my modelling efforts. Regarding spatial bias, it is important to recognize that the removal of 'uncorroborated' occurrences as described earlier may have slightly reinforced the spatial bias in the dataset. However, because of the potential risk of translocated or misidentified turtles in the dataset causing unsuitable areas to be considered occupied by *G. insculpta*, I considered this trade-off appropriate.

3.2.4 Environmental Data

Based on discussions with regional experts, information within the literature on *G. insculpta* ecology, and the previous modelling work for this species (Mothes et al., 2020; Willey et al., 2022), I identified three habitat features important to *G. insculpta* that I sought to account for either directly or through proxies: 1) the presence of watercourse habitat (Greaves & Litzgus, 2007; Jones & Willey, 2015; COSEWIC, 2018); 2) favourable watercourse characteristics including higher sinuosity, lower gradient and flowrate, and higher flow accumulation (Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018); and 3) the presence of terrestrial habitat with either spatially or temporally variable canopy closure (Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019). Many of the covariates I selected to account for these habitat features reflect those that were found to be important in the modelling studies by Mothes et al. (2020) and Willey et al. (2022), including ‘Distance to Watercourse’, ‘Elevation’, ‘Size of Nearest Watercourse’, and ‘Gradient of Nearest Watercourse’. I developed these covariate data from publicly available data layers (Table 3.1) and using ArcGIS Pro (ESRI Inc., 2020). To account for the habitat requirements of *G. insculpta*, my initial suite of variables consisted of 15 covariate layers (Table 3.2), which included a mix of watercourse-based covariates (e.g., ‘Distance to Watercourse’, ‘Size of Nearest Watercourse’, ‘Gradient of Nearest Watercourse’, ‘Watercourse Density’, etc.) and terrestrial habitat covariates (e.g., ‘Distance to Alder’, ‘Distance to Brush’, ‘Distance to Hardwood’, ‘Distance to Mixedwood’, etc.). As *G. insculpta* depend upon sinuous and meandering watercourses, I attempted to produce a covariate layer that represented watercourse sinuosity but was unable to feasibly produce a viable layer across the extent of the province (Jones & Willey 2015; COSEWIC, 2018). Detailed descriptions of how I extracted each covariate can be found in Appendix B.

Table 3.1 Spatial data layers used for the extraction of environmental covariates for Maxent modelling of *G. insculpta* in Nova Scotia, Canada in 2022

Spatial Data Layer	Description	Data Year	Resolution (m)	Rights	Citation
NS Forest Inventory	Forest inventory of NS, interpreted from aerial photography	1992 - 2018	Vector (Polygon)	NS Department of Natural Resources and Renewables (NS NRR)	(NS NRR, 2021)
NS Hydrographic Network	Hydrographic network of NS, updated and maintained from aerial photography	2020	Vector (Line)	Service Nova Scotia and Internal Services	(Service Nova Scotia and Internal Services, 2020)
NS Enhanced Digital Elevation Model	Hydrologically correct digital elevation model for NS	2006	20 m	NS NRR	(NS NRR, 2006, A)
Nature Conservancy of Canada (NCC) Stream Classification v2.0 layer	Hydrographic network of the Northern Appalachian-Acadian Region of Canada and cross-border watersheds of the United States	2019	Vector (Line)	NCC	(Atlantic Science, 2019)
Chain Pickerel Global Biodiversity Information Facility (GBIF) Dataset for NS	Compilation of all INaturalist observations of chain pickerel (*Esox niger*) in NS	2021	Vector (Point)	GBIF	GBIF.org, 2021
NS Freshwater Species Dataset	Compiled list of freshwater fish species presences from survey data	2019	Vector (Point)	NS Department of Fisheries and Aquaculture (NS DFA)	(NS DFA, 2015)

Table 3.2 Environmental covariates used in Maxent modelling of *G. insculpta* in Nova Scotia, Canada in 2022. Descriptions of how each covariate was extracted from the source data are available in Appendix *B*.

Covariate	**Brief Rationale**	**Literature or Expert Support**	**Extraction Method**	**Source Data**	**Data Type**
Presence of Watercourse Habitat					
Distance to Watercourse	Watercourse Habitat Presence	(Greaves & Litzgus, 2007; Jones & Willey, 2015; COSEWIC, 2018)	Euclidean Distance	Hydrographic Network	Continuous
Watercourse Density	Watercourse Habitat Presence; Favourable Watercourse Characteristics (Proxy)	(Gilhen, 1984; Greaves & Litzgus, 2007; Jones & Willey, 2015; COSEWIC, 2018)	Line Density	Hydrographic Network	Continuous
Standard Deviation of Elevation	Watercourse Habitat Presence (Proxy)	(Mothes et al., 2020; Willey et al., 2022)	Focal Statistics	Digital Elevation Model	Continuous
Standard Deviation of Slope	Watercourse Habitat Presence (Proxy)	(Mothes et al., 2020; Willey et al., 2022)	Focal Statistics	Digital Elevation Model	Continuous
Favourable Watercourse Characteristics					
Elevation	Favourable Watercourse Characteristics (Proxy); Climate (Proxy)	(Mothes et al., 2020; Willey et al., 2022)	No Extraction Required	Digital Elevation Model	Continuous
Size of Nearest Watercourse	Favourable Watercourse Characteristics	(Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018; Willey et al., 2022)	Euclidean Allocation	NCC Stream Classification Layer	Categorical
Gradient of Nearest Watercourse	Favourable Watercourse Characteristics	(Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018; Willey et al., 2022)	Euclidean Allocation	NCC Stream Classification Layer	Categorical
Terrestrial Habitat with Variable Canopy Closure					
Distance to Alder	Terrestrial Habitat with Variable Canopy Closure; Watercourse Habitat	(Kaufmann, 1992; Compton et al., 2002; Arvisais et al., 2004; Dubois et al., 2009a; Hughes & Litzgus, 2019;	Euclidean Distance	Forest Inventory	Continuous

Covariate	Brief Rationale	Literature or Expert Support	Extraction Method	Source Data	Data Type
	Presence (Proxy); Favourable Watercourse Characteristics (Proxy)	M. Parker, pers. comm., May 2020; M. Pulsifer, pers. comm., 2020)			
Distance to Brush	Terrestrial Habitat with Variable Canopy Closure	(Kaufmann, 1992; Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019)	Euclidean Distance	Forest Inventory	Continuous
Distance to Hardwood	Terrestrial Habitat with Variable Canopy Closure	(Kaufmann, 1992; Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019)	Euclidean Distance	Forest Inventory	Continuous
Distance to Mixedwood	Terrestrial Habitat with Variable Canopy Closure	(Kaufmann, 1992; Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019)	Euclidean Distance	Forest Inventory	Continuous
Other Habitat Features					
Temperature of Nearest Watercourse	Climate (Proxy)	(Greaves & Litzgus, 2007; COSEWIC, 2018)	Euclidean Allocation	NCC Stream Classification Layer	Categorical
Alkalinity of Nearest Watercourse	General Watercourse Characteristic	(COSEWIC, 2018)	Euclidean Allocation	NCC Stream Classification Layer	Categorical
Distance to Estuary	Hypothesized Avoidance of Brackish Water	(COSEWIC 2018)	Euclidean Distance	NCC Stream Classification Layer	Continuous
Distance to Chain Pickerel	Hypothesized Predation from Invasive Species	(Coastal Action, 2015; Loeza-Quintana et al., 2021	Euclidean Distance	GBIF Dataset & Freshwater Species Dataset	Continuous

All covariate layers were projected to NAD 1983 CSRS UTM Zone 20N, transformed into rasters with 250 m cells, and clipped to the extent of NS, excluding Sable Island. Before running Maxent with the entire initial suite of covariates, I computed a correlation matrix (Appendix C) for them using the ArcGIS Pro tool ‘Band Collection Statistics’. For absolute covariate pairs with correlation values above 0.7, I removed the covariate that was presumed to be less ecologically relevant (Kershaw et al., 2013; Gomez et al., 2014; Kurnaz & Şahin, 2021). Absolute correlation values of 0.4 – 0.7 prompted consideration but did not automatically warrant removal of either covariate. The only covariate pair with a correlation value above 0.7 was ‘Standard Deviation of Elevation - Standard Deviation of Slope’. I chose to omit the ‘Standard Deviation of Slope’ covariate as its extraction requires two processing steps (i.e., Elevation to Slope, then Slope to Standard Deviation of Slope), potentially making it less accurate compared to ‘Standard Deviation of Elevation’ which is extracted directly from ‘Elevation’. The covariates pairs ‘Distance to Hardwood - Distance to Mixedwood’, ‘Distance to Chain Pickerel - Temperature of Nearest Watercourse’, ‘Standard Deviation of Elevation - Gradient of Nearest Watercourse’, ‘Distance to Brush - Distance to Alder’, and ‘Distance to Watercourse - Watercourse Density’ all had correlation values above 0.4. Of these, ‘Distance to Hardwood - Distance to Mixedwood’ had the highest correlation value (>0.6) and I decided to omit the ‘Distance to Mixedwood’ covariate as hardwood stands have higher proportions of broad-leaved deciduous trees than mixedwood stands, and it is these trees which may provide *G. insculpta* open canopy basking sites during the cooler spring and dense canopy shade during the hotter summer. Additionally, mixedwood stands are prolific throughout NS and thus the less common hardwood stands may represent a more discriminatory variable. The correlation between the ‘Standard Deviation of Elevation’ covariate and the ‘Gradient of Nearest Watercourse’ covariate implied that they may be reflecting similar topographic phenomena, and so I chose to retain the continuous variable (‘Standard Deviation of Elevation’) over the categorical variable (‘Gradient of Nearest Watercourse’).

3.2.5 Modelling Procedure

Maxent can model the relationship between an environmental covariate and species presence using one or a combination of different mathematical approaches, referred to as feature types, including an untransformed continuous covariate (Linear), the square of a continuous covariate (Quadratic), the product of two continuous covariates (Product), and thresholds applied to a continuous covariate (Threshold and Hinge; Phillips et al., 2006). To reduce the risk of overfitting to the training data, a regularization multiplier can be applied to Maxent to force the software to limit the number of features used and theoretically produce a more parsimonious distribution (Phillips et al., 2006). To select the most optimal set of feature types and the best regularization multiplier, I used the R package 'ENMevaluate' (Muscarella et al., 2014; Kass et al., 2021; for script see Appendix A). I chose the combination of feature types and regularization multiplier that produced the lowest corrected Akaike Information Criterion (AICc) value. AICc is a statistical measure of how much information has been lost in the construction of a model, with lower AICc values indicating less information lost and thus a theoretically more complete model (Akaike, 1974).

As Maxent is a presence-only modelling approach, it uses pseudo-absences or background points meant to represent the average environmental variables in the study area and stand-in for true absence data (Phillips et al., 2004; Phillips et al., 2006). While rigorously-collected absence data are preferable, the performance of pseudo-absences can be improved if their selection is guided by the spatial bias in the presence data, rather than being scattered randomly throughout the study extent (Zaniewski et al., 2002; Phillips et al., 2006). To accomplish this, I used R to create a spatial bias file, a measure of the kernel density of the occurrence data, which the Maxent software can use to weight its placement of pseudo-absences (as recommended in a Maxent tutorial by J. Banta; for script see Appendix A; the resulting spatial bias file is visualized in Appendix D).

I ran 10 cross-validated model replicates, in that the software randomly split the occurrence data into training (90%) and test (10%) data 10 times and ran separate models for each randomly split set. For each replicate, Maxent uses the test data to assess the predictive power of the model by computing the area under the receiver operating characteristic curve (AUC) for both the training and test data. AUC represents the probability that a random positive instance (i.e., a presence) and a random negative instance (i.e., an absence or pseudo-absence) are correctly identified by the model and this value ranges from 0 to 1 (Phillips et al., 2006). For each Maxent run, the final model was the arithmetic average of these 10 replicates, with averaged Test AUC and Training AUC. I used a process of reverse stepwise elimination to narrow the initial list of covariates to a final suite which was most important for the modelling. This was accomplished by sequentially removing the covariate that had the lowest permutation importance (PI) in the last-run model and running a new model without it. The permutation importance of each covariate is derived from the final model, not the path that Maxent took to produce it, so is a superior measure of covariate importance compared to percent contribution (Phillips, 2017). I evaluated all the models produced and chose a final model based on Test AUC. However, as AUC has been shown to reward overfit models, I also assessed the 10th Percentile Training Presence Test Omission value and the difference between the Training and Test AUC values to evaluate the risk of overfitting (Lobo et al., 2008; Warren & Seifert, 2011; Radosavljevic & Anderson, 2014). Rates of 10th Percentile Training Presence Test Omission close to 0.1 and small differences between Training and Test AUCs both indicate low levels of overfitting (Warren & Seifert, 2011; Radosavljevic & Anderson, 2014).

3.2.6 SDM Application

Using the “maximum training sensitivity plus specificity threshold”, I created a binary raster that identifies the extent of NS that the model predicts to be the potential distribution of *G. insculpta* (Liu et al., 2013). I then intersected the potential distribution with the NS Protected Area System layer and the provincially identified *G. insculpta* core habitat layer to generate estimates of how much of the potential distribution may currently be protected or conserved in NS (Province of Nova Scotia, 2020). The NS Protected Area System layer represents the formal protected area network in NS including properties owned and managed by both government and NGOs, while the *G. insculpta* core habitat layer represents any land, including protected areas, non-protected Crown land, and private land, where core habitat for *G. insculpta* has been identified. I also estimated the percentage of the core habitat layer that falls within the potential distribution of *G. insculpta*, as predicted by the model.

The model output was shared with NS NRR to be used by them as a guide for field surveys in 2022. The goal of this field-survey season was to confirm *G. insculpta* presence at sites with no prior recorded occurrences. As such, NS NRR chose to select candidate survey sites using the cells of the model output that had relative predicted probabilities of occurrence over 0.99, and which were at least 2 km away from the nearest known *G. insculpta* occurrence. NS NRR's survey protocols for *G. insculpta* are informed by the work of Flanagan et al. (2013) and involve searching 1-5 km lengths of watercourse, with observers searching out to 20 m from the watercourse.

3.3 Results

3.3.1 SDM Performance and Covariates

The final model shows high averaged AUC (0.84 +/- 0.02) across its 10 replicates for modelling test data (Figure 3.2). Based on the recommendations of the R package 'ENMevaluate', all feature types were used in this model including Linear, Quadratic, Hinge, Product, and Threshold features (Kass et al., 2021). For the regularization multiplier, the default value of 1 was used. The seven covariates used in the model in order of permutation importance (Table 3.3) include 'Elevation' (50%), 'Distance to Alder' (22.3%), 'Distance to Estuary' (11%), 'Watercourse Density' (7.8%), 'Distance to Hardwood' (3.9%), 'Standard Deviation of Elevation' (2.8%), and 'Size of Nearest Watercourse' (2.2%). The relationships between the 'Elevation', 'Distance to Alder', and 'Distance to Hardwood' covariates and the predicted probability of occurrence were negative, indicating higher probabilities of occurrence at lower elevations and closer to alder and hardwood stands (Figure 3.3). In contrast, the 'Distance to Estuary', 'Watercourse Density', and 'Standard Deviation of Elevation' covariates showed positive relationships with higher probabilities of occurrence being predicted farther from estuaries, in areas with higher watercourse densities and greater elevation variability. Higher probabilities of occurrence were predicted near 'Small Rivers', followed by 'Medium Rivers', and lastly 'Headwaters and Creeks'. The difference between the averaged Test AUC and Training AUC values was 0.01 and the 10th Percentile Training Presence Test Omission rate was 0.12, both indicating low levels of overfitting.

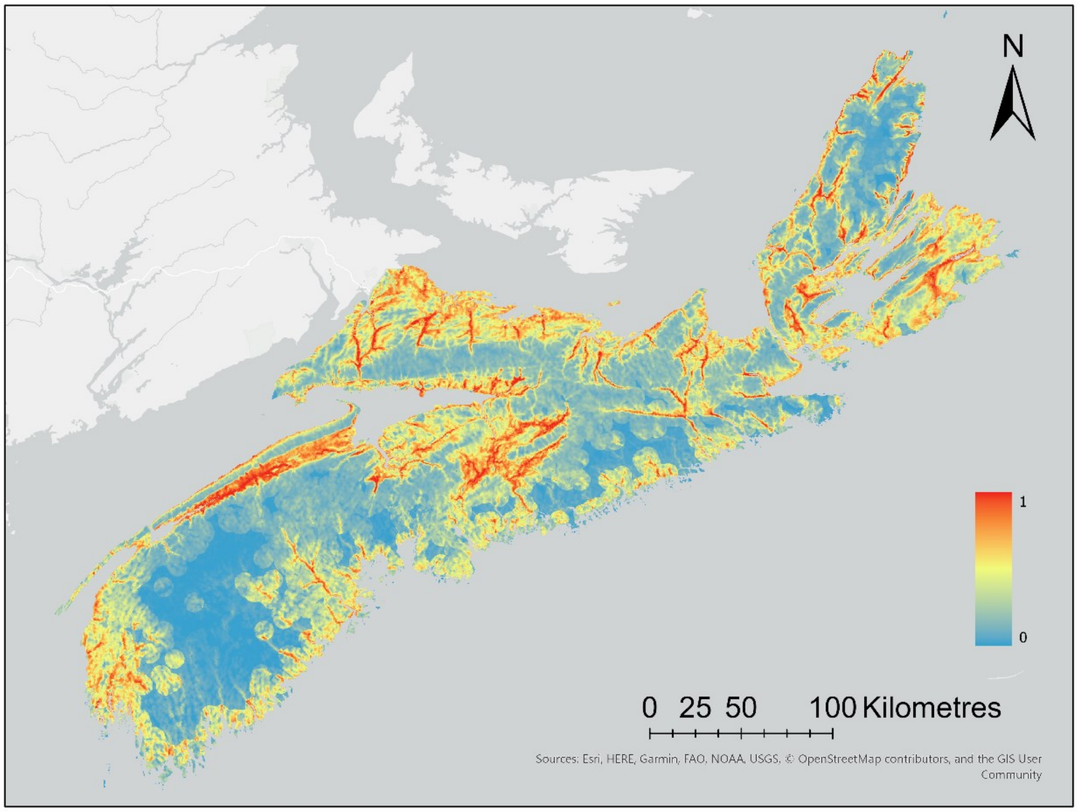
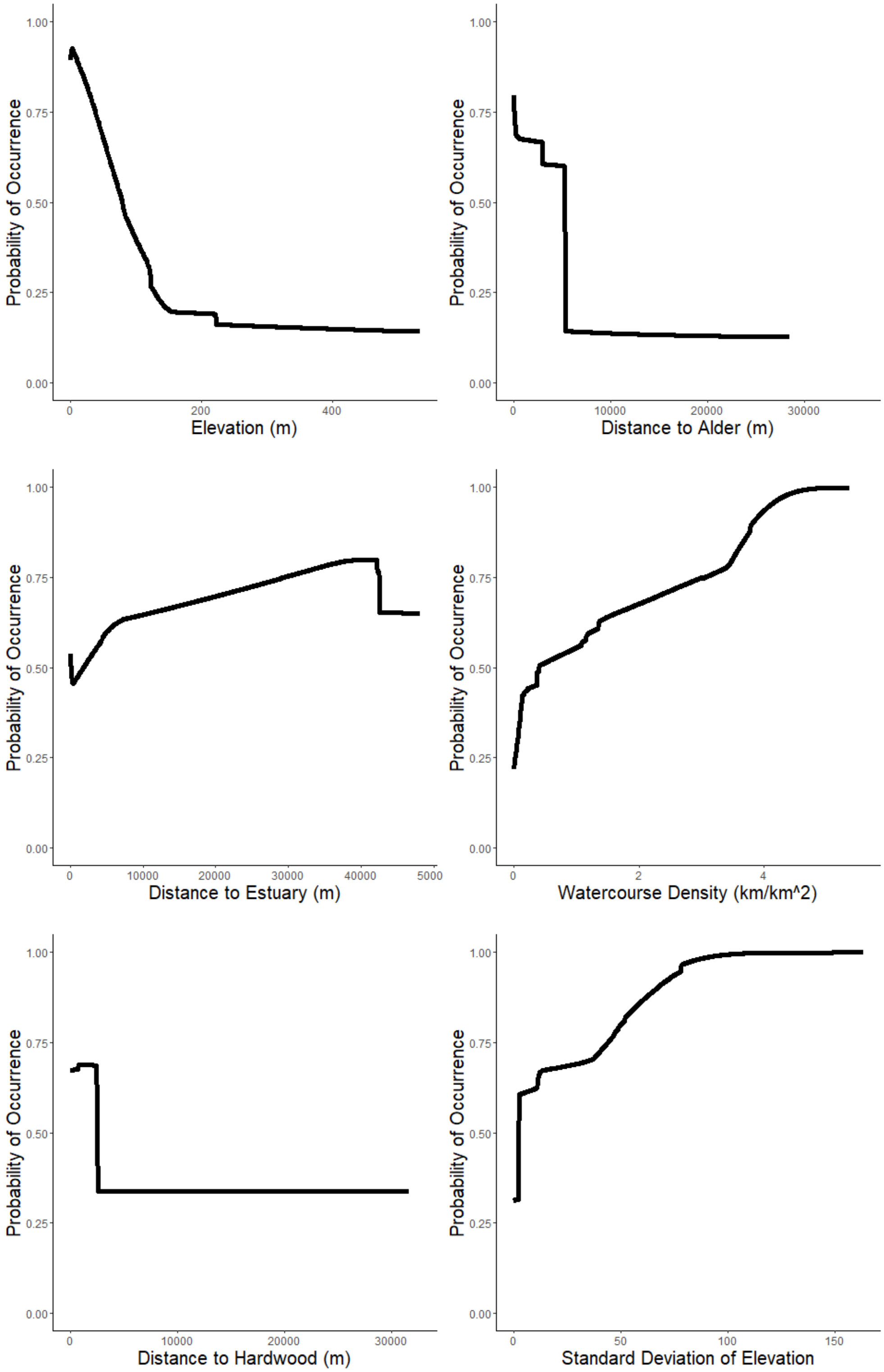


Figure 3.2 Continuous output of predicted relative probability of occurrence of *G. insculpta* in Nova Scotia, Canada, produced by averaging the results of 10 Maxent replicates. The predicted relative probabilities of occurrence range from 0 (representing the lowest probability and displayed in blue) to 1 (representing the highest probability and displayed in red). The averaged test AUC of the 10 replicates is 0.84 +/- 0.02. The difference between the averaged test AUC and training AUC values is 0.01 and the 10th percentile training presence test omission rate is 0.12.

Table 3.3 Average permutation importance and percent contribution to model prediction, across 10 cross-validated Maxent replicates predicting probability of occurrence of *G. insculpta* in Nova Scotia, Canada

Covariate	**Average Permutation Importance**	**Average Percent Contribution**
Elevation	50.0	57.4
Distance to Alder	22.3	8.5
Distance to Estuary	11.0	2.7
Watercourse Density	7.8	10.7
Distance to Hardwood	3.9	0.7
Standard Deviation of Elevation	2.8	1.7
Size of Nearest Watercourse	2.2	18.3


Probability of Occurrence
1.00
0.75
0.50
0.25
0.00
0
200
400
Elevation (m)
Probability of Occurrence
1.00
0.75
0.50
0.25
0.00
0
10000
20000
30000
Distance to Alder (m)
Probability of Occurrence
1.00
0.75
0.50
0.25
0.00
0
10000
20000
30000
40000
5000
Distance to Estuary (m)
Probability of Occurrence
1.00
0.75
0.50
0.25
0.00
0
2
4
Watercourse Density (km/km^2)
Probability of Occurrence
1.00
0.75
0.50
0.25
0.00
0
10000
20000
30000
Distance to Hardwood (m)
Probability of Occurrence
1.00
0.75
0.50
0.25
0.00
0
50
100
150
Standard Deviation of Elevation

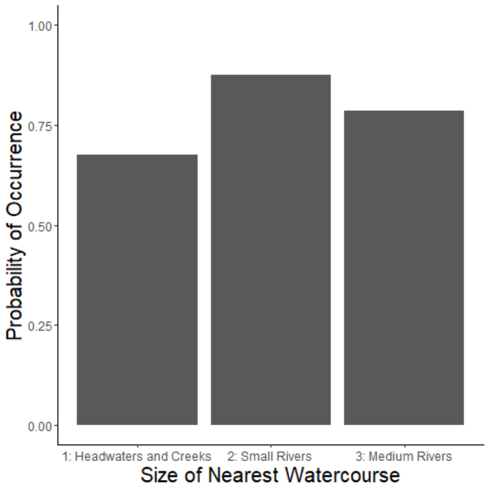


Figure 3.3 Species-response curves indicating the relationship between the predicted probability of occurrence of *G. insculpta* and environmental covariates in the Maxent model for Nova Scotia, Canada

For each environmental covariate used in a model run, Maxent computes how the model would perform if built using only that covariate, as well as if all covariates except that covariate were used. Of the seven environmental covariates retained in the final model, the covariate that produced the highest AUC when run by itself was 'Distance to Alder', with an AUC of 0.74 (Table 3.4). This was followed by 'Watercourse Density' at 0.73 and 'Elevation' at 0.72. In contrast, while not having the highest AUC by itself, the removal of 'Elevation' resulted in the greatest reduction in AUC compared to the full model (Table 3.4).

Table 3.4 Area under the receiver operating characteristic curve (AUC) values for 1) models built with only one covariate; and 2) models built using the final suite of covariates minus one covariate.

Covariate	**AUC of model with only this covariate**	**AUC of full model without this covariate**
Distance to Alder	0.74	0.83
Watercourse Density	0.73	0.83
Elevation	0.72	0.80
Size of Nearest Watercourse	0.64	0.84
Distance to Estuary	0.59	0.82
Distance to Hardwood	0.54	0.84
Standard Deviation of Elevation	0.46	0.84

3.3.2 SDM Application

Maxent determined the “maximum training sensitivity plus specificity threshold” for this model to be 0.627. After applying this threshold to the continuous model output, about 15.1% of NS was identified as the potential distribution of *G. insculpta* (Figure 3.4). Of this distribution, approximately 4.6% occurred within existing protected area boundaries as delineated in the NS Protected Area System layer and about 2.3% occurred within the area identified in the *G. insculpta* core habitat layer. Together, these two layers covered 6.8% of the identified potential distribution, as there is slight overlap between them. Of the area identified as core habitat by NS NRR, 80.4% fell within the potential distribution of *G. insculpta* predicted by the model. Using the model output, NS NRR identified 26 candidate survey sites, of which their staff surveyed 21 between May 3rd to June 7th, 2022. New *G. insculpta* presences were recorded at five of these sites, meaning that 24% of sites surveyed returned occurrence data for this species (L. Doucette, pers. comm., June 2022). On one particular watercourse, NS NRR staff found seven living *G. insculpta*, including two males, one female, and four juveniles.

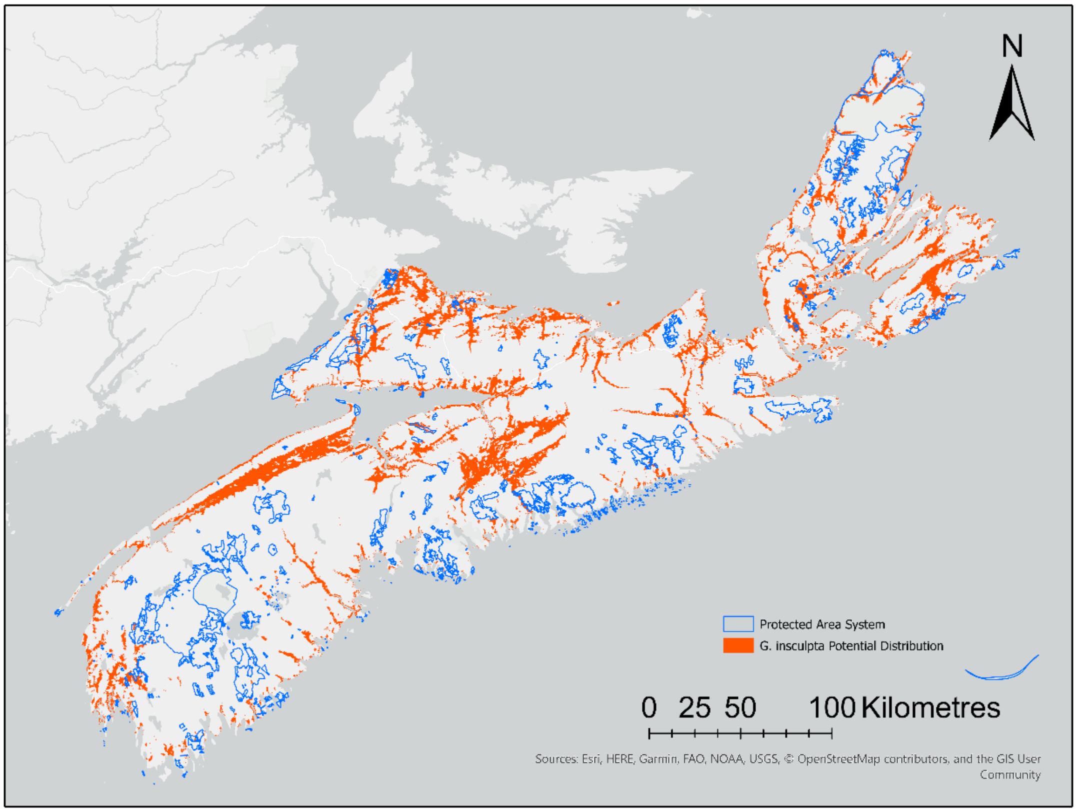


Figure 3.4 Potential distribution of *G. insculpta* (in orange) as predicted by Maxent modelling, overlain with the Nova Scotia protected areas system (in blue). The potential distribution was produced by applying the "maximum training sensitivity plus specificity threshold" of 0.627 to the continuous model output.

3.4 Discussion

3.4.1 SDM Performance and Covariates

As the first species distribution modelling for *G. insculpta* in NS, the model represents a fundamental step forward for conservation and management of this species in Atlantic Canada. The final model appeared to have strong predictive power with a high averaged Test AUC. There also appeared to be little justification for concern about overfitting, as the difference in the averaged Training and averaged Test AUC values was small and the 10th Percentile Training Presence Test Omission rate was close to 0.1 (Warren & Seifert, 2011; Radosavljevic & Anderson, 2014). Additionally, the success of the 2022 field-survey season conducted by NS NRR showed the usefulness of the model to a direct

G. insculpta conservation application. Nonetheless, I recommend that this model be further field-validated by surveys adequate to determine both presence and absence of *G. insculpta*.

The model indicated that 'Elevation' was the strongest predictor for *G. insculpta* presence in NS with a permutation importance of 50% (Table 3.3). The importance of elevation is in agreement with the modelling performed by Mothes et al. (2020), in which elevation was the third most important environmental covariate, after January minimum temperature, and canopy closure. The strength and negative relationship of this predictor covariate likely reflects where the watercourses in the province reach their largest sizes, slowest velocities, and highest sinuosities, which are typically within areas of lower elevation (Figure 3.3; Hohensinner et al., 2018). A simple overlay in ArcGIS Pro between the NS Digital Elevation Model and the watercourses identified in the NCC Stream Classification layer shows that while the average elevation of watercourses of all size types (Creeks, Headwaters, and Rivers) is close to the average elevation of NS, 111.6 m overall vs 94.5 m for watercourses, the average elevation of only the 'Rivers' of NS is much lower at 43.3 m. Lower elevations may also correlate with more sinuous watercourses, as an overlay between the NS Digital Elevation Model and the NS Geological Map reveals the lowest average elevation in the areas of NS with sedimentary bedrock (79.3 m), versus igneous (161.6 m) or metamorphic bedrock (104.7 m); and the softer stone in the sedimentary bedrock of NS has likely allowed for the development of more sinuous watercourses (NS Department of Natural Resources and Renewables, 2006, B; Hohensinner et al., 2018). These findings warrant further investigation into the hydrology of NS and support the testing of the inclusion of bedrock type in future *G. insculpta* SDMs.

The next most important environmental covariate in the final model was 'Distance to Alder'. This represents a previously un-tested covariate for *G. insculpta* modelling as neither Mothes et al. (2020) nor Willey et al. (2022) incorporated it into their models, despite the presence of common alder species such as *Alnus incana* throughout the range of *G. insculpta* in the northeastern United States. My model showed a strong relationship between the presence of *G. insculpta* occurrences and the presence of alder stands, as the highest predicted probabilities of *G. insculpta* occurrence were found in cells that also contain alder stands: the distance to alder for the cell is 0 m (Figure 3.3). Alder stands provide areas of variable canopy closure in close proximity to watercourses and offer abundant terrestrial food sources for *G. insculpta* including slugs, earthworms, and raspberries (Compton et al., 2002; Arvisais et al., 2004). They have been shown to be preferentially selected by *G. insculpta* over adjacent forest habitat, especially during the pre-nesting (late April – early June) and pre-hibernation (late September – early November) periods of their active season (Kaufmann, 1992; Arvisais et al., 2004). Further support for including this covariate came from experts who have years of experience researching *G. insculpta* within NS and who have noted that they have often found *G. insculpta* in association with alder species (M. Parker, pers. comm., May 2020; M. Pulsifer, pers. comm., June 2020). Additionally, as alder species often grow in wet areas including along slow-flowing watercourses in NS, this covariate may not only represent direct *G. insculpta* habitat selection, but also may serve as a proxy for watercourse habitat selected by both *G. insculpta* and alder species (Munro et al., 2014). The ability of the 'Distance to Alder' covariate to account for multiple *G. insculpta* habitat characteristics is further supported by it being the covariate that produced the highest AUC of any of the models built using only one covariate (Table 3.4). Evidently, the 'Distance to Alder' covariate or, more simply, the presence of alder at a site, appears to be a thoroughly useful predictor of *G. insculpta* habitat in NS. This relationship may not be as strong in other parts of the species' range, where other shrub species, such as *Cornus amomum* in the northeastern United States, dominate riparian areas (J. Berkholtz, pers. comm., April 2022).

Another strong covariate-occurrence relationship was that of the 'Distance to Estuary' covariate, which had the third highest permutation importance in the final model. The relationship was positive with higher predicted probabilities of occurrence farther from estuaries (Figure 3.3). This reflects expert opinion in NS which has traditionally believed that *G. insculpta*, as a freshwater species, generally avoid brackish water (M. Pulsifer, pers. comm., 2020). However, it should be noted that because of this traditional opinion, historical surveying for *G. insculpta* in NS may have neglected estuaries, which could be a partial explanation for the lack of recorded occurrences in these areas. Nevertheless, the strength of the effect of this covariate is interesting and may support the traditional viewpoint.

Considering that *G. insculpta* is a riparian species, it was also surprising that a general measure of watercourse presence, ‘Watercourse Density’, scored lower than the preceding three covariates in the final model (COSEWIC, 2018). Perhaps, this was a result of the ‘Elevation’ and ‘Distance to Alder’ covariates acting as proxies for the presence of suitably sinuous and slow-flowing watercourses. Regardless, I tried using two different measures of watercourse presence, a simple ‘Distance to Watercourse’ covariate and the ‘Watercourse Density’ covariate, with the former showing a negative relationship, i.e., lower predicted probabilities as distance increases, and the latter showing a positive one, i.e., higher predicted probabilities as density increases. Of the two, ‘Watercourse Density’ scored higher in terms of permutation importance and was retained in the final model while ‘Distance to Watercourse’ was not. This may be because watercourse density captures not only the availability of watercourse habitat but also some of the watercourse features that *G. insculpta* prefer. A slow-moving, meandering watercourse with high sinuosity and low to moderate gradient will likely produce more metres of watercourse within a raster cell than a fast-flowing and straight watercourse, the former being the kind of watercourse that *G. insculpta* find more suitable (Jones & Willey, 2015; COSEWIC, 2018). A higher watercourse density may also indicate the presence of more tributaries feeding into a river. This may also represent more suitable habitat, as recent radio-tracking research has suggested that *G. insculpta* in NS extensively use the small tributaries that flow into their larger rivers (M. Parker, pers. comm., June 2022). The strength of the ‘Watercourse Density’ covariate indicates that it could potentially be useful for modelling the distributions of other semiaquatic obligate-riparian species, over a simpler ‘Distance to Watercourse’ covariate.

In sum, this model predicted *G. insculpta* distribution in low-elevation areas of NS with high watercourse densities and the presence of alder species, with less suitable habitat near estuaries, where these watercourses meet the ocean. This differs somewhat from the previous *G. insculpta* modelling studies, as Willey et al. (2022) found that the highest predicted probabilities of *G. insculpta* occurrence were preferentially located in watercourses with higher flow accumulation and lower gradient, and which were in areas with warmer July mean temperatures, while Mothes et al. (2020) found the highest predicted probabilities in raster cells that had January minimum temperatures less than about -6 °C, with canopy closure percentages of greater than 80%, and which were located between 200 – 300 m above sea level. These differences are likely in part reflective of the different covariates chosen. For example, I chose to omit temperature variables as, apart from changes in elevation, temperatures are relatively consistent across NS and *G. insculpta* are known to occur across the province (Figure 3.1; Ecological Stratification Working Group, 1995). Despite the differences in covariate selection, my modelling retained the importance of favourable watercourse characteristics emphasized in the modelling of Willey et al. (2022) through the direct variables of 'Watercourse Density' and 'Size of Nearest Watercourse' (correlated with flow accumulation) and the proxy variables of 'Elevation' and 'Distance to Alder'. Additionally, the importance of terrestrial habitat as represented by canopy closure in Mothes et al. (2020) was reflected in my modelling by the variables 'Distance to Alder' and 'Distance to Hardwood', both habitat types that provide variable deciduous canopy closure. While I cannot directly compare the efficacy of my modelling versus that of previous studies, the emergence of both watercourse habitat-based covariates and terrestrial habitat-based covariates in my modelling appear to support the idea that both are important when modelling semiaquatic obligate-riparian species. However, further refinement of my model is likely warranted as it currently predicts the potential distribution of *G. insculpta* as including isolated coastal areas that may not represent real *G. insculpta* habitat, based on our current understanding of its distribution in NS. Before applying survey effort to these coastal areas, I propose that testing a 'Distance to Coast' covariate may be a useful addition to the model.

3.4.2 SDM Application

COSEWIC (2018) estimates that less than 11% of *G. insculpta* habitat in Canada is protected within national parks, provincial parks, and the private conservation properties of non-governmental organizations (NGOs), primarily land trusts such as the Nature Conservancy of Canada, NS Nature Trust, and Sespite'tmnej Kmitkinu Conservancy. When overlain with the NS Protected Area System layer, only about 4.6% of the potential distribution of *G. insculpta* in NS is formally protected within some type of government or NGO-owned protected area, half that of the national estimate. Protections for at-risk species in Canada extend beyond traditional protected areas, however, and to wherever habitat for an at-risk species has been identified (*Species at Risk Act*, 2002). When overlain with the *G. insculpta* core habitat layer for NS, only 2.3% of the potential distribution in NS has been identified as core habitat. Together, the NS Protected Area System layer and the core habitat layer cover only 6.8% of the predicted potential distribution of *G. insculpta*. However, it is important to note that core habitat in NS is not meant to represent all areas inhabited by an at-risk species, but rather only the "specific areas of habitat essential for the long-term survival and recovery" (*Endangered Species Act*, 1998) of the species. Therefore, the more important finding may be that the model's prediction of the potential distribution of *G. insculpta* overlaps with 80.4% of the currently identified core habitat. This level of agreement suggests that the model is sensitive enough to capture most of the important *G. insculpta* habitat in NS. The discrepancy of 19.6% of the core habitat not being predicted as potential distribution could be indicative of error in the model, but likely moreso reflects possible instances of core habitat having been extended too broadly out from known occurrences into adjacent unsuitable habitat. Further analysis of where these two layers do and do not overlap should be conducted to better understand the reason for this misalignment.

One barrier to the protection and restoration of *G. insculpta* habitat in NS may be its overlap with the agriculturally rich parts of the province, much of which are held as private land. The two largest contiguous areas of potential distribution that I have identified using my model were: 1) within the Musquodoboit, Shubenacadie, and Stewiacke River watersheds (Musquodoboit/Shubenacadie/Stewiacke Complex); and 2) in the Annapolis Valley, within the Annapolis and Jijuktu'kwejk watersheds (Annapolis/Jijuktu'kwejk Complex). These watersheds represent some of the most agriculturally developed and important parts of NS and contain relatively few protected areas compared to other parts of the province (Province of Nova Scotia, 2020; The Canadian Encyclopedia, 2021). This presents a challenge and further emphasizes the need for engagement with farmers and other rural landowners, such as with on-going programs like Wood Turtle Strides (NS Federation of Agriculture; Sherren et al., 2020). Agriculture, and principally hay harvesting, poses one of the most significant risks to *G. insculpta* in Canada and modifications to hay harvesting times and mowing protocols can greatly reduce the likelihood of individual mortality (Saumure et al., 2007; Saumure & Bider, 1998; Wallace et al., 2020)

3.4.3 Future Steps

Further *G. insculpta* modelling in NS would benefit from the following recommendations. First, as previously mentioned, the most feasible improvement to my model would be to test the inclusion of a 'Distance to Coast' covariate to determine whether it would reduce the prediction of potential distribution of *G. insculpta* in seemingly inappropriate coastal areas with steep slopes. Some identified areas, such as Pictou Island and the coastal side of North Mountain, seem unlikely to contain *G. insculpta* habitat; and, thus, the testing of this covariate prior to conducting field surveys at those locations would represent an efficient use of resources. The NCC Stream Classification layer, from which I derived the 'Distance to Estuary' covariate, does not classify every watercourse that meets the ocean as an estuary and is subsequently less general than a 'Distance to Coast' covariate would be (Atlantic Science, 2019). As a 'Distance to Coast' covariate would likely be correlated with the 'Distance to Estuary' covariate, its addition may have the effect of reducing the importance of the latter covariate in the model. Second, beyond this small addition, an important goal of further modelling in NS would be to better integrate favourable watercourse characteristics such as sinuosity and flow rate into the model, ideally as directly measured or modelled variables. In the absence of these watercourse variables, geologic and topographic covariates such as bedrock type, slope, and topographic position index could work as additional proxies for them. Third, some of the covariates that I used are produced using adjustable parameters such as 'Watercourse Density', for which I chose to calculate the density within a circle with a radius of 1 km. Sensitivity analysis using 'Watercourse Density' rasters produced using different search areas could be useful in fine-tuning the best suite of covariates for predicting *G. insculpta* occurrence in NS. Finally, further modelling in NS should also investigate the effects of explicitly anthropogenic-influenced features such as land use, impervious surfaces, roads, and agricultural areas to classify the predicted potential distribution into that which still may provide habitat for *G. insculpta* (i.e., opportunities for protection) and that which has been likely compromised by development (i.e., opportunities for restoration).

The field surveys conducted in 2022 provided a strong proof-of-concept for applying SDM to *G. insculpta* survey prioritization in NS, with almost a quarter of sites surveyed returning new occurrence data. Besides further modelling, more survey efforts informed by these modelling results, ideally at levels sufficient to determine both presence and absence, should be pursued to elucidate the full distribution of *G. insculpta* in NS and guide model refinement. In turn, the discoveries of new *G. insculpta* occurrences at previously unconsidered sites such as those found during the 2022 surveys should be used to guide conservation management actions including the identification of habitat, the enforcement of special management practices, and the prioritization of formal land protection.

3.4.4 Conclusion

The species distribution model has strong predictive power, an acceptable level of overfitting, and presents a unique suite of environmental covariates which incorporates both watercourse habitat and terrestrial habitat features that are relevant to *G. insculpta*. Elevation appears to serve as a proxy for those favourable watercourse characteristics that produce *G. insculpta* habitat while the presence of alder stands as modelled using 'Distance to Alder' represents a newly identified covariate that may have potential for capturing multiple *G. insculpta* habitat features in future modelling efforts within the areas of this species' range where alder species dominate riparian areas. My species distribution model for *G. insculpta* provides a guide to conservation managers in NS for directing survey effort. The successful field surveys conducted in 2022 have expanded the understanding of *G. insculpta* distribution in NS and have shown the value of coproducing SDMs for at-risk species with the branches of government that can directly apply the resulting model outputs. Further survey effort will be crucial to locate more *G. insculpta* habitat in NS to ensure its protection under legislation and to possibly contribute to the identification of additional core habitat. As the largest contiguous areas of predicted suitable habitat fall within important agricultural areas of the province, further conservation efforts for this species likely need to consider broader engagement with farmers and other rural landowners to implement best practices for reducing *G. insculpta* mortality.

CHAPTER 4: MODEL COMPARISONS

For this thesis, I developed three SDMs, two for NS—preliminary (2021) and final (2022)—and one for NB. The NB model and the preliminary NS model were developed with similar initial covariate sets, with some differences due to data availability. The final NS model was developed with an expanded suite of covariates, based on reflections from the preliminary modelling and field surveys. The covariate sets retained in these models shared some similarities but also differed in fundamental ways. As described in Chapters 2 and 3, the environmental covariates included in these Maxent models were chosen principally to reflect three habitat features important to *G. insculpta* and to best estimate this species' niche using available spatial data. To recap, these features were the presence of: 1) watercourse habitat (Greaves & Litzgus, 2007; Jones & Willey, 2015; COSEWIC, 2018); 2) favourable watercourse characteristics including higher sinuosity, lower gradient and flowrate, and higher flow accumulation (Gilhen, 1984; Greaves & Litzgus, 2007; COSEWIC, 2018); and 3) terrestrial habitat with either spatially or temporally variable canopy closure (Kaufmann, 1992; Compton et al., 2002; Dubois et al., 2009; Hughes & Litzgus, 2019). The NS modelling in both 2021 and 2022 appeared to capture these three habitat features, while the NB modelling did not retain any strong measure of terrestrial habitat with variable canopy closure.

4.1 Watercourse Habitat with Favourable Characteristics

The primary covariate accounting for the presence of watercourse habitat in the 2021 modelling for both NS and NB was 'Distance to Watercourse', a simple variable that directly measured how far each cell was from the nearest watercourse. This variable was the second most important covariate in the NB modelling (PI = 26.5%), but only the third most important in the 2021 NS modelling with half the permutation importance of its NB counterpart (PI = 11.3%). This difference between the NS and NB models is likely the result of the strength of the 'Distance to Alder' covariate in the 2021 NS model (PI = 31.8%), compared to its counterpart in the NB model (PI = 0.1%). As alder species often grow in wet areas, including along watercourses in NS and NB, the 'Distance to Alder' covariate may act as a proxy for the presence of watercourses in many parts of NS, lessening the importance of the ‘Distance to Watercourse’ covariate (Munro et al., 2014). It is unclear exactly why the ‘Distance to Alder’ covariate was so strong in the NS model while its NB counterpart was so weak. As NS and NB are both within the same Ecozone with broadly similar climates and ecosystem types, this difference likely reflects methodological differences in how the provinces identify alder stands in their respective forest inventories, rather than any biological difference in habitat selection between *G. insculpta* in NS and NB (Ecological Stratification Working Group, 1995). This serves as an example of how important the quality of available geospatial data is to species distribution modelling.

The ‘Distance to Alder’ covariate was retained as an important covariate in the 2022 NS modelling (PI = 22.3%), but the ‘Distance to Watercourse’ covariate was not. The principal reason for this may have been the inclusion of the ‘Watercourse Density’ covariate in the 2022 modelling, which was retained as the fourth-most important covariate (PI = 7.8%). Like the ‘Distance to Watercourse’ covariate, ‘Watercourse Density’ was extracted directly from the NS hydrographic network and thus served as a direct measure of the presence of watercourse habitat, albeit a more complex one (Service Nova Scotia and Internal Services, 2020). However, the greater complexity of this covariate could be why it was retained over ‘Distance to Watercourse’. As described in Chapter 3, the ‘Watercourse Density’ covariate may reflect aspects of the second identified *G. insculpta* habitat feature, that being favourable watercourse characteristics, due to sinuous lower-velocity rivers producing higher line densities than straight ones.

Despite the ‘Watercourse Density’ covariate likely capturing some of the watercourse characteristics favourable to *G. insculpta*¸ the covariate that probably reflected this habitat feature the most was ‘Elevation’, which was consistently the most important covariate in the modelling (NS 2021 PI = 48%; NB PI = 65.7%; NS 2022 PI = 50%). The tendency for the models to predict higher probabilities of *G. insculpta* occurrence at lower elevations corresponds with the fact that watercourses tend to become larger, slower, and more sinuous as they decrease in elevation (Hohensinner et al., 2018). The ‘Distance to Alder’ covariate may also reflect watercourses with lower gradients and flowrates and thus serve as a proxy for them, as riparian vegetation including alders is likely more prevalent along lower-velocity watercourses in Atlantic Canada. The absence of a strong 'Distance to Alder' covariate in the NB model may partially explain why 'Elevation' had such a large permutation importance (PI = 65.7%), the highest PI of any covariate in either the NS or NB models. However, the relatively stronger PI of 'Elevation' in the NB model may also reflect the larger proportion of NB that is covered by highlands and the higher maximum altitudes reached compared to within NS (Ecological Stratification Working Group, 1995). This subsequently leads to the greater prevalence of more boreal ecosystems in NB, which may be climatically unsuitable for *G. insculpta*. This idea is further supported by how the NB model produced using only 'Elevation' had a strong AUC of 0.84 (Table 2.6), on par with the full 2022 NS model and only 0.05 below the AUC of the full NB model (0.89). Clearly, 'Elevation' is a crucial environmental variable for predicting *G. insculpta* occurrence in NB and is seemingly more important than in NS where elevation may be less of a limiting factor.

Another covariate related to favourable watercourse characteristics, 'Size of Nearest Watercourse' was retained in all the models, consistently favouring rivers (categorized as 'Small Rivers' and 'Medium Rivers') over the smaller 'Headwaters and Creeks'. The 'Large Rivers' category, only found in NB and consisting solely of the St. John River and the tidal portions of the Miramichi and Restigouche Rivers, was ranked lower, more akin to 'Headwaters and Creeks' than the other river categories. Both the 'Distance to Alder' and 'Size of Nearest Watercourse' covariates were less important in the 2022 NS model than in the 2021 NS model (Alder PI: 31.8 to 22.3; Watercourse Size PI: 6.8 to 2.2). Perhaps, the inclusion of the 'Watercourse Density' covariate contributed to the lessening importance of the former covariates, due to its contribution to modelling those watercourse characteristics favourable to *G. insculpta*. Regardless, while 'Elevation', 'Distance to Alder', and 'Watercourse Density', appear to all serve as proxies for watercourse characteristics including sinuosity and flowrate, integrating these variables directly will be an important goal for future *G. insculpta* modelling in Atlantic Canada. In the absence of these direct watercourse variables, geologic and topographic covariates including bedrock type, slope, and topographic position index could be explored as alternative proxies.

4.2 Terrestrial Habitat with Variable Canopy Closure

The third *G. insculpta* habitat feature that I identified, the presence of terrestrial habitat with variable canopy closure, was partially accounted for in the 2021 NS model by the covariates 'Distance to Alder' and 'Distance to Brush'. Past research into habitat selection has shown that *G. insculpta* will preferentially use alder stands over adjacent terrestrial habitat, and local experts in Atlantic Canada have corroborated this association (Kaufmann, 1992; Arvisais et al., 2004; M. Parker, pers. comm., May 2020; M. Pulsifer, pers. comm., May 2020). The 'Distance to Brush' covariate on the other hand represents more general scrub-shrub habitat, which is also used by *G. insculpta* (Kaufmann, 1992; Compton et al., 2002; Tingley et al., 2009). For NB, I did not find an equivalent measure of scrub-shrub habitat in the NB forestry data and the 'Distance to Alder' covariate did not have a strong impact in the model. As a result, the NB model does not appear to account for the presence of terrestrial habitat with variable canopy closure. Further research should be conducted to determine how terrestrial habitat features may be integrated into *G. insculpta* modelling in NB in the future.

In 2022, I included an additional terrestrial habitat covariate, 'Distance to Hardwood', in the NS modelling to capture the seasonally variable canopy closure provided by deciduous broadleaf trees which may provide *G. insculpta* with sun in the cooler spring and shade in the hot summer. This new covariate was determined to be important in the 2022 model, while the previously included 'Distance to Brush' covariate was not. This indicates perhaps that hardwood stands are a more useful terrestrial habitat type than non-alder scrub-shrub areas are for predicting *G. insculpta* occurrence in NS, at least based on how they are identified in the NS forest inventory (NS Department of Natural Resources and Renewables, 2021). While it is true that *G. insculpta* spend more of their active time in non-forested areas as opposed to forested areas, they have been shown to select for watersheds with moderate forest cover and to use forest edges, leading to them being described as 'edge species' (Kaufmann, 1992; Compton et al., 2002; COSEWIC, 2018). Future *G. insculpta* modelling should explore the effects of modelling forest edges, the mosaic landscapes that create an abundance of them, or some other measure of vegetation heterogeneity. I attempted to include a variable like this in the 2021 NS modelling, with the 'Standard Deviation of Crown Closure' covariate, but I do not believe my method of extraction was suitable given the quality of forestry data available. This was possibly borne out by this covariate not being retained within the 2021 NS model.

4.3 Visual Comparison Between 2021 and 2022 NS Models

As described above, there were clear differences between the final covariate suites used in the 2021 and 2022 NS models. To visually evaluate how these covariate differences translated into different spatial predictions of *G. insculpta* occurrence, I used the Raster Calculator within ArcGIS Pro to subtract the 2021 model output from the 2022 model output (Figure 4.1). The largest positive and negative changes between the two models were 0.87 and -0.80 respectively, meaning that the 2022 model predicted high probabilities of occurrence in some areas where the 2021 model had predicted low probabilities of occurrence and vice versa. The areas with the greatest decreases in predicted probability of occurrence between the models, as depicted in dark red, appear to predominantly be along the Atlantic coast of NS. Smaller decreases, in lighter red, can be seen across many of the other coastal areas of the province. It is likely that the inclusion of the 'Distance to Estuary' covariate in the 2022 model was the primary reason for the decrease in predicted probabilities along the coastlines.

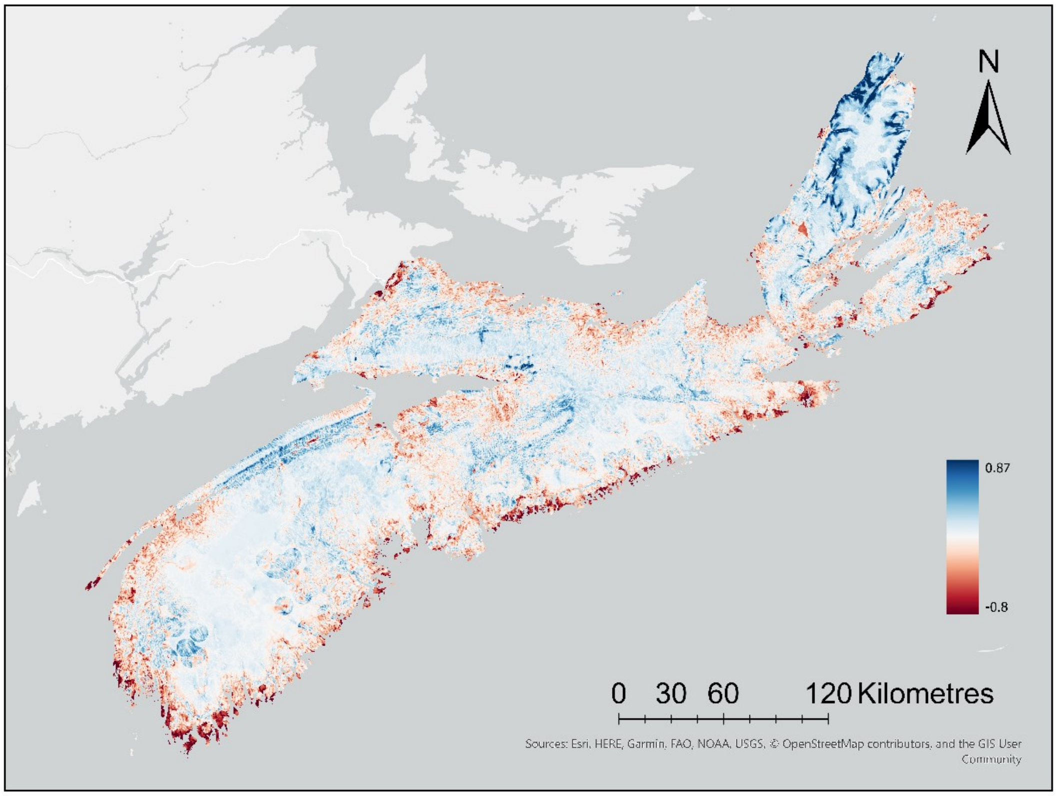
Accordingly, many of the areas with increases in predicted probability of occurrence between the models, shown in varying shades of blue, are located inland and primarily in river valleys such as the Annapolis River and St. Mary's River valleys. However, the greatest increases, depicted in dark blue, are concentrated in the river valleys of the Cape Breton Highlands. It is unclear exactly why these areas are so much more highly predicted by the 2022 model than by the 2021 model. A visual examination of the 'Watercourse Density' covariate raster shows relatively high-density values in the river valleys of the Cape Breton Highlands. Perhaps, the topography of the Cape Breton Highlands, characterized by high plateaus and steep river valleys, produces more small creeks than other parts of NS, leading to higher calculated watercourse densities and thus higher predicted probabilities in the 2022 model. If this is true and the higher watercourse densities in these areas are driven by a high frequency of steep creeks that likely do not provide suitable *G. insculpta* habitat, then this result may support the testing of a 'River Density' covariate in which the smallest watercourses (i.e., creeks and headwaters) do not contribute to the density measure. Alternatively, the addition of a ‘Slope’ covariate may be useful for reducing the predicted probabilities of occurrence in these steep river valleys.

As the above observations were based solely on a visual comparison, further analysis of the differences between the two model outputs could be conducted if desired. Possible additional analyses could include; 1) testing the statistical significance of the difference between the continuous rasters; and 2) producing a binary version of the 2021 model output using the “maximum training sensitivity plus specificity threshold” and calculating the percent overlap between it and its 2022 counterpart.

4.4 Conclusion

The 2022 NS model appeared to capture some of the complexity of *G. insculpta* habitat selection in ways that the 2021 NS model and the NB model did not, while producing a model output with notable differences compared to the former NS model. It had a seemingly richer set of covariates than the 2021 NS model, with only the modelled effect of non-alder scrub-shrub habitat being lost which was a relatively weak relationship in the preliminary model (PI = 2.1%). In contrast, the NB model appeared to be relatively simple as it only contained three covariates that contributed strongly to its prediction, though this may reflect reality, in that ‘Elevation’ may limit *G. insculpta* occurrence more greatly in NB than in NS. Regardless of the various models’ relative effectiveness, additional covariates should be explored and included in future modelling for both NB and NS, including the sinuosity and flow accumulation of watercourses, additional geologic and topographic covariates (bedrock type, slope, topographic position index, etc.), covariates that capture variability in canopy closure such as forest edges, mosaic landscapes, and vegetation heterogeneity, and potentially a ‘River Density’ covariate.

4.5 Figures



Sources: Esri, HERE, Garmin, FAO, NOAA, USGS, © OpenStreetMap contributors, and the GIS User Community

Figure 4.1 Difference between the 2021 and 2022 NS models of *G. insculpta* probability of occurrence, produced by subtracting the 2021 model output from the 2022 model output. The symbology applied to the raster is a stretch using the 'Percent Clip' stretch type. Areas in blue represent where the 2022 model predicts higher probabilities of occurrence than the 2021 model, while areas in red represent where the 2022 model predicts lower probabilities of occurrence than the 2021 model. The white areas represent where there was little to no difference between the model outputs.

CHAPTER 5: DISCUSSION AND CONCLUSION

The newly developed species distribution models for *G. insculpta* in Atlantic Canada have strong predictive power, low levels of overfitting, and at least in the case of NS, appear to model some of the complexity of *G. insculpta* habitat requirements. Furthermore, these models have been directly applied to survey effort in both provinces, yielding new *G. insculpta* occurrences at twelve separate sites in total, broadening our knowledge of this species' distribution in Atlantic Canada. The success of the field-survey seasons in NS and NB in 2021 and 2022 has illustrated the practical applicability of the SDMs to *G. insculpta* conservation in Atlantic Canada. The NS field-survey season in 2022 was especially successful with 24% of surveys conducted returning occurrence data, and seven individual *G. insculpta* being found at one of these sites. This success rate is notable, considering the difficulty of surveying for *G. insculpta* (Biggar, 2008; Flanagan et al., 2013). Biggar (2008) estimated that a single survey of a watercourse, given ideal conditions including little to no seasonal foliage growth, is likely to only detect 14% of the available *G. insculpta* population. Given this estimate of detectability, Site B could be home to a sizable subpopulation of *G. insculpta* (i.e., ~50 individuals). In total, the 2021 and 2022 surveying in NS added seven sections of watercourses to those identified as significant *G. insculpta* habitat in NS and these sites can now be considered for core habitat designation under the NS *Endangered Species Act*. Initial feedback from both NS NRR and NB NRED indicates that this method of using SDMs to prioritize *G. insculpta* survey effort is useful for field-survey planning and decision-making. However, further work remains to refine the underlying models and to explore other possible applications of them.

5.1 Challenges and Limitations

The specific biology of and threats faced by *G. insculpta* have produced multiple challenges and limitations for this thesis that may not have existed or been as prominent with a different study species. These include concerns about data sensitivity, the lack of available presence-absence data, and the underlying difficulty of surveying for this species. One of the risks that threatens the persistence of *G. insculpta* is the illegal collection of this species for the pet trade (Environment Canada, 2016; COSEWIC, 2018). Due to their colourful and appealing appearance, *G. insculpta* make for attractive pets, and their need to bask along watercourses on spring mornings make them relatively easy targets for poachers. While not the sole reason, the threat of poaching is an important factor as to why this species is classified as “data sensitive”, meaning that access to its occurrence data is restricted. Due to the associated restrictions, I was required to sign data sharing agreements with NS NRR and NB NRED to gain access to the provincial datasets. However, the main effect of this data sensitivity on my thesis has been the constant attention needed to ensure I am not revealing locational information while presenting my work. To prevent my work from being used for purposes that further threaten the species (e.g., illegal hunting/poaching, nest/habitat disturbance, etc.), I have had to refrain from showing certain model outputs or referencing specific sites while presenting at conferences and seminars. As a result, it has been challenging to effectively communicate my research at times, which has limited the feedback I have been able to receive from experts concerning my modelling.

As described in Chapter 2, *G. insculpta* is difficult to survey for and thus multiple surveys must be conducted at a site before it can be determined that no *G. insculpta* inhabit that location (Biggar, 2008; Flanagan et al. 2013). This produced two limitations for my thesis, the first being the lack of *G. insculpta* presence-absence data in Atlantic Canada, which meant that I was limited to presence-only SDM techniques. While presence-absence techniques are often preferable, they require rigorously collected absences for the study species (Zaniewski et al., 2002; Phillips et al., 2004; Elith et al., 2006; Austin, 2007; Miller, 2010). In fact, using inaccurate absence data can be more problematic for model calibration than entirely omitting absences, as inaccurate absences will suggest that species-occupied sites (and their associated environmental variables) are unfit for the species in question, simply because the individuals at those sites were not located. Nevertheless, with the presence-only data I had available to use, Maxent provided a robust technique for producing SDMs with good predictive power. However, future *G. insculpta* survey efforts in Atlantic Canada should aim to produce high-quality presence-absence data for use in further modelling, as well as for informing conservation policy and practice.

The other limitation that the difficulty of *G. insculpta* surveying imposed on my thesis regarded the ability to validate the models using field surveys. To properly field-validate the models, government staff or other trusted/responsible representatives would have had to collect presence-absence data and visit sites that had lower predicted probabilities of occurrence in the models. This would have meant staff would be revisiting sites with low predicted probabilities of occurrence multiple times, possibly frustrating persons who were not involved in the study design and thus decreasing their interest in future surveys. With limited staff hours at hand and a short timing window within which surveys are feasible (late-April to late-June), a decision was needed on whether to prioritize field-validating the models or searching for new *G. insculpta* occurrences. In discussion with my provincial partners, we decided that my modelling work would be more immediately useful to *G. insculpta* conservation, and that it would be better received and supported by conservation managers in Atlantic Canada, if we prioritized visiting highly predicted sites to search for as many new occurrences as possible. If statistical field-verification of the models is desired in the future, a focused team of researchers who are knowledgeable about *G. insculpta* habitat should be organized to undertake an intensive field-survey season.

Additional unforeseen challenges arose during my extraction of environmental covariates and my use of R packages. Early on, I identified watercourse sinuosity as an important covariate for modelling the distribution of *G. insculpta.* Accordingly, I attempted to produce a representative covariate layer but I was unable to separate the hydrological networks of NS and NB into the standardized segments needed for the sinuosity calculations to be comparable across the study areas. I also encountered data consistency problems with both the NB forestry data and the NCC Stream Classification layer, which hindered my efforts to produce covariate layers based on them. Concerning the R packages, beyond the expected difficulties when first learning how to use them, 'spThin' and 'ENMevaluate' presented additional roadblocks to my modelling progress (Muscarella et al., 2014; Aiello-Lammens et al., 2015; Kass et al., 2021; R Core Team, 2021). For 'spThin', the tutorials I accessed spoke of the negligible processing time of the package, but these tutorials used datasets containing only a few hundred occurrences. In contrast, the NS dataset that I was attempting to thin contained 14,472 occurrences. My initial attempt at using 'spThin' on this dataset resulted in the script running for over 24 hours with no indication of how much progress had been made. I allowed the script to run for over a day as I was concerned that I might cancel it right before completion. However, not knowing how much longer it would take to complete, I decided to cancel the script on the second day of continuous processing. To shorten the processing time and solve this problem, I tried splitting the dataset roughly in half based on latitude and ran ‘spThin’ separately on each half. I then combined the two thinned halves and ran ‘spThin’ a third and final time. Even with this adapted approach, the ‘spThin’ runs for the dataset halves each took multiple hours to complete. It should be noted that the lengthy processing time of the ‘spThin’ package likely could have been mitigated if I had had access to a more powerful computer than my personal laptop, which only has a 6-core 2.6 GHz processor and 16 GB of RAM. For ‘ENMevaluate’, running the script often took approximately an hour to complete, but this was not the primary issue I encountered with this package. Instead, the problems arose during the 2022 modelling. Sometime between when I conducted modelling in 2021 and 2022, the creator of ‘ENMevaluate’ reworked much of the syntax of the package, rendering much of the original script I had used obsolete. I thus had to troubleshoot multiple new errors and relearn how to use the package before continuing with my 2022 modelling. While each of the above challenges did not create significant roadblocks by themselves, the accumulation of these small challenges resulted in both the 2021 and 2022 modelling taking longer to conduct than initially expected.

It would be remiss not to mention that the entire duration of my project occurred during the COVID-19 pandemic. While the desktop-based nature of my project meant that I was able to continue despite restrictions, there were some limitations that were imposed by the pandemic. First, the vast majority of meetings with both my supervisory committee and my partners in the provincial governments of NS and NB were held virtually over video-conferencing software. While I was still able to communicate effectively during these virtual meetings, the few in-person sessions I had with my committee felt more organic and creative in a way that is difficult to put into words. Second, as a student of Dalhousie University, I was not permitted to undertake travel for field surveys due to provincially mandated travel restrictions and concurrent Dalhousie University research restrictions during the 2021 field season. While my participation in the field surveys was not necessary for their success, it would have been beneficial to develop a deeper understanding of *G. insculpta* ecology and the on-the-ground application of my models.

5.2 Future Research

As described in Chapters 3 and 4, there are multiple environmental covariates that should be tested in future *G. insculpta* modelling in Atlantic Canada. To better capture the watercourse characteristics favourable to *G. insculpta*, further modelling in Atlantic Canada should prioritize including direct variables such as sinuosity and flowrate, and if that is not feasible as it was for my project, the inclusion of additional proxies including bedrock type, slope, and topographic position index should be tested. To include terrestrial habitat with variable canopy closure in the NB modelling, the available forestry data for this province should be more thoroughly explored to identify potentially extractable variables. For both NS and NB, the development of covariates that address forest edges, mosaic landscapes, or vegetation heterogeneity could be useful. Additionally, while not addressing any of the key *G. insculpta* habitat features, the inclusion of a 'Distance to Coast' covariate could serve to reduce the prediction of potential *G. insculpta* distribution in coastal areas of NS that are likely unsuitable. Anthropogenic geospatial data such as land use, impervious surfaces, roads, and agricultural areas should also be employed in the future either as covariates or secondary data for analysis, to determine areas where *G. insculpta* habitat has likely been compromised by development. Historic habitat modification is likely partly reflected by covariates such as 'Distance to Alder' and 'Distance to Hardwood', but variables that explicitly concern anthropogenic land use as listed above would be useful for elucidating more direct relationships.

Future *G. insculpta* SDMs in Atlantic Canada could likely also use environmental covariate rasters with smaller cell sizes than I used to increase the resolutions of the resulting model outputs. My choice to use covariate rasters with 250 m cells was fairly arbitrary and based on a flawed assumption that a larger cell size would make the model outputs easier to apply to field surveys. As the coarsest environmental data used in the NS and NB modelling were the NS Enhanced DEM (cell size = 20 m) and the USGS EROS SRTM DEM (cell size = ~30 m) respectively, the models I produced could have used these cell sizes for all their covariates.

Another methodological change that should be considered for future modelling involves the choice of Maxent feature types. As described in Chapter 2, Maxent models the relationship between an environmental covariate and species presence using one or a combination of different mathematical functions and transformations, referred to as feature types (Phillips et al., 2006; Elith et al., 2011). I relied on the R package ‘ENMevaluate’ to analyze the environmental data and the occurrence data and determine which suite of feature types would be best to use for each round of modelling (Muscarella et al., 2014; Kass et al., 2021). For the NS modelling, ‘ENMevaluate’ recommended that I use all five feature types, including ‘Linear’, ‘Quadratic’, ‘Product’, ‘Hinge’, and ‘Threshold’ feature types for the continuous covariates. The result of using ‘Threshold’ feature types is evident in the response curve of the second-most important covariate in the final NS modelling, ‘Distance to Alder’, which has the appearance of three descending ‘steps’, with the largest ‘step’ (i.e., drop in predicted probability) being at roughly 5000 m from the nearest alder stand. These abrupt drops in predicted probability are the result of ‘Threshold’ features being applied to the ‘Distance to Alder’ covariate. This feature type applies one or more binary thresholds to the environmental covariate resulting in the associated ‘steps’ in predicted probability of occurrence (Phillips et al., 2006; Elith et al., 2011). The application of ‘Threshold’ features appears to have dominated the modelling of the relationship between the ‘Distance to Alder’ covariate and the probability of *G. insculpta* occurrence, with the model seemingly suggesting that *G. insculpta* can likely be found only within 5000 m of alders. However, this exact threshold appears arbitrary, produces noticeable circles with radii of 5000 m in the model output around alder stands, and is difficult to interpret as an ecologically meaningful interaction between *G. insculpta* and alders. In fact, Phillips et al. (2017) recommend not using ‘Threshold’ features as omitting them results in smoother, simpler, and likely more ecologically realistic models. Thus, based on the results of the NS modelling, I recommend that future Maxent modelling of *G. insculpta* should omit the ‘Threshold’ feature type.

Concerning the occurrence data used in the NS modelling, it is important to recognize the existence of occurrences from radio-tracking studies in the combined dataset that I used. As individual *G. insculpta* were tagged and their locations repeatedly recorded throughout the studies, there are many occurrences in the combined NS dataset that represent the same individuals. As *G. insculpta* exhibit strong site fidelity, there is cause for concern that the use of radio-tracking occurrences in the NS modelling may have biased the model calibration towards the habitat selection of particular *G. insculpta* individuals (COSEWIC, 2018). The thinning of the occurrence data using ‘spThin’ likely reduced some of this bias, but it may still be worthwhile to produce a *G. insculpta* SDM for NS with an occurrence dataset that omits all radio-tracking occurrences. This new model could then be compared to the models already produced to evaluate the impact of including multiple occurrences of the same individuals.

Beyond using the models to plan future field-survey seasons in NS and NB, other management applications of these models could be explored. Within NS, SDMs have already been applied to the management of the endangered boreal felt lichen (*Erioderma pedicellatum*), in that an approved lichen surveyor must inspect any publicly owned forests that are proposed for harvest if they have a predicted probability of *E. pedicellatum* occurrence greater than 0.5 in the provincial SDM (Cameron & Neily, 2008; Cameron & Bayne, 2020). If *E. pedicellatum* is found within a proposed harvest site, a 500-m radial buffer is applied to each individual occurrence, within which forestry operations are restricted. Similarly, 100-m and 200-m radial buffers are applied to other species of at-risk lichen (NS Department of Natural Resources, 2018). In NB, a similar approach is currently being explored for white-rimmed shingle lichen (*Fuscopannaria leucosticta*; Setchell & Haughian, 2021). For *G. insculpta,* a comparable buffer approach is applied to forestry operations around known occurrences in NS, with a 200-m seasonal buffer applied to watercourses for 2 km up and downstream of a recorded occurrence (NS Department of Natural Resources, 2012). However, such an approach requires knowledge of *G. insculpta* presence. Unfortunately, given the difficulty of surveying for this species and the small time-window (late-April to late-June) within which surveys are feasible (Flanagan et al., 2013), it would be difficult to mandate a survey requirement for all watercourses in proximity to a proposed harvest. Further discussions and work with NS NRR and NB NRED could potentially determine how the *G. insculpta* models could be applied in similar ways as the *E. pedicellatum* model to ensure increased protection of *G. insculpta* and their habitat. It would likely be beneficial to produce newer, more refined models first using the recommendations detailed above before SDM-based management guidelines are developed.

5.3 Conclusion

As clades, Reptiles and *Testudines* in particular are in grave danger of losing much of their global biodiversity as a result of anthropogenic threats (Sinervo et al., 2010; Böhm et al., 2013). The overarching goal of this thesis was to contribute to the conservation of one *Testudines* species, *Glyptemys insculpta*, by improving understanding of its distribution in Atlantic Canada via the development and application of species distribution models. In developing these models, I had four objectives: 1) to identify and assess the most robust environmental variables for predicting the occurrence of *G. insculpta* in Atlantic Canada; 2) to share the resulting model outputs with NS NRR and NB NRED to guide survey efforts designed to discover new *G. insculpta* subpopulations; 3) for NS, to quantify the portion of areas predicted to have a high probability of *G. insculpta* occurrence that are currently identified as important habitat or are formally protected; and 4) to undertake this research using a coproduction-based approach to improve the actionability of the results and conclusions.

As detailed in Chapters 2 and 3, I identified multiple potential covariates to capture the key habitat features that together provide suitable habitat for *G. insculpta*. Through the development of the SDMs, I determined that 'Elevation', and some measure of watercourse habitat (either 'Distance to Watercourse' or 'Watercourse Density') were important covariates for both NS and NB. In NS, the covariates 'Distance to Alder' and 'Distance to Estuary' were also retained as highly important. While the 'Elevation' and 'Distance to Watercourse' covariates are reflected in previous *G. insculpta* modelling (Mothes et al., 2020; Willey et al., 2022), 'Distance to Alder', 'Distance to Estuary' and 'Watercourse Density' are all newly identified covariates that could be useful for further modelling of the distribution of *G. insculpta*, both in Atlantic Canada and other areas of this species' range. The 'Watercourse Density' covariate in particular could also be useful for modelling the distribution of other semiaquatic obligate-riparian vertebrate species. However, there are other covariates that should be tested in future *G. insculpta* modelling in Atlantic Canada, including ones representing proximity to the coast, favourable watercourse characteristics (sinuosity, flow accumulation, etc.), geologic and topographic features (bedrock type, slope, topographic position index), and measures of canopy variation (forest edges, mosaic landscapes, vegetation heterogeneity).

The model outputs of the SDMs I have developed have been directly applied to survey efforts in both provinces to notable success. Over the course of the model-informed field surveys in 2021 and 2022, *G. insculpta* occurrences were found at seven new sites in NS and five new sites in NB. This has resulted in a direct expansion of our understanding of *G. inscultpa* distribution in Atlantic Canada and has shown the applicability of the SDMs. The success of these field-survey seasons has also shown the benefit of taking a coproduction-based approach to this research, as it was through involving NS NRR and NB NRED in the research design that these field surveys were possible. Furthermore, because of the involvement of these government departments throughout this project, it is likely that the newly identified *G. insculpta* habitat will be integrated into recovery planning in both NS and NB, thus narrowing the knowledge-action gap.

I have also spatially assessed the output of the 2022 NS model to estimate that the potential distribution of *G. insculpta* comprises 15.1% of NS. Of this area, only 4.6% is currently protected within national parks, provincial parks, wildlife reserves and properties owned by conservation-focused NGOs, less than half the national estimate of formal protection of *G. insculpta* habitat (~11%). Additionally, much of the identified potential distribution fell within the agricultural and populated river valleys of NS, underlining the importance to *G. insculpta* conservation of initiatives aimed at engaging with farmers and other rural landowners, like the Wood Turtle Strides program (NS Federation of Agriculture; Sherren et al. 2020). The potential distribution overlapped with 80.4% of the currently identified core habitat for this species, suggesting that the 2022 NS model is sensitive enough to identify much of the known important *G. insculpta* habitat in NS. However, while this agreement with the core habitat layer indicates good sensitivity, the potential distribution appears to have poorer specificity, as it includes some coastal areas that are unlikely to constitute suitable *G. insculpta* habitat.

Through the modelling I have conducted in NS and NB, this research has improved our understanding of *G. insculpta* distribution in Atlantic Canada by identifying important environmental covariates, creating model outputs that can provide maps of potential distribution, and most directly, by informing field surveys which have discovered new *G. insculpta* habitat. This project has been undertaken in partnership with the NS Department of Natural Resources and Renewables and the NB Department of Natural Resources and Energy Development whose staff conducted the model-informed surveys. As such, there is hope that my models' outputs and findings, as well as any future model iterations will continue to be used as tools for promoting *G. insculpta* recovery in NS and NB. More broadly, this work has contributed to the currently expanding application of SDMs to at-risk species conservation in Atlantic Canada. The models developed for NS represent the first SDMs for *G. insculpta* in that province. Additionally, this thesis provides a template for future modellers who wish to predict the distribution of *G. insculpta* using Maxent software at a more locally applicable scale than that used by Mothes et al. (2020). As *G. inscultpa* are semiaquatic obligate-riparian species, this thesis also contributes to the distribution modelling of species with similar characteristics. Considering the current biodiversity crisis, it is essential that novel techniques be explored and employed to prevent the irreversible loss of species. My work is an example of applying the ideas of species distribution modelling and coproduction of knowledge to contribute to the conservation of an at-risk species, both in scholarship and practice.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Atlantic Science. (2019). NCC Stream Classification v2.0 [Dataset]. Nature Conservancy of Canada. Retrieved from https://2c1forest.databasin.org/datasets/3fa5eb769b99496fad0c05c838c8823d/

Arvisais, M., Bourgeois, J. C., Lévesque, E., Daigle, C., Masse, D., & Jutras, J. (2002). Home range and movements of a wood turtle (*Clemmys insculpta*) population at the northern limit of its range. Canadian Journal of Zoology, 80(3), 402–408. https://doi.org/10.1139/z02-013

Arvisais, M., Lévesque, E., Bourgeois, J. C., Daigle, C., Masse, D., & Jutras, J. (2004). Habitat selection by the wood turtle (*Clemmys insculpta*) at the northern limit of its range. Canadian Journal of Zoology, 82(3), 391–398. https://doi.org/10.1139/z04-012

Austin, M. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling, 157(2–3), 101–118. https://doi.org/10.1016/S0304-3800(02)00205-3

Austin, M. (2007). Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling, 200(1–2), 1–19. https://doi.org/10.1016/J.ECOLMODEL.2006.07.005

Bale, S., Beazley, K. F., Westwood, A., & Bush, P. (2020). The benefits of using topographic features to predict climate-resilient habitat for migratory forest landbirds: An example for the Rusty Blackbird, Olive-sided Flycatcher, and Canada Warbler. *Condor*, *122*(1), 1–19. https://doi.org/10.1093/condor/duz057

Banta, J. Tutorials: How to decide which settings to use when running Maxent, as well as how to make a bias file. The Banta Lab. Retrieved from https://sites.google.com/site/thebantalab/tutorials?authuser=0#h.c1h6hsh4lsn

Beier, P., Hansen, L. J., Helbrecht, L., & Behar, D. (2017). A how-to guide for coproduction of actionable science. Conservation Letters, 10(3), 288–296. https://doi.org/10.1111/CONL.12300

Biggar, K. (2008). Factors affecting detection on a population of wood turtles (*Glyptemys insculpta* LeConte) from the St. Mary’s watershed, Nova Scotia [Unpublished honours thesis]. St. Francis Xavier University. 118p.

Böhm, M., Collen, B., Baillie, J. E. M., Bowles, P., Chanson, J., Cox, N., Hammerson, G., Hoffmann, M., Livingstone, S. R., Ram, M., Rhodin, A. G. J., Stuart, S. N., van Dijk, P. P., Young, B. E., Afuang, L. E., Aghasyan, A., García, A., Aguilar, C., Ajtic, R., … Zug, G. (2013). The conservation status of the world's reptiles. Biological Conservation, 157, 372–385. https://doi.org/10.1016/j.biocon.2012.07.015

Bouchard, C., Tessier, N., & Lapointe, F. J. (2019). Watersheds influence the wood turtle's (*Glyptemys insculpta*) genetic structure. Conservation Genetics, 20(3), 653–664. https://doi.org/10.1007/s10592-019-01169-1

Browne, C. (2020). Road mortality hotspot model for wood turtle in the Chignecto Isthmus study area within New Brunswick.

Cameron, R. P., Neily T. (2008). Heuristic model for identifying the habitats of *Erioderma pedicellatum* and other rare cyanolichens in Nova Scotia, Canada. Bryologist. 111(4):650–658.

Cameron, R. P., Bayne DM. (2020). Identifying lichen-rich areas in Nova Scotia. Proc Nova Scotian Inst Sci. 50(2):227–231.

Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015). Accelerated modern human-induced species losses: Entering the sixth mass extinction. Science Advances, 1(5), e1400253. https://doi.org/10.1126/sciadv.1400253

Coastal Action. (2015, July 1). Snapping turtles found inside the stomach of a chain pickerel! Retrieved from https://www.coastalaction.org/old-coastal-chronicles/snapping-turtles-found-inside-the-stomach-of-a-chain-pickerel

Compton, B. W., Rhymer, J. M., & McCollough, M. (2002). Habitat selection by wood turtles (*Clemmys insculpta*): An application of paired logistic regression. Ecology, 83(3), 833–843. https://doi.org/10.2307/3071885

Cooke, S. J., Nguyen, V. M., Chapman, J. M., Reid, A. J., Landsman, S. J., Young, N., Hinch, S. G., Schott, S., Mandrak, N. E., & Semeniuk, C. A. D. (2021). Knowledge co-production: A pathway to effective fisheries management, conservation, and governance. Fisheries, 46(2), 89–97. https://doi.org/10.1002/FSH.10512

COSEWIC. (2018). COSEWIC assessment and status report on the Wood Turtle *Glyptemys insculpta* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. Xiii + 51pp.

Cvitanovic, C., Hobday, A. J., van Kerkhoff, L., Wilson, S. K., Dobbs, K., & Marshall, N. A. (2015). Improving knowledge exchange among scientists and decision-makers to facilitate the adaptive governance of marine resources: A review of knowledge and research needs. Ocean & Coastal Management, 112, 25–35. https://doi.org/10.1016/J.OCECOAMAN.2015.05.002

Daigle, C., & Jutras, J. (2005). Quantitative Evidence of Decline in a Southern Québec Wood Turtle (*Glyptemys insculpta*) Population. Source: Journal of Herpetology, 39(1), 130–132.

Dubois, Y., Blouin-Demers, G., Shipley, B., & Thomas, D. (2009). Thermoregulation and habitat selection in wood turtles *Glyptemys insculpta*: Chasing the sun slowly. Journal of Animal Ecology, 78(5), 1023–1032. https://doi.org/10.1111/j.1365-2656.2009.01555.x

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., … E. Zimmermann, N. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography, 29(2), 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x

Endangered Species Act, 1998 (NS) c. 11 (Can.)

Enneson, J. J., & Litzgus, J. D. (2008). Using long-term data and a stage-classified matrix to assess conservation strategies for an endangered turtle (*Clemmys guttata*). Biological Conservation, 141(6), 1560–1568. https://doi.org/10.1016/J.BIOCON.2008.04.001

Environment Canada. (2016). Recovery Strategy for the Wood Turtle (Glyptemys insculpta) in Canada [Proposed]. *Species at Risk Act Recovery Strategy Series*, v + 48 pp.

Esri Inc. (2020). ArcGIS Pro (Version 2.5). Esri Inc. https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

Fernández, R. J. (2016). How to be a more effective environmental scientist in management and policy contexts. Environmental Science & Policy, 64, 171–176. https://doi.org/10.1016/J.ENVSCI.2016.07.006

Ferraz, K. M. P. M. B., Peterson, A. T., Scachetti-Pereira, R., Vettorazzi, C. A., & Verdade, L. M. (2009). Distribution of capybaras in an agroecosystem, Southeastern Brazil, based on ecological niche modeling. Journal of Mammalogy, 90(1), 189–194. https://doi.org/10.1644/07-MAMM-A-338.1

Flanagan, M., Roy-Mcdougall, V., Forbes, G., & Forbes, G. (2013). Survey methodology for the detection of wood turtles (*Glyptemys insculpta*). Canadian Field-Naturalist, 127(3), 216–223. https://doi.org/10.22621/cfn.v127i3.1486

Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges. Diversity and Distributions, 19(10), 1217–1223. https://doi.org/10.1111/DDI.12125

GBIF.org (21 December 2021) GBIF Occurrence Download https://doi.org/10.15468/dl.tuub6j

Gilhen J. (1984). Amphibians and reptiles of Nova Scotia. Nova Scotia Museum of Natural History, Nova Scotia, CA. 161p.

Gomez, J. J., Túnez, J. I., Fracassi, N., & Cassini, M. H. (2014). Habitat suitability and anthropogenic correlates of Neotropical river otter (*Lontra longicaudis*) distribution. Journal of Mammalogy, 95(4), 824–833. https://doi.org/10.1644/13-MAMM-A-265

Greaves, W. F., & Litzgus, J. D. (2007). Overwintering ecology of Wood Turtles (*Glyptemys insculpta*) at the species' northern range limit. Journal of Herpetology, 41(1), 32–40. https://doi.org/10.1670/0022-1511(2007)41[32:OEOWTG]2.0.CO;2

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. Ecological Modelling, 135(2–3), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hamilton, H., Smyth, R. L., Young, B. E., Howard, T. G., Tracey, C., Breyer, S., Cameron, D. R., Chazal, A., Conley, A. K., Frye, C., & Schloss, C. (2022). Increasing taxonomic diversity and spatial resolution clarifies opportunities for protecting US imperiled species. Ecological Applications, 32(3), e2534. https://doi.org/10.1002/EAP.2534

Haughian, S. R., Claydenb, S. R., & Cameron, R. (2018). On the distribution and habitat of Fuscopannaria leucosticta in New Brunswick, Canada. *Https://Doi.Org/10.1080/11956860.2018.1526997*, *26*(2), 99–112. https://doi.org/10.1080/11956860.2018.1526997

Heppell, S. S., Crowder, L. B., & Crouse, D. T. (1996). Models to evaluate headstarting as a management tool for long-lived turtles. Ecological Applications, 6(2), 556–565. https://doi.org/10.2307/2269391

Hohensinner, S., Hauer, C., Muhar, S., Hohensinner, S., Muhar, · S, & Hauer, C. (2018). River morphology, channelization, and habitat restoration. Riverine Ecosystem Management, 41–65. https://doi.org/10.1007/978-3-319-73250-3_3

Hutchinson, G. E. (1957). Concluding remarks. Cold Spring Harbor Symposium Quantitative Biology 22, pp. 415–427.

Hughes, G. N., & Litzgus, J. D. (2019). Impact of natural resource extraction on thermal properties of wood turtle (*Glyptemys insculpta*) habitat. Journal of Thermal Biology, 84, 469–478. https://doi.org/10.1016/j.jtherbio.2019.07.031

International Union for Conservation of Nature. (2011). Wood Turtle: *Glyptemys insculpta*. IUCN Red List. Retrieved from https://www.iucnredlist.org/species/4965/97416259

IPBES. (2019). Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. https://doi.org/10.5281/ZENODO.6417333

Jaynes, E. T. (1957). Information theory and statistical mechanics. Physical Review, 106(4), 620. https://doi.org/10.1103/PhysRev.106.620

Jones, M. T., & Willey, L. L. (2015). Status and conservation of the Wood Turtle in the Northeastern United States: Draft. Northeastern Wood Turtle Working Group, 290.

Kass, J. M., Muscarella, R., Galante, P. J., Bohl, C. L., Pinilla-Buitrago, G. E., Boria, R. A., Soley-Guardia, M., and R. P. Anderson (2021). ENMeval 2.0: Redesigned for customizable and reproducible modeling of species' niches and distributions. Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210X.13628

Kaufmann, J. H. (1992). Habitat use by wood turtles in central Pennsylvania. Journal of Herpetology, 26(3), 315–321. https://doi.org/10.2307/1564887

Kershaw, F., Waller, T., Micucci, P., Draque, J., Barros, M., Buongermini, E., Pearson, R. G., & Mendez, M. (2013). Informing conservation units: Barriers to dispersal for the yellow anaconda. Diversity and Distributions, 19(9), 1164–1174. https://doi.org/10.1111/DDI.12101

Kurnaz, M., & Şahin, M. K. (2021). A contribution to the biogeography and taxonomy of two anatolian mountain brook newts, *Neurergus barani* and *N. strauchii* (Amphibia: Salamandridae) using ecological niche modeling. Turkish Journal of Zoology, 45(1), 54–64. https://doi.org/10.3906/ZOO-2007-37

La Forest, G.V. (1957). Rights of landowners in New Brunswick respecting water in streams on or adjoining their lands. University of New Brunswick Law Journal. 10: 21-31.

Liu, C., White, M., & Newell, G. (2013). Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography, 40(4), 778–789. https://doi.org/10.1111/jbi.12058

Lobo, J. M., Jiménez-valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17(2), 145–151. https://doi.org/10.1111/J.1466-8238.2007.00358.X

Loeza-Quintana, T., Crookes, S., Li, P. Y., Reid, D. P., Smith, M., & Hanner, R. H. (2021). Environmental DNA detection of endangered and invasive species in Kejimkujik national park and historic site. Genome, 64(3), 172–180. https://doi.org/10.1139/GEN-2020-0042/SUPPL_FILE/GEN-2020-0042SUPPLA.XLSX

Lovich, J. E., Ennen, J. R., Agha, M., & Whitfield Gibbons, J. (2018). Where have all the turtles gone, and why does it matter? BioScience, 68(10), 771–781. https://doi.org/10.1093/BIOSCI/BIY095

Makowsky, R., Marshall, J. C., McVay, J., Chippindale, P. T., & Rissler, L. J. (2010). Phylogeographic analysis and environmental niche modeling of the plain-bellied watersnake (*Nerodia erythrogaster*) reveals low levels of genetic and ecological differentiation. Molecular Phylogenetics and Evolution, 55(3), 985–995. https://doi.org/10.1016/J.YMPEV.2010.03.012

Meadow, A. M., Ferguson, D. B., Guido, Z., Horangic, A., Owen, G., & Wall, T. (2015). Moving toward the deliberate coproduction of climate science knowledge. Weather, Climate, and Society, 7(2), 179–191. https://doi.org/10.1175/WCAS-D-14-00050.1

Miller, J. (2010). Species distribution modeling. Geography Compass, 4(6), 490–509. https://doi.org/10.1111/J.1749-8198.2010.00351.X

Mothes, C. C., Howell, H. J., & Searcy, C. A. (2020). Habitat suitability models for the imperiled wood turtle (*Glyptemys insculpta*) raise concerns for the species' persistence under future climate change. Global Ecology and Conservation, 24. https://doi.org/10.1016/j.gecco.2020.e01247

Mullin, D. I., White, R. C., Lentini, A. M., Brooks, R. J., Bériault, K. R., & Litzgus, J. D. (2020). Predation and disease limit population recovery following 15 years of headstarting an endangered freshwater turtle. Biological Conservation, 245, 108496. https://doi.org/10.1016/j.biocon.2020.108496

Munro, M. C., Newell, R.E., & Hill, N.M. (2014). Nova Scotia Plants. Nova Scotia Museum of Natural History, Nova Scotia, CA

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecology and Evolution, 5(11), 1198–1205. https://doi.org/10.1111/2041-210X.12261

Mynsberge, A. R., Strager, M. P., Strager, J. M., & Mazik, P. M. (2009). Developing predictive models for freshwater mussels (Mollusca: Unionidae) in the appalachians: Limitations and directions for future research. Ecoscience, 16(3), 387–398. https://doi.org/10.2980/16-3-3221

Needham, J.L., K.F. Beazley, V. Papuga. (2020). Accessing local tacit knowledge as a means of knowledge co-production for effective wildlife corridor planning in the Chignecto Isthmus, Canada. Land 9 (9): 332. DIO: 10.3390/land9090332

Nel, J. L., Roux, D. J., Driver, A., Hill, L., Maherry, A. C., Snaddon, K., Petersen, C. R., Smith-Adao, L. B., Van Deventer, H., & Reyers, B. (2016). Knowledge co-production and boundary work to promote implementation of conservation plans. Conservation Biology, 30(1), 176–188. https://doi.org/10.1111/COBI.12560

Nguyen, V. M., Young, N., Brownscombe, J. W., & Cooke, S. J. (2019). Collaboration and engagement produce more actionable science: quantitatively analyzing uptake of fish tracking studies. Ecological Applications, 29(6), e01943. https://doi.org/10.1002/EAP.1943

Nova Scotia Department of Fisheries and Aquaculture. (2015). Nova Scotia Freshwater Fish Species Distribution Records [Dataset]. Retrieved from https://data.novascotia.ca/Fishing-and-Aquaculture/Nova-Scotia-Freshwater-Fish-Species-Distribution-R/jgyj-d4fh

Nova Scotia Department of Lands and Forestry. (2020). Recovery Plan for the Wood turtle (*Glyptemys insculpta*) in Nova Scotia [Final]. Nova Scotia Endangered Species Act Recovery Plan Series.

Nova Scotia Department of Natural Resources. (2012). Vulnerable Wood Turtle (*Glyptemys insculpta*) Special Management Practices.

Nova Scotia Department of Natural Resources. (2018). At-risk lichens – Special Management Practices

Nova Scotia Department of Natural Resources and Renewables (A). (2006). Enhanced Digital Elevation Model, Nova Scotia, Canada (Version 2) [Dataset]. Retrieved from https://novascotia.ca/natr/meb/download/dp055.asp

Nova Scotia Department of Natural Resources and Renewables (B). (2006). Geological Map of the Province of Nova Scotia [Dataset]. Retrieved from https://novascotia.ca/natr/meb/download/dp043.asp

Nova Scotia Department of Natural Resources and Renewables. (2021). Forest Inventory – Geographic Information Systems [Dataset]. Retrieved from https://novascotia.ca/natr/forestry/gis/forest-inventory.asp

Nova Scotia Federation of Agriculture. Wood Turtle Strides: Financial incentives to make farms wood turtle friendly. Biodiversity Landowner's Guide. Retreived from http://www.farmbiodiversity.ca/strides/#:~:text=Wood%20Turtle%20Strides%20is%20a,their%20critical%20habitat%20on%20farms.

Pearson, K., Cameron, R., & Troy McMullin, R. (2018). Habitat associations and distribution model for *Fuscopannaria leucosticta* in Nova Scotia, Canada. Lichenologist, 50(4), 487–497. https://doi.org/10.1017/S0024282918000300

Phillips, S. J., Anderson, R. P., Dudík, M., Schapire, R. E., & Blair, M. E. (2017). Opening the black box: an open-source release of Maxent. Ecography, 40(7), 887–893. https://doi.org/10.1111/ECOG.03049

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudik, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. Proceedings of the Twenty-First International Conference on Machine Learning, 655-662. https://doi.org/10.1145/1015330.1015412

Phillips, S. J. (2017). A Brief Tutorial on Maxent. Retrieved from http://biodiversityinformatics.amnh.org/open_source/maxent/

Province of Nova Scotia. (2020) The Nova Scotia Protected Areas System [Dataset]. Retrieved from https://nsgiwa.novascotia.ca/arcgis/rest/services/ENV/ENV_NS_Prot_Area_Sys_UT83/MapServer

Pulliam, H. R. (2000). On the relationship between niche and distribution. Ecology Letters, 3(4), 349–361. https://doi.org/10.1046/J.1461-0248.2000.00143.X

Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. Journal of Biogeography, 41(4), 629–643. https://doi.org/10.1111/JBI.12227

Reyes, R. P., Firn, J., Nicol, S., Chadès, I., Stratford, D. S., Martin, T. G., Whitten, S., & Carwardine, J. (2019). Building a stakeholder-led common vision increases the expected cost-effectiveness of biodiversity conservation. PLOS ONE, 14(6), e0218093. https://doi.org/10.1371/JOURNAL.PONE.0218093

Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A. M., Bernard, R., Böhm, M., Castro-Herrera, F., Chirio, L., Collen, B., Colli, G. R., Dabool, L., Das, I., Doan, T. M., Grismer, L. L., Hoogmoed, M., Itescu, Y., Kraus, F., Lebreton, M., … Meiri, S. (2017). The global distribution of tetrapods reveals a need for targeted reptile conservation. Nature Ecology & Evolution. https://doi.org/10.1038/s41559-017-0332-2

RStudio Team (2022). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Saumure, R. A., Herman, T. B., & Titman, R. D. (2007). Effects of haying and agricultural practices on a declining species: The North American wood turtle, *Glyptemys insculpta*. Biological Conservation, 135(4), 565–575. https://doi.org/10.1016/j.biocon.2006.11.003

Saumure, R. a, & Bider, J. R. (1998). Impact of agricultural development on a population of wood turtles (*Clemmys insculpta*) in southern Québec, Canada. In Chelonian Conservation and Biology (Vol. 3, Issue 1, pp. 37–45).

Service Nova Scotia and Internal Services. (2020). Hydrographic Network, Nova Scotia [Dataset]. Retrieved from https://www.arcgis.com/home/item.html?id=a1a810df3bde46caa5aa8cde306164e1

Setchell, A., & Haughian, S. R. (2021). Translating a species distribution model into management recommendations for an at-risk lichen. Https://Doi.Org/10.1080/11956860.2021.1949821, 1–7. https://doi.org/10.1080/11956860.2021.1949821

Sherren, K., Tourangeau, W., Lamarque, M., & Greenland-Smith, S. (2020). Exploring motivation crowding around farmer incentives for riparian management in Nova Scotia. Canadian Geographer, 64(2), 250–265. https://doi.org/10.1111/cag.12572

Sinervo, B., Méndez-de-la-Cruz, F., Miles, D. B., Heulin, B., Bastiaans, E., Cruz, M. V. S., Lara-Resendiz, R., Martínez-Méndez, N., Calderón-Espinosa, M. L., Meza-Lázaro, R. N., Gadsden, H., Avila, L. J., Morando, M., De La Riva, I. J., Sepúlveda, P. V., Rocha, C. F. D., Ibargüengoytía, N., Puntriano, C. A., Massot, M., Lepetz, V., Oksanen, T. A., Chappie, D. G., Bauer, A. M., Branch, W. R., Clobert, J., and Sites, J. W. (2010). Erosion of lizard diversity by climate change and altered thermal niches. Science, 328(5980), 894–899. https://doi.org/10.1126/SCIENCE.1184695/SUPPL_FILE/SINERVO.SOM.PDF

Species at Risk Act, 2002 (Can.) c. 29

Species at Risk Act, 2012 (NB) c. 6 (Can.)

Spencer, R. J., Van Dyke, J. U., & Thompson, M. B. (2017). Critically evaluating best management practices for preventing freshwater turtle extinctions. Conservation Biology, 31(6), 1340–1349. https://doi.org/10.1111/COBI.12930

Staicer, C., C. Ferrari, and A. Westwood. (2015). Habitat modeling for landbird species at risk in southwestern Nova Scotia final report for the Nova Scotia Habitat Conservation Fund (year 2, April 2013-Oct 2014) and final report for the Nova Scotia Species at Risk Conservation Fund (June 2013-Oct 2014). 24pp

Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O., & Ludwig, C. (2015). The trajectory of the Anthropocene: The Great Acceleration. The Anthropocene Review, 2(1), 81–98. https://doi.org/10.1177/2053019614564785

The Canadian Encyclopedia. (2021, March 30). Geography of Nova Scotia. Retrieved from https://www.thecanadianencyclopedia.ca/en/article/geography-of-nova-scotia

Tingley, R., McCurdy, D. G., Pulsifer, M. D., & Herman, T. B. (2009). Spatio-temporal differences in the use of agricultural fields by male and female wood turtles (*Glyptemys insculpta*) inhabiting an agri-forest Mosaic. Herpetological Conservation and Biology, 4(2), 185–190.

United States Geological Survey. (2000). USGS EROS Archive – Digital Elevation – Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global [Dataset]. Retrieved from https://www.usgs.gov/centers/eros/science/usgs-eros-archive-digital-elevation-shuttle-radar-topography-mission-srtm-1#overview

Wallace S. (2020). Wood Turtle ecology and management strategies in a landscape under active agriculture [Unpublished masters thesis]. University of New Brunswick. 140p.

Wallace, S. D., Forbes, G. J., & Nocera, J. J. (2020). Experimental assessment of the impact of agricultural machinery on Wood Turtles (*Glyptemys insculpta*). Chelonian Conservation and Biology, 19(1), 78–84. https://doi.org/10.2744/CCB-1404.1

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications, 21(2), 335–342. https://doi.org/10.1890/10-1171.1

Wesley P.A. (2006). Local and regional scale habitat selection by Wood Turtles (*Glyptemys insculpta*) in Ontario [Unpublished masters thesis]. University of Guelph. 97p.

Westwood, A. R., Barker, N. K., Grant, S., Amos, A., Camfield, A. F., Cooper, K., Dénes, F. V, Jean-Gagnon, F., Mcblane, L., Schmiegelow, F. K. A., Simpson, J. I., Slattery, S. M., Sleep, D. J. H., Sliwa, S., Wells, J. V, Whitaker, D. M., Westwood, A. R., Barker, N. K., Grant, S., … Canada, A. (2020). Toward actionable, coproduced research on boreal birds focused on building respectful partnerships. Avian Conservation and Ecology, Published Online: Jun 30, 2020 | Doi:10.5751/ACE-01589-150126, 15(1), 1–28. https://doi.org/10.5751/ACE-01589-150126

Westwood, R., Westwood, A. R., Hooshmandi, M., Pearson, K., LaFrance, K., & Murray, C. (2019). A field-validated species distribution model to support management of the critically endangered Poweshiek skipperling (*Oarisma poweshiek*) butterfly in Canada. Conservation Science and Practice, 2(3), 1–15. https://doi.org/10.1111/csp2.163

White T.B. (2013). In the cold and near the edge: Hypoxia tolerance, seasonal site fidelity, and communal hibernation in a peripheral population of Wood Turtles (*Glyptemys insculpta*) [Unpublished masters thesis]. Acadia University. 137p.

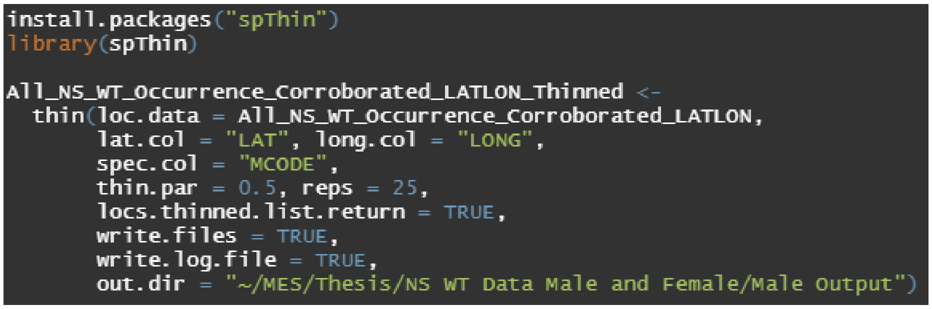
Wildlife Act, 1989 (NS) c. 504 (Can.)

Willey, L. L., Jones, M. T., Sievert, P. R., Akre, T. S. B., Marchand, M., Demaynadier, P., Yorks, D., Mays, J., Dragon, J., Erb, L., Zarate, B., Kleopfer, J. D., Gipe, K. D., Parren, S., Andrews, J., Roberts, H. P., Tamplin, J. W., Raithel, C., Johnson, L., Ross, A. M., Breisch, A., Johnson, G., Angus, S., Drasher, J., and Wicklow, B. (2022). Distribution models combined with standardized surveys reveal widespread habitat loss in a threatened turtle species. Biological Conservation, 266, 109437. https://doi.org/10.1016/j.biocon.2021.109437

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth edition. Springer.

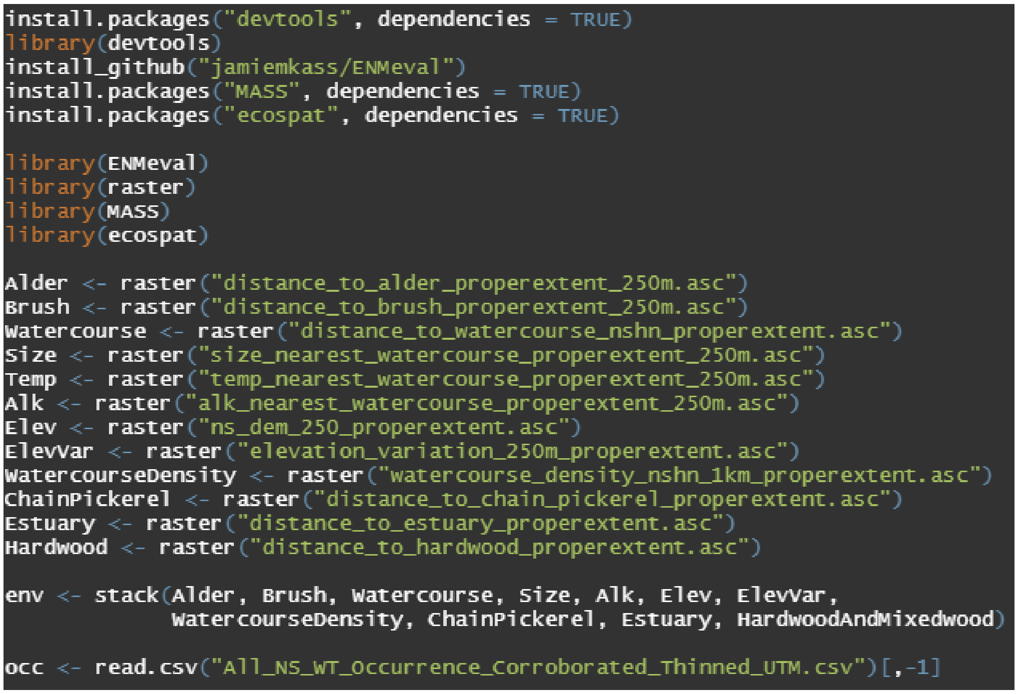
Zaniewski, A. E., Lehmann, A., & Overton, J. M. C. (2002). Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling, 157(2–3), 261–280. https://doi.org/10.1016/S0304-3800(02)00199-0

APPENDIX A: R SCRIPTS



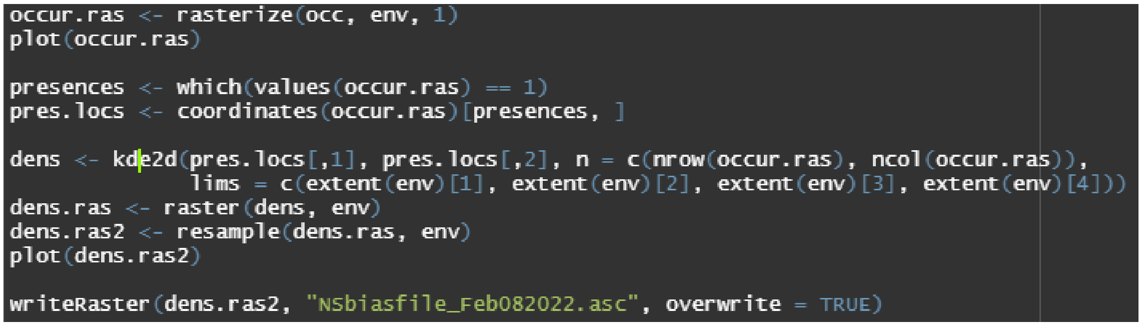
```r
install.packages("spThin")
library(spThin)

All_NS_WT_Occurrence_Corroborated_LATLON_Thinned <-
  thin(loc.data = All_NS_WT_Occurrence_Corroborated_LATLON,
       lat.col = "LAT", long.col = "LONG",
       spec.col = "MCODE",
       thin.par = 0.5, reps = 25,
       locs.thinned.list.return = TRUE,
       write.files = TRUE,
       write.log.file = TRUE,
       out.dir = "~/MES/Thesis/NS WT Data Male and Female/Male Output")
```

Figure A.1 The R script used for the R package 'spThin' (Aiello-Lammens et al., 2015; R Core Team, 2021). The example given is that of thinning the corroborated NS occurrence data. The parameter named "thin.par" specifies the thinning distance used in kilometres (km). In this case, the "thin.par" is 0.5, meaning that the data was thinned so that in the resulting dataset, all remaining occurrences were separated by distances of at least 0.5 km or 500 metres (m).



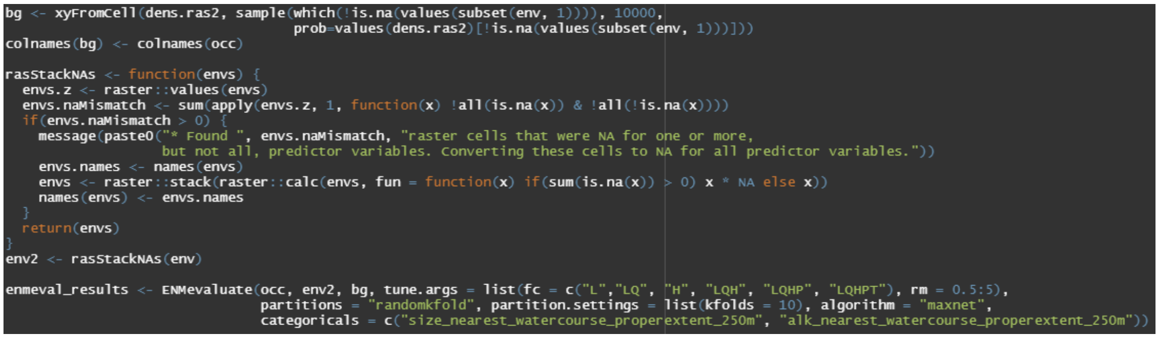
```r
install.packages("devtools", dependencies = TRUE)
library(devtools)
install_github("jamiemkass/ENMeval")
install.packages("MASS", dependencies = TRUE)
install.packages("ecospat", dependencies = TRUE)

library(ENMeval)
library(raster)
library(MASS)
library(ecospat)

Alder <- raster("distance_to_alder_properextent_250m.asc")
Brush <- raster("distance_to_brush_properextent_250m.asc")
Watercourse <- raster("distance_to_watercourse_nshn_properextent.asc")
Size <- raster("size_nearest_watercourse_properextent_250m.asc")
Temp <- raster("temp_nearest_watercourse_properextent_250m.asc")
Alk <- raster("alk_nearest_watercourse_properextent_250m.asc")
Elev <- raster("ns_dem_250_properextent.asc")
ElevVar <- raster("elevation_variation_250m_properextent.asc")
WatercourseDensity <- raster("watercourse_density_nshn_1km_properextent.asc")
ChainPickerel <- raster("distance_to_chain_pickerel_properextent.asc")
Estuary <- raster("distance_to_estuary_properextent.asc")
Hardwood <- raster("distance_to_hardwood_properextent.asc")

env <- stack(Alder, Brush, Watercourse, Size, Alk, Elev, ElevVar,
             WatercourseDensity, ChainPickerel, Estuary, HardwoodAndMixedwood)

occ <- read.csv("All_NS_WT_Occurrence_Corroborated_Thinned_UTM.csv")[,-1]
```

Figure A.2 The R script used to prepare the occurrence and environmental data for both the production of the spatial bias files and the running of the R package 'ENMevaluate' (Banta; Kass et al., 2021; R Core Team, 2021). The example given is that of the preparation of the occurrence and environmental data used in the 2022 NS modelling. This script produces two objects: 'env' (a stack of the environmental covariate rasters); and 'occ' (a dataframe containing the occurrence data). Also provided is the code for installing the packages required for both the production of the spatial bias files and the running of ‘ENMevaluate’.



```
occur.ras <- rasterize(occ, env, 1)
plot(occur.ras)

presences <- which(values(occur.ras) == 1)
pres.locs <- coordinates(occur.ras)[presences, ]

dens <- kde2d(pres.locs[,1], pres.locs[,2], n = c(nrow(occur.ras), ncol(occur.ras)),
              lims = c(extent(env)[1], extent(env)[2], extent(env)[3], extent(env)[4]))
dens.ras <- raster(dens, env)
dens.ras2 <- resample(dens.ras, env)
plot(dens.ras2)

writeRaster(dens.ras2, "NSbiasfile_Feb082022.asc", overwrite = TRUE)
```

Figure A.3 The R script used to produce the spatial bias files (Banta, R Core Team, 2021). This example follows the previous in showing the production of the spatial bias file for the 2022 NS modelling. The ‘kde2d’ function estimates the two-dimensional kernel density of the input occurrences (Venables and Ripley, 2002).



```
bg <- xyFromCell(dens.ras2, sample(which(!is.na(values(subset(env, 1)))), 10000,
                                   prob=values(dens.ras2)[!is.na(values(subset(env, 1)))]))
colnames(bg) <- colnames(occ)

rasStackNAs <- function(envs) {
  envs.z <- raster::values(envs)
  envs.naMismatch <- sum(apply(envs.z, 1, function(x) !all(is.na(x)) & !all(!is.na(x))))
  if(envs.naMismatch > 0) {
    message(paste0("* Found ", envs.naMismatch, "raster cells that were NA for one or more,
                   but not all, predictor variables. Converting these cells to NA for all predictor variables."))
    envs.names <- names(envs)
    envs <- raster::stack(raster::calc(envs, fun = function(x) if(sum(is.na(x)) > 0) x * NA else x))
    names(envs) <- envs.names
  }
  return(envs)
}
env2 <- rasStackNAs(env)

enmeval_results <- ENMevaluate(occ, env2, bg, tune.args = list(fc = c("L","LQ", "H", "LQH", "LQHP", "LQHPT"), rm = 0.5:5),
                               partitions = "randomkfold", partition.settings = list(kfolds = 10), algorithm = "maxnet",
                               categoricals = c("size_nearest_watercourse_properextent_250m", "alk_nearest_watercourse_properextent_250m"))
```

Figure A.4 The R script used for the R package ‘ENMevaluate’ (Banta; Kass et al., 2021; R Core Team, 2021). This example follows the previous two in showing the use of ‘ENMevaulate’ to determine the best parameters for the 2022 NS modelling. The parameter, ‘tune.args = list(fc = c(“L”, “LQ”, “H”, “LQH”, “LGHP”, “LQHPT”)’ determines the sets of feature types for ‘ENMevaluate’ to test (i.e., ‘L’ = Linear, ‘Q’ = Quadratic, ‘H’ = Hinge, ‘P’ = Product, and ‘T’ = Threshold). The ‘rm’ parameter determines the range of regularization multipliers for ‘ENMevaluate’ to test, in this example, ranging from 0.5-5.

APPENDIX B: CREATION OF ENVIRONMENTAL COVARIATE LAYERS

Elevation:

For NS, I used the ‘Resample’ tool to convert the NS Digital Elevation Model (DEM) from 20 m^2 cells to 250 m^2 cells using the ‘NEAREST’ resampling technique (Nova Scotia Department of Natural Resources and Renewables, 2006). As I did not find a publicly available DEM for NB, I used the ‘Mosaic to New Raster’ tool to combine all the United States Geological Survey (USGS) EROS SRTM DEM tiles that cover NB into one ~30 m^2 raster (United States Geological Survey, 2000). I then used the ‘Resample’ tool to convert this raster to a 250 m^2 cell raster.

Distance-based Layers:

For ‘Distance to Watercourse’ in NS, I selected all features in the NS hydrographic network that had the following Feature Codes: WACORV59 (COAST RIVER spine), WARV50 (RIVER/STREAM SINGLE LINE spine), WARV55 (RIVER/STREAM SINGLE indefinite/approximate line), and WARV59 (RIVER spine; Service Nova Scotia and Internal Services, 2020). I exported these features into a new layer that only represented watercourses, as the NS hydrographic network also contains lakes. I then ran the ‘Euclidean Distance’ tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest watercourse. For NB, I selected all the features in the NCC Stream Classification layer that had ‘Size_Simp’ categories of either ‘Headwaters and Creeks’, ‘Small Rivers’, ‘Medium Rivers’, or ‘Large Rivers’, and exported this as a new layer (Atlantic Science, 2019). This excluded the features that were categorized as ‘Lakes’ and thus only represented watercourses. I then ran the ‘Euclidean Distance’ tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest watercourse.

For ‘Distance to Alder’ in NS, I selected all the NS Forest Inventory polygons that had ‘FORNON’ codes of 38 (‘Alders less than 75% cover’) and 39 (‘Alders 75% or greater cover’) and exported them as a new layer, representing alder stands (Nova Scotia

Department of Natural Resources and Renewables, 2021). I then ran the 'Euclidean Distance' tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest alder stand. For NB, I selected all the NB Landbase polygons that had 'VT' categories of 'AW' (Alder) and exported them as a new layer, representing alder stands. I then rand the 'Euclidean Distance' tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest alder stand.

For 'Distance to Brush' in NS, I selected all the NS Forest Inventory polygons that had 'FORNON' codes of 33 ('Brush') and exported them as a new layer, representing areas of brush (Nova Scotia Department of Natural Resources and Renewables, 2021). I then ran the 'Euclidean Distance' tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest area of brush.

For 'Distance to Hardwood' in NS, I selected all the NS Forest Inventory polygons that had 'COVER_TYPE' codes of 8 ('Hardwood') and exported them as a new layer, representing forest stands with less than 25% softwood species by basal area (Nova Scotia Department of Natural Resources and Renewables, 2021). I then ran the 'Euclidean Distance' tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest hardwood stand.

For 'Distance to Mixedwood' in NS, I selected all the NS Forest Inventory polygons that had 'COVER_TYPE' codes of 5 ('Mixedwood') and exported them as a new layer, representing forest stands with 26-74% softwood species by basal area (Nova Scotia Department of Natural Resources and Renewables, 2021). I then ran the 'Euclidean Distance' tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest mixedwood stand.

For 'Distance to Estuary' in NS, I selected all features in the NCC Stream Classification layer that had the 'Tidal' category 'Yes', and exported them as a new layer, representing watercourse segments that are tidally influenced (i.e., estuaries; Atlantic Science, 2019). I then ran the 'Euclidean Distance' tool on this layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest estuary.

For ‘Distance to Chain Pickerel’ in NS, I selected all occurrences in the NS Freshwater Species Dataset that corresponded to chain pickerel (*Esox niger*) occurrences and exported them as a new layer (Nova Scotia Department of Fisheries and Aquaculture, 2015). I then used the ‘Merge’ tool to combine this chain pickerel occurrence layer with the Global Biodiversity Information Facility Dataset for chain pickerel in NS (GBIF.org, 2021). I ran the ‘Euclidean Distance’ tool on this combined layer to create a 250 m^2 cell raster, with values representing how far each cell was to the nearest chain pickerel occurrence.

Watercourse Density:

Using the layer I extracted from the NS hydrographic network that only contains watercourse and the ‘Line Density’ tool (with a search radius of 1000 m), I created a 250 m^2 cell raster representing ‘Watercourse Density’ (Service Nova Scotia and Internal Services, 2020).

Standard Deviation of Elevation and Slope:

For ‘Standard Deviation of Elevation’, I ran the ‘Focal Statistics’ tool (with a search radius of 1000 m) on the NS Digital Elevation Model to calculate the standard deviation of this raster (Nova Scotia Department of Natural Resources and Renewables, 2006). I then used the ‘Resample’ tool and the ‘CUBIC’ resampling technique to convert the resulting 20 m^2 cell raster to a 250 m^2 cell raster.

For ‘Standard Deviation of Slope’, I first created a ‘Slope’ raster by using the ‘Slope’ tool on the NS Digital Elevation Model (Nova Scotia Department of Natural Resources and Renewables, 2006). I then ran the ‘Focal Statistics’ tool (with a search radius of 1000 m) to calculate the standard deviation of this raster and used the ‘Resample’ tool and the ‘CUBIC’ resampling technique to convert the resulting 20 m^2 cell raster to a 250 m^2 cell raster.

Categorical Watercourse Variables:

For both NS and NB, I extracted categorical watercourse variables from the NCC Stream Classification layer using the ‘Euclidean Allocation’ tool (Atlantic Science, 2019).

To use this tool, I first had to convert the text-based categories (Nominal data) into numeric categories (Ordinal data). For ‘Gradient of Nearest Watercourse’, I converted the ‘Gradient_Simp’ categories; ‘Low’ to 1, ‘Moderate’ to 2, and ‘High’ to 3. For ‘Size of Nearest Watercourse, I converted the ‘Size_Simp’ categories; ‘Headwaters and Creeks’ to 1, ‘Small Rivers’ to 2, ‘Medium Rivers’ to 3, and ‘Large Rivers’ to 4. For ‘Temperature of Nearest Watercourse’, I converted the ‘Temp’ categories; ‘Cold’ to 1, ‘Cool’ to 2, and ‘Warm’ to 3. For ‘Alkalinity of Nearest Watercourse’, I converted the ‘Alk’ categories; ‘Low’ to 1, ‘Moderate’ to 2, and ‘High’ to 3. For ‘Tidal Influence of Nearest Watercourse’, I converted the ‘Tidal’ categories; ‘No’ to 1, and ‘Yes’ to 2. For each variable, I then used the ‘Euclidean Allocation’ tool to extract these categorical values and create 250 m^2 cell rasters.

Standard Deviation of Crown Closure:

I created a new field in the NS Forest Inventory called ‘Tot_CRNCL’ (representing total crown closure) by adding together the ‘CRNCL’ (1st story crown closure) and the ‘SS_CRNCL’ (2nd story crown closure; Nova Scotia Department of Natural Resources and Renewables, 2021). I then used the ‘Polygon to Raster’ tool with the ‘MAXIMUM_AREA’ cell assignment option to create a ‘Tot_CRNCL’ raster with 10 m^2 cells. I ran the ‘Focal Statistics’ tool to calculate the standard deviation of this raster over 25 by 25 cell windows, then used the ‘Resample’ tool and the ‘NEAREST’ resampling technique to convert the resulting 10 m^2 cell raster to a 250 m^2 cell raster.

APPENDIX C: CORRELATION MATRICES

Nova Scotia 2021:

Table C.1 Correlations between the environmental covariates in the initial suite for Maxent modelling the distribution of *G. insculpta* in Nova Scotia, Canada in 2021. Covariate pairs with absolute correlation values above 0.7 would have been removed but none were found for this set of covariates. Absolute correlation values of 0.4 – 0.7, represented in yellow, prompted consideration but did not immediately warrant removal of either covariate. The numbers represent environmental covariates (see legend in Notes, below).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	1.00	0.43	-0.03	0.19	0.02	0.05	-0.20	0.30	-0.18	-0.10
2	0.43	1.00	-0.06	0.11	0.23	-0.02	-0.04	0.08	-0.15	-0.11
3	-0.03	-0.06	1.00	-0.06	-0.14	0.02	-0.08	0.03	0.02	0.04
4	0.19	0.11	-0.06	1.00	-0.03	0.01	-0.07	0.17	-0.02	0.10
5	0.02	0.23	-0.14	-0.03	1.00	-0.11	0.28	-0.35	-0.22	-0.25
6	0.05	-0.02	0.02	0.01	-0.11	1.00	-0.19	0.21	0.02	0.01
7	-0.20	-0.04	-0.08	-0.07	0.28	-0.19	1.00	-0.38	-0.05	0.02
8	0.30	0.08	0.03	0.17	-0.35	0.21	-0.38	1.00	-0.05	0.14
9	-0.18	-0.15	0.02	-0.02	-0.22	0.02	-0.05	-0.05	1.00	0.17
10	-0.10	-0.11	0.04	0.10	-0.25	0.01	0.02	0.14	0.17	1.00

Notes

Environmental Covariate	**Corresponding Number**
Distance to Alder	1
Distance to Brush	2
Standard Deviation of Crown Closure	3
Distance to Watercourse	4
Elevation	5
Size of Nearest Watercourse	6
Gradient of Nearest Watercourse	7
Temperature of Nearest Watercourse	8
Alkalinity of Nearest Watercourse	9
Tidal Influence of Nearest Watercourse	10

New Brunswick 2021:

Table C.2 Correlations between the environmental covariates in the initial suite for Maxent modelling the distribution of *G. insculpta* in New Brunswick, Canada in 2021. Covariate pairs with absolute correlation values above 0.7 would have been removed, but none were found for this set of covariates. Absolute correlation values of 0.4 – 0.7, represented in yellow, prompted consideration but did not immediately warrant removal of either covariate. The numbers represent environmental covariates (see legend in Notes, below).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1.00	-0.02	-0.09	-0.03	0.15	-0.05	0.08	-0.11
2	-0.02	1.00	0.13	-0.03	-0.15	-0.05	-0.08	0.28
3	-0.09	0.13	1.00	-0.31	-0.35	-0.02	-0.08	0.42
4	-0.03	-0.03	-0.31	1.00	0.25	0.06	0.00	-0.13
5	0.15	-0.15	-0.35	0.25	1.00	-0.06	0.08	-0.54
6	-0.05	-0.05	-0.02	0.06	-0.06	1.00	0.13	-0.05
7	0.08	-0.08	-0.08	0.00	0.08	0.13	1.00	-0.20
8	-0.11	0.28	0.42	-0.13	-0.54	-0.05	-0.20	1.00

Notes

Environmental Covariate	**Corresponding Number**
Distance to Alder	1
Distance to Watercourse	2
Gradient of Nearest Watercourse	3
Size of Nearest Watercourse	4
Tidal Influence of Nearest Watercourse	5
Temperature of Nearest Watercourse	6
Alkalinity of Nearest Watercourse	7
Elevation	8

Nova Scotia 2022:

Table C.3 Correlations between the environmental covariates in the initial suite for Maxent modelling the distribution of *G. insculpta* in Nova Scotia, Canada in 2022. For covariate pairs with absolute correlation values above 0.7, represented in red, I removed the covariate that was presumed to be less ecologically relevant. Absolute correlation values of 0.4 – 0.7, represented in yellow, prompted consideration but did not immediately warrant removal of either covariate. The numbers represent environmental covariates (see legend in Notes, below).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	1.00	0.85	-0.18	0.30	0.36	-0.15	-0.14	-0.10	0.35	-0.02	-0.33	0.00	0.39	-0.01	-0.18
2	0.85	1.00	-0.15	0.27	0.35	-0.15	-0.18	-0.12	0.31	-0.06	-0.37	-0.02	0.44	-0.04	-0.19
3	-0.18	-0.15	1.00	-0.56	-0.16	0.11	0.15	0.14	-0.04	-0.07	0.24	0.03	-0.11	0.11	0.23
4	0.30	0.27	-0.56	1.00	0.28	-0.14	-0.13	-0.05	0.09	0.16	-0.34	0.04	0.14	-0.14	-0.30
5	0.36	0.35	-0.16	0.28	1.00	-0.32	0.12	0.17	0.20	0.17	-0.47	-0.09	0.27	-0.07	-0.20
6	-0.15	-0.15	0.11	-0.14	-0.32	1.00	-0.03	-0.04	0.32	-0.18	0.15	0.05	-0.16	0.27	0.32
7	-0.14	-0.18	0.15	-0.13	0.12	-0.03	1.00	0.63	0.01	0.02	0.13	-0.03	-0.07	0.08	0.23
8	-0.10	-0.12	0.14	-0.05	0.17	-0.04	0.63	1.00	0.11	0.06	0.01	-0.03	-0.03	0.04	0.11
9	0.35	0.31	-0.04	0.09	0.20	0.32	0.01	0.11	1.00	-0.22	-0.35	-0.11	0.28	0.23	0.02
10	-0.02	-0.06	-0.07	0.16	0.17	-0.18	0.02	0.06	-0.22	1.00	-0.05	0.02	-0.05	-0.15	-0.18
11	-0.33	-0.37	0.24	-0.34	-0.47	0.15	0.13	0.01	-0.35	-0.05	1.00	0.21	-0.38	0.08	0.30
12	0.00	-0.02	0.03	0.04	-0.09	0.05	-0.03	-0.03	-0.11	0.02	0.21	1.00	-0.19	-0.02	0.05
13	0.39	0.44	-0.11	0.14	0.27	-0.16	-0.07	-0.03	0.28	-0.05	-0.38	-0.19	1.00	-0.04	-0.20
14	-0.01	-0.04	0.11	-0.14	-0.07	0.27	0.08	0.04	0.23	-0.15	0.08	-0.02	-0.04	1.00	0.43
15	-0.18	-0.19	0.23	-0.30	-0.20	0.32	0.23	0.11	0.02	-0.18	0.30	0.05	-0.20	0.43	1.00

Notes

Environmental Covariate	**Corresponding Number**
Standard Deviation of Slope	1
Standard Deviation of Elevation	2
Distance to Watercourse	3
Watercourse Density	4
Distance to Chain Pickerel	5
Distance to Estuary	6
Distance to Hardwood	7
Distance to Mixedwood	8
Elevation	9
Alkalinity of Nearest Watercourse	10
Temperature of Nearest Watercourse	11
Size of Nearest Watercourse	12
Gradient of Nearest Watercourse	13
Distance to Brush	14
Distance to Alder	15

APPENDIX D: SPATIAL BIAS FILES

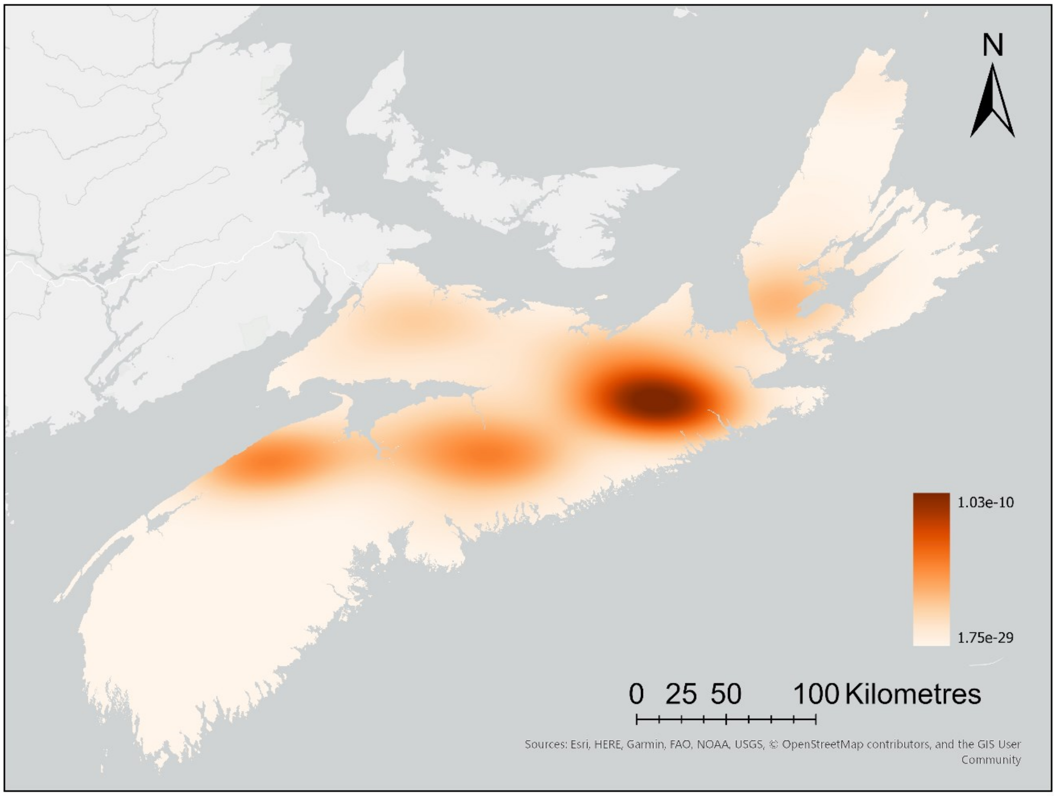


Figure D.1 Spatial bias of thinned (using the R package *'*spThin*'*) *G. insculpta* occurrences used in the Maxent modelling for *G. insculpta* in Nova Scotia, Canada. This file was produced by calculating the kernel density of the occurrences *(See Appendix A for R script)*. The highest concentration of occurrences represents the most well-researched *G. insculpta* subpopulation in NS, with concentrated survey efforts over a 30 year period and multiple radio tracking studies (Biggar, 2008; White, 2013).

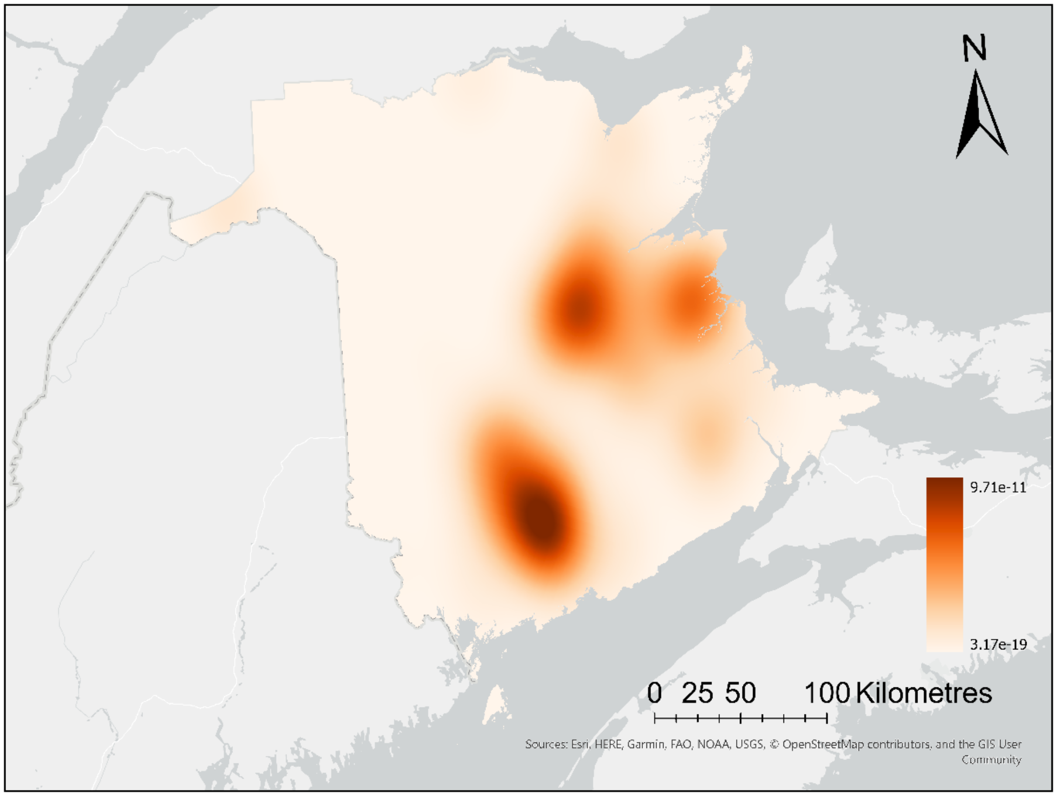


Figure D.2 Spatial bias of thinned (using the R package 'spThin') *G. insculpta* occurrences used in the Maxent modelling for *G. insculpta* in New Brunswick, Canada. This file was produced by calculating the kernel density of the occurrences *(See Appendix A for R script)*.