The spatial ecology of Blanding’s turtles (*Emydoidea blandingii*): from local movement patterns, home ranges and microhabitat selection to Ontario-wide habitat suitability modelling.

Catherine Stéphanie Millar

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the
Ottawa-Carleton Institute of Biology

Thèse soumise à la
Faculté des études supérieures et postdoctorales
Université d’Ottawa
En vue de l’obtention de la maîtrise ès sciences
Institut de biologie d’Ottawa-Carleton

ABSTRACT

The Blanding's turtle, a semi-aquatic freshwater species found in North America, is listed as a species at risk in 17 of the 18 state/provincial jurisdictions across its range (NatureServe, 2009). Furthermore, approximately 20% of the global range of the Blanding's turtle is contained in Ontario (COSEWIC, 2005). The purpose of this study was to characterise the spatial ecology of Blanding's turtles in Ontario at multiple spatial scales, from a single island population to Ontario-wide habitat suitability modeling. I followed 38 Blanding's turtles (20 males, 13 gravid females, and 5 non-gravid females) in 2008 and 2009 on Grenadier Island in the St. Lawrence River, Ontario via radio-telemetry. Furthermore, I built habitat suitability models using historical Blanding's turtle records in Ontario and two machine learning algorithms: maximum entropy modeling (MAXENT) and boosted regression trees (BRTs).

At the local scale, Blanding's turtles selected colder water with more submerged and floating vegetation and avoided open water. Reproductive class and month did not have a significant effect on the mean daily movement of turtles in May, July, and August. In June, however, gravid females moved significantly more than males. Gravid females also had significantly larger home ranges than both males and non-gravid females. At the landscape level, Blanding's turtle habitat suitability was best predicted by air temperature, wetland area, open water area, cropland area, and road density. Habitat suitability increased with increasing air temperature, wetland area, forested area, alvar area, bedrock outcrop area, and decreased with increasing cropland area, pastures and field area, precipitations, terrain ruggedness, and settlements and developed land. Mean area under the operating curve (AUC) values for habitat suitability models tested on independent data ranged from 0.878 to 0.912. My results highlight the importance of stratifying spatial analyses by reproductive class and time and of including terrestrial habitat in management plans for Blanding’s turtles.

RÉSUMÉ

Le but de cette étude était de caractériser l'écologie spatiale des tortues mouchetées en Ontario à plusieurs échelles spatiales, d'une population à la modélisation de la qualité de l’habitat à travers l’Ontario. J'ai suivi 38 tortues mouchetées par radio-télémétrie (20 mâles, 13 femelles gravides, et 5 femelles non-gravides) en 2008 et 2009 sur l'île Grenadier dans le Fleuve Saint-Laurent, Ontario. J'ai aussi créé des modèles de la qualité de l’habitat pour les tortues mouchetées en Ontario à plusieurs échelles spatiales en utilisant des données historiques de présence et deux algorithmes d'apprentissage automatisés: modélisation avec entropie maximale (MAXENT) et des arbres de régression rehaussés (BRTs).

Au niveau du microhabitat, les tortues mouchetées préféraient les eaux plus froides avec une plus grande abondance de végétation submergée et flottante et évitaient l’eau libre. Le statut reproducteur et le mois n'ont pas exercé un effet significatif sur la moyenne des distances parcourues quotidiennement par les tortues en mai, juillet, et en août. En juin, cependant, les femelles gravides se sont déplacées significativement plus loin quotidiennement que les mâles. Les femelles gravides avaient aussi des domaines vitaux significativement plus grands que les mâles et les femelles non-gravides. Au niveau du paysage, les meilleurs indices de la qualité de l'habitat étaient: la température de l'air, la densité des routes, la superficie des terres humides, des terres agricoles et de l’eau libre. La qualité de l'habitat augmentait avec la température de l’air, la superficie des terres humides, des forêts, des alvars et des affleurements rocheux. Par contre, la qualité de l'habitat diminuait lorsqu’il y avait plus de terres agricoles, de précipitations, de développement urbain et un terrain plus accidenté. Les aires sous la courbe caractéristique du fonctionnement du récepteur (AUC) des modèles de qualité de l'habitat, testé sur des données indépendantes, variaient entre 0.878 et 0.912. Mes résultats démontrent l’importance de stratifier les données écologiques par classes reproductives et dans le temps et offre davantage de soutien pour la protection des corridors terrestres entourant les terres humides habitées par les tortues mouchetées.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis supervisor, Dr. Gabriel Blouin-Demers. Between his lightning fast e-mail replies and vast knowledge of reptile ecophysiology, he has been an inspiring, motivating, and knowledgeable mentor. In times of modeling crisis and statistical despair he was there to lend an ear, an e-mail, and a book or article. He has spent several hours proof-reading the many copies of my thesis, and always directed me in the right direction when I was stuck in a loop. Thank you for believing in me and making this opportunity possible!

I would also like to thank Dr. Lenore Fahrig (Carleton University) and Dr. Scott Findlay (University of Ottawa) for their guidance, suggestions, and constructive criticism during my Master's project. My thesis is much stronger because of your inputs. I would also like to thank Dr. Mike Sawada (University of Ottawa) for initial help in troubleshooting my GIS analyses, Nancy Lemay (University of Ottawa) for acquiring rasters from the LIO database, and Rachelle Desrocher (University of Ottawa) for introducing me to species distribution modelling. And last but not least, thanks to all my laboratory colleagues for their advice and feedback throughout my Master's thesis: Dr. Grégory Bulté, Véronique Juneau, Sarah Larocque, M.Sc. Alana Plummer, and Laura Robson.

Over the course of two years, I have had the distinct honour to work with several students and volunteers in the field. A big thank you has to go to Nikki Reshke and Jeffrey Graham for their blood, sweat, and tears over the course of two years of field work. You were field assistants extraordinaire! I would also like to thank Marie-Andrée Carrière and Gabriel Picard for showing me the ropes and helping to set up the

Blanding's turtle project down at St- Lawrence Islands National Park. Also, Karyne Bellehumeur, Liesbet Colson, Andréaz Dupoué, and Ronald Pringadi offered their bog-walking services for several days/weeks on Grenadier Island. Thank you to all the other volunteers and friends who have joined me out on the field for a day or two. You know who you are and your enthusiasm for the outdoors and for turtles kept me going during the hard times.

Logistical support for my fieldwork was provided by St. Lawrence Islands National Park. For all their help preparing and maintaining our island abode, our boats and our turtle traps, I would like to thank Kent Dowsley, Kevin Hodges, Lindsay Shannon, and Mike Ronan. I would also like to thank Josh VanWeiren, Dr. Emily Gonzales, Brian Farkas, Chris Bellemore, Greg Saunders, and Jeff Leggo for their help throughout this study. And finally, I would like to thank all of the SLINP staff and students for their unwavering enthusiasm, loaning of extra canoes, team spirit and putting up with our runaway turtles. Without the generous loaning of motor boats, canoes, laboratory space, summer residence, telemetry equipment, GPS, and gasoline, this project could never have been possible.

All the work I conducted would not have been possible without the assistance of several funding agencies. The bulk of my research expenses were covered by Dr. Blouin-Demers' research grant from the Natural Sciences and Engineering Research Council of Canada. My own funding came from two main sources: a postgraduate research scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC), and various teaching assistant placements.

I could not have gotten to where I am today without the unconditional support, love and guidance of my family. Maman et papa, merci! Vous avez toujours été là pour me motiver, m'inspirer et me guider. I would also like to thank my partner, Ronald Pringadi, for his love, support and understanding. Whenever I was in a bind and needed someone to help me in the field, you were there without question. The technical, logistical and emotional support my parents, family and friends have provided throughout the years is immeasurable.

Like all researchers and students, I would not have gotten to where I am, and this thesis would not have been possible, if it were not for the advances, discoveries and bright minds that came before me. I am but a dwarf standing on the shoulders of giants and any discoveries that I have made were possible because of the wealth of knowledge at my disposal. I can only hope that the work I have done will please the great scientists of the past…

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

GENERAL INTRODUCTION

Biodiversity Loss and Conservation Laws

Species diversity is waning globally and, unlike past extinction events (Raup and Sepkoski, 1984), this decline is due to anthropogenic phenomena (Ehrlich, 1995; Aitken, 1998). The primary threat for all groups is human related habitat loss and degradation (Gibbons et al., 2000; Hoekstra, 2005). As human population and resource consumption continue to increase, so does the pressure on endangered species, thus diminishing the prospects of recovering these species. For conservation efforts to be fruitful, we must understand the relationship between animals and their environment, as conservation of declining species directly entails habitat management (Boersma et al., 2001).

In an effort to protect biodiversity, several nations have enacted legislations. In Canada, the Species at Risk Act (SARA, 2003) affords federal legal protection of species at risk on federal land. Similarly, Ontario's Endangered Species Act (OESA, 1971, 2007) provides provincial legal protection of species at risk on provincial land. Under both SARA and OESA, all species at risk must be given a recovery strategy (SARA, 2003; OESA, 1971, 2007) and this recovery strategy is the mechanism used to identify the critical habitat of a species at risk. A similar process is triggered in the USA under the Endangered Species Act (ESA, 1973, 1978). The critical habitat must be defined "to the extent possible" (SARA s. 41(1) (c) 2003) and all critical habitat will be afforded federal, and if in Ontario, provincial protection. Recovery strategies thus become the main tools for the management and protection of listed species and must be appropriately suited to each species' situation and biology to be effective (Boersma et al., 2001).

Spatial Ecology

Spatial ecology is an ecological study which centers upon understanding how landscape configurations influence the community and population dynamics of organisms (Collinge, 2001). There are three important elements that can be studied in spatial ecology: movement patterns, home range, and habitat selection. Thus, empirical studies in spatial ecology provide a clear link between conservation biology research and practical mechanisms for species management and conservation planning.

Habitat selection can occur at multiple spatial scales, from geographical distribution (landscape), to the composition of home ranges within a landscape (macrohabitat), to specific areas within the home range (microhabitat). These three scales relate to the third, second, and first order of selection of Johnson (1980). Differences in selection pressures and limiting factors can sometimes lead to differing (Orians and Wittenberger, 1991; Luck, 2002) and conflicting (Compton et al., 2002) patterns of selection at multiple scales. Johnson (1980) observed that selection need not occur at all scales and that different species respond to different scales of selection, thus the increased incidence of multiple-scale habitat selection studies. It is only when all elements of selection are examined at multiple scales that management actions will clearly reflect the needs of the species (Morin et al., 2005).

In the last twenty years there has been a rise in the use of habitat suitability models (Rushton et al., 2004). The goal of these models is to predict the likelihood of occurrence of species based on environmental variables. Landscape-level habitat suitability models have proven especially useful for predicting quality habitat for organisms that are endangered, rare, or have a patchy distribution over space or time (Wu and Smeins, 2000; Gibson et al., 2004). There are several frameworks available for habitat suitability modeling (HSM) including Resource Selection Functions (Boyce et al., 2002), Ecological Niche Modeling (Peterson, 2001), Gradient Analysis (Whittaker, 1967), and Species Distribution Modeling (Guisan and Thuiller, 2005; Elith and Leathwick, 2009). The latter of these HSM frameworks, species distribution models (SDMs), are numerical tools that combine observations of species occurrence or abundance with environmental estimates (Elith and Leathwick, 2009).

The data available for modeling the geographic distribution of a species can vary in amount, type and quality, as do the type of models available. Logistic models are typically used to estimate species distribution, ideally based on high quality presence-absence data sets (Elith et al., 2006; Ward et al., 2009). While vast stores of presence-only data exist (occurrence records generally collected by herbaria and museums), absence data are rarely available or reliable (Anderson et al., 2003). Therefore, modeling techniques that require presence data only are very valuable (Graham et al., 2004). The modeling techniques that do not require explicit absence occurrences can be divided into two classes: presence-only (i.e. HABITAT; Walker and Cocks, 1991) and presence/pseudo-absence (Elith et al., 2006; Ward et al., 2009). The presence/pseudo-absence models require a background sample consisting of random samples of locations taken from the region or landscape of interest (Ward et al., 2009). In a review of modeling techniques that do not require explicit absences, modeling algorithms using presence/pseudo-absence data consistently outperformed modeling algorithms using presence-only data. Furthermore, machine learning algorithms consistently outperformed more traditional modelling techniques (Elith et al., 2006).

Declining Reptilian Populations

According to the IUCN (the World Conservation Union), 51% of reptilian species worldwide and evaluated by the IUCN are considered to be at risk. In Canada, 8 of the 12 freshwater turtle species are considered at risk and, thus, turtles are proportionally the most at risk vertebrate group in Canada. Habitat loss and degradation remains the single most important factor responsible for their demise (Gibbons et al., 2000; Shine, 1991), closely followed by road mortality (Gibbons et al., 2000).

The Blanding's turtle, *Emydoidea blandingii*, is a medium-sized freshwater turtle largely confined to the Great Lakes Basin (COSEWIC, 2005). The Blanding's turtle is the only representative of the genus *Emydoidea*, it has one of the smallest global ranges of any North American turtle (COSEWIC, 2005), and is listed as a species at risk in 17 of the 18 state/provincial jurisdictions across its range (NatureServe, 2009). In Ontario, the Blanding's turtle is listed as threatened and thus requires critical habitat delineation. A large portion of its global range (~20%) is contained within southern and south-central Ontario (COSEWIC, 2005) and, as such, this province has a significant responsibility towards the conservation of this species (COSEWIC, 2005).

Objectives

The purpose of my study was to characterise the spatial ecology of Blanding's turtles in Ontario and help in the development of future management plans. I worked at multiple spatial scales: from movement patterns, home ranges, and microhabitat selection for a single island population to Ontario-wide habitat suitability modeling using sighting data. First, I looked at the spatial ecology of a population of Blanding's turtles located on Grenadier Island in St. Lawrence Islands National Park, Ontario, Canada. I compared movement patterns, behaviour, and home range size and length between three reproductive classes: males, gravid females, and non-gravid females. I also discussed microhabitat selection of this population of Blanding's turtles. Second, I conducted habitat suitability modeling for Ontario's Blanding's turtles using sighting data, at three different scales. By identifying potentially important habitat features, at both the microhabitat and landscape level, future management plans can better identify and protect Blanding's turtle habitat.

CHAPTER 1

Movement patterns, home range, and microhabitat selection of Blanding's turtles on Grenadier Island, St. Lawrence-Islands National Park, Ontario, Canada.

Abstract

The purpose of this study was to describe the spatial ecology of a population of Blanding’s turtles on Grenadier Island, in St. Lawrence Islands National Park. I radio-tracked 38 Blanding’s turtles (20 males, 13 gravid females, and 5 non-gravid females) in 2008 and 2009 to determine daily movement patterns, home ranges, seasonal activity, and microhabitat use. Reproductive class and month did not have a significant effect on the mean daily movement of turtles in May, July, and August. In June, however, gravid females moved significantly more than males. Reproductive class had a significant effect on turtle home range size, although high individual variation was observed (1.6 - 40.6 ha). Gravid females had significantly larger home ranges (20.3 ± 3.5 ha) than both males (8.5 ± 1.7 ha) and non-gravid females (7.3 ± 3.2 ha). Finally, at the microhabitat scale, Blanding’s turtles selected colder water with more submerged and floating vegetation and avoided open water. My results highlight the importance of maintaining intact and diverse aquatic habitat in suitable wetlands, creating buffer zones of approximately 1.9 km around wetlands known to harbour Blanding’s turtles, and including terrestrial habitat in home range analysis. Clearly, reproductive class and season greatly influence the behaviour in this species and data must be analysed accordingly. Range-wide concerns for Blanding’s turtles coupled with the fact that Ontario comprises a significant portion of the global distribution indicate that the species should be closely monitored in the province.

Introduction

Species diversity is waning globally and, unlike past extinction events (Raup and Sepkoski 1984), this decline is due to anthropogenic phenomena (Ehrlich, 1995; Aitken, 1998). The primary threat for all groups is human related habitat loss and degradation (Gibbons et al., 2000; Hoekstra, 2005). As human population and resource consumption continue to increase, so does the pressure on endangered species, thus diminishing the prospects of recovering these species. For conservation efforts to be fruitful, we must understand the relationship between animals and their environment, as conservation of declining species directly entails habitat management (Boersma et al., 2001).

In an effort to protect biodiversity, several nations have enacted legislations. In Canada, the Species at Risk Act (SARA, 2003) affords federal legal protection of species at risk on federal land. Similarly, Ontario's Endangered Species Act (OESA, 1971, 2007) provides provincial legal protection of species at risk on provincial land. Under both SARA and OESA, all species at risk must be given a recovery strategy (SARA, 2003; OESA, 1971, 2007) and this recovery strategy is the mechanism used to identify the critical habitat of a species at risk. A similar process is triggered in the USA under the Endangered Species Act (ESA, 1973, 1978). The critical habitat must be defined "to the extent possible" (SARA s. 41(1)(c) 2003) and all critical habitat will be afforded federal, and if in Ontario, provincial protection. Recovery strategies thus become the main tools for the management and protection of listed species and must be appropriately suited to each species' situation and biology to be effective (Boersma et al., 2001).

According to the IUCN (the World Conservation Union), 51% of reptilian species worldwide and evaluated by the IUCN are considered at risk. In Canada, eight of the twelve species of freshwater turtles are listed as species at risk by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC). According to these COSEWIC statistics, reptiles are proportionally the most at risk vertebrate group in Canada. Habitat loss and degradation is the single most important factor responsible for reptile loss (Gibbons et al., 2000; Shine, 1991), closely followed by road mortality (Gibbs and Steen, 2005; Aresco, 2005; Gibbs and Shriver, 2002; Forman and Alexander, 1998), but the problem is multifaceted and other causes include introduction of exotic species, nest predation, the pet trade/commercial use, diseases/parasitism, pollution, and global climate change (Gibbons et al., 2000). Even in cases where the jurisdictional wetland is protected, the surrounding terrestrial habitat needed by semi-aquatic reptiles for nesting, hibernation, and refuge may not (Burke and Gibbons, 1995).

Spatial Ecology

Spatial ecology is an ecological study which centers upon understanding how landscape configurations influence the community and population dynamics of organisms (Collinge, 2001). Whitaker and Shine (2003) stated that studying spatial ecology can lead to three important contributions: (1) a better understanding of movement and habitat selection by the animal, (2) vital information regarding animal-human interactions, and (3) aids in assessing the response and role of an animal within its habitat. Empirical studies in spatial ecology beneficially link conservation biology research to practical mechanisms for species management and conservation planning.

Movement patterns and home range estimates are important elements of spatial ecology. For example, examining movement patterns enables the identification of key portions of critical habitat for listed species, such as nesting and hibernation sites. Many reptiles necessitate more than one habitat for their activities (food acquisition, aestivation, copulation, nesting, thermoregulation, etc.) and, therefore, movement between various habitats is required (Rutherford and Gregory, 2003). Tracking these movements is critical for two fundamental reasons: (1) the availability of suitable habitat in favourable spatial configurations may be a key factor limiting the distribution and abundance of a species and (2) manipulation of habitat is an important potential tool in conservation and management (Rutherford and Gregory, 2003). In complement to movement patterns, home ranges enable researchers to estimate the size and shape of the area used by an animal (Burt, 1943), which allows for proper delineation of regions that need to be protected.

Another important aspect of spatial ecology is the study of habitat use and selection. Habitat selection studies reveal specific habitat types that are used and needed by the species. Once identified, such critical habitat types can then be protected. Selection can occur at multiple spatial scales, from geographical distribution (landscape), to the composition of home ranges within a landscape (macrohabitat), to specific areas within the home range (microhabitat). These three scales relate to the third, second, and first order of selection of Johnson (1980). Differences in selection pressures and limiting factors can sometimes lead to differing (Orians and Wittenberger, 1991; Luck, 2002) and conflicting (Compton et al., 2002) patterns of selection at multiple scales. When all elements of selection are examined at multiple scales, management actions will clearly reflect the needs of the species (Morin et al., 2005).

Blanding's Turtle

The Blanding's turtle, *Emydoidea blandingii*, is a medium-sized freshwater turtle largely confined to the Great Lakes Basin (COSEWIC, 2005). The Blanding's turtle is the only representative of the genus *Emydoidea*, it has one of the smallest global ranges of any North American turtle (COSEWIC, 2005), and a large portion of its global range (20%) is contained within southern and south-central Ontario and the extreme southwest edge of Québec (COSEWIC, 2005). With Ontario containing such a large portion of the global range of this species, this province has a significant responsibility towards the conservation of this species (COSEWIC, 2005).

According to COSEWIC's most recent status report (2005), despite being fairly numerous and widespread, the St. Lawrence population of this species is declining. Urbanization has lead to more extensive road networks that criss-cross the turtles' habitat, effectively limiting movement between subpopulations. Furthermore, development and alteration of wetlands and their surrounding areas has reduced the amount of available and suitable habitat for adult and juvenile turtles and has destroyed potential and existing nesting grounds for females. The primary threat to Blanding's turtles is habitat loss and degradation (Van Dam, 1993; Harding, 1997) and several life-history characteristics of turtles enhance this group's vulnerability to anthropogenic stressors. For Blanding's turtles, these life-history traits include delayed sexual maturity, low recruitment caused by low fecundity, high juvenile and egg mortality, and high levels of egg and nest failure (Congdon et al., 1983; Herman et al., 1995; Linck and Moriarty, 1997; Petokas, 1986). Not only must the residential wetlands and their immediate surroundings be protected (Burke an Gibbons, 1995), but also nesting areas as far as 1620 m from such waterbodies (Joyal et al., 2001).

Adult Blanding's turtles use a variety of wetland habitats. These include vernal pools, bogs, marshes, buttonbush shrub marshes, and impoundments dominated by cattails, water lilies, and duckweed (Hartwig and Kiviat, 2007). Emergent plants such as sedges (*Carex* spp.) and cattails (*Typha* spp.) are often associated with typical Blanding's turtle habitat: shallow, stagnant wetlands with mucky substrate (Congdon et al., 1983; Rowe and Moll, 1991; Pappas and Brecke, 1992; Hamernick, 2000; Sajwaj et al. 1998; Piepgras and Lang, 2000; Sajwaj and Lang, 2000). Terrestrial habitats are frequently used as travel corridors between wetlands (Joyal et al., 2001) and upland habitats that are not fragmented by roads and developments are essential (Grgurovic and Sievert, 2005). Furthermore, these upland habitats are also used for aestivation (Ross and Anderson, 1990; Joyal et al., 2001). This turtle is highly vagile and uses relatively large areas and there is considerable variation in home range length and area across this species' range. Furthermore, Grgurovic and Sievert (2005) and Schuler and Thiel (2008) found little overlap between annual turtle home ranges from year to year, indicating that short-term studies on habitat use underestimate the lifetime home range of an individual.

Previous studies in Northeastern Illinois have demonstrated that Blanding's turtles are active from late March to late October, with frequent basking in April and May (Rowe and Moll 1991). Duration, onset, and termination of the nesting season for Blanding's turtles vary between all studies (26 May-22 June: Rowe and Moll, 1991; 23 May-9 June: Congdon et al., 1983; 28 May-24 June: Linck et al., 1989) and is believed to be dependent upon early spring temperatures. Blanding's turtle males are either slightly bigger or equal in size to females and thus sexual size dimorphism is not apparent in this species (Barry and Shine, 1980).

Objectives

Regional differences in temporal and spatial movement patterns and habitat use are apparent in this specie's (Rowe and Moll, 1991; Herman et al., 1995; Piepgras and Lang, 2000) and these differences are directly related to local management techniques. To facilitate the recovery of Blanding's turtle populations in Ontario, more information is needed on the ecology of this turtle in this province. The purpose of my study was to describe the movement patterns, seasonal activity, microhabitat use, and home range size of Blanding's turtles on Grenadier Island in St. Lawrence Islands National Park, Ontario, Canada.

Due to long distance nesting forays in June, I expected that the mean daily distance moved (DDM) by gravid females in June would be longer than that of males and non-gravid females. I also expected that the mean DDM by males in May and August should be longer than that of all females, as peaks of Blanding's turtle mating activity have been observed during these months (Innes et al., 2008; Jenkins and Babbitt, 2003; Herman et al., 1995). Finally, I expected to see a significant difference in the area and length of home ranges between gravid females and non-gravid females and males, as gravid females must make long distance journeys outside of their respective wetlands to reach their preferred nesting sites.

Reptiles are ectothermic and, as such, they must obtain heat from their environment. Since nearly all physiological processes are temperature dependant, variation in the body temperature of reptiles will greatly affect their development, physiology and behaviour (Dawson, 1975; Huey, 1982; Peterson et al., 1993). Turtles, and other reptiles, thermoregulate mostly by adjusting habitat selection, body posture, and timing of activity (Huey et al., 1989). Thus, I expected turtles to select sites of high thermal quality: sites with high solar radiation (as measured by air and water temperatures). I also expected turtles to select higher percentages of emergent vegetation and floating vegetation as Blanding's turtles are extremely wary and these cover types provide refuge (Sajwaj and Lang, 2000; Hamernick, 2000).

Methods

Study Area

I conducted this study from May 2008 to August 2009 on Grenadier Island in the Thousand Islands Ecosystem. Grenadier Island is located in the St. Lawrence River directly south of Mallorytown, Ontario, Canada. It is one of largest islands in St. Lawrence Islands National Park (ca. 600 ha). In May 1999, 28.4 ha of ecologically sensitive wetland along Brooker's Creek, on the north side of Grenadier Island, became part of St. Lawrence Islands National Park. This site, which I will refer to as Brooker's Pond, is the home of a small population of Blanding's turtles that was the focus of this study.

Brooker's Pond is part of the Grenadier Island wetland complex which is maintained by an active beaver population. There are several other ponds on the island, all part of the same watershed, but beaver dams, matted marshes, or small patches of land make access by water between each difficult or impossible.

Radio-Telemetry and Data Collection

I captured Blanding's turtles by hand, dip-net, and submerged hoop nets baited with canned sardines. All individuals captured were brought back to the laboratory at the Park's Canada station in Mallorytown for weighing, measuring, marking, and transmitter fitting (if required). Firstly, turtles were weighed in a bucket to the nearest 5 g with a spring scale. Secondly, sex was determined by plastron concavity, head and upper neck markings, and preanal tail length. Thirdly, any injuries or scars were recorded. Fourthly, turtles were measured with forestry calipers to the nearest 1 mm for carapace length, carapace width, carapace height, and plastron length. Finally, all turtles were marked with a unique combination of holes in the marginal scutes of their carapace made with a hand-held high-speed drill (Cagle, 1939). No holes were drilled in the marginal scutes attached to the bridge.

I fitted 38 Blanding's turtles (20 males, 13 gravid females, and 5 non-gravid females) with radio-transmitters (Holohil SI-2FT 17 g, battery life of 36 mo) bolted to the rear marginal scutes of the carapace using stainless steel screws, washers, and nuts. I used marine silicone to cover screws and transmitter edges to aid in adhesion and to seal any openings between the shell and transmitter, thus preventing detritus and macrophytes from catching on the transmitter. Transmitters (including screws, nuts, and silicone) represented at most 5% of the turtle's body mass and were removed at the end of the study. Handling of the turtles and use of telemetry was approved by the University of Ottawa Animal Care Committee (protocol BL-221).

The following morning, all turtles were released at their site of capture and tracked every day or every other day using a telemetry receiver (TR-4 Tracking Receiver, Telonics Inc., Mesa, AZ) and a directional antenna (3 Element Yagi Antenna). At each location, I recorded the Universal Transverse Mercator (UTM) coordinates (3D differential receiver status, NAD83 datum) with a GPSmap72 (Garmin International Inc, Olathe, KS) with an estimated accuracy of ≤ 6 meters. I also recorded the behaviour (swimming, basking, at surface, underwater, in forest, or other) of the turtles when located.

Movement Patterns

I derived movement statistics from 19 individuals (6 gravid females, 4 non-gravid females, 9 males) (1099 relocations) tracked in 2009 every 1-2 days, for at least three consecutive months during the active season (April-October). Movement data for individuals tracked in 2008 were not used because of large difference in the time between relocations (2-5 days) in 2008, compared to 2009.

I used the recorded UTM coordinates to calculate the distances moved between locations, with the help of ArcView 3.2 (ESRI, 1995-2010) outfitted with the Animal Movements Extension 2.0 (Hooge and Eichenlaub, 2000). Movements were measured as straight-line distances between relocations. I calculated the daily distance moved by dividing the distance moved between relocations by the number of days elapsed since the last location. I calculated the average daily distance moved by each individual, for each month, before proceeding with any analysis.

To determine whether movement patterns varied by month throughout the active season, I examined the mean daily distance moved per month, while controlling for reproductive class. I used data from 13 individuals (4 gravid-females, 3 non-gravid females, 6 males) (800 relocations) tracked for four consecutive months (May 2009 to August 2009) during the active season (April to early October). I used each individual as a factor to produce a repeated measures design for my two-way ANOVA.

Home Range

I calculated the home range for each individual turtle using the minimum convex polygon method (MCP). MCP's are calculated by creating the smallest possible convex polygon that encompasses all known locations for an individual. MCP's are the most commonly used method to estimate animal home ranges (Powell, 2000) and according to a study by Row and Blouin-Demers (2006), MCP's provide a robust estimate of maximum home range size for herpetofauna. MCP's were calculated using Hawth's Tools (Beyer, 2004), an animal movement extension for ArcMAP 9.0 (ESRI, 1995-2010).

Furthermore, to access home range robustness an asymptote analysis was run, using the ABODE.mxt tool (Laver, 2005). Past studies have shown that the estimates of home range area converge as the number of relocations increases, until the home range estimate reaches an asymptote (Swihart and Slade, 1985; Seaman et al., 1999). Similar to Carrière et al. (2009), home range area as a function of the number of successive relocations showed a series of plateaus, instead of asymptotes. Turtles often stayed in one location for long periods of time or returned to previous locations, as observed with females returning from their nesting sites. Due to non-asymptotic home range curves, I only calculated home ranges for individuals tracked for at least three consecutive months during the active season (April to October). I also calculated home range length (HRL) by measuring the straight-line distance between the two furthest locations for each individual (Pluto and Bellis, 1988). In all, I calculated home ranges for 36 “turtle years”, from 2008 to 2009, based on 35 individual Blanding’s turtles (20 males, 12 gravid females, 5 non-gravid females). To examine the effect of reproductive class on both home range size and home range length, I used one-way ANOVAs.

Microhabitat Analysis

I quantified habitat directly at the locations used by the turtles in 2008 and at the paired random locations. It is important to note that I did not quantify habitat during nesting forays and, to keep sampling manageable, I only characterized the habitat at every second relocation.

A total of seven habitat variables (Table 1-1.) were quantified within a 1 m diameter. All variables were uncorrelated ($r < 0.15$), except for open water and emergent vegetation ($r = -0.83$). I determined air temperature using a mercury thermometer and measured water depth using a meter stick. I measured water temperature 30 cm from the surface of the water with a digital thermometer. Finally, I visually determined the percentage of emergent vegetation, submerged vegetation, floating vegetation, and open water within a 50cm radius around each turtle location (Table 1-1.).

To measure habitat availability, I also characterized paired random locations. A random location was identified by randomly selecting a direction from the characterized turtle location and moving a distance of 40 m (an approximation of the mean daily distance traveled by Blanding's turtles in Brooker's Pond in 2007). Such a distance ensures that the random location selected was indeed available to the turtle. I quantified each random location immediately after the paired turtle location to ensure that no temporal or environmental changes affected the measured variables.

To examine habitat selection within the turtles' home ranges, I used matched-pairs logistic regression. Logistic regression is more robust than MANOVA and discriminant function analysis because it does not require multivariate normality (North and Reynolds, 1996). In the paired design, each turtle location is compared to its paired random location, thus controlling for changes in environmental conditions through time and ensuring that each random location was available to the individual being tracked (Compton et al., 2002).

In the matched-pairs logistic regression, values for each random point are subtracted from the values of each paired turtle location. A standard logistic regression with the constant term removed is then used to fit a response (all 1's) to the differences in habitat values between used and random locations. The resulting estimated coefficients, β_i, are interpreted the same way as with a standard logistic regression. For a coefficient βi, an n-unit increase in the habitat variable corresponds to an $e^{n\beta i}$ increase in the odds ratio. Because the presence of a Blanding's turtle is a low probability event, the odds ratio estimates the relative risk, which is the ratio of the probability of x (the presence of a Blanding's turtle) given A (variables measured at the turtle's location) to the probability of x given B (variables measured at the random location) (Breslow and Day, 1980). Also, because the variables are differences in values between turtle and random observations, the model is interpreted as differences in the habitat and not as absolute measured values (Hosmer and Lemeshow, 2000; Compton et al., 2002). Positive coefficients represent increased selection toward that particular habitat type, whereas negative coefficients represent decreased selection toward that particular habitat type.

Another assumption of the logistic regression is that each observation is independent. Although using radio-telemetry as the sample unit causes pseudoreplication (Aebischer et al., 1993), it is difficult to avoid when few locations are taken for each individual. Microhabitat use data were collected for 11 adult females and 12 adult males. Although it would have been better to fit models separately for each individual, in many cases I had too few locations per individual to allow such analyses. Since no individual represented a large proportion of the total locations (average = 4.3%, max = 6.6%), however, no individual had the opportunity to unduly bias the group means.

Due to the small number of observations for each sex, I pooled the data from both sexes to increase the statistical power of my habitat selection models. I ran univariate analyses for each habitat variable and I selected variables with p-values less than 0.25 as candidates for subsequent multivariate analyses (Hosmer and Lemeshow, 2000). Candidate models were fit using a backward stepwise regression to select the most parsimonious model. I selected the final model based on the lowest Akaike's Information Criterion (AIC) score. AIC scores help to identify which model accounts for the most variation with the fewest variables and they are the most powerful approach currently available for model selection studies (Boyce et al., 2002). Lastly, I evaluated the fit of the each model using the likelihood-ratio statistic (LR) (Hosmer and Lemeshow, 2000).

Statistical Analyses

All statistical analyses were performed with JMP version 8.0 (SAS Institute, 2005-2010) and R version 2.7.2 (R Foundation for Statistical Programming, 2008). All means are reported ± SE and I considered tests significant at $\alpha = 0.05$. Data were transformed as needed to meet assumptions of homogeneity of variance, normal distribution, and sphericity (compound symmetry).

Results

I captured 91 Blanding's turtles (51 males and 40 females) from 2008-2009 and 38 individuals received radio-transmitters (1589 relocations). A one-way ANOVA revealed that males had significantly longer carapace lengths (23.0 ± 0.1 cm) than females (21.7 ± 0.2 cm; $R^2 = 0.29$, $F_{1,90} = 36.76$, $p < 0.001$). Males were also found to be significantly heavier than females (male = 1698 g ± 28, female = 1504 g ± 33; $R^2 = 0.18$, $F_{1,90} = 36.76$, $p < 0.001$).

Movement Patterns and Nesting Activity

A one-way ANOVA revealed that reproductive class had no effect ($R^2 = 0.07$, $F_{2,18} = 0.61$, $p = 0.55$) on the annual mean of the daily distance moved by turtles tracked in 2009. A two-way repeated-measures ANOVA revealed a significant interaction of month and reproductive class ($F_{6,30} = 2.533$, $p = 0.042$) on the monthly mean daily distance moved. Mauchly's test indicated that the assumption of sphericity was met ($X^2_5 = 9.08$, $p = 0.106$).

Four separate ANOVAs testing for the effect of reproductive class on mean DDM during each month revealed no significant effect of reproductive class on the DDM in May ($R^2 = 0.22$, $F_{2,12} = 1.37$, $p = 0.30$), July ($R^2 = 0.17$, $F_{2,12} = 1.01$, $p = 0.40$), and August ($R^2 = 0.11$, $F_{2,12} = 0.64$, $p = 0.55$). In June, however, reproductive class had a marginally significant effect on the DDM ($R^2 = 0.44$, $F_{2,18} = 5.4$, $p = 0.05$). A Tukey-Kramer HSD test indicated that gravid females moved significantly more than males in June 2009. However, there were no significant difference between non-gravid females and gravid females and between males and non-gravid females.

A total of sixteen nesting forays, defined here as movements to and from nesting areas, occurred between June 4th and June 25th in 2008 and between May 29th and June 27th in 2009. Thirteen females were radio-tracked to their nesting sites and four of the six gravid female in 2008 were also gravid in 2009. Furthermore, two non-gravid females in 2008 were gravid in 2009. Nesting sites were located 150 to 1850 m (mean of 919 ± 129 m) from the home wetland. During these nesting forays, females travelled a total distance of 650 to 3479 m (mean of 2095 ± 285 m), staying up to 15 days around the nesting area before travelling back to their respective wetlands. Several nesting locations were used by multiple females and females returned to the same nesting sites year after year, using the same upland corridors to reach their nesting sites.

Home Ranges

Home range size varied between individuals, ranging from 1.6 to 40.6 ha with a mean of 12.2 ± 1.8 ha. A one-way ANOVA revealed that reproductive class had a significant effect on the size of the annual home range of Blanding's turtles ($R^2 = 0.31$, $F_{2,36} = 7.64$, $p = 0.0018$). A Tukey-Kramer HSD test indicated that gravid females had significantly larger home ranges than both males and non-gravid females (Figure 1-5). There was no significant difference between home range size of non-gravid females and males (Figure 1-5). The mean home range size was 8.5 ± 1.7 ha for males, 7.3 ± 3.2 ha for non-gravid females, and 20.3 ± 3.5 ha for gravid females. The largest female home range, 40.6 ha, belonged to a nesting female in 2008 and the largest male home range was 32.9 ha.

Home range length also varied between individuals, ranging from 258.0 to 2046.0 m with a mean of 812.9 ± 78.1 m. A one-way ANOVA revealed that reproductive class had a significant effect on the length of turtle home ranges ($R^2 = 0.36$, $F_{2,36} = 9.55$, $p = 0.0005$). A Tukey-Kramer HSD test indicated that gravid females had significantly longer home ranges than both males and non-gravid females (Figure 1-6.). There was no significant difference between home range lengths of non-gravid females and males (Figure 1-6.). The mean home range length was 630.8 ± 79.7 m for males, 586.0 ± 130.5 m for non-gravid females, and 1210.9 ± 138.9 m for gravid females.

Microhabitat Analysis and Behaviour

I used data from 187 turtle locations (76 female locations, 111 male locations), with the 187 corresponding random paired locations for microhabitat characterization and analysis. The model with the lowest AIC value (AIC = 231.7509) had the variables H2O_TEMP, %FLOAT_VEG, %SUB_VEG and %OPEN_H2O and was statistically significant (LR = 35.5, $R^2 = 0.091$, $p < 0.0001$) (Table 1-2.). Based on the odd ratios, Blanding's turtles selected colder waters where both floating and submerged vegetation are abundant. A 1^{o}C increase in water temperature resulted in a 16% decrease in the probability of selection, a 25% increase in submerged vegetation resulted in a 70% increase in the probability of selection, a 25% increase in floating vegetation resulted in a 28% increase in the probability of selection and a 25% increase in open water resulted in a 31% decrease in the probability of selection (Table 1-3., Figures 1-7. and 1-8.).

Pearson Chi-square analysis revealed a significant difference in behaviour at telemetry location between reproductive classes ($X^2_{10,\,1561} = 64.36$, $p < 0.0001$) and between months ($X^2_{15,\,1561} = 158.11$, $p < 0.0001$). Turtles were most often found immobile underwater, usually under bog mats or floating vegetation (≥45%). Gravid females were located underwater less often than non-gravid females and males (Figure 1-3.) and the frequency of underwater turtle observation was higher in August, than any other month (Figure 1-4). Furthermore, gravid females were found basking at the surface of the water and in the forest more often than males or non-gravid females (Figure 1-3.). Basking behaviour was most often observed in May, thereafter decreasing in frequency (Figure 1-4.).

Discussion

Body Size

My population of Blanding's turtles is at the northern periphery of the species' range. Previous studies of freshwater turtles have shown that turtle body size may increase with latitude (Ashton and Feldman, 2003; Litzgus et al., 2004) and this is supported by my results. The mean body sizes of the Grenadier Island population of Blanding's turtles fall slightly above those of other more southerly populations (McMaster and Herman, 2000; Pappas et al., 2000; Rubin et al., 2004).

Several studies have also found that male Blanding's turtles have a significantly longer carapace than females (Graham and Doyle, 1979; Germano et al., 2000; Pappas et al., 2000). This held true for my population of Blanding's turtles, although the difference in mean carapace length (CL) between males and females was quite small (14 mm). This is consistent with past studies on Blanding's turtles, the difference in mean size between male and female CL ranged from 11.3 to 17.6 mm (Graham and Doyle, 1979; Germano et al., 2000; Pappas et al., 2000). This marginal difference between male and female CL puts into question the biological relevance of sexual size dimorphism in Blanding's turtles (Pappas et al. 2000), especially when compared to other freshwater turtle genera with true dimorphism (e.g., *Graptemys, Malaclemys, Pseudemys*; Ernst et al., 1994).

Nesting

Local information regarding nesting activity and nesting sites is crucial for the continuing protection of declining populations of Blanding's turtles. Nest fidelity is common in Blanding’s turtles (Congdon et al., 1983; Joyal et al, 2000; Standing et al., 1999, 2000; Innes et al., 2008) and nests are generally found closer to other bodies of water than to their residential wetland (Congdon and Rosen, 1983; Piepgras and Lang, 2000, Rowe and Moll, 1991). In my study, two gravid females returned to the same nesting sites in 2008 and 2009, using similar upland corridors. This return to previous nesting sites, some quite a distance from the residential wetland, could be explained by many different hypotheses, the two main ones being that these sites are simply good nesting sites or that these turtles are philopatric, they return to areas at or near where the gravid turtle itself hatched (Bernstein et al., 2007; Emrich, 1991).

Similar to past studies, females were found in small wetlands (i.e. staging areas) near to or adjacent to nesting areas prior to nesting (Innes et al., 2008; Grgurovic and Sievert, 2005; Congdon et al., 2000; Emrich, 1991). These adjacent wetlands provide moisture, protection (from both thermal stress and predators), and food (Emrich, 1991; Burger and Montevecchi, 1975). Furthermore, these areas could be critical to the survival of emerging hatchlings by providing a proximal source of food and protection from predators and further study on this subject is required.

Almost all nesting sites were on private land, often in or around the front lawns, gardens, or mulch piles of a cottage. Nest site fidelity, upland corridor fidelity, the use of wetlands adjacent to nesting sites, and the private ownership of most nesting sites raises concern for the conservation of Blanding’s turtle habitat. It is clear that gravid females extensively use unprotected habitat during nesting forays, the time during which they are the most vulnerable, and it is only by protecting these areas from further urbanization and fragmentation that we can assure the continued survival of this species.

Interestingly, two nesting females laid their clutch in at least two nesting episodes 1 day apart: they excavated and laid eggs in at least two different nests. This is the first observation of multiple laying of the same clutch of eggs. Furthermore, several females used the same nesting areas. One site originally discovered by Petokas (1986) was used by seven females, four of which did not have transmitters in 2008 and 2009. This phenomenon of aggregated nesting sites is common to other Blanding's turtle populations (Emrich, 1991; Congdon et al., 1983). The observation of communal nesting areas only increases the importance of protecting these areas.

Daily Movements

The daily movements of Blanding's turtles on Grenadier Island were almost three times longer than those documented in past studies (Table 1-4.; Innes et al., 2008; Piepgras and Lang, 2000; Sajwaj et al., 1998; Rowe and Moll, 1991; Ross and Anderson, 1990; Edge et al., 2010). Inter-study differences could be attributed to difference in tracking patterns, methodology, individual variation, sample size, variable climatic conditions, and local environmental factors. The greater vagility of gravid females in my study could be associated with the inclusion of nesting forays in the analysis, however all three reproductive classes had much higher than average daily movements. I found that differences in time between telemetry locations dramatically influenced the accuracy of daily movement estimates. In 2008, the average DDM by turtles in my study was 50.92 m per day and turtles were tracked 1-7 times per week. In 2009, however, when turtles were tracked 5-7 times per week, the average DDM was 180.48 m per day. These results suggest that, to estimate the daily movement patterns of Blanding's turtles properly, individuals must be tracked every day of the week, otherwise researchers run the risk of grossly underestimating the daily movement of this species. This same pattern is evident in the literature, the highest observed DDM were obtained in studies where turtles were tracked at least 4 times per week (Ross and Anderson, 1990; Rowe and Moll, 1991; Edge et al., 2010).

I found that the effect of reproductive class on the DDM varies depending on the month. The significant interaction I observed can be attributed to the significant difference in the DDM by gravid females, compared to males and non-gravid females in June. Unsurprisingly, gravid females traveled longer distances daily during the nesting season than any other time and they traveled significantly more than males. This follows the predictions derived from the reproductive strategies hypothesis. Furthermore, females spent a considerable amount of time around the nesting areas, spending up to 15 days around the nesting site before laying their last clutch and heading back to their residential wetland. These results suggest that inclusion of nesting forays in individual home range analysis is essential. Classifying nesting excursions and the area used during nesting excursion as outlying habitat, not part of the home range of the turtle, greatly underestimates the area needed by gravid females during a nesting year. Furthermore, most gravid females in 2008 were gravid again in 2009 and exhibited high nesting fidelity, suggesting that not only is the nesting habitat is required yearly, the same nesting habitat is used.

Sexual differences in movement are usually attributed to different reproductive strategies (Morreale et al., 1984; Gibbons et al., 1990; Ernst, 1986). Males tend to maximize their reproductive success by maximizing the number of eggs they are able to fertilize. Hence, increased movement should increase chances for copulation with several females and maximize fertilizations (Ernst, 1986; Morreale et al. 1984; Parker, 1984). Equivalent mean DDM between non-gravid females and males on Grenadier Island over the course of the study could be a result of spring and fall gatherings in overwintering areas. These biannual aggregations may decrease the benefit for males to travel farther and more frequently as they have access to several females ready to mate in a relatively confined area. In Minnesota, Blanding's turtle breeding behaviour began as soon as the ice started to melt and was observed in every month of the active season with peaks occurring in early spring and fall when turtles were concentrated at hibernacula (Congdon et al., 2008). Similar peaks in breeding behaviour prior and directly following hibernation have been observed in other populations of Blanding's turtles, including a population in Algonquin Park, Ontario (Edge et al., 2009, Newton and Herman, 2009). Several other studies observed a similar trend and when turtles aggregated during mating season no change in male movement rates were observed during the mating season (Jones, 1996; Litzgus and Mousseau, 2004; Carrière et al., 2009).

Home Ranges

In comparison with several other studies, the Grenadier Island population exhibited intermediate home ranges sizes and average home range lengths (Table 1-5.; Rowe and Moll, 1991; Piepgras and Lang, 2000; Hamernick, 2000; Grgurovic and Sievert, 2005; Schuler and Thiel, 2008). The large variation in home range size estimates among studies can be attributed to several factors including, but not limited to, methodological differences, turtle age and reproductive class, sample size, population density, and yearly climatic fluctuations. Furthermore, local differences in home range sizes may reflect habitat composition, availability, and variability (Bury, 1979), or resource distribution (Cagle, 1939). In a study by Rowe (1987) in Illinois, most Blanding's turtles established activity centers in "highly productive" ponds. In Minnesota, Hamernick (2001) attributed larger Blanding's turtle home ranges to the extensiveness of the aquatic habitat in Weaver Dunes, a characteristic shared with two other study sites (Ross and Anderson, 1990; Piepgras, 1998). Ross and Anderson (1990) reported that Blanding's turtles in central Wisconsin appeared to have larger activity centers in vicinities consisting of more available aquatic habitats of preference. Piepgras (1998) also concluded that a population of Blanding's turtles in Minnesota had large home ranges due to low population density, patchy resources, and widely dispersed, extensive habitat. Also, a previous study on the spatial ecology of the eastern musk turtle demonstrated that individual home ranges increased in areas of fragmented habitat versus contiguous habitats (Edmonds, 1998). The wetland complex on Grenadier Island has an extensive area of available contiguous aquatic habitat; this area is highly productive and is protected by provincial laws. It is important to note, however, that all turtle movements are constrained to the island, thereby limiting the maximal area available to each individual. These two factors combined help to explain why Blanding's turtles on Grenadier Island have intermediate sized home ranges, despite a relatively small study area.

Like past studies, I found no significant difference in home range size and length between males and non-gravid females and I attribute the significantly larger home range size and home range length of gravid females to their long distance nesting forays. It is unsurprising that home range length and size were not significantly different between males and non-gravid females as I also did not find a significant difference in the mean DDM between reproductive classes.

Microhabitat and Behaviour

My data indicated that, during the active season, Blanding's turtle habitat consisted of muck substrate, water depths of 1- 220 cm, cattails, sedges, grasses (predominantly wild rice), water-lilies, swamp loosestrife, and various submerged vegetation. This is consistent with past studies showing that Blanding's turtles inhabit productive, clean, shallow waters with abundant aquatic vegetation and soft muddy bottoms over firm substrates (Ernst et al., 1994). Blanding's turtles on Grenadier Island favoured colder water with abundant floating and submerged aquatic vegetation and avoided open water.

Blanding's turtles may be associated with filamentous algae and other submerged vegetation because they provide food and/or cover from predators (Hartwig and Kiviat, 2007). Submerged vegetation and filamentous algae are known to harbour high densities of macroinvertebrates (Evans et al., 1999, Yozzo and Diaz 1999), which are a primary food for Blanding's turtles (Lagler, 1943; Rowe, 1992; Sajwaj and Lang, 2000; Ernst and Lovich, 2009).

Blanding's turtles in Brooker's Pond also selected for greater floating vegetation and were often found basking at the surface of the water, surrounded by neuston and lilypads. Turtles in Minnesota (Sajwaj and Lang, 2000) and New York (Hartwig and

Kiviat, 2007), USA, exhibited a similar pattern. Hartwig and Kiviat (2007) proposed that Blanding’s turtles may conserve energy by simultaneously basking and foraging while in neuston. Neuston may provide camouflage for Blanding’s turtles (Hartwig and Kiviat, 2007; Ross and Lovich, 1992) and is an important source of food. Lilypads and duckweeds are important habitats for macroinvertebrates (Gaston, 1999) and crystalwort may also host macroinvertebrates (Hartwig and Kiviat, 2007).

Unlike the Blanding's turtles of Weaver Dunes, Minnesota (Hamernick, 2000), my individuals did not select for emergent vegetation. Because Blanding's turtles spent most (› 70%) of their time underwater (Figure 1-3), it is not surprising that selection occurred for water column properties (e.g. floating vegetation, submerged vegetation, and water temperature). Furthermore, I agree with Carter et al. (1999) that the selection of a particular habitat type by turtles may be more related to its associated structural characteristics, instead of vegetation type. In their study, Carter et al. (1999) found that bog turtles in southwestern Virginia responded more to structural habitat components versus actual vegetation type.

In association with the preference for submerged and floating vegetation, Blanding's turtles did avoid some habitat types, most notably open water. These results are consistent with the findings of Hamernick (2000) in Minnesota. I believe that the lack of basking sites and refuge from predators when in open water would make this habitat unattractive and potentially dangerous for turtles. Furthermore, open water typically lacks the abundance of potential food sources and complex structural characteristics that habitats with more vegetative biomass contain. My data suggest that the presence of vegetation at the edges of open deep-water areas increases the probability of a Blanding’s turtle occurrence.

Finally, I expected Blanding's turtles to select for warmer shallower water due to their ectothermic nature. Contrary to my predictions, Blanding's turtles selected cooler waters than the paired random locations. As the active season progresses, water levels fall and surface water temperatures dramatically increase. During the hot summer months, turtles may retreat below the surface to cooler wet areas and tunnels to reduce their body temperatures maintaining them within an optimal range that maximizes performance and fitness (Christian and Tracy, 1981; Huey and Kingsolver, 1989). This observed selection for colder closed waters could also represent a trade-off between less productive warm open waters and more productive colder closed waters. Forthcoming studies on the thermal ecology of this population of turtles should help to elucidate the reasons for this response (Graham, pers. comm.).

My microhabitat model for Blanding’s turtles did not show strong predictive power. This may be a consequence of Blanding’s turtles selecting different microhabitat features based on their current activity (e.g., foraging or thermoregulation) or their reproductive state (e.g., gravid or searching for mates). Stratification by activity and/or reproductive state could presumably increase the predictive power of this model (Belleau, 2008; Compton et al. 2002). Also, the predictor variables themselves might not be the right predictors for Blanding’s turtle habitat. Finally, Brooker’s pond is a relatively homogenous wetland that remains undisturbed by human activities and thus it remains possible that little selection is needed at the local scale in such a high quality habitat

(Edge et al., 2010). In this case, all habitats are likely represented in sufficient amounts and the preferred habitat is not a limiting resource (Edge et al., 2010).

Similar to Sajwaj and Lang (2000), as the active season progressed, turtles spent more time underwater and less time basking. Sajwaj and Lang (2000) observed that the proportion of thermoregulating turtles on sunny days dropped from >90% in April and May to 33% in July and August. As water temperatures increased, the thermoregulatory constraints on the turtles decreased. Thus, turtles could spend more time eating and searching for mates than thermoregulating in the latter half of the active season.

Interestingly, gravid females were more often found basking on floating vegetation and basking at the surface of the water than males and non-gravid females. This trend could be the manifestation of the higher energetic needs of gravid females. In a study on black rat snakes, another oviparous reptile, Blouin-Demers and Weatherhead (2001) found evidence that gravid females thermoregulated more carefully than other snake groups (non-gravid females and males) when females are developing and incubating eggs. My results contrast those of Piepgras and Lang (2000); they found little difference in the thermal behaviour and body temperature of male and female Blanding's turtles in the spring and summer.

The majority of the Grenadier Island wetland complex is afforded provincial protection; however, very little of the upland forested area is protected and an important percentage of the land on Grenadier Island is privately owned. I found that Blanding’s turtle, regardless of reproductive class, extensively used all regions of the study area, from the cattail strands on the outside perimeter of the island to a medium-sized fen 1.5 km from the main study site.

Terrestriality not associated with nesting is common among freshwater turtles (Gibbons, 1970). In our study, the adjacent upland forest was used on a few occasions by males and non-gravid females during the first two months of the active season (April to May). The motives for this behaviour are unknown. Terrestrial aestivation has been observed in several populations of Blanding's turtles and has been attributed to unfavourable environmental conditions, such as dropping water levels and extreme summer temperatures (Innes et al., 2008; Linck and Moriarty, 1997; Rowe and Moll, 1991; Ross and Anderson, 1990). Similar to the Blanding's turtles in Minnosota (Piepgras and Lang, 2000), I did not observe any terrestrial aestivation during the warm summer months, probably because summer temperatures were not uncomfortably warm in Ontario in 2008 and 2009.

Blanding's turtles were often found in dense cattail strands along Brooker's creek, a 400 m long creek that feeds into the St-Lawrence River. This is similar to the observations of Ross and Anderson (1990) in Wisconsin: wetlands covered with cattail stands were not used, but cattail stands with open water spaces were used. Rowe and Moll (1991) also observed sojourns into surrounding cattail areas, although they described this behaviour as transient and did not attribute a cause to these forays. This is surprising, as I found that turtles stayed in various cattails strands throughout the wetland complex for several weeks at a time, establishing well defined activity centers in these areas, both in the stands and in open water areas adjacent to the stands. I suggest that these areas are important foraging and basking sites for Blanding's turtles, although further studies are required.

Conservation Implications

These results have implications for the conservation of Blanding's turtles and, more broadly, for conservation planning in general. Wetlands with diverse and abundant floating and submerged vegetation and the terrestrial habitat up to 1.9 km away from such wetlands must be protected. When determining the size and shape of a protected area, conservation practitioners should keep in mind that individuals followed throughout this study required approximately 12 ha (with a maximum of 40.6 ha) to complete their annual activity cycle.

Interestingly, past studies on Blanding's turtles excluded nesting forays and other terrestrial movements from home range analysis (Schuler and Thiel, 2008; Grgurovic and Sievert, 2005; Hamernick, 2000; Rowe and Moll, 1991; Ross and Anderson, 1990). Although studying activity centers within a home range might reveal information about a turtle's daily activity patterns; I believe that all terrestrial movements should be included in home range analysis as these behaviours are integral parts of the annual life cycle of these animals and are essential to the survival of the population as a whole. Furthermore, as we have shown above, reproductive class and season greatly influences the behaviour and spatial ecology of this species. Thus, future studies should take into account these two separate effects when analyzing data (i.e. it is not enough to separate males from females and juveniles from adults, gravid females must be separated from non-gravid females). Although the population of Blanding's turtles on Grenadier Island inhabits, for the majority of the active season, a wetland complex that is protected by the St-Lawrence Island National Park, the majority of the upland forested areas, all of the nesting areas and the wetlands adjacent to the nesting sites are unprotected.

This study offers a small glimpse into the spatial ecology of Blanding’s turtles in Ontario. Given the large percentage of the global Blanding’s turtle range confined to Ontario, it is essential that the spatial ecology of this species is better defined in this province.

Table 1-1. Variables used to quantify microhabitat for Blanding's turtles on Grenadier Island, St-Lawrence Islands National Park, Ontario, Canada.

Variable	Classes[1]	Description
%EMERG_VEG	*Low* *Medium-Low* *Medium-High* *High*	Percentage of area with emergent vegetation.
%SUB_VEG	*Low* *Medium-Low* *Medium-High* *High*	Percentage of area with submerged vegetation.
%FLOAT_VEG	*Low* *Medium-Low* *Medium-High* *High*	Percentage of area with floating vegetation.
%OPEN_H_2O	*Low* *Medium-Low* *Medium-High* *High*	Percentage of area with open water. The area of open water within the sampling circle had to be continuous with an adjacent open water area of equal or greater width than 1m, creating a clear passage way for turtles.
H_2O_DEPTH		Distance (cm) between the water surface and bottom.
H_2O_TEMP		Water temperature (^oC) one foot below the surface of the water.
AIR_TEMP		Air temperature (^oC) thirty centimeters above the surface of the water.

[1] Low = 0-25%, Medium-Low = 26-50%, Medium-High = 51-75%, High = 76-100%.

Table 1-2. Matched-pairs logistic regression models explaining microhabitat use by Blanding's turtles on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada. The models with the lowest AIC scores and highest Akaike weights (ω) have more substantial support. The number of parameters (k) is also indicated.

Model	k	AIC	ΔAIC	ω
H_2O_TEMP, H_2O_OPEN	2	235.7767	4.0260	0.0517
H_2O_TEMP, H_2O_OPEN, %FLOAT_VEG	3	231.7882	0.0373	0.3801
H2O_TEMP, H2O_OPEN, %FLOAT_VEG, %SUB_VEG	4	231.7509	0.0000	0.3873
H2O_TEMP, H2O_OPEN, %FLOAT_VEG, %SUB_VEG, %EMERG_VEG	5	233.2741	1.5230	0.1809

Table 1-3. Coefficients ± SE and odds ratios for matched-pairs logistic regression model explaining microhabitat use by Blanding's turtles on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada.

Variables	**Coefficients**	**SE**	**Increase**	**Odds Ratio**	**95% C.I.***
H2O_TEMP	-0.171	± 0.072	1^{o}C	0.84	(0.73, 0.97)
%SUB_VEG	0.532	± 0.433	25% (1 level)	1.7	(0.73, 3.97)
%FLOAT_VEG	0.246	± 0.102	25% (1 level)	1.28	(1.05, 1,56)
%OPEN_H2O	-0.365	± 0.104	25% (1 level)	0.69	(0.57, 0,85)

*95% C.I. from odds ratios

Table 1-4. Review of literature on the daily distance moved ± SE (n) by male (M), gravid female (GF) and non-gravid female (F) Blanding's turtles. Results from statistical analyses looking at the effects of month and reproductive class (RC) on the mean daily distance moved are reported.

Investigators	Mean Daily Distance Moved (m/day) ± SE (n)			RC Effect	Month Effect
	M	F	GF		
Current study	199.42 ± 9.76 (9)	195.32 ± 15.57 (4)	249.50 ± 16.96 (6)	Interaction significant	
Edge et al., 2010	83.5 ± 39.9 (5)	91.0 ± 37.4 (16)		?	?
Innes et al., 2008[*]	27.55 ± 6.89 (4)	30.08 ± 14.15 (3)		?	?
Sajwaj et al., 1998	22.75 ± 8.34 (4)	41.75 ± 8.83 (4)		?	?
Rowe and Moll, 1991[**]	48.9 ± 20.85 (4)	32.4 ± 22.66 (3)		Yes	Yes
Ross and Anderson, 1990[***]	48.4 ± 8.99 (21)	95.1 ± 10.56 (56)		Yes	?

[*] The DDM was averaged over April-August in order to compare with other studies.

[**] The DDM was only calculated within activity centers.

[***] The minimum DDM was calculated.

Table 1-5. Review of Blanding's turtle home range sizes and lengths ± SE. (N) for males (M), non-gravid females (F), and gravid females (GF).

Investogators	Location	Method	Mean HR Size (Ha) ± Std. Err (N)				Mean HR Length (m) ± Std. Err (N)				Sign. Diff.?
			M	*F*	*GF*	*ALL*	*M*	*F*	*GF*	*ALL*	
Millar and Blouin-Demers, unpub.	Ontario	MCP	8.5 ± 1.7 (20)	7.3 ± 3.2 (5)	20.3 ± 3.5 (12)	12.2 ± 1.8 (37)	630.8 ± 79.7 (20)	586.0 ± 130.5 (5)	1210.9 ± 138.9 (12)	812.9 ± 78.1 (37)	Y
Egde et al., 2010	Ontario	MCP	57.1 ± 15.3 (5)	61.2 ± 30.4 (16)							
Schuler and Thiel, 2008	Wisconsin	MCP	28.1 (9)	20.7 (9)		25.5 (18)					N
Innes et al., 2008***	New Hampshire	MCP		6.8**(3)		3.3** (10)					
Innes et al., 2008***	New Hampshire	MCP	3.7** (4)	1.5**(3)							N
Grgurovic and Sievert, 2005	Massachussetts	Fixed Kernel (95%	27.5 ± 0.10 (14	19.9 ±0.07 (27)		22 ± 0.06 (41)	866 ±0.05 (19)	852 ± 0.04 (31)		856 ± 0.03 (50)	N
Hamernick, 2001	Minnesota	MCP	94.9 ± 58.4 (8)	60.7 ± 12.6 (16)		72.1 ± 20.6 (24)	1794.0 ± 547.7 (8)	1472.0 ± 191.3 (16)		1579 (24)	N
Piepgras and Lang, 2000	Minnesota	MCP	38.4	35.4						906	N
Rowe and Moll, 1991	Illinois	MPM*	1.4**(4)	1.2**(3)		1.3 ± 0.64 (7)	630 ± 304.9 (4)	800 ± 545.8 (3)			N
Ross and Anderson, 1990	Wisconsin	MPM*	0.76 ± 0.19** (2	0.56 ± 0.15** (4)							N

*MPM is equivalent to the MCP method

**Measure of activity center size: excludes nesting forays and long distance movements

***Median values instead of averages

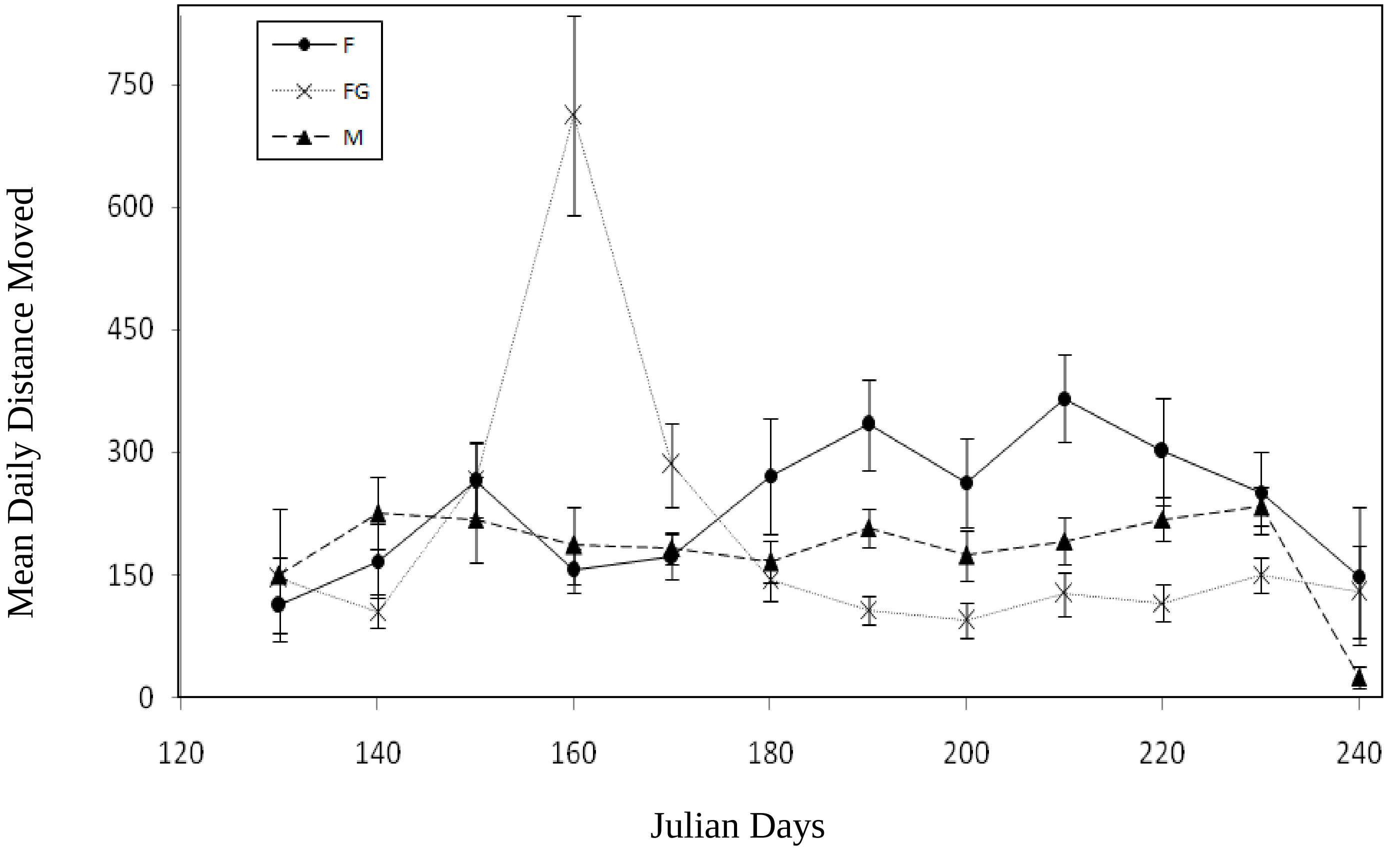

Figure 1-1. Mean distance moved per day at 10-day intervals ± SE for male (M), non-gravid female (F), and gravid female (FG) Blanding's turtles radio-tracked from May 2009 to August 2009 on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada. (n = 13: 4 FG, 3 F, and 6 M)

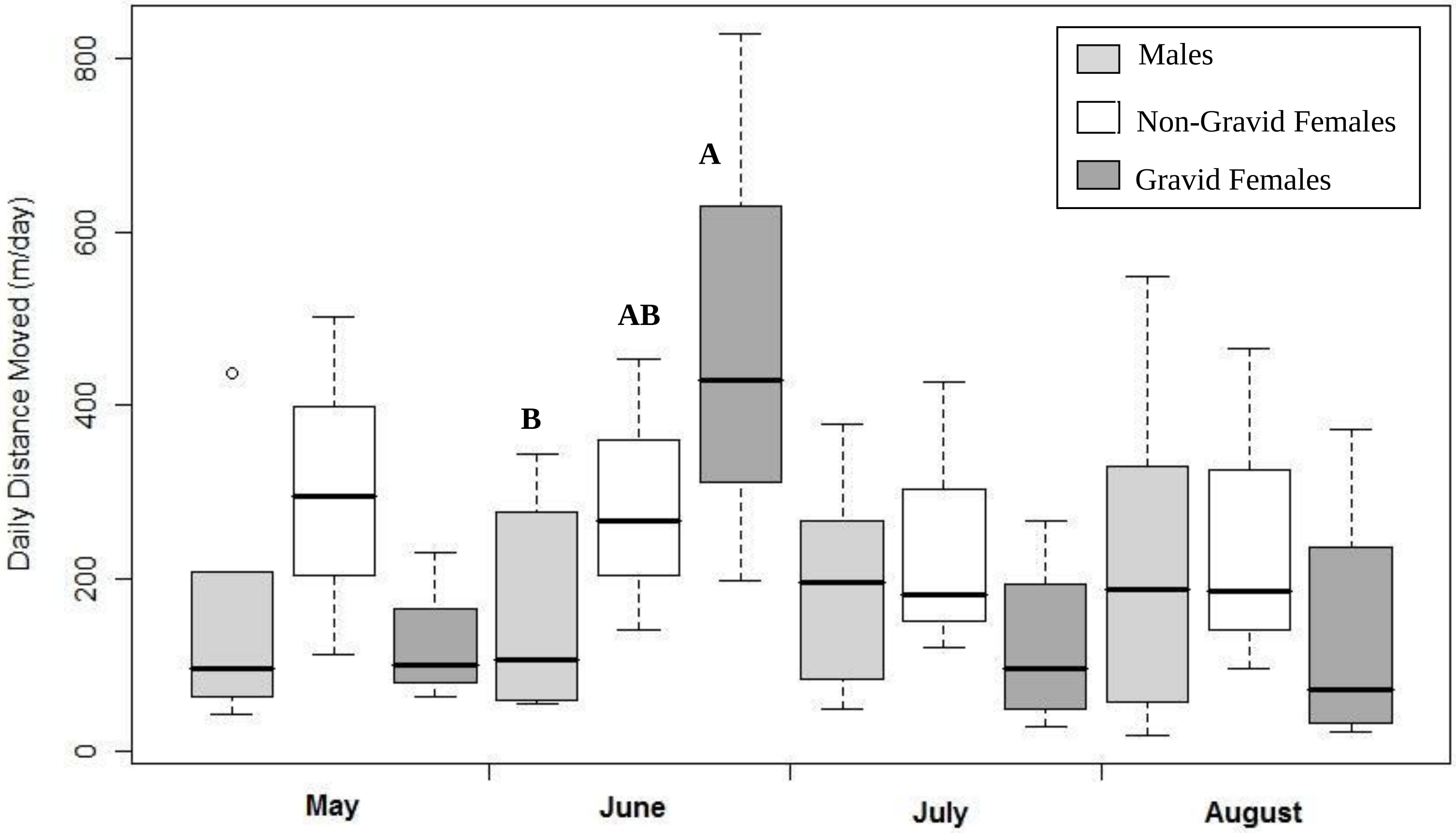

Figure 1-2. Box and whisker plots of the montly averages of the daily distance moved ± SE by male (M), non-gravid female (F), and gravid female (FG) Blanding's turtles radio-tracked from May 2009 to August 2009 on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada. The mean daily distance moved by gravid females in June 2009 was significantly greater than that of males in June 2009, paired letters indicate no significant difference. (n = 13: 4 FG, 3 F, and 6 M)

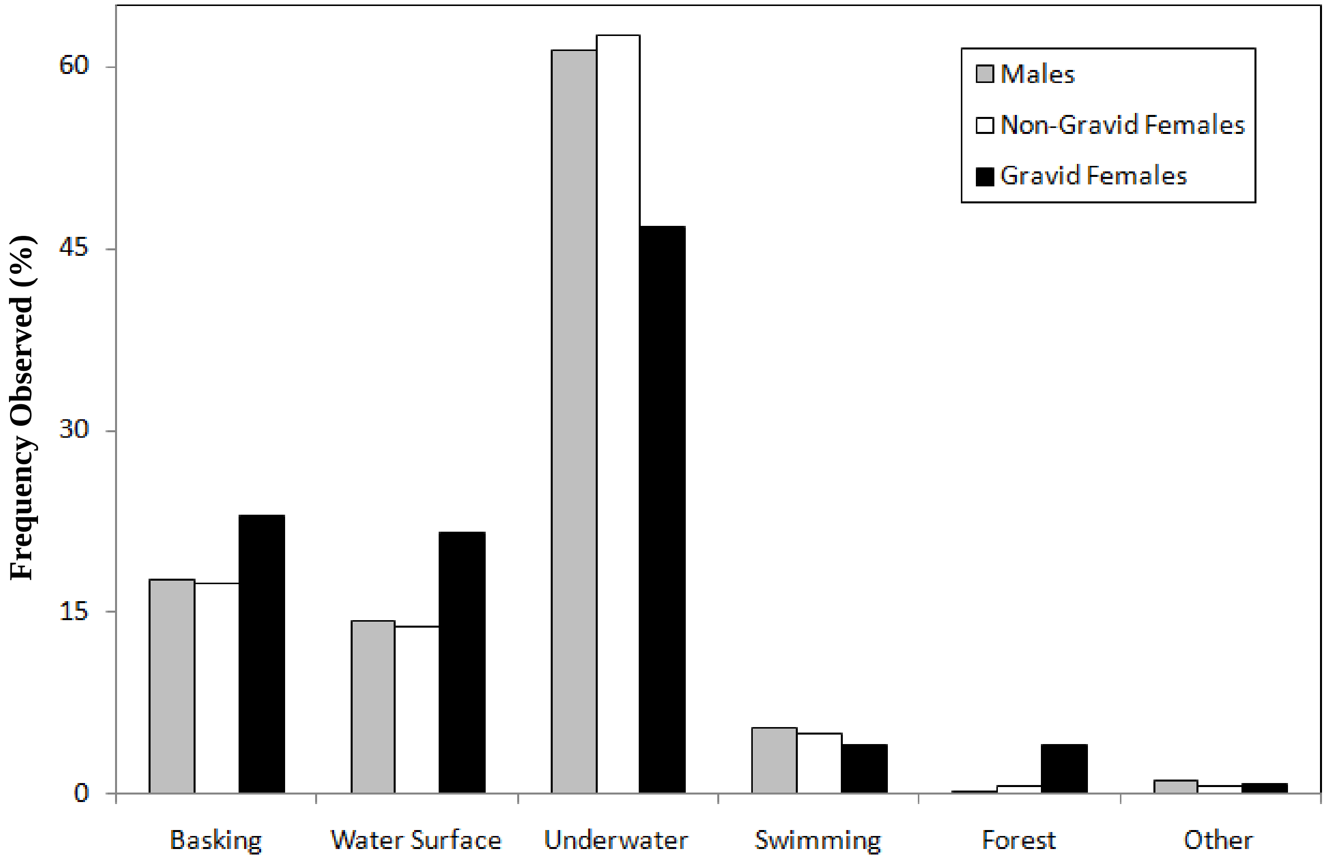

Figure 1-3. Frequency of observed turtle behaviour, by reproductive class, exhibited at relocation points (n = 1566) on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada. Behaviours described as “other” include copulating (n = 3), captured in hoop net (n = 6) or climbing a beaver lodge (n = 4).

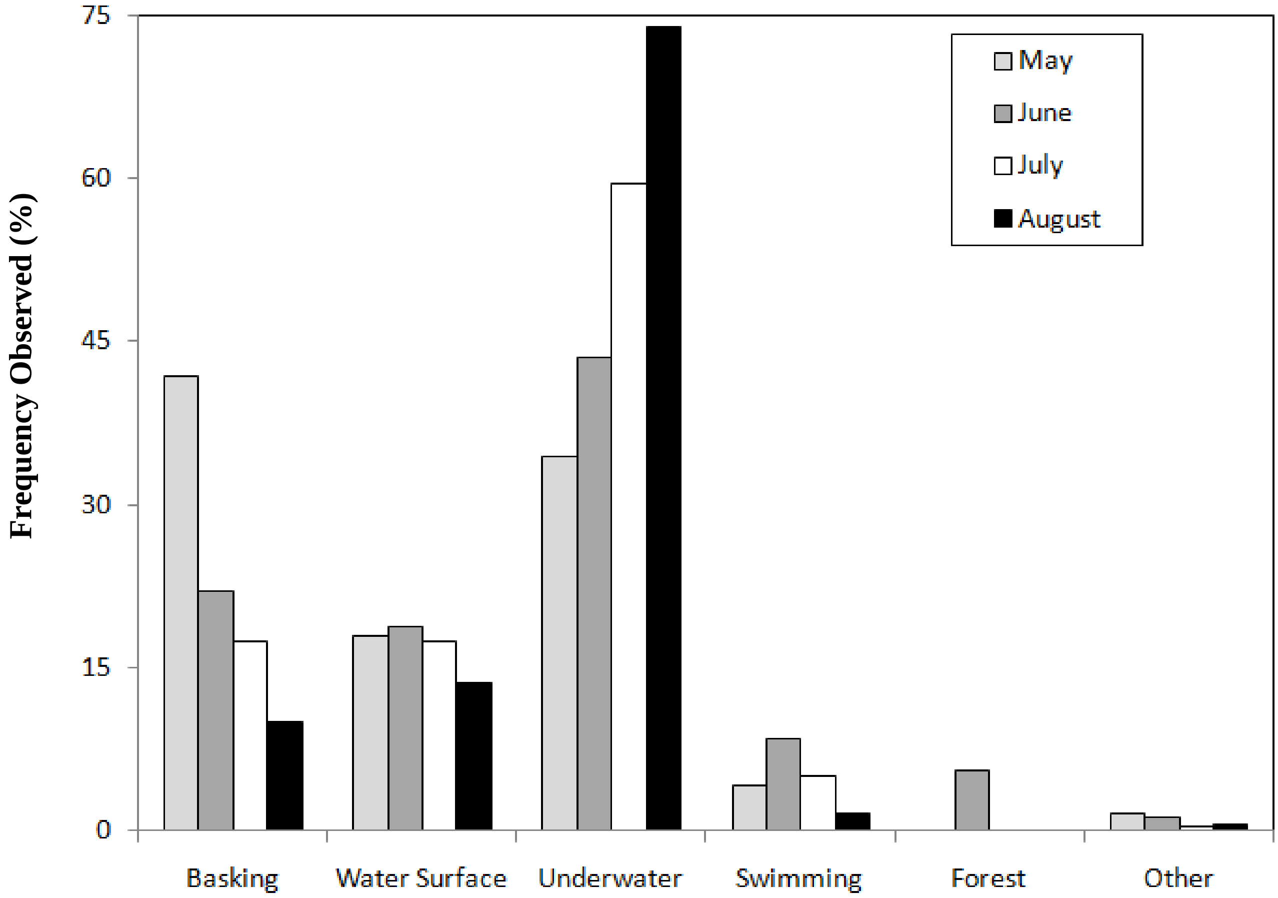

Figure 1-4. Frequency of observed turtle behaviour exhibited at relocation points (n = 1566) in May, June, July and August 2008 and 2009 on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada. Behaviours described as "other" include copulating (n = 3), captured in hoop net (n = 6) or climbing a beaver lodge (n = 4).

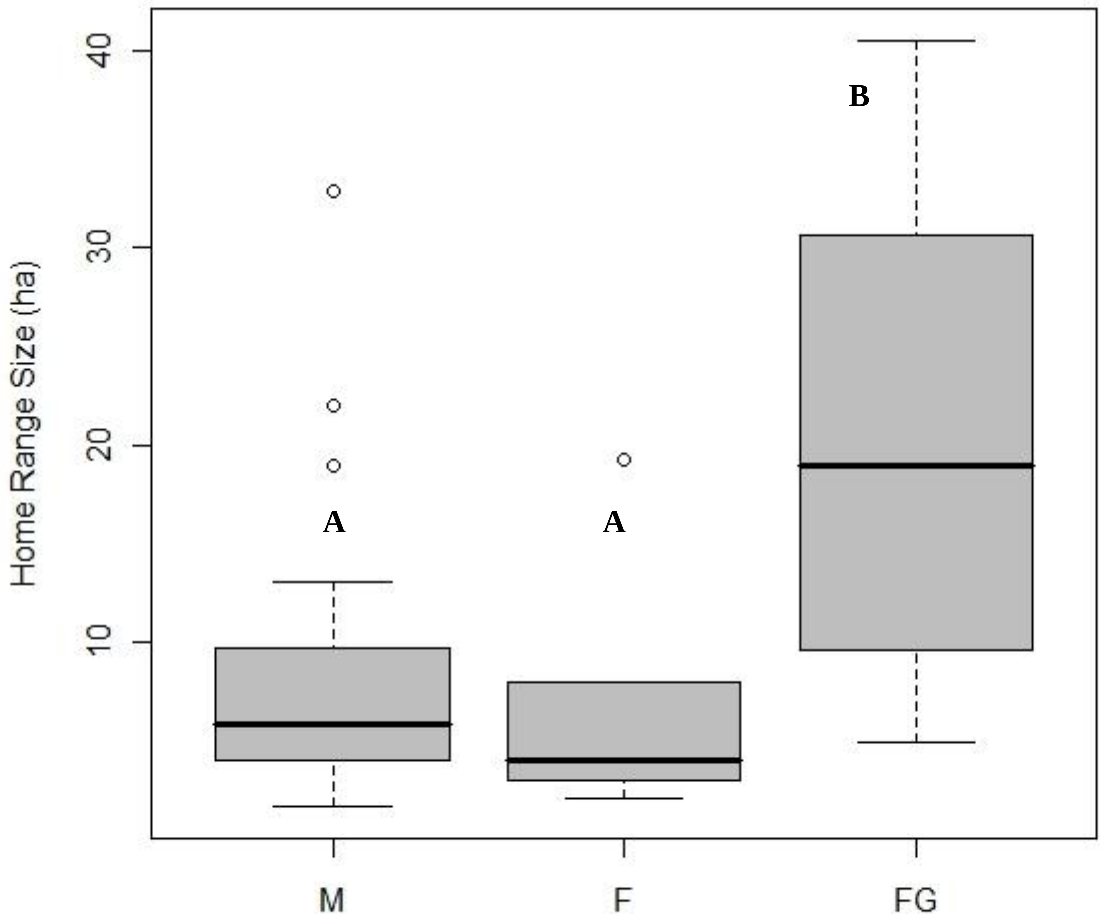

Figure 1-5. Box and whisker plot of the average home range size of male (M), non-gravid female (F) and gravid female (FG) Blanding's turtles tracked in 2008-2009 on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada. Reproductive classes with the same letters are not significantly different. (n = 36: 15 GF, 5 NFG, 20 M)

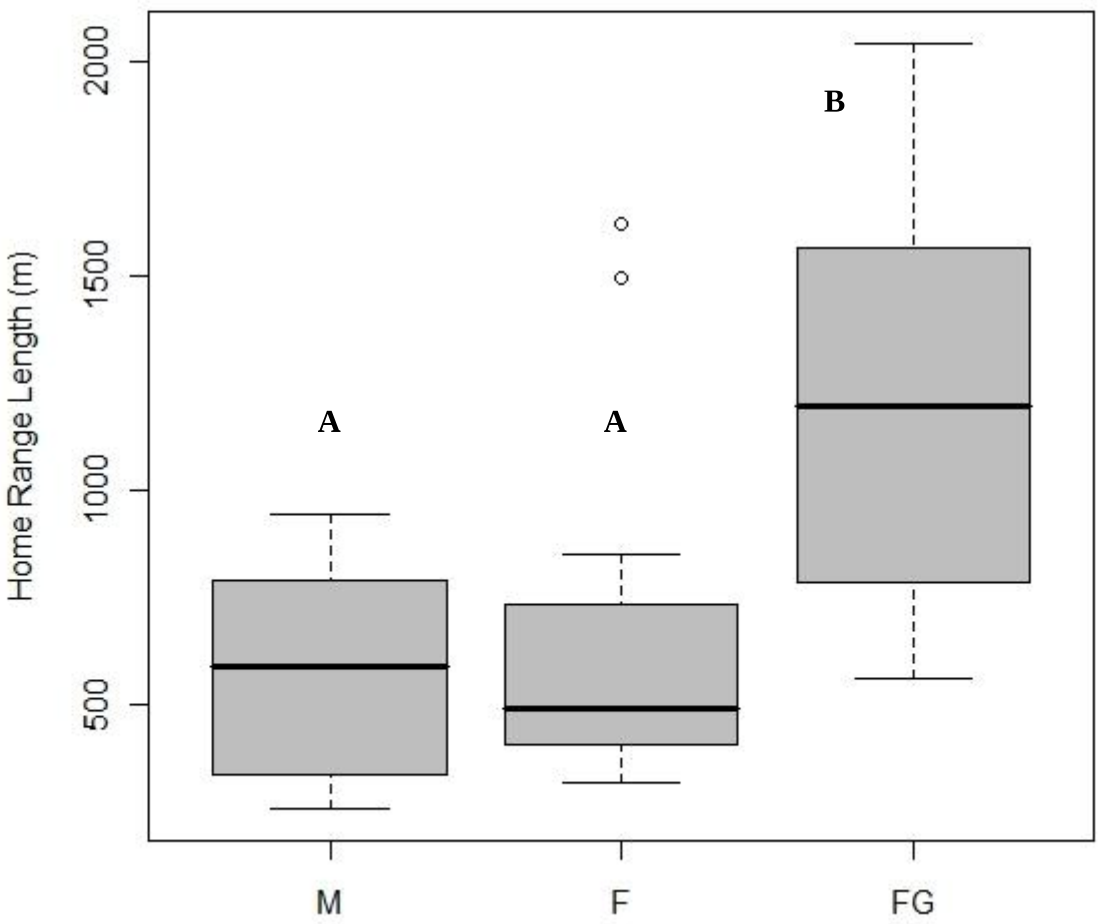

Figure 1-6. Box and whisker plot of the average home range length of male (M), non-gravid female (F) and gravid female (FG) Blanding's turtles tracked in 2008-2009 on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada. Reproductive classes with the same letters are not significantly different. (n = 36: 15 GF, 5 NFG, 20 M)

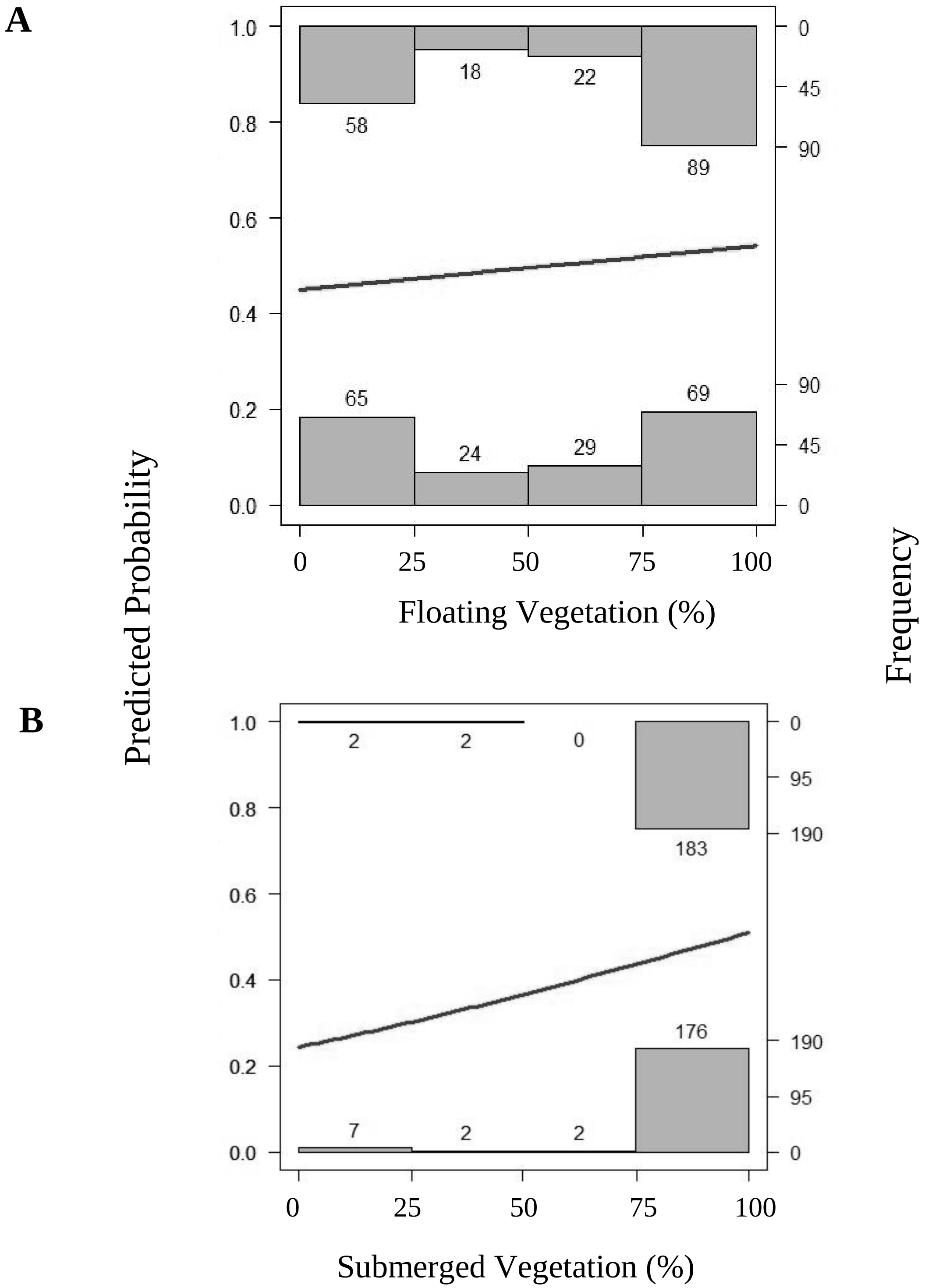

Figure 1-7. Frequency of observed data (histograms) and predicted probability of selection as the percentage of **A)** floating vegetation increases and **B)** submerged vegetation increases for Blanding's turtles (1 = habitat used by turtle, 0 = habitat at random locations) followed by radio-telemetry on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada.

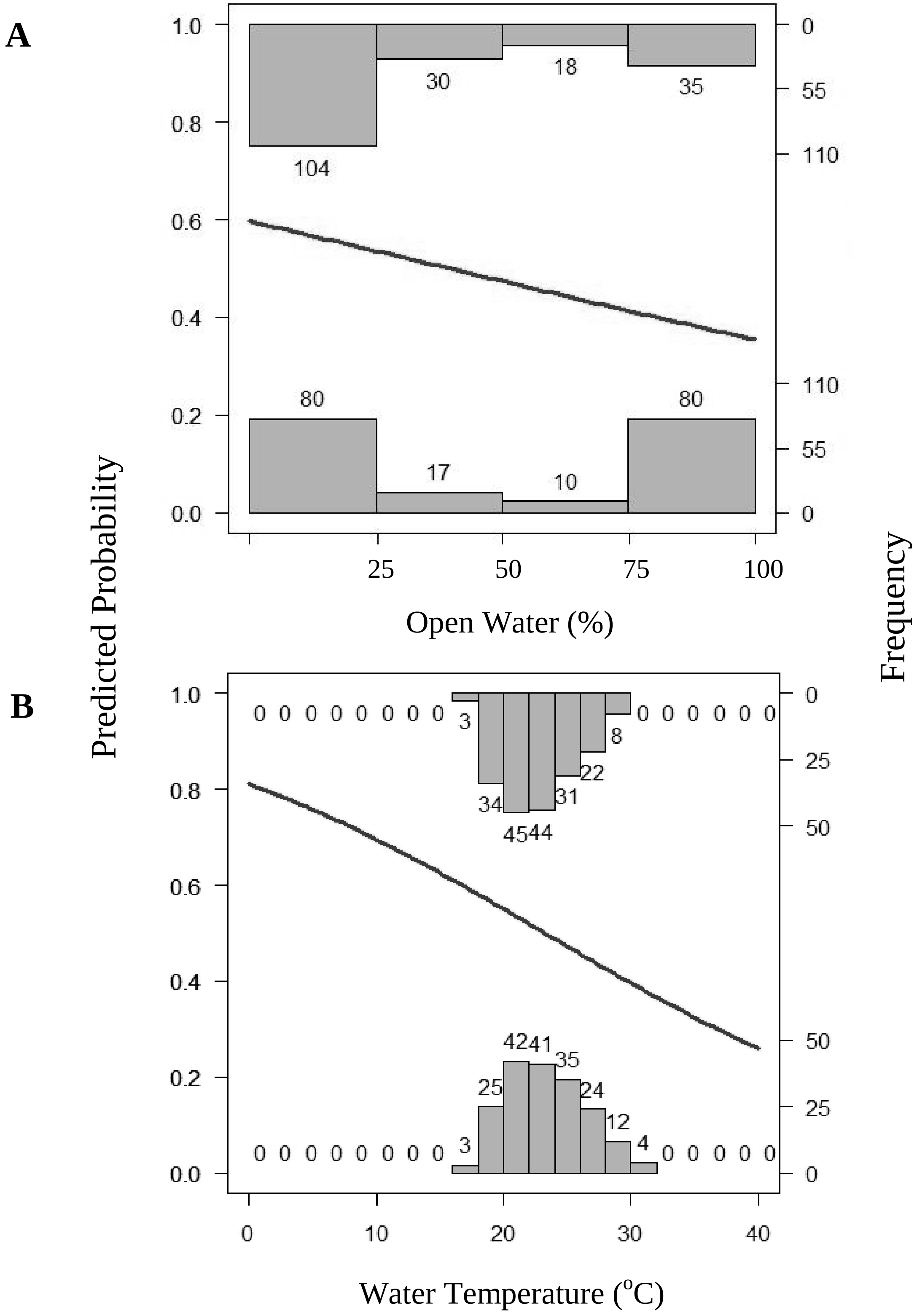

Figure 1-8. Frequency of observed data (histograms) and predicted probability of selection as **A)** the percentage of open water increases and **B)** water temperature increases for Blanding's turtles (1 = habitat used by turtle, 0 = habitat at random locations) followed by radio-telemetry on Grenadier Island, St. Lawrence Islands National Park, Ontario, Canada.

CHAPTER 2

Modeling habitat suitability for Blanding's Turtle at multiple spatial scales in Ontario, Canada.

Abstract

Species distribution models (SDMs) can be very helpful tools in the arsenal of conservation biologists because they can: 1) provide information on the ecological requirement of an organism, 2) facilitate fieldwork by predicting areas of potential occurrence, and 3) assess risk for invasive species or disease vectors. The goals of my study were fivefold: to determine what environmental factors best predict habitat suitability for Blanding’s turtles in Ontario, to determine if habitat suitability is hierarchical, to see if I could use historical sighting records to build strong predictive models, to compare and contrast the results of two different modeling algorithms, and to create high resolution habitat suitability maps for conservation practitioners. Thus, I built habitat suitability models for Blanding’s turtles in Ontario at three different spatial scales using historical Blanding’s turtle occurrence records and two SDM machine learning algorithms: boosted regression trees (BRTs) and maximum entropy modeling (Maxent). At all spatial scales, Blanding’s turtle habitat suitability was best predicted by mean monthly maximum air temperatures during the active season, wetland area, open water area, road density, and cropland area. Habitat suitability increased with increasing air temperature, wetland area, forested area, alvar area, bedrock outcrop area and decreased with increasing cropland area, pastures and field area, precipitations, terrain ruggedness, and settlements and developed land. Mean area under the receiver operating characteristic curve (AUC) values for habitat suitability models, tested on independent data, ranged from 0.878 to 0.912. By using multiple scales I was able to capture hierarchical landscape influences and produce accurate predictive models. These models provide more information on contributing factors for Blanding’s turtle habitat suitability in Ontario.

Introduction

“Where do species occur?” is a fundamental ecological question and prediction of occurrence is indispensable for much conservation and population management (Ferrier, 2002; Funk and Richardson, 2002; Rushton et al., 2004). For this reason, scientists have devoted decades of work to quantifying the relationship between the distributions of species and their abiotic and biotic environments. Over the past twenty years, predictive modeling of species distribution has become a central tool in ecology, biogeography, evolution, epidemiology, and conservation science (Corsi et al., 1999; Peterson and Shaw, 2003; Peterson et al., 1999; Scott et al., 2002; Welk et al., 2002; Yom-Tov and Kadmon, 1998). This is particularly true in the case of species of conservation concern. To help prepare schemes for mitigation of species’ decline or to create new populations through re-introduction, conservation officials need to know what biotic and abiotic variables determine species’ distribution within a landscape. Examples of predictive modeling used for conservation include reserve design (Araújo and Williams, 2000), invasive-species management (Peterson, 2003; Goolsby, 2004), effect of climate change on species distribution (Peterson et al., 2004; Araújo et al., 2004; Thuiller et al., 2005), population viability analysis (Boyce et al., 1994; Akçakaya et al., 2004), and species or resource management (Johnson et al., 2004).

SARA and Critical Habitat

In an effort to protect biodiversity, several nations have enacted legislations. In Canada, the Species at Risk Act (SARA, 2003) affords federal legal protection of species at risk on federal land. Similarly, Ontario's Endangered Species Act (OESA, 1971, 2007) provides provincial legal protection of species at risk on provincial land. Under both SARA and OESA, all species at risk must be given a recovery strategy (SARA, 2003; OESA, 1971, 2007) and this recovery strategy is the mechanism used to identify the critical habitat of a species at risk. A similar process is triggered in the USA under the Endangered Species Act (ESA, 1973, 1978). The critical habitat must be defined "to the extent possible" (SARA s. 41(1)(c) 2003) and all critical habitat will be afforded federal, and if in Ontario, provincial protection. Recovery strategies thus become the main tools for the management and protection of listed species and must be appropriately suited to each species' situation and biology to be effective (Boersma et al., 2001). In general, studies on species at risk focus on specific populations, thus requiring a great expenditure on behalf of the inquiring bodies for little gain in understanding of the spatial/environmental needs of the species as a whole (see chapter 1). Therein lays the attractiveness of habitat suitability models for species at risk: for comparatively little expense, range wide information on the habitat needs of a species can be defined and areas of suitable, perhaps critical, habitat delineated. Habitat suitability modeling does have its shortcomings, however, and in no way can it replace detailed thorough analysis at the population and individual level. Habitat suitability modelling is simply a new, accessible, and very powerful tool in the arsenal of conservation biologists.

Orders of selection

Selection can occur at multiple spatial scales, from geographical distribution (landscape), to the composition of home ranges within a landscape (macrohabitat), to specific areas within the home range (microhabitat) (Mayor et al., 2009). These three scales relate to the third, second, and first order of selection of Johnson (1980). Differences in selection pressures and limiting factors can sometimes lead to differing (Luck, 2002; Orians and Wittenberger, 1991) and conflicting (Compton et al., 2002) patterns of selection at multiple scales. Johnson observed that selection need not occur at all scales and that different species respond to different scales of selection, thus the increased incidence of multiple-scale habitat selection studies (Morin et al., 2005; Compton et al., 2002; Luck et al., 2002). Furthermore, because the influence of each environmental variable in determining a species' niche is scale dependent, different degrees of ecological niche variation can arise among populations, depending on the spatial resolution of analyses (Wiens, 1989). It is only when all elements of selection are examined at multiple scales and resolutions that management actions will clearly reflect the needs of the species (Mayor et al., 2009; Morin et al. 2005).

Habitat Suitability Modeling and Presence-Only Data

There are several frameworks available for habitat suitability modeling (HSM) including Resource Selection Functions (Boyce et al., 2002), Ecological Niche Modeling (Peterson, 2001), Gradient Analysis (Whittaker, 1967), and Species Distribution

Modeling (Guisan and Thuiller, 2005; Elith and Leathwick, 2009). The latter of these HSM frameworks, species distribution models (SDMs), are numerical tools that combine observations of species occurrence or abundance with environmental estimates (Elith and Leathwick, 2009). All SDM studies have three basic components: (1) a data set describing the incidence or abundance of the species of interest and a data set of putative explanatory variables, (2) a mathematical model that relates the species data to the explanatory variables, and (3) an assessment of the utility of the model developed in terms of a validation exercise or an assessment of model robustness.

The data available for modeling the geographic distribution of a species can vary in extent, type, and quality, as do the type of models available. Logistic models are typically used to estimate species distribution, ideally based on high quality presence-absence data (Elith et al., 2006; Ward et al., 2009). While vast stores of presence-only data exist (occurrence records generally collected by herbaria and museums), absence data are rarely available or reliable (Anderson et al., 2003). Therefore, modeling techniques that require presence data only are very valuable (Graham et al., 2004). The modeling techniques that do not require explicit absences can be divided into two classes: presence-only (i.e. HABITAT; Walker and Cocks, 1991) and presence/pseudo-absence (Elith et al., 2006; Ward, 2009). The presence/pseudo-absence models require a background sample consisting of random samples of locations taken from the region or landscape of interest (Ward et al., 2009). In a review of modeling techniques that do not require explicit absences, modeling algorithms using presence/pseudo-absence data consistently outperformed modeling algorithms using presence-only data (Elith et al., 2006).

Status of Reptilian Populations

According to the IUCN (the World Conservation Union), 51% of reptilian species evaluated by the IUCN worldwide are considered to be at risk. In Canada alone, eight of the twelve species of freshwater turtles are listed as species at risk by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC). According to these COSEWIC statistics, reptiles are proportionally the most at risk vertebrate group in Canada. Habitat loss and degradation is the largest single factor responsible for reptile loss (Gibbons et al., 2000; Shine 1991), closely followed by road mortality (Gibbs and Steen, 2005; Aresco, 2005; Gibbs and Shriver, 2002; Forman and Alexander, 1998). Even in cases where the jurisdictional wetland is protected, the surrounding terrestrial habitat needed by semi-aquatic reptiles for nesting, hibernation, and other refugia may not (Burke and Gibbons, 1995).

The Blanding’s turtle is a medium-sized freshwater turtle, it has one of the smallest global ranges of any North American turtle (COSEWIC, 2005), and is listed as a species at risk in 17 of the 18 state/provincial jurisdictions across its range (NatureServe, 2009). Ontario has a significant responsibility towards the conservation of this turtle as approximately 20% of its global range is contained in this province (COSEWIC, 2005). Unfortunately, the Great Lakes population of Blanding’s turtles is listed as a threatened and thus requires critical habitat delineation.

Objectives

The purpose of my study was to determine what environmental factors best predict habitat suitability for Blanding's turtles in Ontario, the direction of selection for these variables, and how these relationships change at different spatial scales. Using Blanding's turtle sighting data from the National Heritage Information Center in combination with two machine learning modeling methods, boosted regression trees (BRTs) and maximum entropy modelling (Maxent), I created habitat suitability models for Blanding's turtles in Ontario at three different spatial scales. These three spatial scales are represented by buffers of different sizes around each sighting point: 1 km (maximum mean annual home range radius), 500 m, and 250m (maximum mean daily movement). I also produced several maps, at a resolution of 100 m, predicting habitat suitability for Blanding's turtle across Ontario.

Methods

Data Sources and Study Area

Sighting Data

I compiled all occurrence records, with an accuracy better than 100 m, of Blanding's turtles in Ontario from the Natural Heritage and Information Center (Oldham, and Weller, 2010). These historical sighting records were collected over the past 40 years and are voluntarily reported to the NHIC on an ongoing basis. Within the subset of records with an accuracy better than 100 m, peaks in sightings occur in 1985-1990 and 2002-2003. From this subset of 1035 sightings points, I randomly selected the maximum number of non-overlapping points within a 1 km radius using the Focus tool (Holland et al., 2004) in an effort to reduce spatial autocorrelation. This left me with a total of 616 sightings points, 60% of which were used to build the models and 40% were kept for independent evaluation of the models.

Environmental Predictor Variables

The environmental predictor variables I used fall into four categories: (1) ecological, (2) climatic, (3) topographical, and (4) anthropological (Table 2-1.). Multicolinearity should not be a problem as all 14 environmental predictors were not highly correlated (all $r < 0.8$, mean $|r| = 0.125$) at all spatial scales examined.

I obtained all my land cover data from the Ontario Land Cover Dataset at a resolution of 25 meters. The Ontario Land Cover data were derived from digital, multispectral LANDSAT Thematic Mapper data recorded on a range of dates between 1986 and 1997, but the majority of the satellite data frames were recorded in the early 1990s. The database was produced in nine segments under three separate programs of the Ontario Ministry of Natural Resources (OMNR) between 1991 and 1998 and was rationalized in 1999 into one consistent land cover dataset with 28 classes covering the entire province (OMNR, 1998). The variables derived from the 28 land cover types can be found in Table 2-2. In an effort to reduce overfitting of models to training data, wetland categories were combined.

I measured terrain ruggedness using the terrain ruggedness index (TRI) developed by Riley et al. (1999). Using a moving-window routine, the TRI script calculates the sum of changes in elevation between a grid cell and its eight neighborhood grid cells (3 x 3 neighborhoods) (Riley et al., 1999). Index values are low in flatter areas and high in both steep areas and in steep and rugged areas (Riley et al., 1999; Sappington et al., 2007). I ran the TRI script in ArcInfo Workstation (ESRI, 1982-1996) using the Ontario Provincial Digital Elevation Model (DEM) v.2.0.0 at 25m resolution (re-sampled from a finer resolution using bilinear interpolation) (OMNR, 2006). Thus terrain ruggedness was calculated for each $5625 m^2$ neighbourhood.

I obtained my road layers from the Ontario Road Network (OMNR, 2006) and I calculated road density as the length of road segments (m) within a given buffer area divided by the area of land contained in the buffer (m^2).

I downloaded mean maximal monthly temperature and mean monthly precipitation rasters from the WorldClim online database (Hijmans et al., 2005). These data have a spatial resolution of 30 arc s (approximately 1 km^2) and were generated through interpolation of average monthly climate data from weather stations around the world from 1950-2000. For both temperature and precipitation, I averaged the values over the active season (April - October) to get an average of the monthly maximal temperature during the active season and an average of the monthly precipitations during the active season for Blanding's turtles.

When selecting environmental variables for model building, it is important that several implicit ecological assumptions are met. Firstly, there should be a temporal correspondence between the occurrence localities and the environmental variables (Anderson and Martínez -Meyer, 2004). Since I am using data that spans several decades, the environmental variables I selected must span the same time frame. For this reason, variables such as terrain ruggedness and broad scale climatic data are relevant as these conditions are not likely to change markedly in thirty years. The use of land cover data is a little trickier as it is subject to greater change. The land cover dataset I used, however, was derived from satellite images taken from the 1980s to late 1990s, and thus should reflect conditions similar to when the sightings took place.

Secondly, the variables should affect the species' distribution at the relevant scale, determined by the geographic extent, resolution of the data, and the type of modeling technique used (Pearson et al., 2004; Elith and Leathwick, 2009). As Rushton et al. (2004) put it: "In an ideal world, the target species would be sedentary at a fixed point in space, and its ecological requirements well known and measurable at the same spatio-temporal scales". In reality, species vagility is highly variable, ecological requirements are generally unknown, and some variables are simply not measurable.

To take into account the vagility of Blanding's turtles, I created three buffers around each sighting point. The environmental data were thus read as the percentage of *x* cover within the buffer area or the mean value of *x* within the buffer area, depending on the nature of the variable analyzed. The radii of these three circular buffers corresponded to a rounded up value of the highest known value for the mean daily distance moved by Blanding's turtles (approximately 250 m per day; Chapter 1), an intermediate value of 500 m, and a rounded up value of the highest published value of the average length of the annual home range of Blanding's turtles, divided by two (approximately 1 km; Hamernick, 2000). I interpreted these three scales of selection as local/daily, intermediate, and annual habitat requirements.

Study Area

The spatial extent selected for modeling can heavily impact a model's predictive power. If presence/abundance records are drawn from too small a geographic area, the spatial variation in the data might be insufficient for the model to detect selection for key ecological variables. In a larger study region, spatial variation exits in community composition as well as in environmental conditions available to the species. Therefore, modeling in a large study region, given sufficient sampling effort, is preferable (Phillips et al., 2006). I believe there should be sufficient spatial variation in community composition and environmental conditions in Ontario for selection to be observed.

It is important to note that I clipped the provincial data 1 km north of the most northerly sighting point for Blanding's turtles and that I restricted all the data to the same geographic area. I removed areas that I judged inaccessible to Blanding's turtles from all environmental datasets. These sets of disregarded pixels included all open water areas in the Great Lakes further than 300 m from land. These areas were also removed from the habitat suitability maps. All environmental data layers had, or were re-sampled to, a pixel resolution of 25 m for model building and 100 m for predictions to grids (habitat suitability mapping). This made for a total of 743 495 780 pixels for model building (25 m resolution) and 46 467 100 pixels for predicting habitat suitability across Ontario (100 m resolution). For both Maxent and BRTs, the model predictions are given in logistic format and can be interpreted as the predicted habitat suitability for Blanding's turtles, ranging from 0 to 1.

Model Building

In a study by Elith et al. (2006) comparing 16 modeling methods over 226 species from 6 regions of the world, machine-learning methods consistently outperformed more established methods. Modelling algorithms that characterized the background environment and differentially weighed contributing variables vastly outperformed those that use presence-data alone (i.e. BIOCLIM, LIVES) (Elith et al., 2006). Furthermore, the machine learning algorithms shared a high level of flexibility in fitting complex responses (i.e. they could model complex interactions among variables) and this high degree of expressiveness was controlled by different “regularization” techniques (Elith et al. 2006; Elith et al., 2008; Phillips and Dudík, 2008). This balance between high expressivity/complexity and parsimony is believed to be the main reason for the high performance of these modelling techniques when using presence-only data (Elith et al., 2006). For this reason, I chose to use two machine learning methods which were among the top performing models in the Elith et al. (2006) review: maximum entropy modeling (Maxent) and boosted regression trees (BRTs). Since both of these modeling techniques require data akin to absences, I generated a random sample of 16 666 background sites (pseudo-absences) from the environmental data (minimum of 8000 background sites; Phillips, 2008). I then used sixty percent of these points for model building (10 000 points) and forty percent for model testing (6 666 points).

BRT combines two algorithms: the boosting algorithm that iteratively calls the regression-tree algorithm to construct a combination of trees (Elith et al., 2008). Multiple regression trees are constructed by iteratively fitting new trees to the residual errors of the existing tree assemblage (De’ath, 2007). Existing trees are not changed through iterations, and the final model is a linear combination of all the trees in the assemblage (Elith et al., 2008). Regression trees are used because they are good at selecting relevant variables and can model interactions and boosting is used to overcome the inaccuracies inherent in a single tree model (Elith et al., 2008).

Maxent estimates species’ distributions by finding the distribution of maximum entropy (i.e. closest to uniform) subject to the constraint that the expected value of each environmental variable (or its transform and/or interaction) under this estimated distribution matches its empirical average (Phillips et al., 2006). This is achieved through an iterative algorithm, which begins with a uniform distribution with zero gain. The gain is a measure of likelihood of the samples and it expresses how much higher the average sample likelihood is compared to that of a random background pixel. The algorithm sequentially updates the weights of predictor variables until it converges to the optimum potential distribution and the gain of the final model can be interpreted as a measure of how much better the predicted distribution fits the sample points compared to the theoretical uniform distribution (Phillips et al., 2004; 2006; Phillips and Dudík, 2008).

For a more detailed introduction to these two statistical models, see Appendix I.

BRTs

All BRT models were fitted in R version 2.7.2 (R Development Core Team, 2008), using the gbm package version 1.5-7 (Ridgeway, 2006) and additional codes (Elith et al., 2008). Based on preliminary tests and past studies, I built my models using a tree complexity of 5 and a bag fraction of 0.5 (Elith et al., 2008) (Appendix I). For all models, I used 10-fold cross validation on my training data to determine the optimal number of trees (*nt*) and picked the fastest learning rate (*lr*) that reached at least 1000 trees (Elith et al., 2008).

Using a method analogous to backward stepwise selection in regression (Miller, 1990), recursive feature elimination, I dropped unimportant variables from my models using test AUC values as an evaluation metric (Appendix II). I progressively removed each of the least important variables, re-ran my model after each removal, and tested my simplified model's predictive power on independent data. Using test AUC values, I picked the simplified model that had the best predictive power when tested on independent data.

Maxent

All Maxent models were run in Maxent version 3.3.1 (Phillips et al., 2006; http://www.cs.princeton.edu/~schapire/Maxent) using the default parameters. These default parameters have been shown to be well suited for a wide range of presence-only datasets (Phillips and Dudík, 2008), most notably datasets with 11 to 13 environmental variables. Similarly to the datasets used in the Phillips and Dudík (2008) study, my dataset contains 14 environmental variables and > 100 occurrences.

Models were built and tested on 10 replicate random partitions of the presence and pseudo-absence data. As before, these sets consisted of 60% training data and 40% testing data. To facilitate comparison between the two modeling methodologies, the same presence and pseudo-absence localities used for BRT models were used in Maxent. I again used recursive feature elimination to drop unimportant variables from my Maxent models (Appendix II).

Model evaluation

ROC and AUC

A common approach to evaluate and compare models of species distributions is the receiver operating characteristic curve (ROC) (Hanley and McNeal, 1983). The main advantage of ROC analysis is that the area under the ROC curve (AUC) provides a single measure of model performance, independent of any particular choice of a threshold. The ROC curve is obtained by plotting sensitivity on the y-axis and 1-specificity on the x-axis for all possible thresholds. Here, the sensitivity of a classifier for a particular threshold is the fraction of all positive instances that are classified as positive, while specificity is the fraction of all negative instances that are classified as negative. Sensitivity is also known as the true positive rate, and represents the absence of omission error. The value of 1-specificity is also known as the false positive rate, and represents commission error. Thus, the area under the ROC curve is the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one (Fielding and Bell, 1997). With presence-only data, however, we do not have true absences and must consider the AUC statistics in another light. In this case, negative instances represent random locations within the study area. As in Wiley et al. (2003) and Phillips et al. (2006), we used pseudo-absences in place of absences in the ROC analysis. In other words, AUC can be interpreted, when using presence/pseudo-absence data, as the probability that when we randomly pick one positive and one random example, the classifier will assign a higher score to the positive example than to the random (Phillips et al., 2006). Models with AUC values above 0.75 are considered useful and models with AUC exceeding 0.9 are considered excellent (Elith, 2002; Swets, 1988). It is important to note that with presence-only data, the maximum achievable AUC is less than 1 (Wiley et al., 2003); however random prediction still corresponds to an AUC of 0.5.

Biserial Correlation Coefficient

The point biserial correlation is similar to AUC but provides additional information on the distribution of predictions (Elith et al., 2006; Murphy and Winkler, 1992). This value represents the correlation between the occurrence records and the model predictions and can be calculated as the Pearson correlation coefficient (COR; (Zheng and Agresti, 2000). The higher the correlation value, the more discriminative the model is.

Response to Predictor Variables

One of the most important steps of model interpretation is determining which variables are important for prediction, and understanding their joint effect on the response. For this reason, I determined the relative contribution of each input variable in predicting the response for all models.

With BRTs, the importance of each input variable is based on the number of times that variable was selected for splitting in the tree weighted by the squared improvement to the model as a result of each of those splits. This importance measure is then averaged over all trees (Freidman & Meulman, 2003). The relative influence (contribution) of each variable is then scaled so that the sum adds to 100, with higher numbers indicating stronger influence over the response.

With Maxent, the percent contributions are determined by the increase in the gain of the model due to the modification of a coefficient for a single feature. Each step of the Maxent algorithm increases the gain of the model by modifying the coefficient for a single feature; the program then assigns the increase in the gain to the environmental variable(s) that the feature depends on, converting to percentages at the end of the training process. To integrate information provided by both models at each geographic scale, I also averaged the overall contributions of each variable at all three spatial scales.

To investigate the effect of each environmental variable on the response after accounting for the average effect of all other variables in the model, I examined the partial dependence plots. These partial dependency plots provide a useful basis for interpretation; however if strong interactions are present or if predictors are strongly correlated, these plots might not accurately reflect the true impact of each predictor alone on the response. For this reason, I also created precise dependence plots for the top five contributing variables by running univariate models for each variable at each spatial scale, using both algorithms.

Results

Model Performance

All models performed very well and closely fitted the presence points of Blanding's turtles in the study area, as suggested by both the training and test AUC values (Table 2-3.). These AUC values ranged from 0.878 to 0.912 on test data and 0.904 to 0.960 on training data. Test AUC values in BRT models and Maxent models were similar across all spatial scales. The difference between training AUCs and test AUC, however, was greatest in BRT models.

Explanatory variables

The top five explanatory variables, at all three spatial scales using both modeling methods were: (1) maximal air temperature during the active season, (2) road density, (3) total wetland area, (4) total open water area, and (5) total cropland area (Figure 2-1). At all three spatial scales, Blanding's turtle habitat suitability increased with increasing maximal air temperatures during the active season, increasing road density up to a maximum of 0.0025 m/m^2 - decreasing thereafter, increasing wetland area, increasing open water area up to a maximum of 40% - decreasing thereafter, and decreasing cropland area (Figures 2-2 to 2-5). Following these top five predictors, terrain ruggedness, mean monthly precipitations during the active season, total forest cover, and total rock cover were of medium importance. Alvar, pastures and fields, settlements and developed lands, and cutovers and burns cover types were of less importance. Overall habitat suitability increased slightly with forest, rock, and alvar cover types and decreased slightly with terrain ruggedness and crop, pasture, urban, and cuts and burns cover types.

Furthermore, I observed two overall trends relating to explanatory variable contributions across the three spatial scales studied (Figure 2-1). Firstly, the overall importance of road density increased as the buffer size decreased. Secondly, the overall importance of maximal temperature decreased as the buffer size decreased.

Finally, I observed several differences between BRT and Maxent results. On average, fewer variables were retained in Maxent models during recursive feature elimination (Appendix II). Furthermore, mean precipitation and forested area were of little importance in Maxent models despite being of high to medium importance in BRT models. Water area, wetland area, and cropland area were of greater importance in Maxent models than in BRT models. Finally, the increase in the overall importance of road density at the local scale was much more apparent in Maxent models.

Habitat suitability predictions in Ontario

I observed a marked difference in habitat suitability maps produced by BRTs and Maxent models. Overall, BRT models were much more site specific when identifying highly-suitable habitat and vast areas were deemed to be low quality habitat (Figure 2-6 to 2-8). Maxent models offered more gradation in habitat suitability predictions and a much greater extent of Ontario was deemed to have high-quality Blanding’s turtle habitat. The relationship between BRT habitat suitability predictions and Maxent habitat suitability predictions was exponential (Figure 2-9) and the grids had a correlation of 0.66 at the 1 km scale, 0.65 at the 500 m scale and 0.61 at the 250 m scale.

As the spatial scale analyzed decreased, the extent of highly suitable habitat decreased for both modeling methods. The only noticeable difference between habitat suitability models built using different spatial scales is the effect of roads in southern Ontario at the smaller spatial scale. The overall trends describing which areas had greater suitability were consistent across all spatial scales and both modeling methods.

In both BRT and Maxent HS maps, most areas with highly suitable habitat had historical Blanding's turtle sightings (e.g. Leeds-Grenville, Kawartha Lakes-Brock, Long Point, Point Pelee). However, a few key areas with highly suitable habitat had no historical sightings (e.g. Stormont-Dundas-South Glengarry and Trinity-Spadina). Finally, Blanding's turtles have been confirmed in areas of very low habitat suitability such as the Bruce-Gray-Owen Sound (Bruce Peninsula) and Markham-Unionville.

Discussion

Species distribution models are tools that can provide understanding about the ecological requirement of an organism, facilitate fieldwork by the prediction of potential occurrence areas for rare species, assess risk for invasive species or disease vectors, and help overall conservation actions using spatial data (Elith and Leathwick, 2009; Araújo and Guisan, 2006; Guisan et al., 2006). As such, my results offer new information on the ecological factors related to the distribution of Blanding's turtles in Ontario and provide habitat suitability maps. Furthermore, additional examinations offer insight on model optimization procedures (i.e. recursive feature elimination using maximum AUC values; Appendix II) and sampling bias (Appendix III and IV).

Biological interpretations

Predictor variables contributions and dependency plots

As revealed by both Maxent and BRT modeling, temperature, habitat type, and road density are important environmental factors when describing the geographical distribution of Blanding's turtles in Ontario. This is not surprising given that several other studies, using various modeling algorithms, looking at reptilian species, have shown that climatic conditions and land cover characteristics are highly relevant indicators of habitat suitability for reptiles (Tingley and Herman, 2009; Kaliontzopoulou et al., 2008; Fisher et al., 2004).

Unsurprisingly, I observed a positive relationship between mean monthly maximal air temperature and Blanding's turtle habitat suitability. Mean monthly maximal air temperature remained, on average, the best overall predictor variable for habitat suitability at all spatial scales. This can be expected since Ontario's Blanding's turtles are ectotherms at the northern extreme of their global range. Precipitations were of lesser importance in my models, although habitat suitability did increase with decreasing precipitations. Similar to a study by Frazer et al. (1991) looking at painted turtles, Blanding's turtle habitat suitability was greatest in drier hotter areas. Climate is typically though to govern species distributions at broad spatial scales, whereas the effects of land cover become more apparent at finer scales within the range dictated by climatic variables (Pearson and Dawson, 2003). Thus, the thermal environment plays a crucial role in Blanding's turtle habitat selection at all spatial scales, effectively defining the range of this species in Ontario.

I also observed a positive relationship between wetland area and habitat suitability. Once again, this is unsurprising as Blanding's turtles are freshwater turtles known to occupy a variety of eutrophic wetlands such as swamps, marshes, beaver dams, permanent and temporary ponds, and slow flowing streams (Kofron and Schreiber, 1985; Petokas, 1986; Rowe, 1987; Ross and Anderson, 1990; Rowe and Moll, 1991; Pappas and Brecke, 1992; Joyal et al., 2001). A study by Attum et al. (2008) demonstrated that Blanding's turtles were significantly more likely to be seen in larger wetlands. Thus, as the total wetland area in a region increases, so does that region's suitability for Blanding's turtles. Similar to maximal temperature, the overall importance of wetlands did not change with buffer size, which is indicative of its status as primary habitat.

Open water had a more complex effect on habitat suitability. As the percentage of open water increased, so did habitat suitability. After reaching a peak of 40% at all spatial scales, however, habitat suitability then decreased with increasing open water habitat. This non-linear relationship could be explained by the transient use of open water by turtles. Open water is often used as a travel corridor by Blanding's turtles to reach new wetlands, aestivation sites, or nesting sites, but is not the primary habitat for this species. Open water is an essential component of their habitat and greatly facilitates movement between areas of residence, diminishing the energetic requirements of long distance movements.

Blanding's turtles are vagile and often use upland corridors for movement between wetlands, aestivation, and thermoregulation. Although used less than its availability (Ross and Anderson, 1990), terrestrial habitat is a vital component for turtles (Gibbons, 2003). Similar to past studies (Attum et al., 2008; Gibbons, 2003), my results strongly suggest that the nature of these corridors will influence habitat quality for Blanding's turtles. Attum et al. (2008) found that forested area within buffer zones of up to 250 m was a significant predictor of Blanding's turtle habitat use. Forested area within a buffer zone can be considered an index of connectivity. Connectivity increases with natural habitat cover types and decreases when the habitat is disturbed or altered, as seen with cropland and other anthropogenic cover types. My results suggest that undisturbed terrestrial habitat, such as forested, alvar, or rock outcrop, increases overall habitat suitability for Blanding's turtles, probably by increasing connectivity. While protecting extensive forests, or other undisturbed natural habitat, surrounding wetlands may not be practical because of financial, logistical or political considerations, I suggest prioritizing the protection of frequently used upland corridors to maintain connectivity quality between wetlands, nesting and aestivations sites and minimizing potentially deleterious edge effects. Although two wetlands may appear well connected due to their close proximity, if the upland habitat between the wetlands prevents movement or increases mortality risk for migrating animals, then the two wetlands are more isolated than simple distance would indicate (Rickets, 2001). The use of buffer zones to protect forest/alvar area near small wetlands would improve connectivity between wetlands and aid in the recovery efforts for Blanding's turtles. However, forest restoration and prevention of further agricultural use alone may not be enough to assist in their recovery without addressing the impacts of roads (Kramer-Schadt et al., 2004; Roe et al., 2006).

Road density was one of the top five predictor variables for Blanding's turtle habitat suitability at all three spatial scales. The positive relationship between road density and habitat suitability reversed itself after reaching a maximum value of approximately 0.0025m/m^2. This non-linear relationship could potentially reflect several key factors. The first of these factors is sampling bias in the sighting records. The sighting records from the NHIC have been collected haphazardly throughout the past 30-40 years and no structured sampling methodology was used. As such, areas that are more accessible to humans tend to have more sightings than remote areas, whether or not Blanding`s turtles are more abundant. Areas along roads are simply better sampled. For these reasons, I ran a separate analysis using target background data which would share this sampling bias (NHIC sightings of all turtle species in Ontario). The results from this analysis, where sampling bias is theoretically accounted for, suggest that the effect of roads on Blanding turtle location is less important than observed in non-target background models (Appendix III). In fact, the importance of roads was considerably reduced (approximately 10-15 percentage points) in target background models, although the importance increased with decreasing spatial scale, as before. However, caution should be taken when interpreting my target background results. I also observed a decrease in the importance of open water at all spatial scales when using target background. This trend could be explained by the fact that all freshwater turtle species used for the target background are primarily found in riparian habitat. In fact, all the turtle species used (snapping, spiny softshell, red eared slider, painted, map, and stinkpot turtles) were primarily residents of both open water habitat and wetland habitat, whereas Blanding’s turtles are residents of wetland habitats only. Thus, almost like roads, open water habitat was being targeted. In fact, since all turtles are attracted to roads, any actual selection for roads by turtles could be overshadowed by the sampling bias and the use of target background. In the future, although I recommend the use of target background, I would suggest that taxonomic groups with similar sampling bias and life-history traits, but with different habitat characteristics, be used. In this case, the use of historical snake sightings instead of turtle sightings might have been more revealing.

In addition to sampling bias, turtles have a documented affinity for roads. Roadsides create artificially disturbed and open habitats that may be attractive for thermoregulation or nesting (Joyal et al., 2001 ; Wood and Herlans, 1997; Haxton, 2000, Carr, 1952). Furthermore, roads built between adjacent wetlands, through natural parks or along shorelines will fragment Blanding's turtle habitat (Evink, 1980; Johnston 1994, Mitsch and Gosselink 2000). Thus, turtles are often found crossing roads to reach adjacent wetlands, aestivation sites, or nesting sites. Although my results suggest that Blanding's turtles select for areas with low road densities, it is unclear whether this selection is due to sampling bias, attraction, or habitat fragmentation.

Finally, the reversal of this positive relationship at high road densities could be attributed to the negative effects of roads on turtle populations as a whole. The second biggest threat to reptile populations is road mortality, especially in long lived species with poor recruitment, such as the Blanding's turtle (Congdon et al., 1993; Herman et al., 2003). Due to the attractiveness of road shoulders for egg laying, females are more susceptible to road mortality and past studies have shown that increased road densities can lead to male-biased sex ratios (Marchand & Litvaitis, 2004; Steen & Gibbs, 2004). Large populations with highly skewed sex ratios can experience sudden and potentially unrecoverable population crashes because of sex specific mortality (Milner-Gulland et al., 2003). Furthermore, for long-lived species with long generation times, there may be a lag time between current road mortality rates and observable population declines (Findlay and Bourdage, 2000).

In an effort to determine the effect of road mortalities on habitat suitability models, I also ran models using only sightings where turtles were found dead on the road (DOR) and only sightings where turtles were alive (NODOR), irrespective of turtle placement (Appendix IV). These two sets of models enabled me to determine whether roadkill was artificially increasing the importance and positive effect of road density on habitat suitability. Results from these analyses suggest that the inclusion of DOR sightings did not alter the results of my models as NODOR models were almost identical to the models presented in this chapter. Furthermore, the removal of DOR sightings weakened the predictive ability of my models. These results suggest that, as previously discussed by Phillips et al. (2006) and Elith et al. (2008), BRT and Maxent algorithm are robust to outliers and that models using all available data perform better than those where data are removed in an effort to reduce spatial-autocorrelation (Kaliontzopoulou et al., 2008)

Blanding’s turtles commonly encounter graded terrain during their seasonal movements, however the vertical component of this movement is rarely considered in turtle habitat modeling. Extreme topographical and landscape features may form barriers to movement and the need to circumvent such obstacles may lead to deviations from a straight line homing course (Metcalf and Metcalf, 1970; Carroll and Ehrenfeld, 1978; Strang, 1983). Thus, increased terrain ruggedness could lead to increasing energetic expenditure, predation risks, and may determine the boundaries of the home range itself (Muegel and Claussen, 1994; Claussen et al., 2002; Stickel, 1950; Strang, 1983). In two similar studies, Muegel and Claussen (1994) and Claussen et al. (2002) demonstrated that inclines of $\geq 60^{o}$ were insurmountable to turtles. Thus, although the overall importance of terrain ruggedness was intermediate, Blanding's turtle habitat suitability did decrease with increasing terrain ruggedness. Natural terrain is often graded and our results suggest that the effects of terrain ruggedness on the distribution of turtles deserve more attention.

Tingley and Herman (2009) demonstrated that accounting for the effect of land cover can significantly improve the explanatory and predictive power of bioclimatic models for turtles at a regional scale. High resolution land cover data may allow refinement of habitat suitability predictions by identifying areas that are climatically suitable, but that are inhospitable owing to the effects of habitat loss and degradation (Pearson et al., 2004). By including environmental features such as cropland area and road density, I was better able to characterize the habitat suitability for Blanding`s turtles. It is clear from my results that these two environmental predictors are crucial in delimiting areas that are inhospitable to turtles.

Effect of spatial scale

The observed differences in variable contributions in habitat suitability models at differing spatial scales support the affirmation that habitat selection is hierarchical (Johnson, 1980). The importance of maximal air temperature decreased as the spatial scale analyzed decreased, thus suggesting that local site selection is less impacted by broad-scale climatic conditions than is the placement of the home range within the landscape. At the local scale, the importance of road density more than doubled. These results suggest that although road density is important at all spatial scales examined, the effect of road density on turtles increases at the local level. Finally, I also observed a decrease in the importance of open water with decreasing spatial scale in BRT models, suggesting that turtle home range placement is more affected by open water area than is daily site selection. On a daily basis, turtles will spend much less time in travel corridors than in their residential wetlands, however home-ranges with more open water enable easier movement between adjacent wetlands by diminishing the energetic requirements related to long distance movements and reduce predation risks, thereby increasing overall habitat suitability.

Potential Distribution

The Blanding's turtle is located throughout southern and south-central Ontario as far north as the Chippewa River (COSEWIC, 2005). For the most part, historical Blanding's turtle sightings have been in areas of predicted high habitat suitability. According to the Ontario Herpetofaunal Summary (Oldham and Weller, 2010), however, there are significant gaps in the range of the Blanding's turtle. The most apparent occurs in extreme southeastern Ontario in the districts of Prescott-Russell-Glengarry and Stormont-Dundas-South Glengarry. These areas could represent either potentially undiscovered populations, populations that have been studied but the sighting were not reported to the NHIC (i.e. Minnow Lake, Sudbury (Litzgus pers. comm.); Region of Halton (RMH, 2009); Greater Cataraqui Marsh, Kingston (Blancher, 1984)), or areas that are inaccessible to Blanding's turtles.

On the other hand, there are several districts with very low habitat suitability that do have historical Blanding's turtle sightings, some as recent as 2002. These areas could potentially represent population sinks due to habitat alteration and destruction over the past few decades, although further research on each population would be necessary to determine current population trends. Abundance data and confirmed absences are not available for this species at the spatial extent of analysis and thus inferences about population declines or increases cannot be made. This is the first attempt at modeling Blanding's turtle habitat in Ontario using species distribution modeling methodologies and presence-only data.

The apparent extirpation of the spotted turtle (*Clemmys guttata*) and the declining populations of Blanding's turtles in the protected habitats of Point Pelee National Park in Ontario, an area with high habitat suitability in all my models, is further evidence that factors other than habitat loss can be responsible for freshwater turtle declines (Browne and Hecnar, 2007). Already there are changes in the adult sex ratios and the number of younger Blanding's turtles is declining (Browne and Hecnar, 2007). It seems that even in areas of high habitat suitability, Blanding's turtle populations are declining.

Methodological considerations

The modelling techniques that I have used assess the suitability of habitat for Blanding`s turtles in Ontario and do not reflect actual population demographics or distributions. Areas of predicted presence will typically be larger than the species' realized distribution (Phillips et al., 2006). This can be due to several factors, such as geographic barriers to dispersal (Peterson et al., 1999), biotic interactions (Anderson et al., 2002), and human modification of the environment (Anderson and Martínez -Meyer, 2004). Although I have removed areas where the Blanding's turtles are inferred not to inhabit, thereby increasing the reliability of the habitat suitability estimate, there are many ecologic factors that are not modeled. These include, but are not limited to, population dynamics, meta-population dynamics, and biotic-interactions (Araújo and Williams, 2000).

Unfortunately, presence-only data have errors and biases associated with them, reflecting the frequent haphazard manner in which samples were accumulated (Hijman et al., 2000). Firstly, as previously discussed, occurrence localities may be biased: high correlation with the nearby presence of roads, rivers or other access conduits (Reddy and Dávalos, 2003), spatial auto-correlation, and variations with respect to sampling intensity and sampling methods across the study area. To reduce possible spatial autocorrelation, I only used sighting points that were at least 1 km apart. Furthermore, results from an analysis using target background, eliminating the effects of sampling bias and detailed in Appendix III, were qualitatively similar to the results presented here.

Finally, the number of occurrences may be too low to estimate the parameters of the model reliably (Stockwell and Peterson, 2002) and/or the set of environmental variables may not be sufficient to describe all the parameters of the species' fundamental niche which are relevant to its distribution at the grain of the modeling task (Phillips et al., 2006). However, even after considering the potential problems associated with presence-only records, there is evidence that these data can be used effectively, given the right choice of statistical model and environmental variables, for modeling species distribution at various spatial scales (Elith and Leathwick, 2009). In fact, some argue that presence-only data are more appropriate because presence-absence records can introduce confounding information (Elith and Leathwick, 2009). A negative occurrence can indicate that either the habitat is unsuitable or that the habitat is suitable but unoccupied, perhaps because it is inaccessible. Furthermore, absence data can be misleading because the species or environment is not at equilibrium (e.g. invasions, climate change) or the species is not easily detected.

In this study, I used the highest resolution possible for all my datasets. Kaliontzopoulou et al. (2008) demonstrated that fine-resolution models have greater predictive ability than coarser resolution ones. Furthermore, coarser grain models had a lower capacity for detecting which variables were more important for species distribution (Kaliontzopoulou et al., 2008). In my study, I built models using 25 m resolution land cover and DEM data, and 1 km^2 resolution bioclimatic data, and predicted to 100 m resolution grids. This modeling grain size is much finer than that of past studies (Kaliontzopoulou et al., 2008; Philllips et al., 2006) and should increase the ability of my models to establish important predictor variables and resolve complex relationships between the predictor variables and the response. Given the ecological characteristics of Blanding's turtles (small body size, medium-sized home range, and little dispersal), the use of a finer-resolution dataset is warranted (Guisan and Hofer, 2003).

Unlike several SDM studies, I did not use any threshold dependant model evaluation methods. I believe the selection of a threshold in this case would not only be arbitrary, but could mislead conservation practitioners. From an ecological perspective, habitat quality is a continuous rather than a categorical variable, so a statistically valid method that produces a range of occurrence probabilities is likely to be more effective than a binomial classification scheme (Fleishman et al., 2001). Many species have dynamic patterns of occurrence and therefore predicting the suitability of currently unoccupied habitat could be helpful. Furthermore, continuous habitat suitability predictions may help management professionals prioritize locations for different land uses, based on their degree of suitability, and determine whether or not management actions could render a location more (or less) suitable (Fleishman et al, 2001).

Statistical tests and modeling algorithms

Both algorithms consistently performed better than random. Given the differences in predicted habitat suitability maps between the two algorithms, I strongly recommend that future studies employ more than one modeling methodology. The overall importance of some predictor variables varied between the two modeling algorithms suggesting that these two methods can lead to different models. Despite these predictive differences, the top five predictor variables and the shape of the response curves for each variable were almost identical. Furthermore, I observed high fidelity between the models for areas with high and low habitat suitability.

BRT and Maxent share one very important feature: they have a high level of flexibility in fitting complex responses. However this expressiveness, a well-developed ability to express or demonstrate the complex relationships in the data (Elith et al., 2006), needs to be controlled so that models are not overfit, leading to the use of regularization techniques that achieve a balance between complexity and parsimony (Hastie et al., 2001). BRT models had a greater degree of overfitting and retained a greater number of predictor variables during RFE than Maxent models and this could be due to the different regularization techniques utilized by the algorithms.

Maxent predicted the largest distribution of high quality habitat over geographical space, much more than BRTs. Depending on the use of the maps, both algorithms have their advantages. If the objective is to delineate protected areas for rare or endangered species, overestimating areas of high quality habitat might be preferable to underestimating the habitat needed (Zaniewski et al., 2002). Optimistic predictions proved false could however damage the level of confidence and support (Zaniewski et al., 2002). Furthermore, conservation dollars are not unlimited and thus models delimiting smaller patches of core habitat could be more helpful. Also, on average fewer variables were retained in Maxent model, thereby simplifying the set of predictor variables that must be accounted for.

Conservation Implications

The overall contribution of cropland land cover on habitat suitability for Blanding’s turtles and its negative impact on habitat suitability strongly suggests that the terrestrial habitat surrounding wetlands is a critical component of Blanding’s turtle habitat. Furthermore, the negative impact of cropland cover was evident at all spatial scales studied, indicating that agricultural lands within 1 km of residential wetlands are having deleterious effects on Blanding’s turtle habitat. Large buffer zones need to be established around residential wetlands to prevent further development and alteration of terrestrial corridors, thereby maintaining connectivity between wetlands and aiding in recovery efforts. These terrestrial areas are more than just neutral buffers, they are vital habitat in the life cycle of this species’ and are essential to its continued survival (Gibbons, 2003). However, forest restoration and prevention of further agricultural use alone may not be enough to assist in the recovery of Blanding turtle populations. As discussed above, identifying landscapes that have a low density of roads or wetland assemblages that are distant from roads would be an important consideration for securing habitats capable of sustaining this species (Litvaitis and Tash, 2008).

By using multiscale data capable of capturing hierarchical landscape influences, I was able to produce accurate predictive models and provide further insight into contributing factors for Blanding’s turtle habitat suitability in Ontario.

Table 2-1. The fourteen environmental variables used for predicting Blanding's turtle habitat suitability in Ontario.

Variable (units)	Range	Code	Source(s)
Land Cover (%)			OMNR, 1998
Wetlands	0 - 100	WETL	
Open Water	0 - 100	WATER	
Forests	0 - 100	FOREST	
Plantations	0 - 100	PLANT	
Settlements and Developed Land	0 - 100	URBAN	
Cropland	0 - 100	CROP	
Pastures and Fields	0 - 100	PAST	
Mine Tailings, Quarries and Bedrock Outcrops	0 - 100	ROCK	
Alvar	0 - 100	ALVAR	
Cutovers and Burns	0 - 100	CB	
Mean Monthly Precipitations (mm) ***(April-October)***	67.27 – 89.47	PREC	Hijmans et al. , 2005
Mean Monthly Maximal Temperature (°C) ***(April-October)***	18.9 – 25.2	TMAX	Hijmans et al. , 2005
Terrain Ruggedness (TRI Index)	0 – 26.64	TRI	OMNR, 2005; Riley et al., 1999
Total Road Density (m/m^2)	0 – 0.026	RD	OMNR, 2006

Table 2-2. Descriptions of the grouped land cover classes used for model building.

Land Cover Classes	Descriptions
Wetlands	Includes all *marshes*, *treed fens*, *open fens*, *treed bogs*, *open bogs*, and *swamps.*
Open Water	Includes all water bodies that are not categorized as wetlands (e.g. rivers, streams, and lakes).
Forests	Forested areas with greater than 30% forest canopy closure. Includes *dense coniferous forests*, *dense deciduous forests*, *mixed mainly coniferous forest*, *mixed mainly deciduous forests*, *sparse coniferous forest*, and *dense deciduous forests*; excludes *plantations*.
Plantations	Mature conifer plantations, mostly pine, occurring in evenly spaced rows; excludes artificially regenerated cutovers and burns.
Settlements and Developed Lands	Clearings for human settlement and economic activity.
Cropland	Row crops, hay and open soil in areas of agricultural land use.
Pastures and Fields	Open grassland with sparse shrubs mapped in agricultural areas; includes orchard lands.
Mine Tailings, Quarries and Bedrock Outcrops	Clearings for mining activity, aggregate quarries and bedrock outcrops.
Alvar	Homogeneous areas of dry grassland growing on thin soils over a limestone substrate.
Cutovers and Burns	Forest clear-cuts and burns. Includes *new cutovers*, *new burns*, and *old cutovers and burns.*

Table 2-3. Predictive performance of BRT and Maxent models tested on independent data (test), data used to build models (training) and all data (all) for each spatial scale examined. The threshold independent area under the receiver characteristic operating curve (AUC) and biserial correlation (COR) values are reported.

Model	**BRT**			**Maxent**		
	AUC		**COR**	**AUC**		**COR**
	Test	Training	All	Test	Training	All
1km	0.882	0.945	0.518	0.878	0.904	0.360
500m	0.883	0.960	0.530	0.887	0.916	0.389
250m	0.912	0.956	0.547	0.898	0.904	0.379

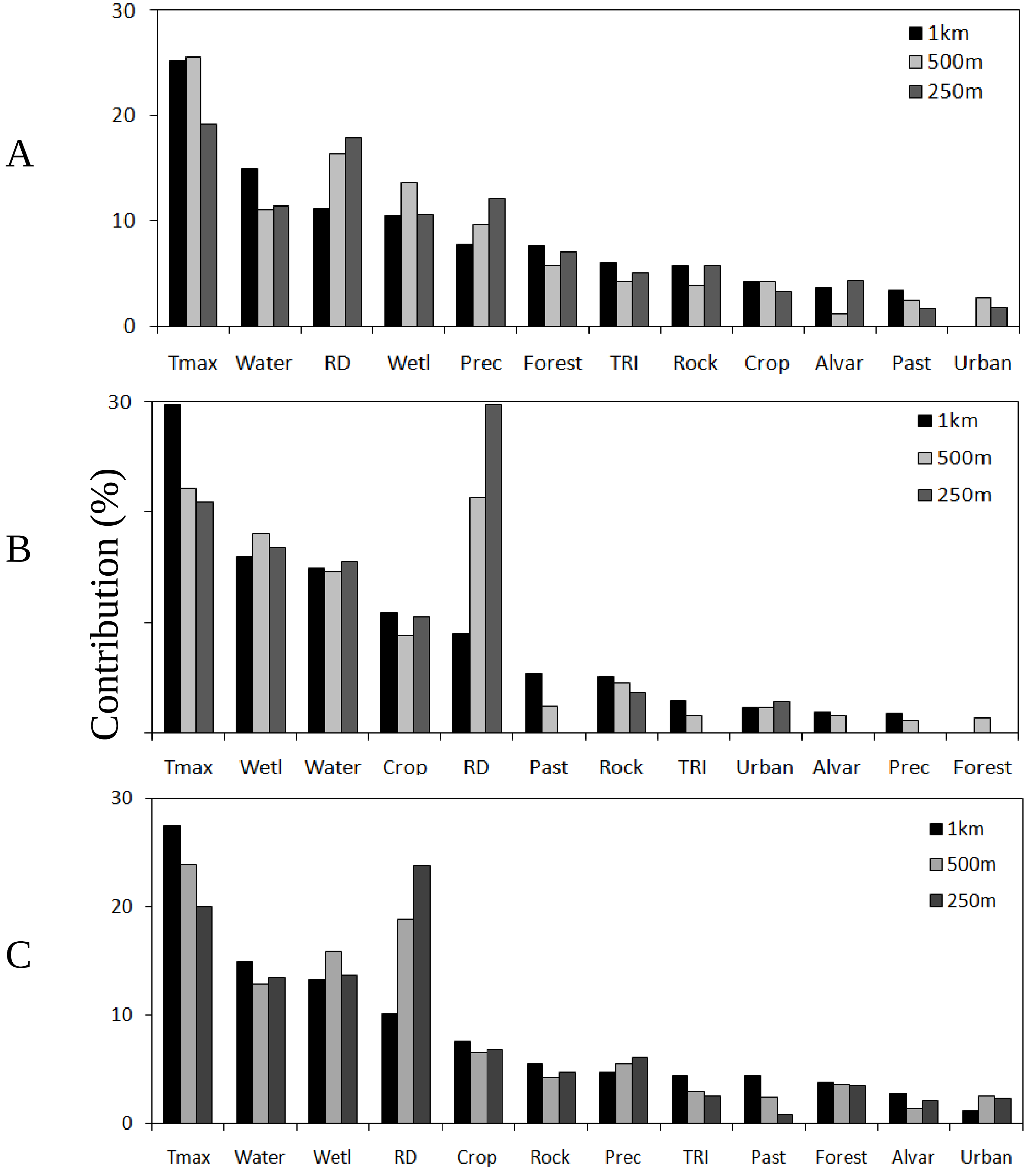

Figure 2-1. Average contribution of predictor variables in habitat suitability models built using (**A**) BRTs, (**B**) Maxent, and (**C**) the average results of both at three different spatial scales.

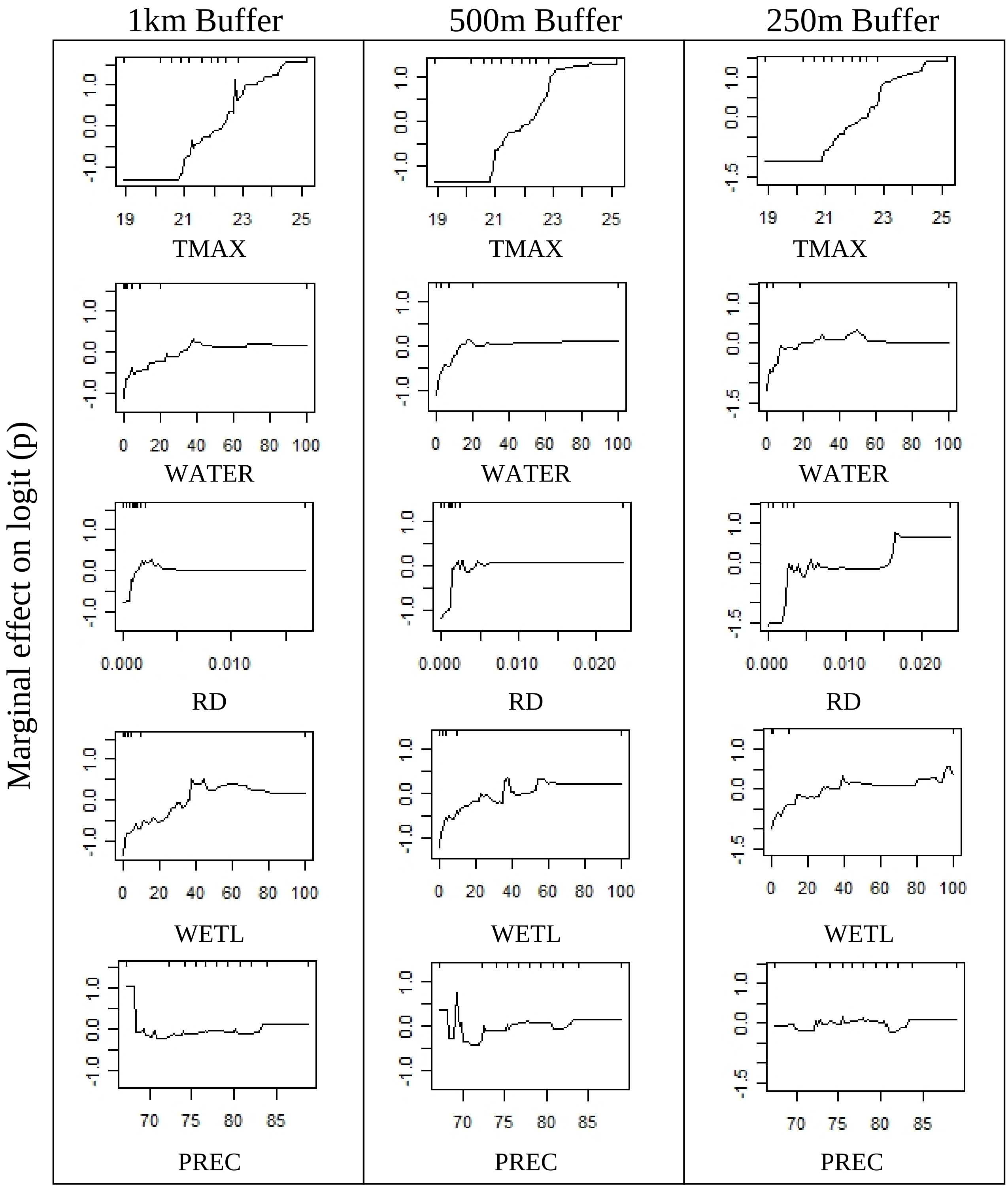

Figure 2-2. Partial dependence plots for the five most influential predictor variables in habitat suitability models built using BRTs at three different spatial scales. Y axes are on the logit scale, are centered to have zero mean over the data distribution, and represent the predicted probability of suitable conditions with all other variables set to their average values over the set of presence localities. Rug plots at inside top of plots show distribution of sites across each variable, in deciles.

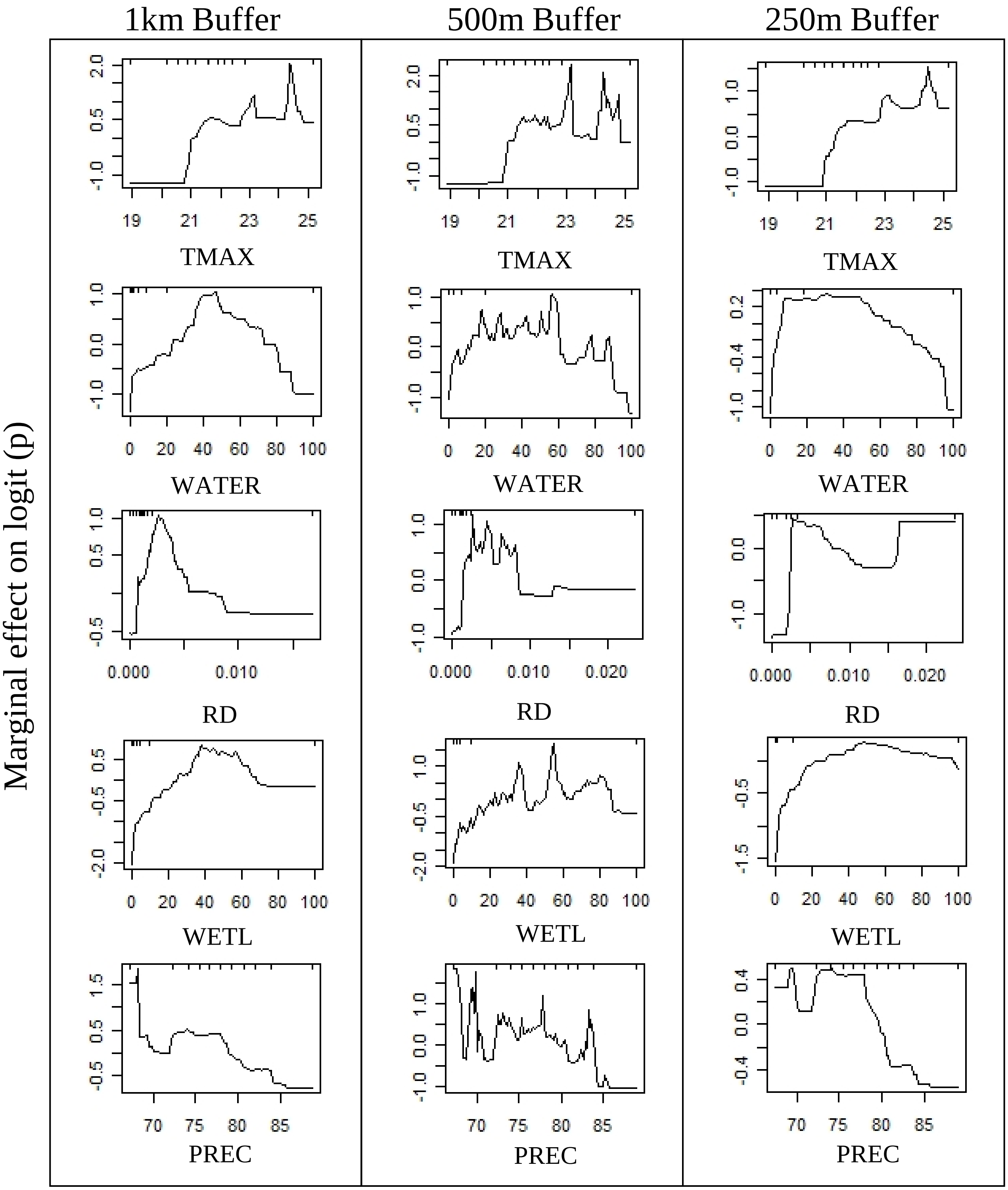

Figure 2-3. Response curves produced by univariate models of the five most important predictor variables in models built using BRTs at three different spatial scales. Y axes are on the logit scale, are centered to have zero mean over the data distribution, and represent the predicted probability of suitable conditions based on each predictor variable independently. Rug plots at inside top of plots show distribution of sites across each variable, in deciles.

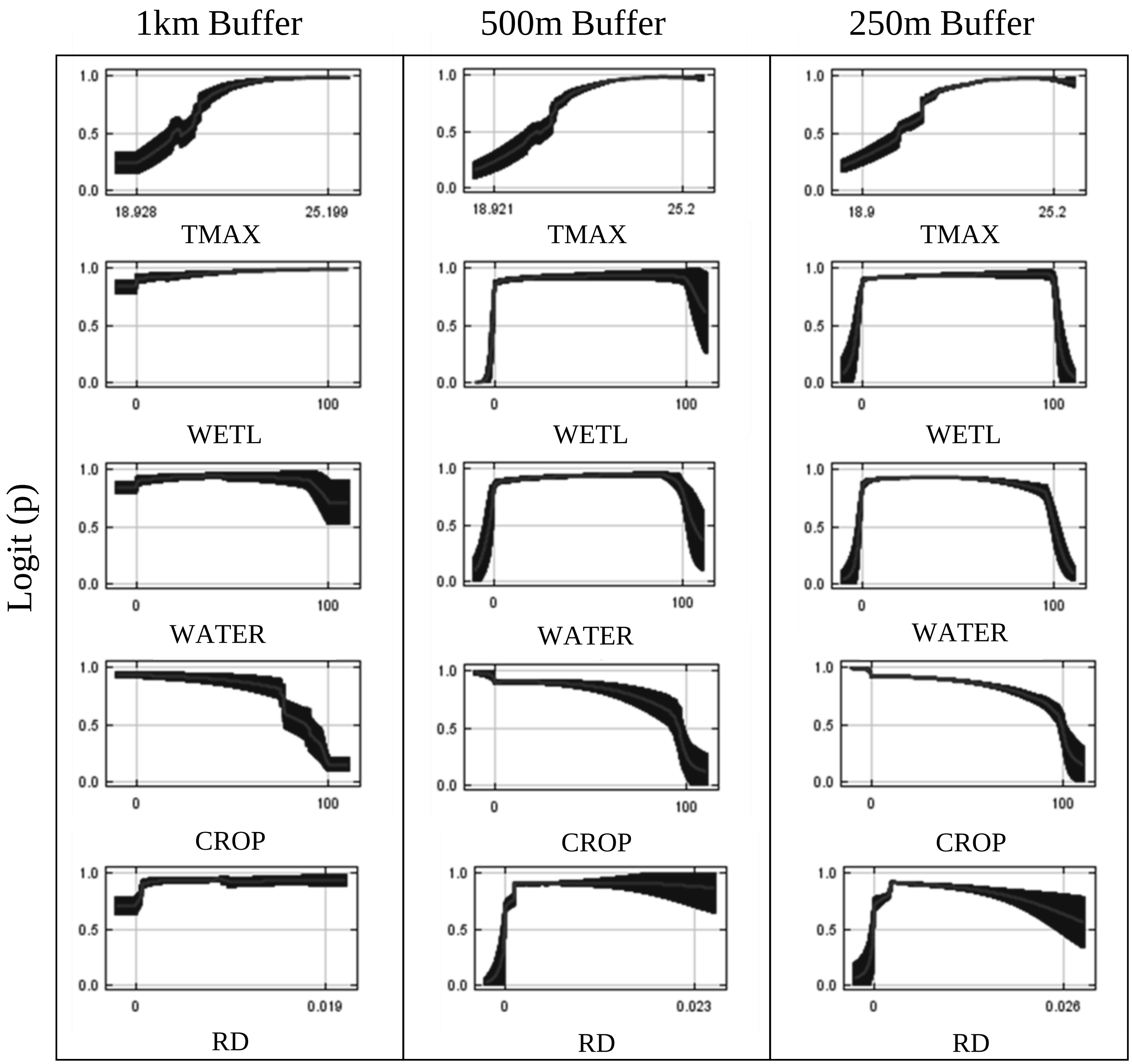

Figure 2-4. Partial dependence plots ± SD for the five most influential predictor variables in habitat suitability models built using Maxent at three different spatial scales. Y axes are on the logit scale and represent the predicted probability of suitable conditions with all other variables set to their average values over the set of presence localities.

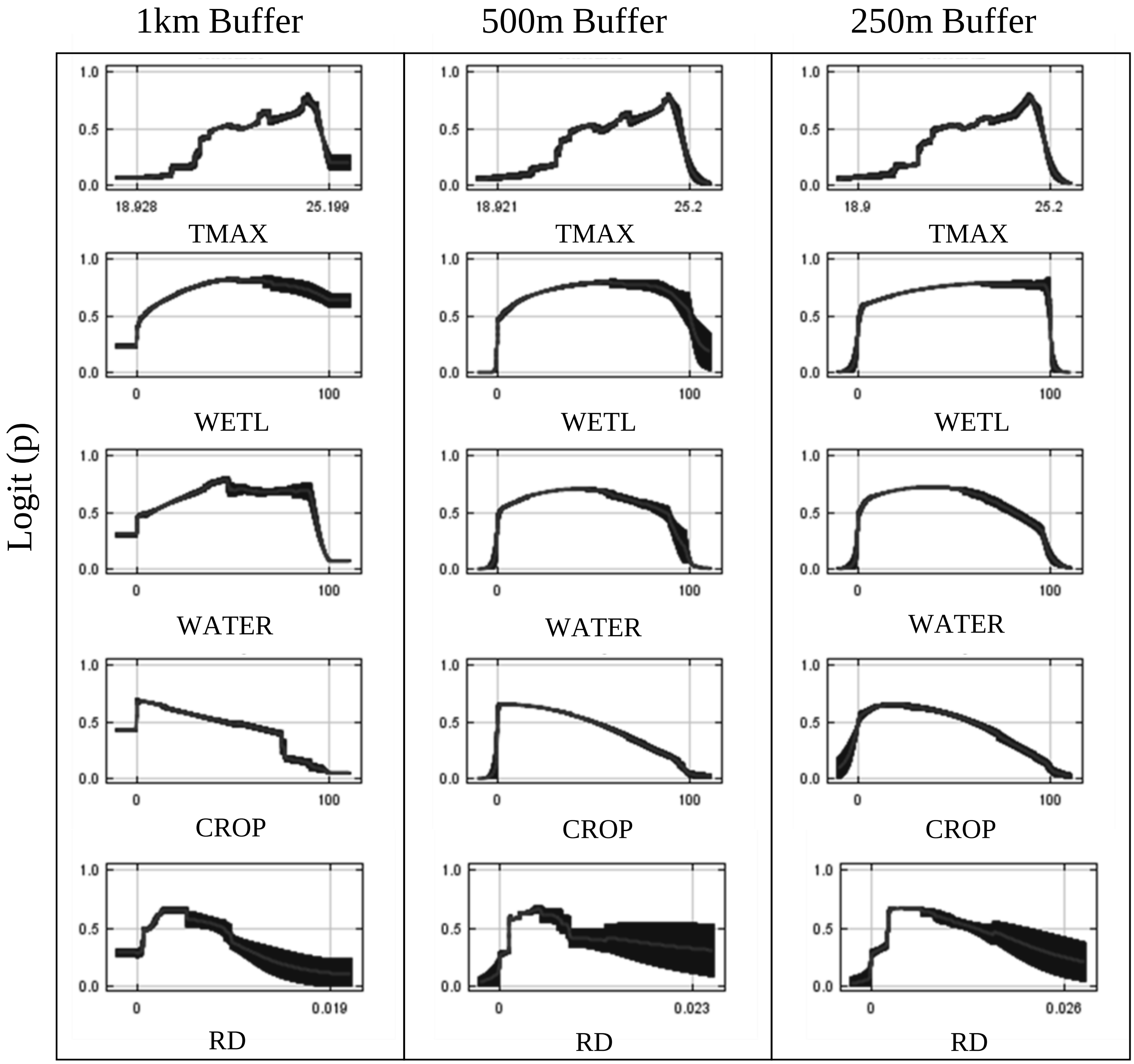

Figure 2-5. Response curves ± SD produced by univariate models of the five most important predictor variables in models built using Maxent at three different spatial scales. Y axes are on the logit scale and represent the predicted probability of suitable conditions based on the effect of each predictor variable, independently.

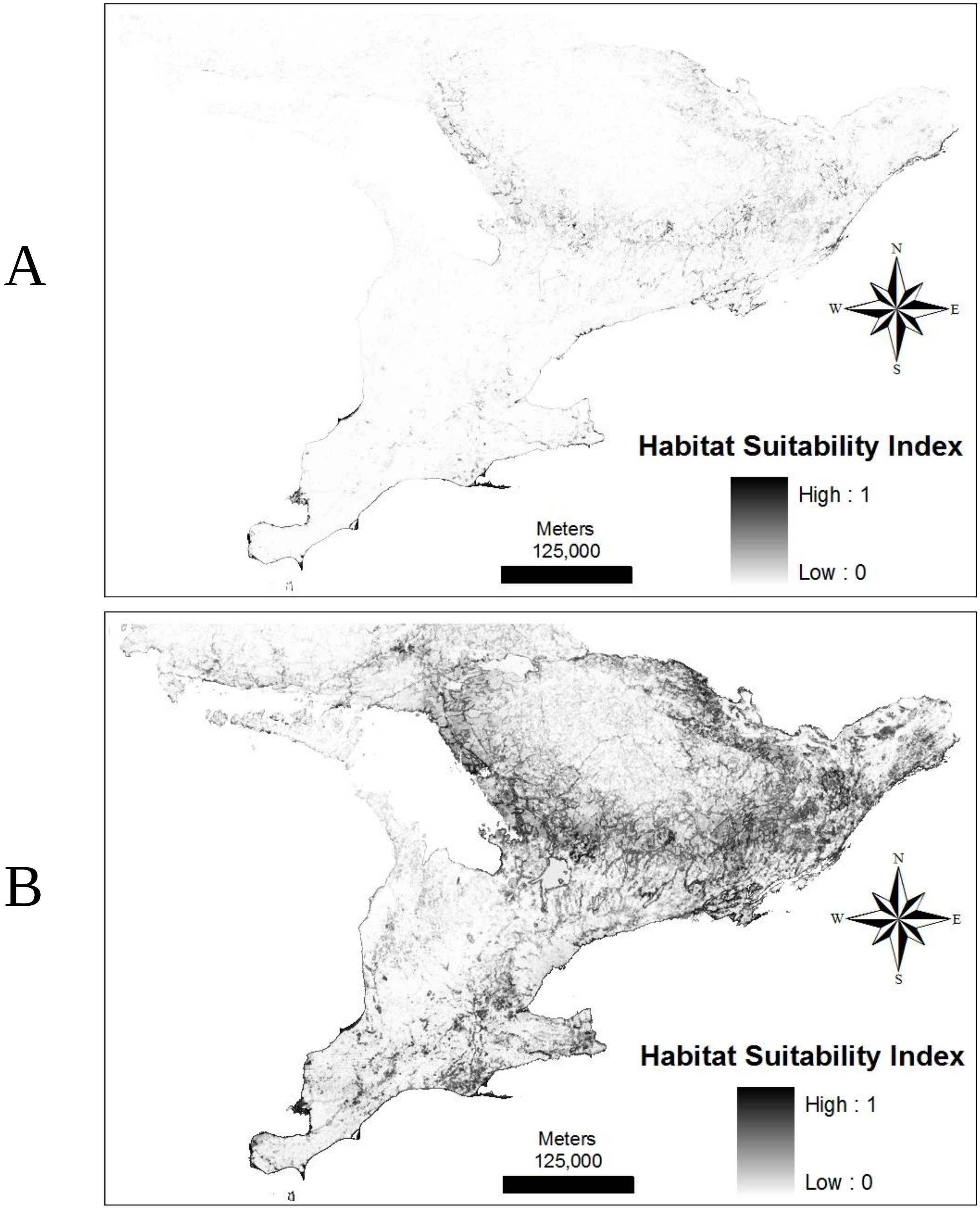

Figure 2-6. Predicted habitat suitability for Blanding's turtles across their range in Ontario using a 1 km buffer and two modeling algorithms: (**A**) BRTs and (**B**) Maxent.

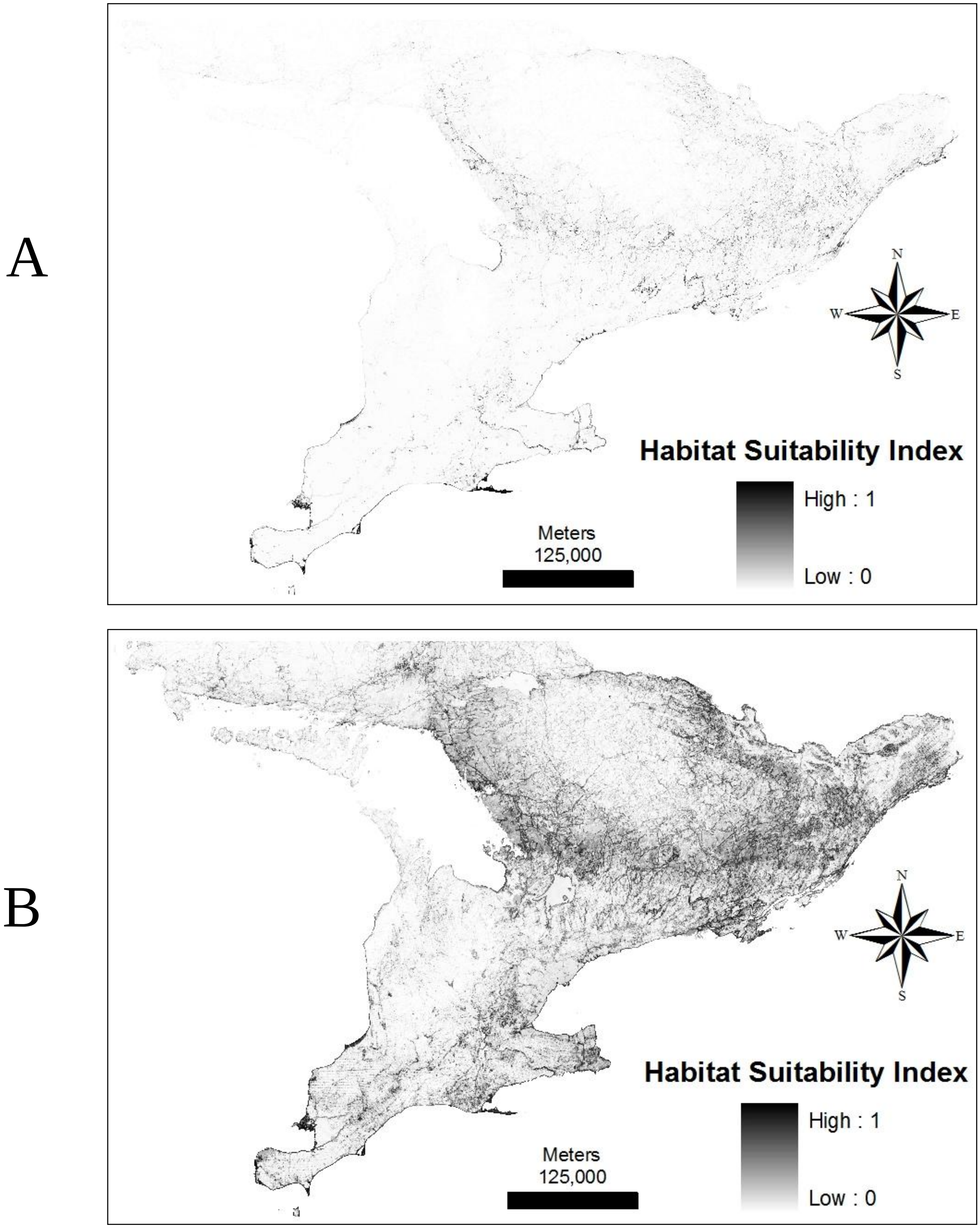

Figure 2-7. Predicted habitat suitability for Blanding's turtles across their range in Ontario using a 500 m buffer and two modeling algorithms: (**A**) BRTs and (**B**) Maxent.

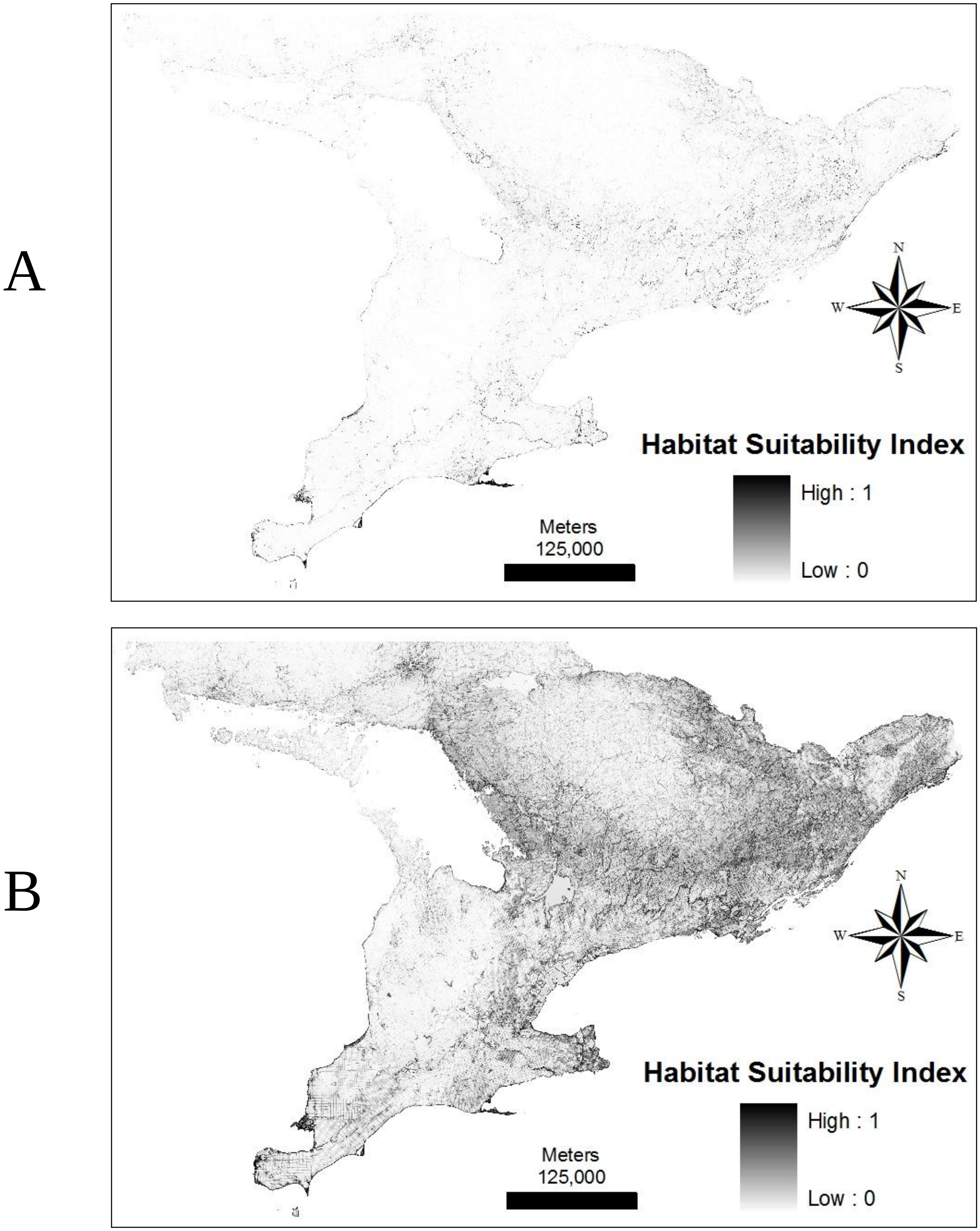

Figure 2-8. Predicted habitat suitability for Blanding's turtles across their range in Ontario using a 250 m buffer and two modeling algorithms: (**A**) BRTs and (**B**) Maxent.

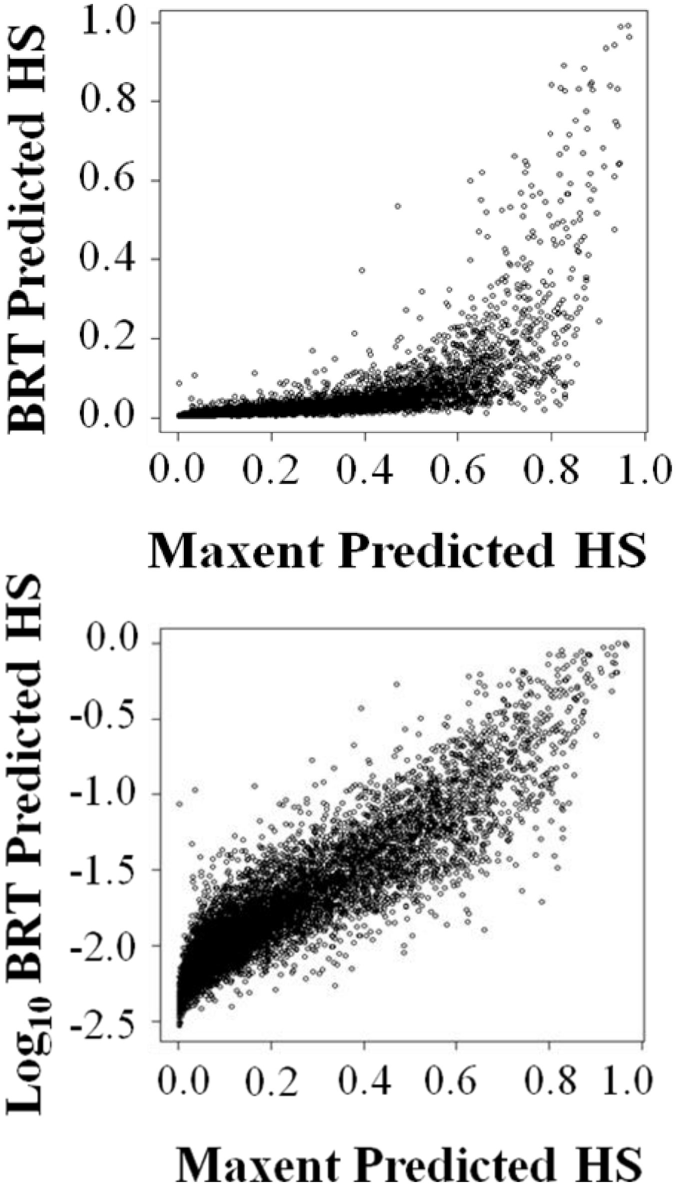

Figure 2-9. Predicted habitat suitability at 10 000 random points across the study area in the 1 km BRT model as a function of predicted habitat suitability at the same points in the 1 km Maxent model.

APPENDIX I

REVIEW: Introduction to two machine learning algorithms for species distribution modeling: boosted regression trees (BRTs) and maximum entropy modeling (Maxent).

In a study by Elith et al. (2006), comparing 16 modeling methods over 226 species from 6 regions of the world, machine-learning methods and community models consistently outperformed more established methods. For this reason, I chose to use two machine learning methods which were among the top performing models in the Elith et al. (2006) review, a maximum entropy model (Maxent) and boosted regression trees (BRTs, also called stochastic gradient boosting). In the following paragraph I will attempt to describe in greater detail the principles behind these two statistical models, how these models function, and how to interpret modeling results. This appendix serves as an introduction only to these statistical modeling methodologies. For a more in-depth discussion of BRTs, I recommend reading the articles written by Elith et al. (2008), De'ath (2007), and De'ath and Fabricius (2000). For a more in-depth discussion of Maxent, I recommend reading the articles written by Phillips and Dudík (2008), Phillips et al. (2006), and Phillips et al. (2004).

Boosted Regression Trees

Boosted regression trees (BRTs) are a relatively new tool that draws insights from both statistical and machine learning techniques. BRTs differ from traditional regression methods that produce a single 'best' model. Instead, using a boosting technique, BRTs adaptively combine large numbers of relatively simple tree models to create one final model with substantially better predictive performance.

BRTs use two separate algorithms, regression trees from the classification and regression tree group of models and boosting builds. Each of these components will be discussed in turn.

Regression Trees

Ecological data are often complex, unbalanced, and may contain missing values. Furthermore, relationships between variables may be strongly nonlinear and involve high-order interactions. Regression trees (RTs) are a modern statistical technique ideally suited for both exploring and modeling such data (Breiman et al., 1984).

RTs have been widely used for the exploration, description, and prediction of ecological data (De'ath and Fabricius, 2000; Vayssières et al., 2000; De'ath, 2002). In the most general terms, a tree-building algorithm will allow researchers to accurately predict cases using a set of *if-then* logical (split) conditions. In ecological terms, a RT explains how a single response will vary based on sequential divisions (splits) of data into increasingly homogenous groups (minimizing the sums of squares within groups), using combinations of explanatory variables that may be categorical and/or numerical. Each split is defined by a simple rule based on a single explanatory variable and at each split the data is partitioned into two mutually exclusive groups. Splitting is continued until an overlarge tree is grown, which is then pruned back to the desired size, determined by cross-validation. The size of the tree equals the number of final groups. Trees are represented graphically with the root node (which represents the undivided data) at the top, and the branches and leaves (each leaf represents one of the final groups) beneath.

RTs have many desirable properties, including (1) their ability to handle various types of predictor variables (numerical, binary, categorical, etc.), (2) trees are invariant to monotonic transformations of the predictors, (3) trees are insensitive to outliers and can accommodate missing data by using surrogates (Breiman et al., 1984), (4) tree methods are nonparametric and nonlinear and therefore well suited for data mining tasks, (5) interactions between predictors are automatically modeled (hierarchical structure of tree), (6) the performance of tree based models are highly resistant to the inclusion of a large number of irrelevant variables, thereby making feature selection much less of an issue, and finally (7) trees present information in a way that is intuitive and easy to visualize.

Despite these benefits, RTs are generally not as accurate as other traditional methods, such as generalized and general linear models (GLMs) and generalized additive models (GAMs). RTs have difficulty in modeling smooth functions and the tree structure depends on the sample of data chosen. In fact, small changes in training data can result in very different series of splits, leading to the creation of very different decision trees (Hastie et al., 2001). Furthermore, large trees can be difficult to interpret. Thankfully, these weaknesses can be overcome through the use of boosted trees; now widely acknowledged as excellent predictors that also render simple graphical and numerical interpretations of complex relationships (Elith et al., 2008)

Boosting

Boosting is a method for improving model accuracy and it originated in the machine learning community. It is based on the idea that it is easier to find and average many rough rules of thumb, than to find a single, highly accurate prediction rule

(Schapire, 2003). There are several techniques that merge results from multiple models; these include bagging, stacking, model averaging, and boosting. Boosting is a forward stagewise procedure where models are fitted iteratively to the training data, gradually, to increase emphasis on observations that are poorly modeled by the existing collection of trees.

For regression problems, boosting is a form of ‘functional gradient descent’. Boosting minimizes the loss function by adding, at each step, a new tree that best reduces (steps down the gradient of) the loss function (Elith et al., 2008). The loss function can be any measure that represents the loss in predictive performance due to a suboptimal model. In my case, binomial deviance was used as the loss function because it is more robust than exponential loss and is likely to perform better against misclassification (e.g. false negative observations; Elith et al., 2008).

The first regression tree in a BRT is the one that maximally reduces the loss function, for the selected tree size. For each following step, the focus is on the residuals: the variation in the response that is not explained by the model. For example, the second tree will be fitted to the residuals of the first tree. The model will then be updated to contain two trees (two terms) and the residuals from this two-term tree model will then be calculated (Elith et al., 2008). This process continues, where the *n*th tree will be fitted to the residuals of the *n*-1 model, the model will be updated to contain *n* trees and the residuals will be re-calculated for this *n*th-term tree. The process is stagewise, not stepwise, thus the existing trees are left unchanged as the model is enlarged. Only the fitted values for each observation are re-estimated at each step to reflect the contribution of the newly added tree. It is important to note that all subsequent trees could contain quite different variables and split points from the first tree. The final BRT model is a linear combination of many trees, in my case 1000-2000 trees, that can be thought of as a regression model where a term is a tree.

One of the downsides of boosting procedures is over-learning, where predictive error reaches a minimum before increasing. This over-learning can be controlled by slowing the learning process (rate at which the model building process moves down the loss gradient). Typically, the contribution of each tree to the model is shrunk by a learning rate (shrinking rate) that is substantially less than one, in my case 0.01 - 0.001. Thus smaller learning rates need to be compensated for by increasing the number of iterations (trees) in the boosting sequence. A 10-fold reduction in learning rate requires an approximately 10-fold increase in iterations (Elith et al., 2008). Fitted values in the final model are then computed as the sums of all the trees multiplied by the learning rate (Elith et al., 2008). These fitted values are much more stable and accurate than those from a single decision tree model. I also used a 10-fold cross-validation method to determine the optimal number of trees for a given learning rate and data set, thus further reducing model overfitting (Elith et al., 2006).

Stochastic gradient boosting

Finally, the performance of gradient boosting is also improved by injecting randomness into the sequential fitting (Freidman, 2002). According to De'ath (2007), injecting randomness has three important benefits: predictive performance improves substantially, computation is reduced (subsamples of the training data are taken without replacement), and over-learning is reduced. However, stochasticity does introduce variance in fitted values and predictions between runs (Elith et al., 2008).

In BRTs, stochasticity is controlled through a “bag fraction” that specifies the proportion of data to be selected, without replacement from the full training set, for each iteration. For example, a bag fraction of 0.5 means that, for each iteration (tree), 50% of the data would be drawn at random, without replacement, from the full training data set. The use of bag fractions in BRTs has shown to increase model performance when values for the bag fraction vary from 0.5 to 0.75 (Elith et al, 2008). Based on preliminary tests (Figure 3-1) and the Elith et al. paper, I used an intermediate bag fraction of 0.5.

Tree Complexity and Interactions

Tree complexity (*tc*) represents the number of nodes in a tree and theoretically should reflect the true interaction order in the response being modeled (Friedman, 2001). Similar to learning rate (*lr*), *tc* also affects the optimal number of trees needed to minimize loss. As *tc* is increased (1-10), *lr* must be decreased if sufficient trees are to be fitted (Elith et al., 2008). In their review using datasets varying from 250 to 6000 localities, Elith et al. (2008) demonstrated that gains from increased *tc* were greater with larger data sets and that decision stumps (*tc* = 1) always had higher predictive deviance. The best strategy thus remained to have higher *tc* and very slow *lr*, however the slower the *lr* the longer the computing time. Thus, Elith et al. (2008) developed a general rule of thumb: *lr* should vary inversely with *tc* and users should use the fastest *lr* that achieves more than 1000 trees.

It is important to note that even if a decision tree has several nodes (*tc*>1), it may not be modeling interaction between predictors. Interactions are only fitted if supported by the data and *tc* controls the maximum level of interaction that can be quantified. Based on preliminary test and the Elith et al. paper (2008), I used an intermediate tree complexity of 5 to reach a balance between minimizing running time, predictive performance, and an interpretable interaction order (Figure 3-1). To extract information on the nature and magnitude of fitted interactions the gbm.interactions function (Elith et al., 2008) will create, for each possible pair of predictors, a temporary grid of variables representing combinations of values at fixed intervals along each of their ranges. Predictions are then formed on the linear predictor scale of this grid, while all other variable values are set to their respective means. A linear model is then used to relate the temporary predictions to the two marginal predictors, fitting the later as factors (Elith et al., 2008). Finally, the residual variance in this linear model indicates the relative strength of interactions fitted by BRT. A residual variance of zero indicates that no interaction effects have been fitted.

Relative Importance of Predictor Variables

One of the most important steps of model interpretation is determining which variables are important for prediction, and understanding their joint effect on the response. For this reason, it is often useful to determine the relative importance (contribution) of each input variable in predicting the response.

For a single tree *T*, Breiman *et al.* (1984) proposed a measure of (squared) relevance I^2_j (T) for each predictor variable x_j, based on the number of times that variable was selected for splitting in the tree weighted by the squared improvement to the model as a result of each of those splits. This importance measure is easily generalized to boosted regression trees; it is simply averaged over all trees (Freidman and Meulman, 2003). The relative influence (or contribution) of each variable is then scaled so that the sum adds to 100, with higher numbers indicating stronger influence over the response.

MAXENT

Application to Species Distribution Modelling

The purpose of species distribution modeling is to predict the distribution of a species of interest. Thus the question arises: "what is the best approximation of this unknown distribution"? The Principle of Maximum Entropy is based on the premise that when estimating the probability distribution, you should select the distribution which leaves you the largest remaining uncertainty (i.e., the maximum entropy) consistent with your constraints (Jayne, 1957). These constraints represent the incomplete information about the target distribution. By following this golden rule, you avoid introducing additional assumptions or biases into your calculation of the probability distribution (Jaynes, 1990). Shannon (1948) described entropy as "a measure of how much 'choice' is involved in the selection of an event". Thus, by utilizing the principle of maximum entropy we are selecting the distribution offering the most choices and no unfounded constraints are placed on the probability distribution (Phillips et al., 2006).

In maximum entropy modelling, the true distribution of a species is estimated using a set of constraints derived from the occurrence data. The constraints for this estimated probability distribution are the empirical averages of each environmental variable at the occurrence sites. For example: for the feature “terrain ruggedness”, the corresponding constraint dictates that the mean terrain ruggedness predicted by the model should be close to the average value of terrain ruggedness at sighting locations. Typically, the set of constraints under-specifies the model. To account for this, one must select the model, among all probability distributions satisfying the constraints, that maximizes entropy (i.e. the most unconstrained one) (Jaynes, 1957). In Maxent models, the entropy is non-negative and is at most the natural log of the number of elements in the set of sites (Phillips et al., 2006).

The Maxent distribution belongs to the family of Gibbs distributions derived from the set of features $f_1, \ldots, f_n$ (Dudík et al., 2004). Gibbs distributions are exponential distributions parameterized by a vector of feature weights $\lambda = (\lambda_1, \ldots, \lambda_n)$ and whose probabilities $q_\lambda(x)$ are normalized by Z_λ, thus ensuring that probabilities $q_\lambda(x)$ sum to one over the study area (Phillips and Dudík, 2008). Therefore, the value of the Maxent model q_λ at a site x depends only on the feature values (environmental values) at x. This means that not only can the Maxent model predict to the study area, it can project predictions to other study sites where the same environmental variables are available (Phillips, 2008). However predictions to new study sites or temporal predictions should be interpreted with caution (Phillips, 2008).

For each Maxent model, the estimated distribution is the Gibbs distribution q_λ that maximizes a penalized log likelihood of the presence sites, the regularization parameter is the width of the error bound for feature f_j, and $x_1, \ldots, x_n$ are the presence sites. Log likelihood gets larger as we obtain a better fit to the data. Thus, log likelihood is larger for Maxent models that give more probability to the presence sites and less to the rest of the sites (i.e. models that can best distinguish the presence sites from the background). Regularization (also known as the lasso penalty; Tibshirani, 1996) gets larger as the weights λ_j get larger. Larger weights typically mean that the model is more complex and is thus more likely to overfit (Phillips and Dudík, 2008). Thus to obtain the best fit to our data we select the Gibbs distribution that maximize the difference between the log likelihood and regularization, but isn't too complex a model. The tradeoff between these three components can be controlled by the selection of regularization parameters. By governing how close the expected values under the estimated probability distribution are required to be to the observed values, the regularization parameters prevent the Maxent algorithm from overfitting to training data (Phillips and Dudík, 2008; Hastie et al., 2001). The Maxent algorithm is forced to focus on the most important features and thereby reduces the number of parameters included in each model (Phillips et al., 2006).

Machine Learning Algorithm

Phillips et al.'s (2006) implementation of Maxent for species distribution modeling utilizes a sequential-update algorithm (Dudík et al., 2004) that iteratively picks a weight λ_j and adjusts it so as to minimize the resulting regularized log loss (the negative log likelihood of the sample points). The regularization parameter for each feature depends on its observed variability and is automatically selected by the algorithm based on the types of features used and the number of sample localities. The algorithm is deterministic, and is guaranteed to converge to the Maxent probability distribution (Phillips et al., 2006). The algorithm reaches its end when either the change in log loss in an iteration falls below a user-specified value or when a user-specified number of iterations has been performed (in this study, 1000 iterations).

Relative Importance of Predictor Variables

With Maxent the percent contributions of predictor variables are determined by the increase in the gain of the model due to the modification of a coefficient for a single feature. Each step of the Maxent algorithm increases the gain of the model by modifying the coefficient for a single feature; the program then assigns the increase in the gain to the environmental variable(s) that the feature depends on, converting to percentages at the end of the training process.

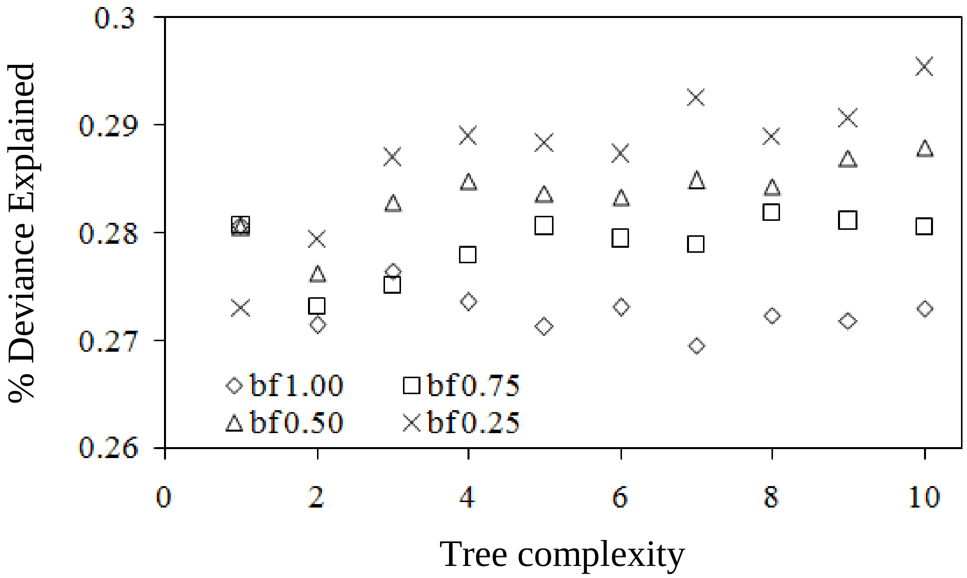

Figure 3-1. Percentage deviance explained as a function of tree complexity and bag fraction (*bf*). The predictive ability of models was tested on independent data. The learning rate that achieved a minimum of 1000 trees was used for each parameter combination.

APPENDIX II

Recursive Feature Elimination using Maximum AUC Values

Feature selection offers a number of advantages, including more powerful classification models by eliminating irrelevant or noisy features (Krishnapuram et al., 2004), more compact and faster models by constructing them using only a small subset of the original set of features (Abeel et al., 2010), and the ability to focus on a subset of relevant features, which can be used for the discovery of new knowledge (Guyon and Elisseeff, 2003). In the context of machine learning, feature selection can help decrease the risk of a classification model over-fitting to training data. Habitat suitability models built using widely available environmental variables can then be extrapolated to regions outside the initial study area, with great care, using similar GIS datasets (Phillips, 2008).

In the machine learning community, feature selection techniques can fall into the filter, wrapper, or embedded class of methods, depending on how they interact with the estimation of the classification model (Saeys et al., 2007; Guyon and Elisseeff, 2003). Wrappers utilize the learning machine of interest as a black box to score subsets of variable according to their predictive power. Filters select subsets of variables as a pre-processing step, independently of the chosen predictor. Embedded methods perform variable selection in the process of training and are usually specific to given learning machines (Guyon and Eliseeff, 2003).

In a procedure similar to backwards stepwise selection in logistic regression, I reduced the number of environmental variables in each model using the area under the operating curve (AUC) as an evaluation metric. In backwards or forwards stepwise logistic regression, the Akaike Information Criterion is often used to determine the stopping point during model simplification. However, because I used machine learning algorithms, I had to use a different evaluation metric than the more commonly used AIC.

The AUC is threshold independent and used in both Maxent and BRT modeling techniques to evaluate model performance. Thus, for those reasons, I used AUC values to determine the stopping point during model simplification.

In the past, several studies (Yang and Hu, 2009; Abeel et al., 2010; Doetsch et al., 2009; Dreiseitl and Osl, 2009; Wang et al., 2007) have used changes in AUC values to determine when to stop removing environmental variables from models. In this study, I used AUC values during backwards feature selection to determine my stopping point and used the ranking given by the modeling algorithm itself (percent contribution of predictor variables) to determine which variables were eliminated first, thus combining the embedded and wrapper type methods for recursive feature elimination (RFE). In all cases, final subset selection corresponded to the model that had the highest AUC values when tested on independent data. Thus, candidate variables were removed in a backwards fashion, beginning with the lowest overall contributors, until test AUC values reached their maximum.

This is the first reported instance where a researcher has used the maximal test AUC value as the threshold stopping criterion. By looking to maximize AUC values on independent testing data I am reducing the risk of selecting a model that is overfit to the training data (Reunanen, 2003).

When using the BRT algorithm, I compared the results of my simplification process with those obtained using the embedded recursive feature elimination tool provided by Elith et al. (2006). This tool simplifies the model by dropping the least important predictor, then re-fits the model and sequentially repeats the process until the average change in predictive deviance exceeds its original standard error. For each new run, a 10-fold cross-validation procedure is used to determine the average CV error. Following analysis, the optimal number of variables that can be removed without affecting predictive performance is identified. Because this tool was only available for BRTs and only used training data to determine variable selection, I did not use it for more than a comparison. However, both techniques yielded similar results using BRTs with respect to the ranking of variables for elimination and the optimal number of variables to be removed (Table 4-1.).

BRTs and Maxent are two very different modeling techniques and it is not surprising that both these techniques reacted differently to RFE (Figure 4-1). When all spatial scales are grouped together, the overall average percent change in AUC values during RFE is greatest when using the BRT modeling algorithm and the optimal number of eliminated features is approximately 2. On the other side, in Maxent models, the optimal number of variables removed is approximately 5 and there is a much smaller overall percent change in AUC values during RFE (Figure 4-1). This trend implies that Maxent models are less likely to overfit to the training data than BRT models. Interestingly, the removal of the %Plantation and %Cutovers and prescribed burns land cover variables increased model performance when tested on independent data at all spatial scales for both modeling algorithms.

Table 4-1. Comparison between the embedded simplification tool (deviance)(Elith et al., 2006) and the maximum test area under the receiver characteristic operating curve technique (max AUC) for recursive feature elimination in BRT models at all spatial scales. The optimal number of variables identified for elimination is reported. Furthermore, the ranking of variables for RFE is examined and compared between both models.

	1 km		500 m		250 m	
	Max AUC	Deviance	Max AUC	Deviance	Max AUC	Deviance
Optimal #	3	1	2	2	2	1
Ranking[1]	same		Same		same	

[1] same = same ranking of variables for elimination.

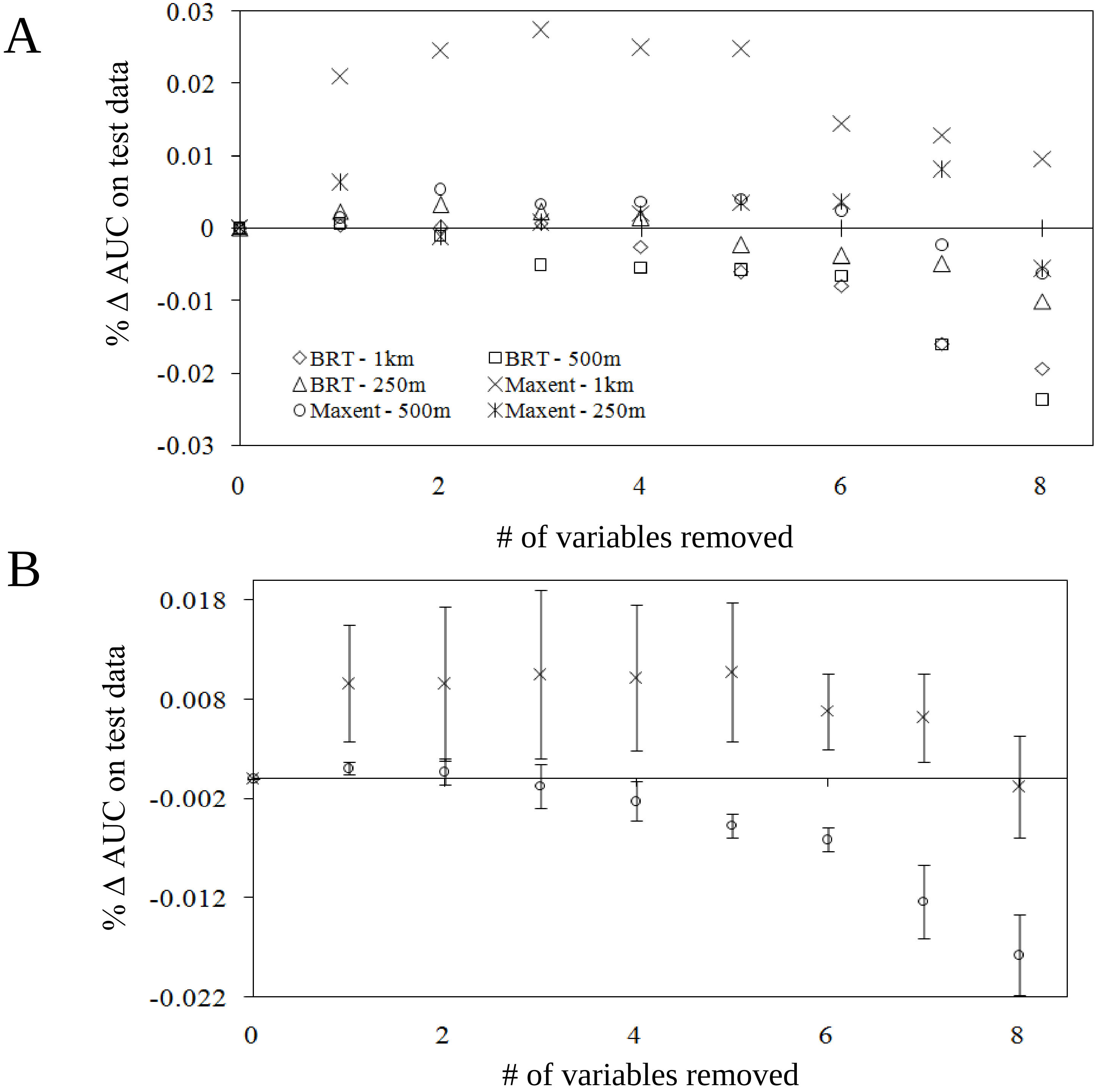

Figure 4-1. A) Percentage change in test AUC values during recursive feature elimination on models built using the BRT and Maxent modeling algorithms at all spatial scales. **B)** Average percent change in test AUC values during recursive feature elimination on models built using BRTs (o) and Maxent (x).

APPENDIX III

Target Background

Occurrence records, usually collected by museum or herbaria, often exhibit strong spatial bias in survey effort (Dennis and Thomas, 2000; Reddy and Dávalos, 2003; Schulman et al., 2007). This bias, referred to as sample selection bias or survey bias (Phillips et al., 2009), can severely impact model quality. To correct estimates of species' distributions, Phillips et al. (2009) proposed choosing background data with the same bias as occurrence data. When the sampling bias is not known, it can be approximated by combining occurrence records for a target group of species that are all collected or observed using the same methods. Phillips et al. (2009) found that target-group background improved the average performance of all the modeling methods examined, these included BRTs and Maxent. In fact, the choice of background data had as large an effect on predictive performance as the choice of modeling method when using presence-only data (Philips et al., 2009).

The goal of this appendix is to determine whether using target background data, sharing in the same sampling bias, would improve model predictive ability by removing sampling bias. To achieve this goal, instead of using random points to populate my pseudo-absences sites, I used sighting data from the NHIC for turtles species found in Ontario (Oldham and Weller, 2010). Thus, occurrence records for the red-eared slider (*Trachemys scripta elegans*), eastern box turtle (*Terrapene carolina carolina*), common musk turtle (*Sternotherus odoratus*), northern map turtle (*Graptemys geographica*), painted turtle (*Chrysemys picta*), snapping turtle (*Chelydra serpentina serpentina*), and spiny softshell turtle (*Apalone spinifera spinifera*) were used as background data. Using the focus tool (Holland, 2004), I choose the biggest subset of target background points that were at least 1 km apart (n=4833). As before, models were created using both the

BRT and the Maxent modeling algorithms since different results were obtained when using these different techniques. Apart from using target background, the modeling methodology is identical to that of Chapter 2.

When using target background, the overall contribution of roads to all models dropped by approximately 10-15 percentage points (Figure 5-4). Furthermore, the dependency plots for univariate models built using only road density as the explanatory variable showed that habitat suitability dramatically decreased with increasing road density (Figure 5-5). The two main possibilities for these observed differences are the following: 1) there is a strong sampling bias towards roads due to the increased frequency with which these areas are used by human populations and thus by using target background we were able to account for this bias, and 2) in addition to this human sampling bias towards roads, all turtle species are equally attracted to roads for thermoregulation and nesting activities and thus by using target background we are also accounting for a natural bias towards roads. Both of these explanations have significant impacts on model interpretation and thus the results from this analysis must be carefully weighed.

As discussed in Chapter 2, the importance of road density increased with decreasing spatial scale. It seems the negative impact of roads on habitat suitability is greater at the local scale than at the larger annual scale.

The overall contribution of water to habitat suitability models also greatly decreased when using target background, approximately 10 percentage points. This could be due to the nature of the taxonomic groups used as target background. All the turtle species used for the target background are primarily found in riparian habitat. These species are primarily residents of both open water habitat and wetland habitat, whereas Blanding's turtles are primarily residents of wetland habitats. Thus, similar to roads, open water habitat was being targeted and any selection for this habitat could not be quantified appropriately. These results suggest that to properly account for human sighting bias taxonomic groups with similar sampling bias and life-history traits should be used, as long as these groups have different habitat characteristics. In this case, the use of historical snake sightings instead of turtle sightings might have been more revealing. Both taxonomic groups would have similar human sighting biases, they both have similar life-history traits, and yet they do not share the same primary habitats.

Cropland area was of much greater importance in models built with target background. Identical to models built with random background points, increasing cropland area greatly decreased habitat suitability for Blanding's turtles whereas increasing forest area increased habitat suitability. These results suggest that the greatest threat to Blanding's turtle habitat suitability is habitat alteration and destruction in Ontario and preservation of undisturbed terrestrial habitat would greatly increase overall habitat suitability (Figures 5-2 to 5-5). Apart from road density, total open water area, and total cropland area, the other predictor variables were of similar importance when compared to models using random background points and had similar partial dependency and univariate plots.

Although the use of target background supported our earlier results using random background points, and effectively eliminated the sampling bias towards roads, the overall predictive performance of models dramatically decreased. When tested on independent data, models build with target background had informative, but not excellent

AUCs (Table 5-1). However, the difference in results between my study and the Phillips et al. (2009) study on the use of target background could be attributed to the fact that I did not use explicit presence-absence records to test my models. Similar to another study using target background (Lütolf et al., 2006), models trained and tested on target background had much lower AUC than models trained and tested on random background. Furthermore, the changes I observed in the overall importance of water area and road density must be interpreted with caution. For these reasons, I decided to present the models built using random background as they had excellent predictive power and any possible natural selection for roads and open water habitat, although clouded by sampling bias, was not removed. There is great potential for further research in the area of target background selection. Using target background has proven to be very effective for accounting for sampling bias (Phillips et al., 2009), however clear guidelines need to be put into place for target background selection. Great differences in study sites, number of sighting occurrences, and the species of interest, could have profound effect on the appropriate selection of target background.

Table 5-1. Predictive performance of BRT and Maxent models tested on independent data (test) and training data (data used to build models) for each spatial scale examined. The threshold independent, area under the receiver characteristic operating curve (AUC) is reported.

Model	BRT		Maxent	
	Test	Training	Test	Training
1km	0.739	0.816	0.741	0.793
500m	0.741	0.819	0.731	0.791
250m	0.728	0.823	0.727	0.778

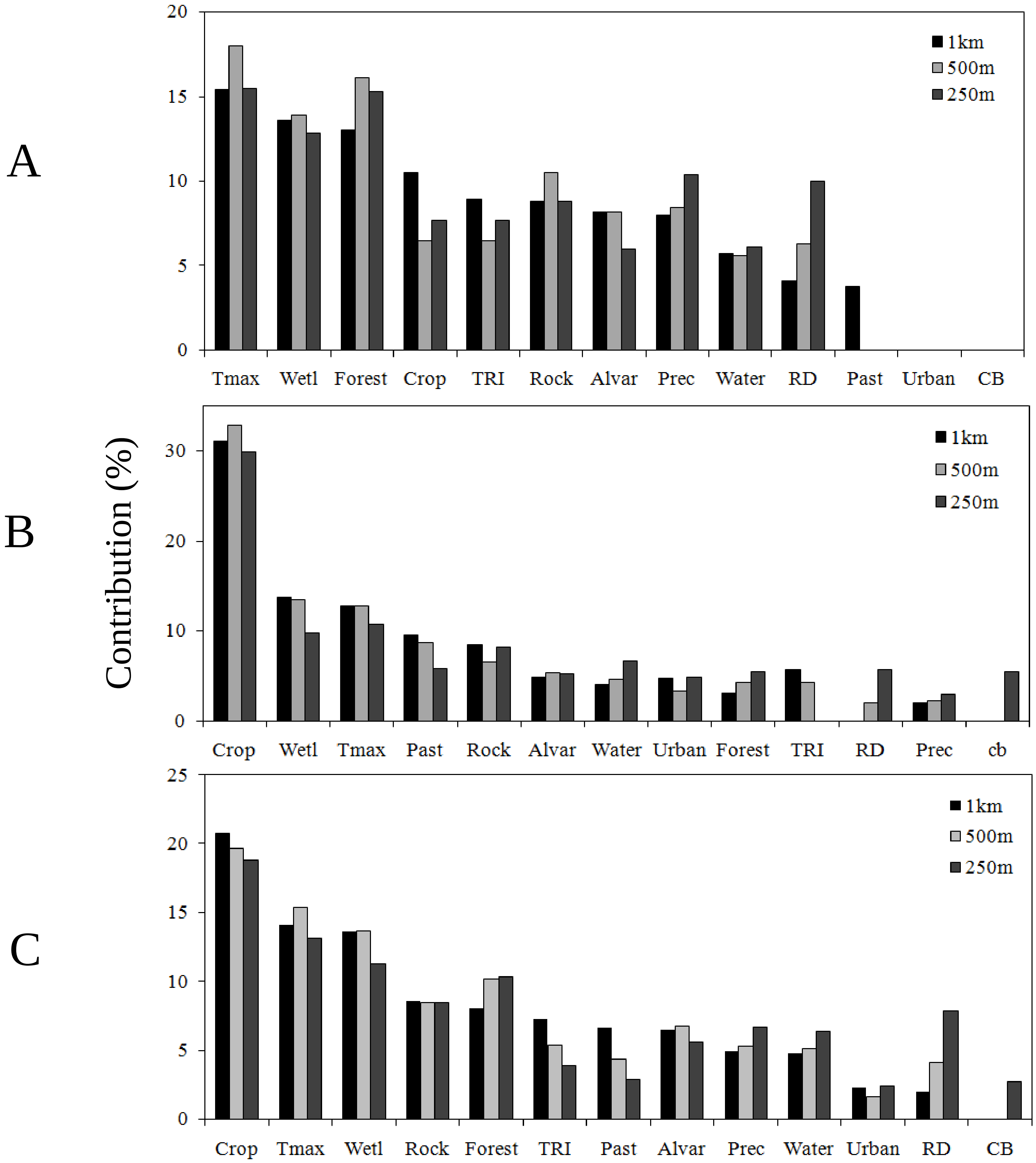

Figure 5-1. Average contribution of predictor variables in habitat suitability models built using target background and (**A**) BRTs, (**B**) Maxent, or (**C**) the average of both, at all three spatial scales. For comparison with models built using random background, see Figure 2-5.

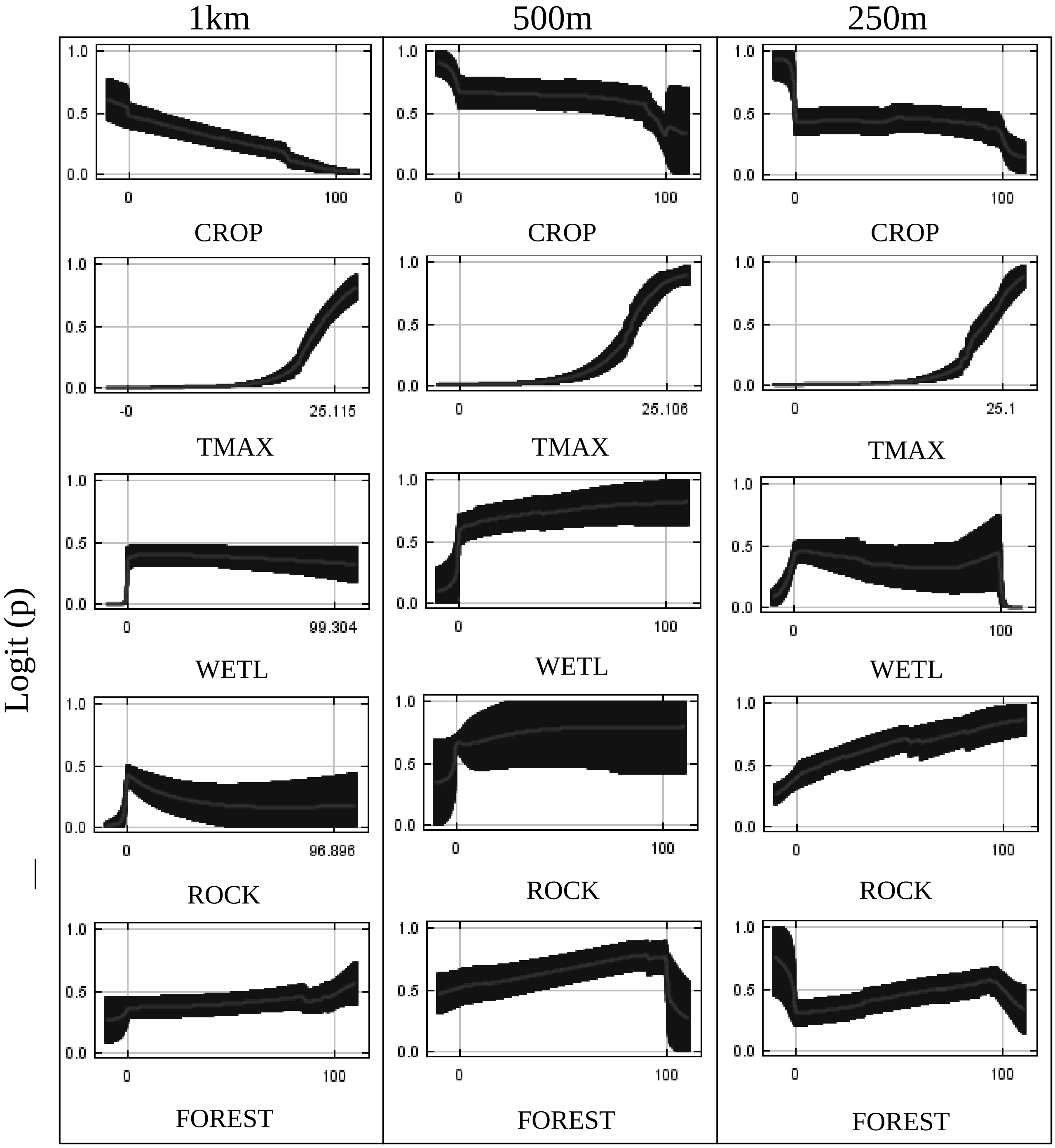

Figure 5-2. Partial dependence plots ± SD for the top five most influential predictor variables in habitat suitability models built using target background, Maxent, and three different spatial scales. Y axes are on the logit scale and represent the predicted probability of suitable conditions with all other variables set to their average values over the set of presence localities.

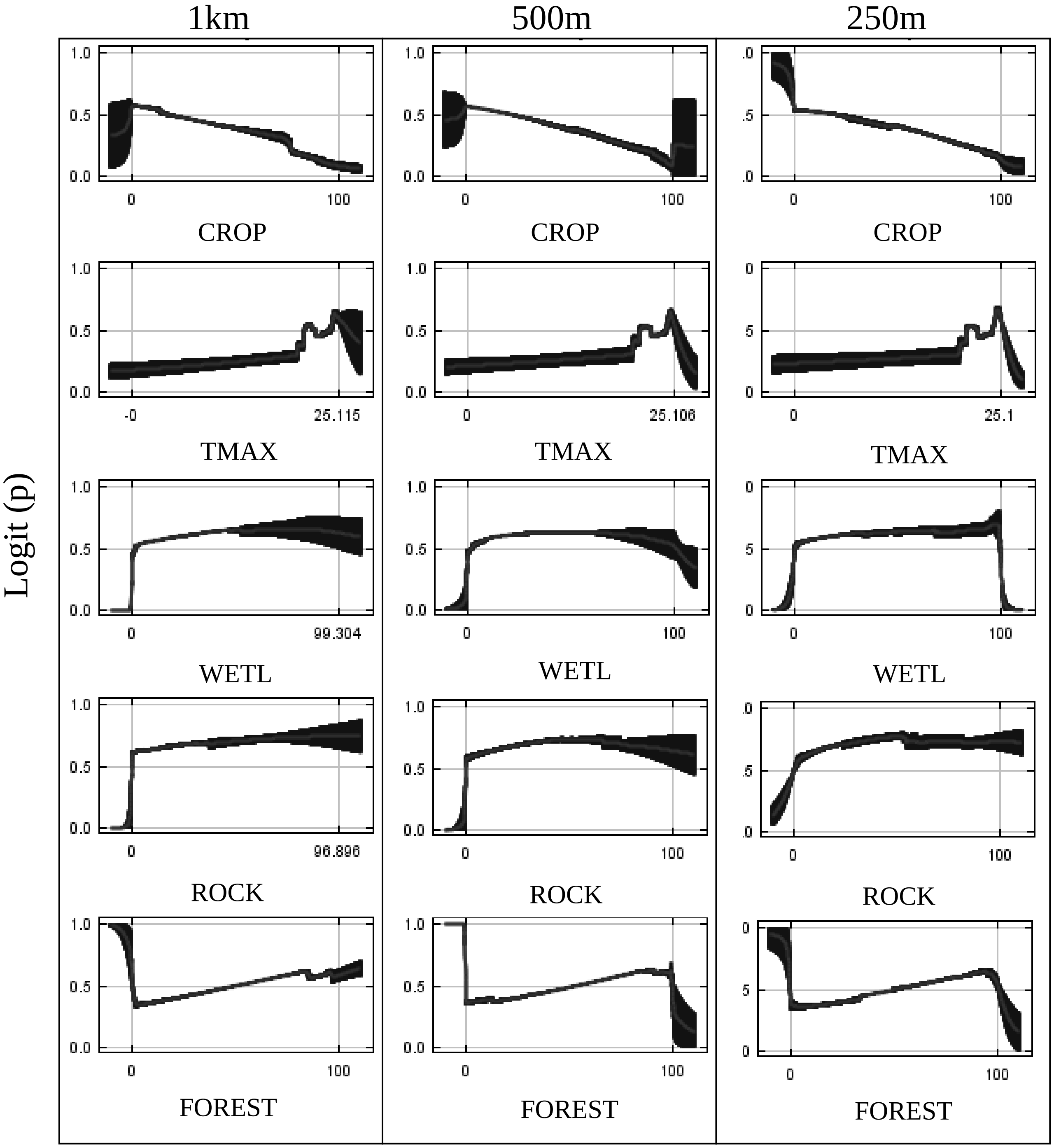

Figure 5-3. Response curves ± SD produced by univariate models of the five most important predictor variables in habitat suitability models built using target background, Maxent, and three different spatial scales. Y axes are on the logit scale and represent the predicted probability of suitable conditions based on the effect of each predictor variable independently.

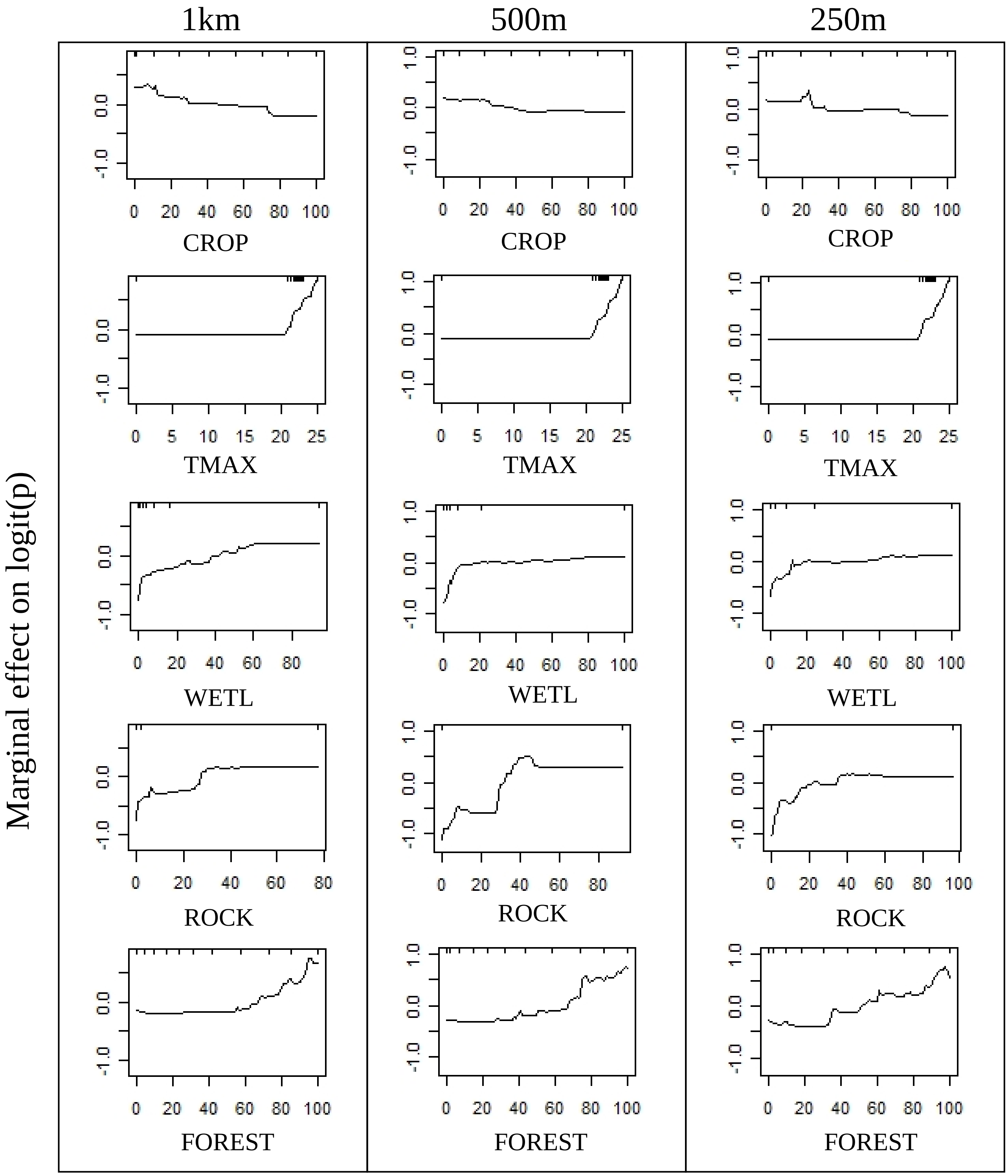

Figure 5-4. Partial dependence plots for the top five most influential predictor variables in habitat suitability models built using target background, BRTs, and three different spatial scales. Y axes are on the logit scale and represent the predicted probability of suitable conditions based on the effect of each predictor variable independently.

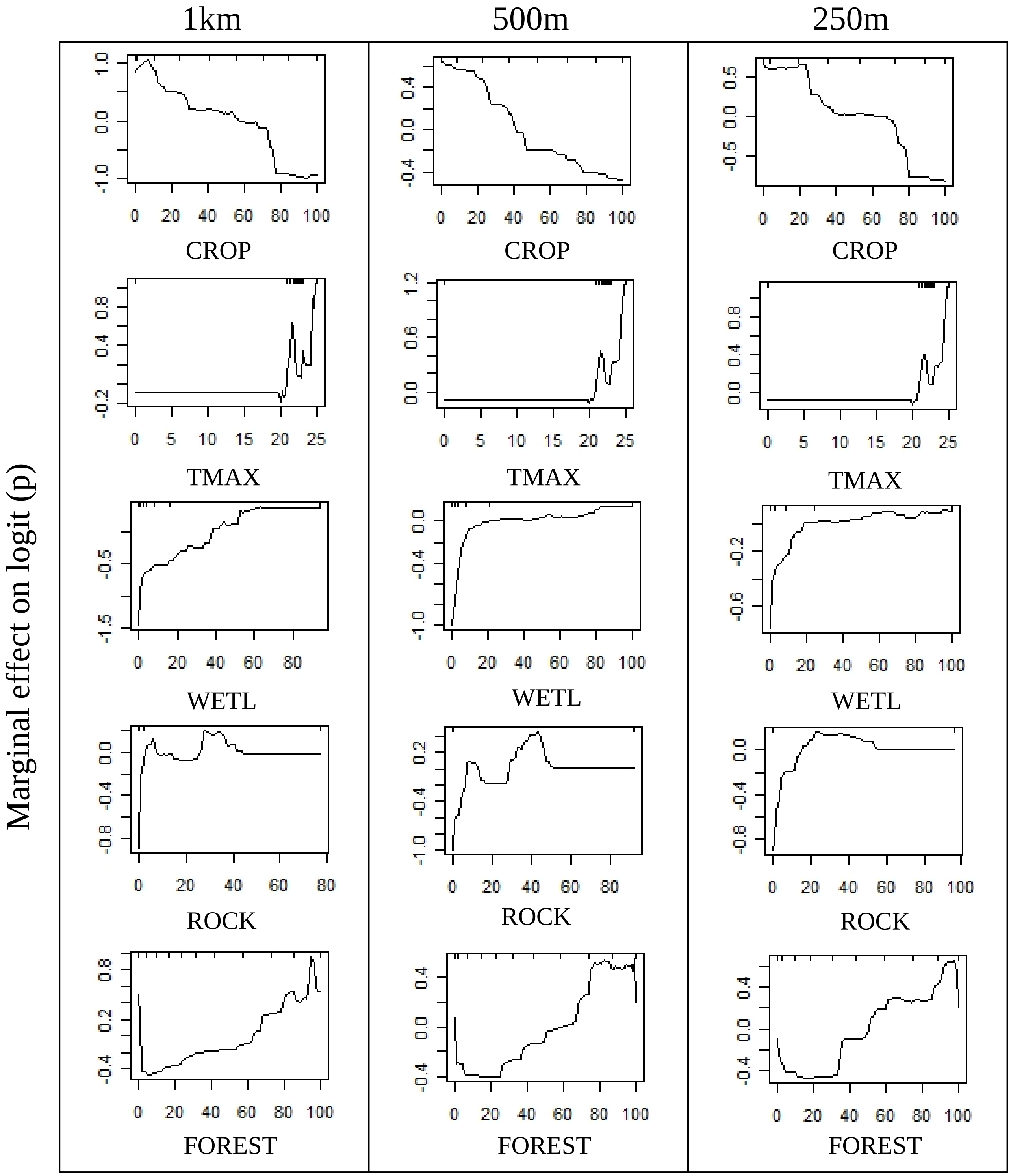

Figure 5-5. Response curves ± SD produced by univariate models of the five most important predictor variables in habitat suitability models built using target background, BRTs, and three different spatial scales. Y axes are on the logit scale and represent the predicted probability of suitable conditions based on the effect of each predictor variable independently.

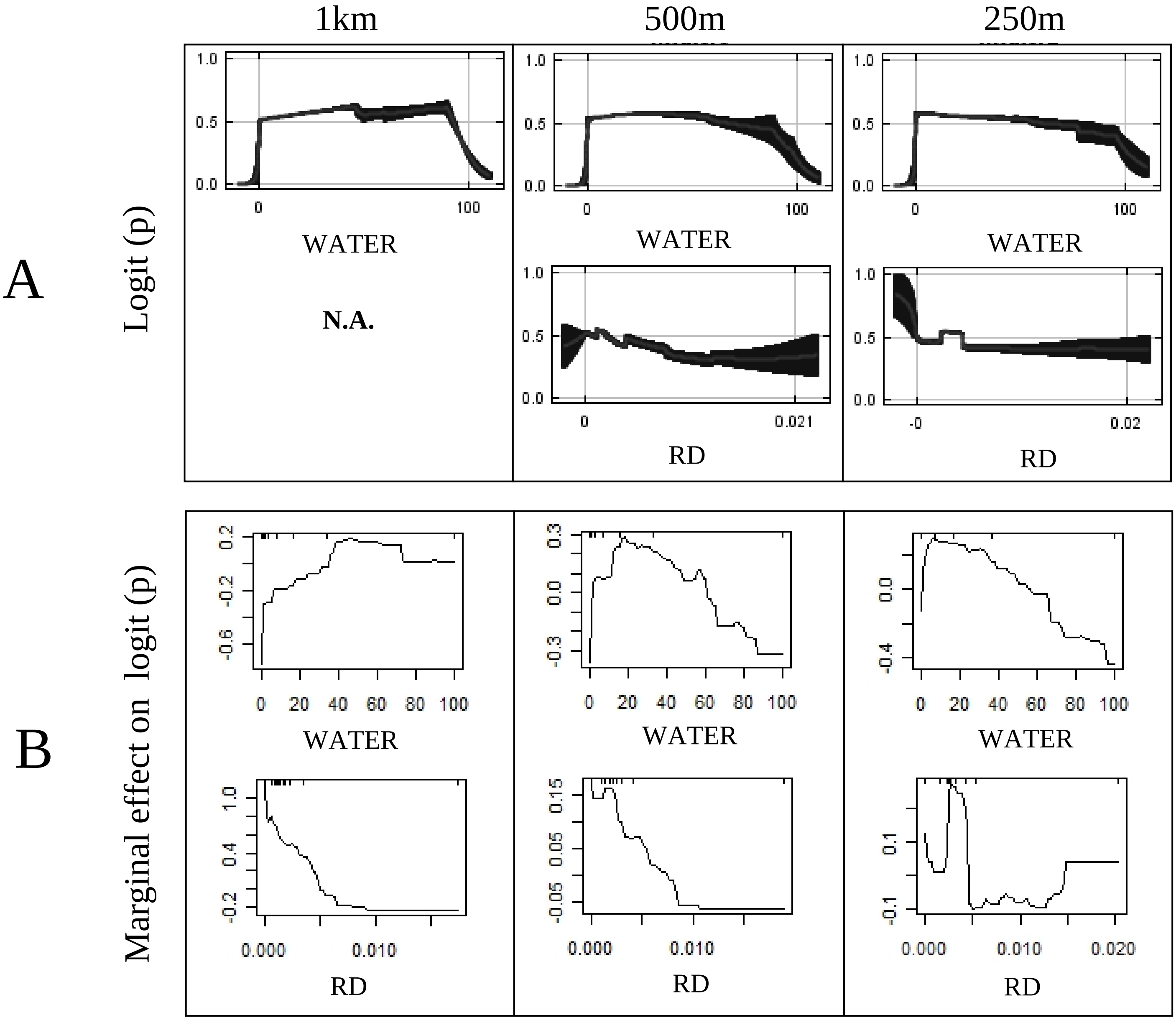

Figure 5-6. Response curves ± SD produced by univariate models of road density and open water area built using target background, BRTs or Maxent, and three different spatial scales. Y axes are on the logit scale and represent the predicted probability of suitable conditions based on the effect of each predictor variable independently.

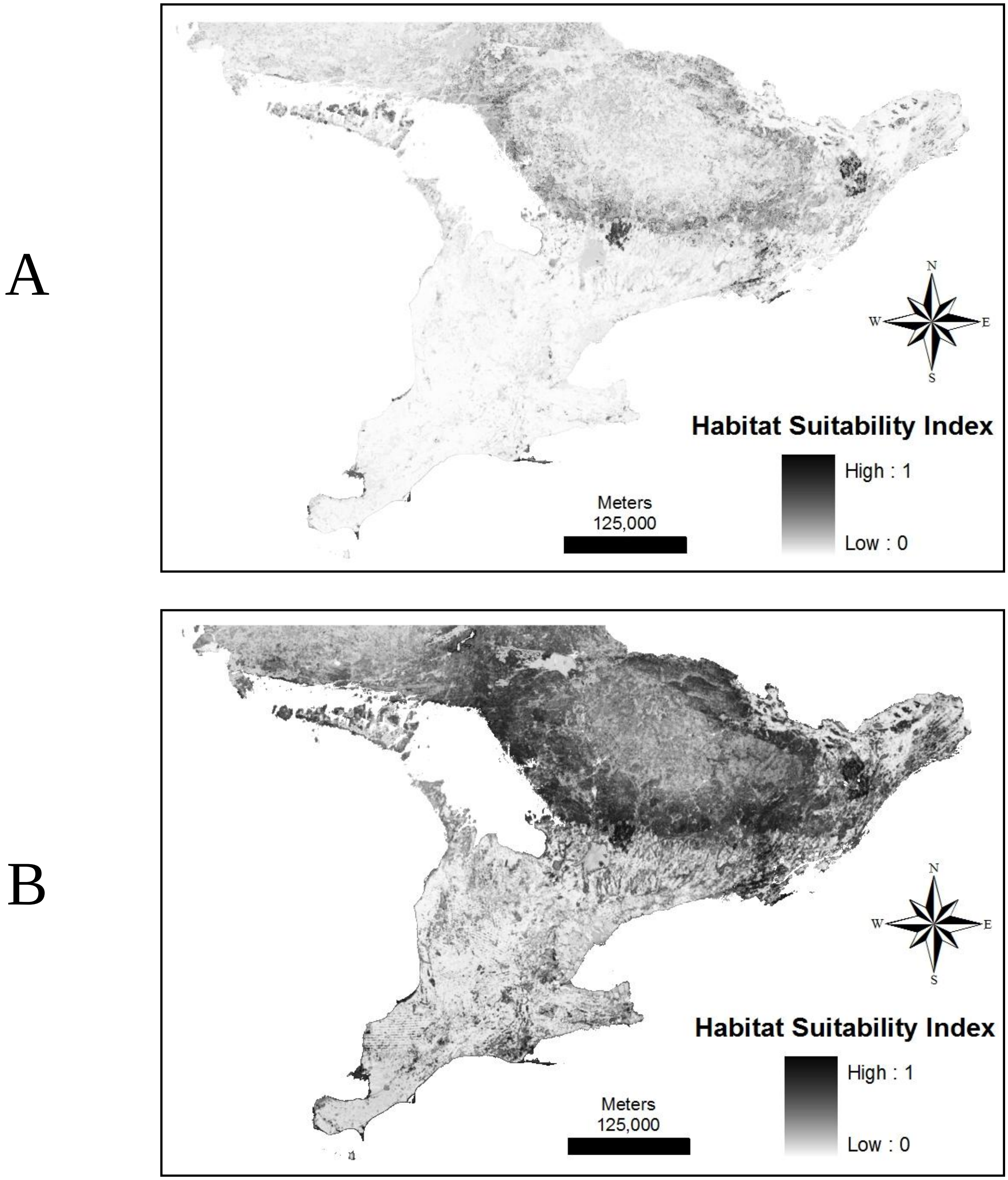

Figure 5-7. Predicted habitat suitability for Blanding's turtles across their range in Ontario using target background, a 250 m buffer, and two modelling algorithms: (**A**) BRTs and (**B**) Maxent.

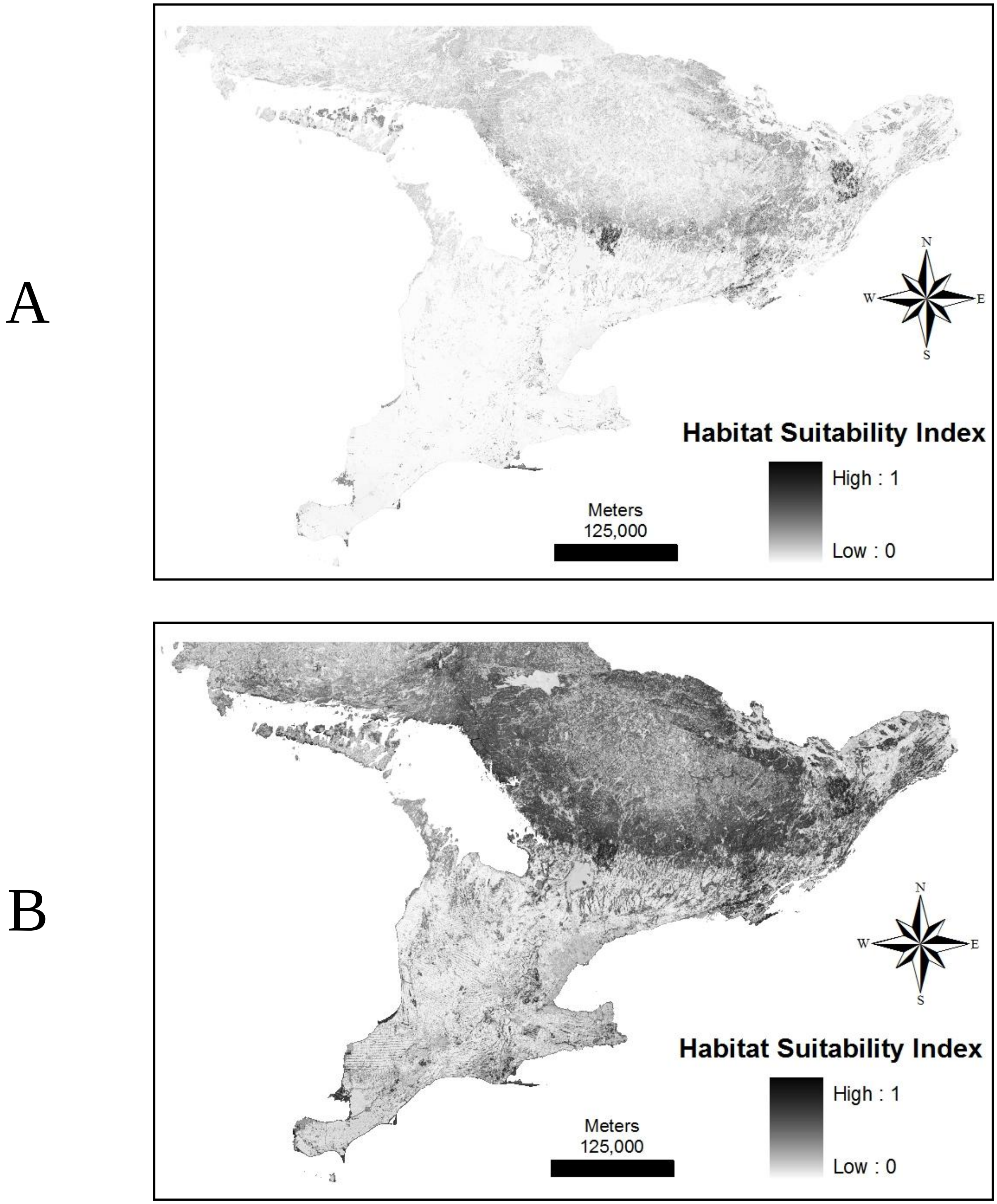

Figure 5-8. Predicted habitat suitability for Blanding's turtles across their range in Ontario using target background, a 500 m buffer, and two modelling algorithms: (**A**) BRTs and (**B**) Maxent.

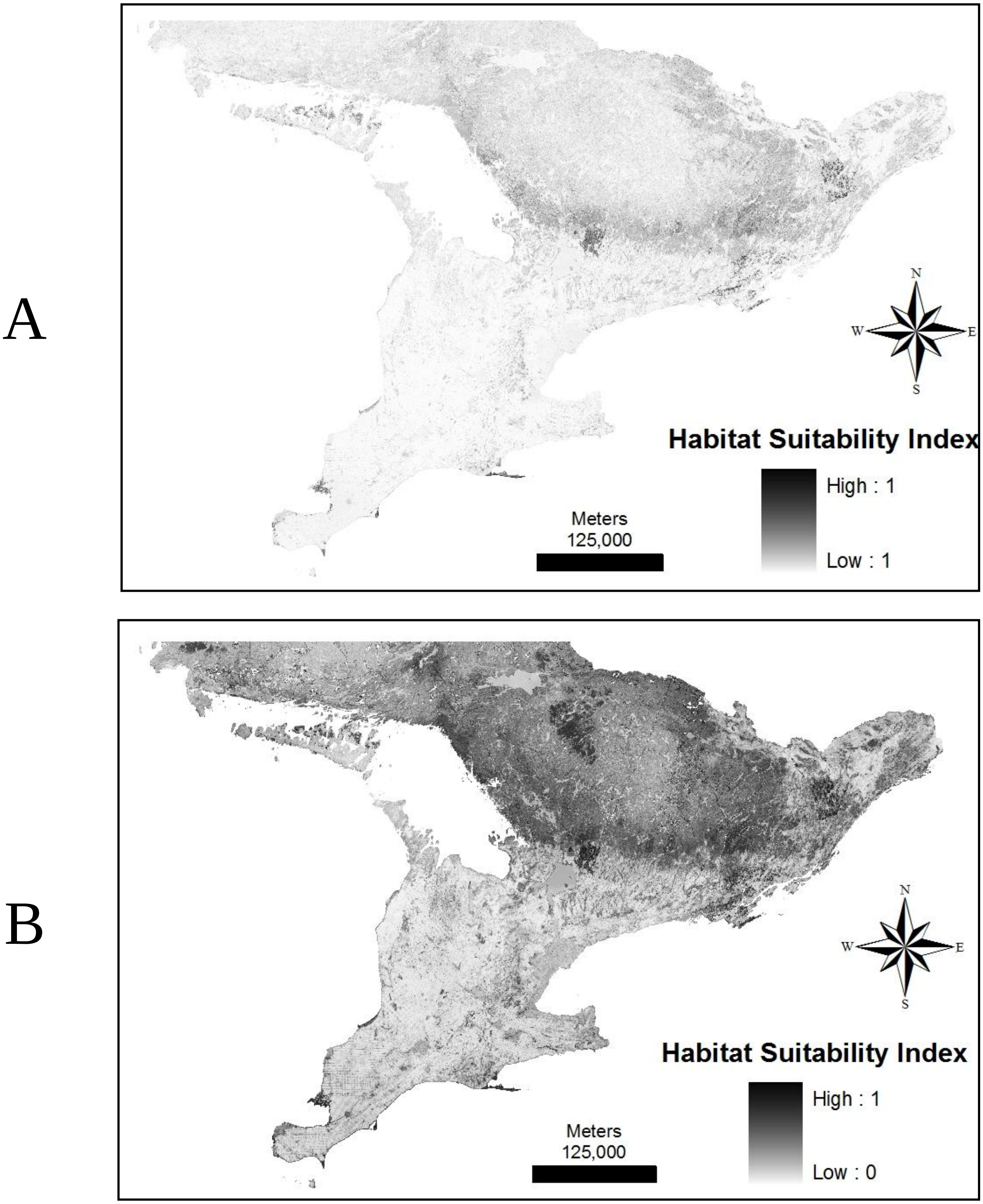

Figure 5-9. Predicted habitat suitability for Blanding's turtles across their range in Ontario using target background, a 250 m buffer, and two modelling algorithms: (**A**) BRTs and (**B**) Maxent.

APPENDIX IV

Effect of roadkill on model performance and predictor variable contributions.

There are several methodological considerations to be made when building SDM models and decision-making becomes an important tool for researchers. Decisions on sample size (Stockwell and Peterson, 2002; Elith et al., 2006; Hernandez et al., 2006), sample clustering and bias (Guisan et al., 2006; Segurado et al., 2006; Phillips et al., 2009), and the geographical accuracy of occurrence records (Hernandez et al., 2006) must be made. In my study, I only used records with the highest accuracy, 100 m, and I studied the effect of sampling bias in Appendix III. In this Appendix, I will discuss the tradeoff between removing sample clustering around roads, due to dead turtle sightings on roads, and sample size.

Dead turtles have an increased likelihood of being observed by passing pedestrians and motorists, compared to live turtles, as they are temporally fixed in a highly visible location. The goal of this appendix is to determine whether including sightings where turtles were found dead on roads unduly influenced model results. To achieve this goal, I subdivided the dataset of 616 sighting points (Chapter 2), into two sets. The first set consisted of occurrence localities where all turtles were found alive (NODOR) and the second set consisted of occurrence localities where turtles were found dead on roads (DOR). Firstly, I wanted to determine whether there were differences in predictor variable contributions and model predictive ability between models built using NODOR data and models built using DOR data. Secondly, I compared models built with NODOR data and those built using all the data available (ALL), to determine whether or not the inclusion of DOR sightings was unduly influencing model results. Furthermore, the overall decrease in sample size caused by the removal of DOR sightings could be more influential on predictive ability than the clustering of DOR record. Models were created using BRTs only and the modeling methodology is the same as that of Chapter 2.

Although predictor variable contributions did change between DOR and NODOR models, road density was 15-20 percentage points higher in DOR models, the differences between NODOR and ALL models were negligible (Figures 6.1-6.8). Models built using all the data available had, on average, greater predictive ability than those built using subsets of the data (Table 6-1). These results are similar to those of Kaliontzopoulous et al. (2008): complete datasets worked better than those reduced to eliminate clustering and higher AUC values and lower AUC standard deviations characterized models produced using all the observations available. Since both BRT and Maxent modeling are not regression-based techniques and automatically incorporate complex interactions between predictor variables, they are less liable to autocorrelation than other methods (Phillips et al., 2006; Elith et al., 2006).

Although both BRTs and Maxent have proven to be robust to variations in sample size and outperformed other modeling algorithms at smaller sample sizes (Elith et al., 2006; Phillips et al., 2006), sample size strongly affects predictive ability (Pearce and Ferrier, 2000; Stockwell and Peterson, 2002). Thus it is no surprise that models built using all the data available had greater predictive ability than those build using only live turtle sightings. For this reason, I decided to include occurrence records where turtles were found dead on roads. My results, like Kaliontzopoulou et al. (2008) suggest that sample size overcomes the importance of data clustering at all spatial scales.

Table 6-1. Predictive performance of BRT models tested on independent data (test) and training data (data used to build models) for each spatial scale examined. DOR models were built using only roadkill sightings, NODOR models were built using only live turtle sightings and ALL models were built using all sighting records (NODOR + DOR). The threshold independent, area under the receiver characteristic operating curve (AUC) is reported.

Model	**Scale**	**AUC**	
		Test	Training
DOR	1km	0.827	0.991
	500m	0.871	0.989
	250m	0.907	0.991
NODOR	1km	0.870	0.964
	500m	0.900	0.945
	250m	0.900	0.957
ALL	1km	0.882	0.945
	500m	0.883	0.960
	250m	0.912	0.956

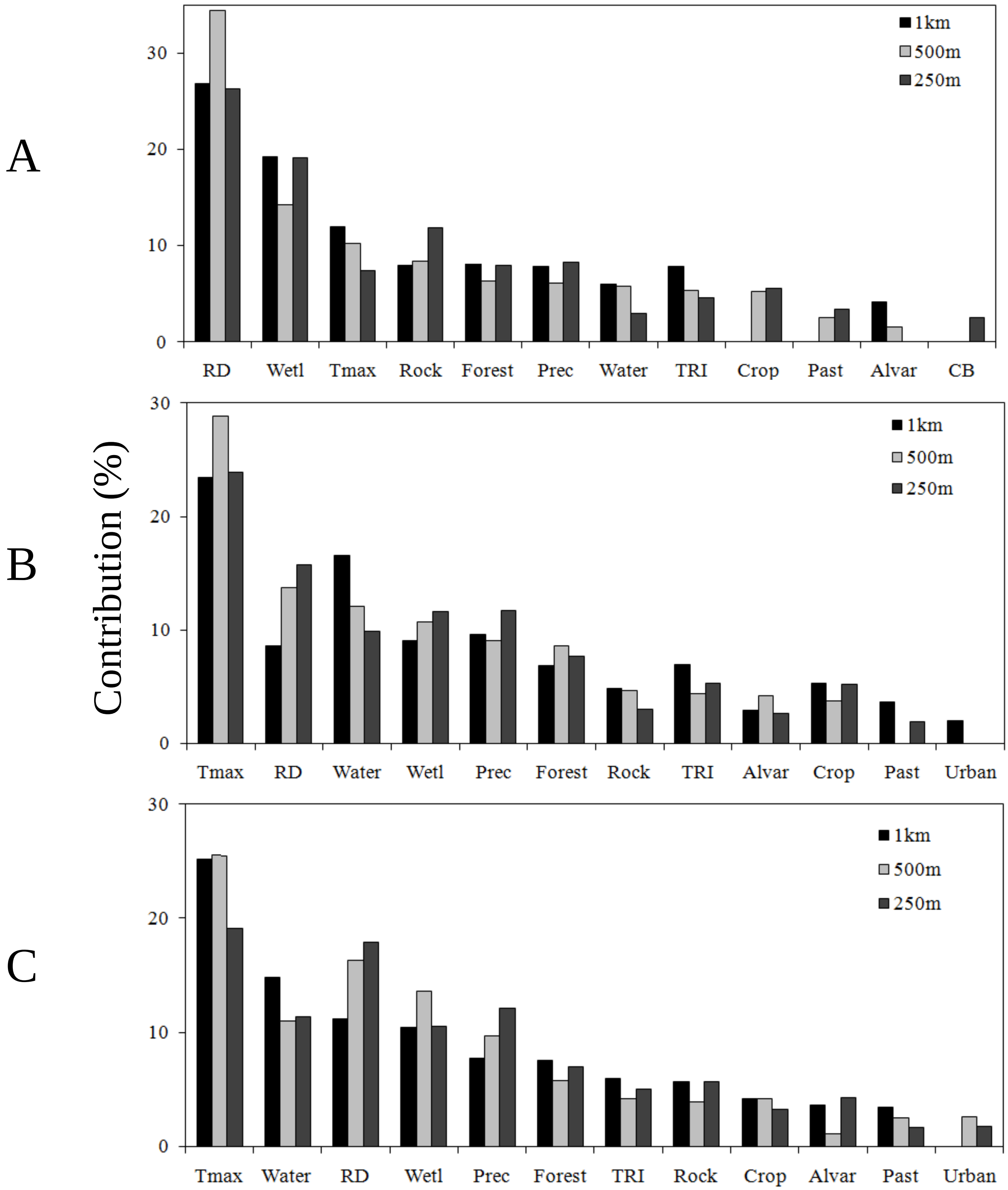

Figure 6-1. Average contribution of predictor variables in habitat suitability models built using BRTs and (**A**) only roadkill sightings (DOR), (**B**) only live turtle sightings (NODOR), or (**C**) all turtle sightings (ALL), at three spatial scales.

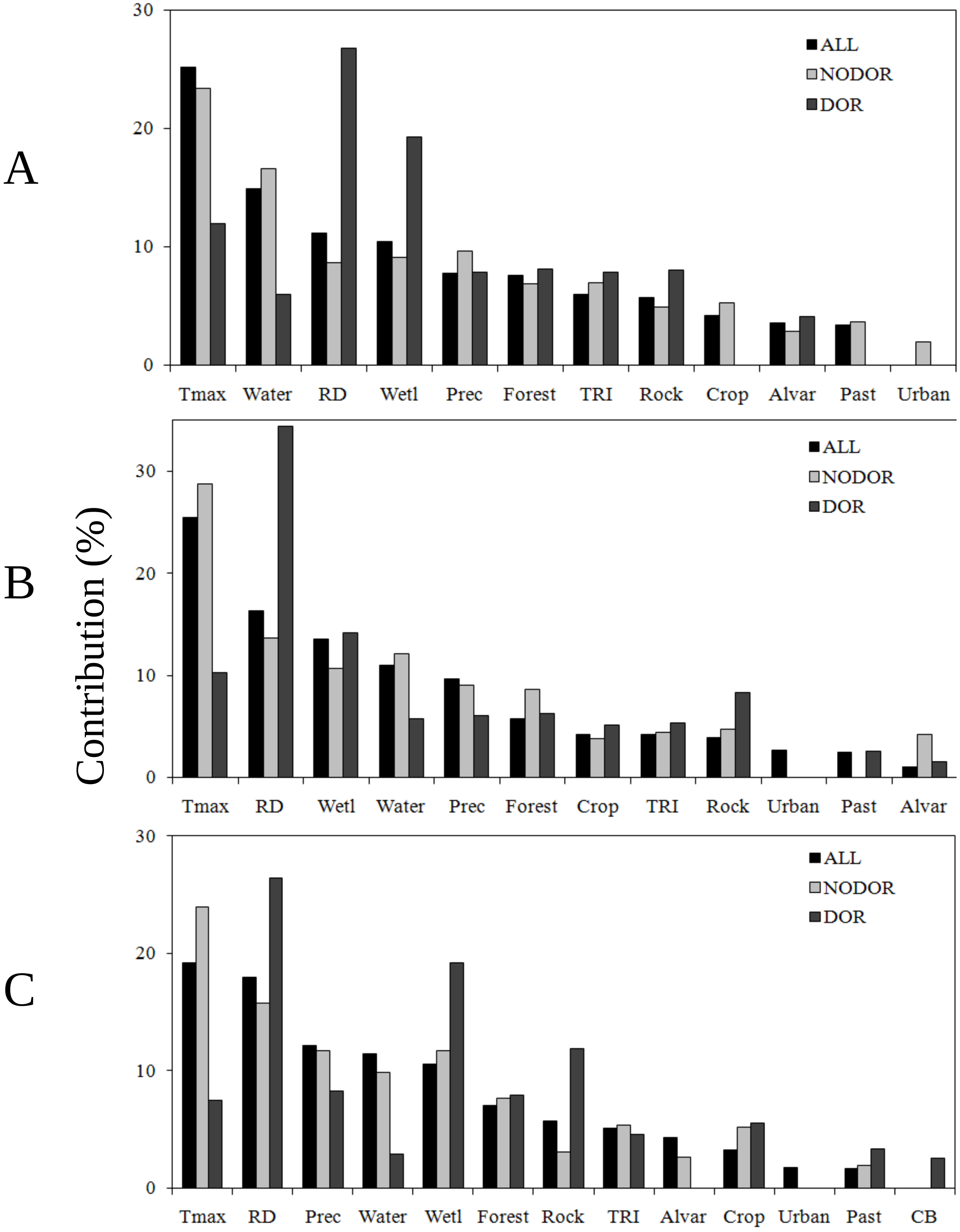

Figure 6-2. Average contribution of predictor variables in habitat suitability models built using BRTs and all Blanding's turtle sighting records (ALL), without any roadkill sightings (NODOR), or with only roadkill sightings (DOR) at all three spatial scales: (**A**) 1 km, (**B**) 500 m, and (**C**) 250 m.

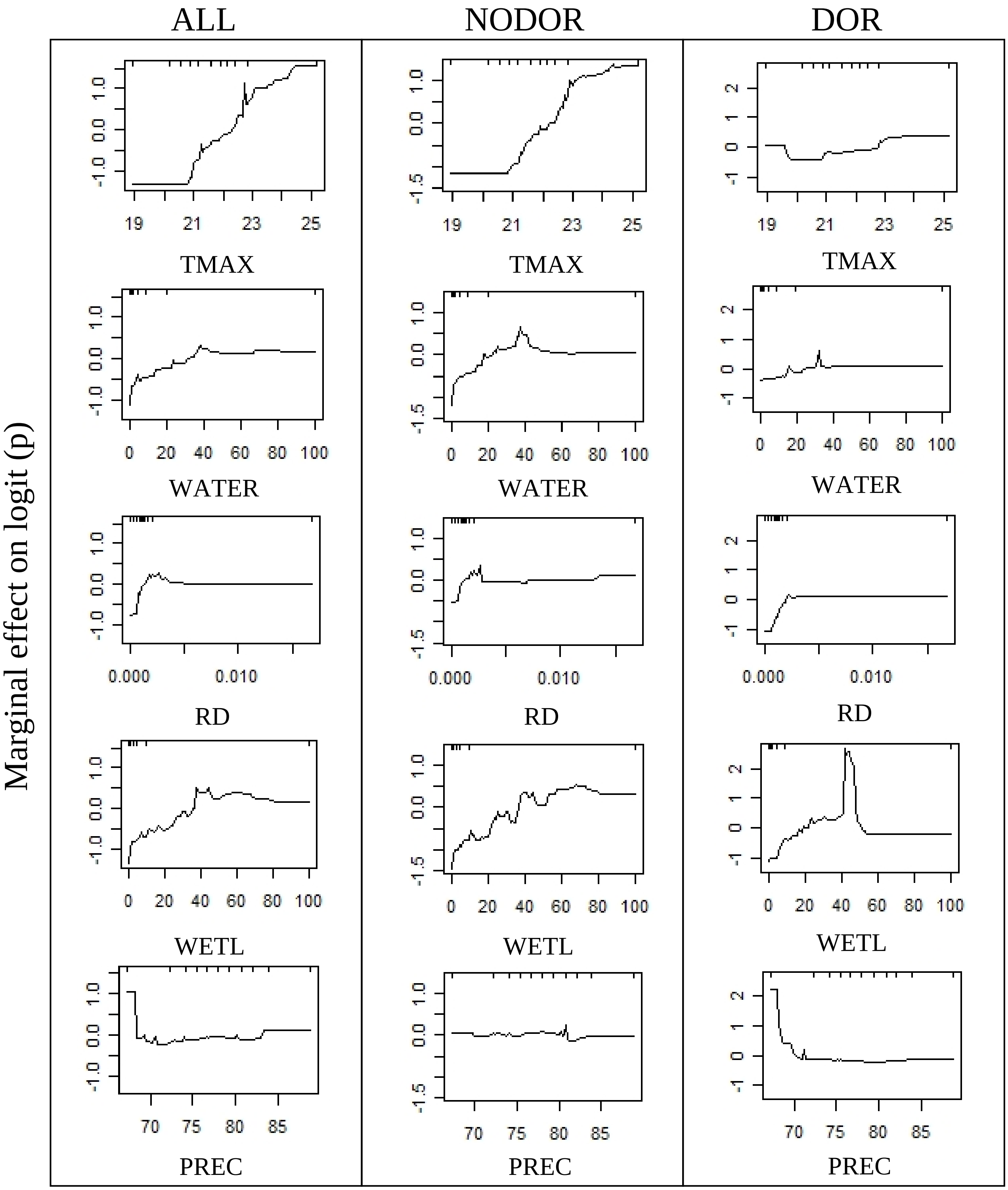

Figure 6-3. Partial dependence plots for the five most influential predictor variables in habitat suitability models built using BRTs, a 1 km buffer, and various subsets of sighting data: all sighting records (ALL), without roadkill sightings (NODOR), or using only roadkill sightings (DOR). Y axes are on the logit scale, are centered to have zero mean over the data distribution and represent the predicted probability of suitable conditions with all other variables set to their average values over the set of presence localities. Rug plots at inside top of plots show distribution of sites across each variable, in deciles.

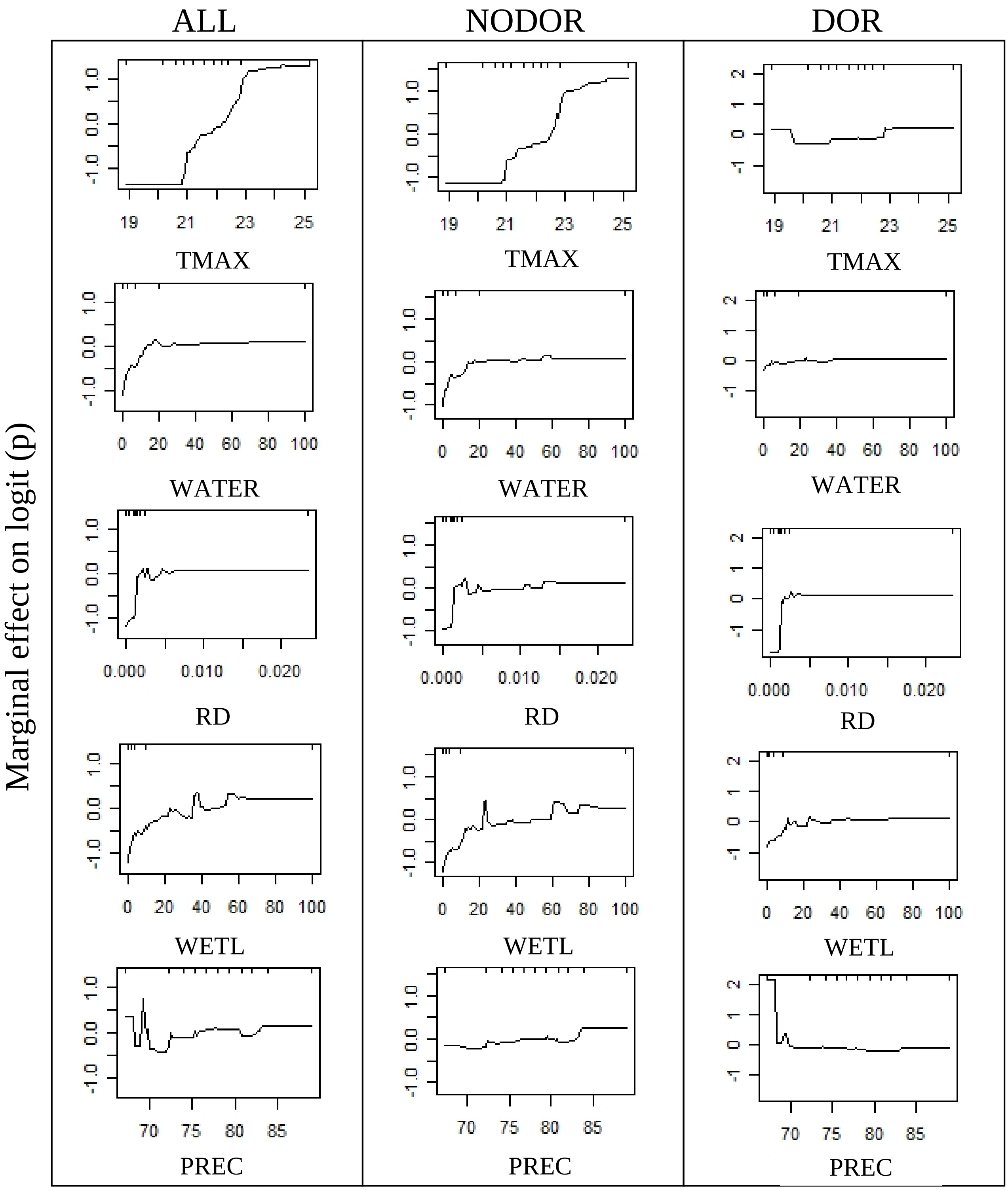

Figure 6-4. Partial dependence plots for the five most influential predictor variables in habitat suitability models built using BRTs, a 500 m buffer, and various subsets of sighting data: all sighting records (ALL), without roadkill sightings (NODOR), or using only roadkill sightings (DOR). Y axes are on the logit scale, are centered to have zero mean over the data distribution and represent the predicted probability of suitable conditions with all other variables set to their average values over the set of presence localities. Rug plots at inside top of plots show distribution of sites across each variable, in deciles.

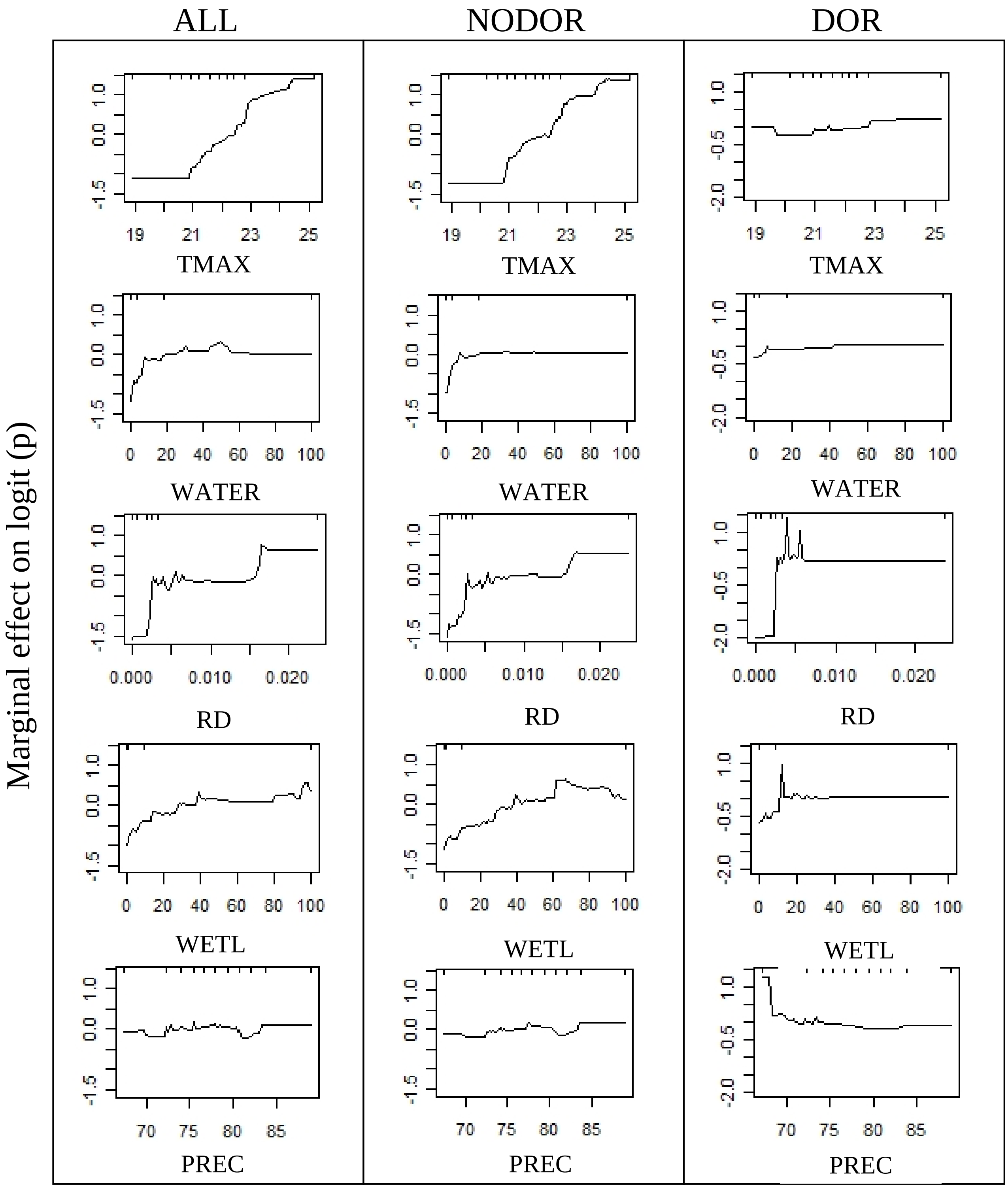

Figure 6-5. Partial dependence plots for the five most influential predictor variables in habitat suitability models built using BRTs, a 250 m buffer, and various subsets of sighting data: all sighting records (ALL), without roadkill sightings (NODOR), or using only roadkill sightings (DOR). Y axes are on the logit scale, are centered to have zero mean over the data distribution and represent the predicted probability of suitable conditions with all other variables set to their average values over the set of presence localities. Rug plots at inside top of plots show distribution of sites across each variable, in deciles.

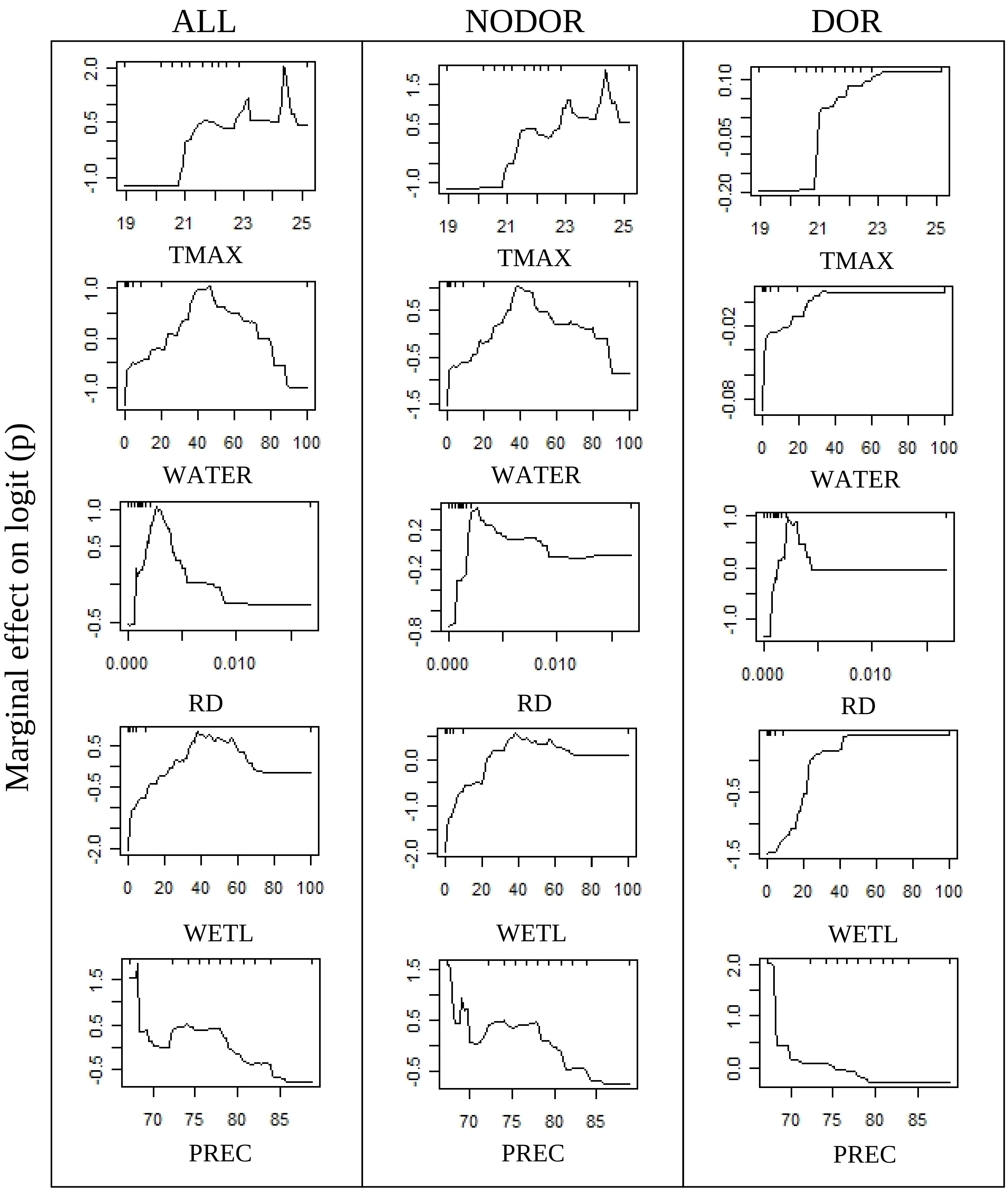

Figure 6-6. Response curves produced by univariate models of the five most important predictor variables in habitat suitability models built using BRTs and a 1 km buffer. Y axes are on the logit scale and are centered to have zero mean over the data distribution and represent the predicted probability of suitable conditions based on each predictor variable independently. The top most rug plots on the inside top of each plot shows the distribution of sites, in deciles, across each variable.

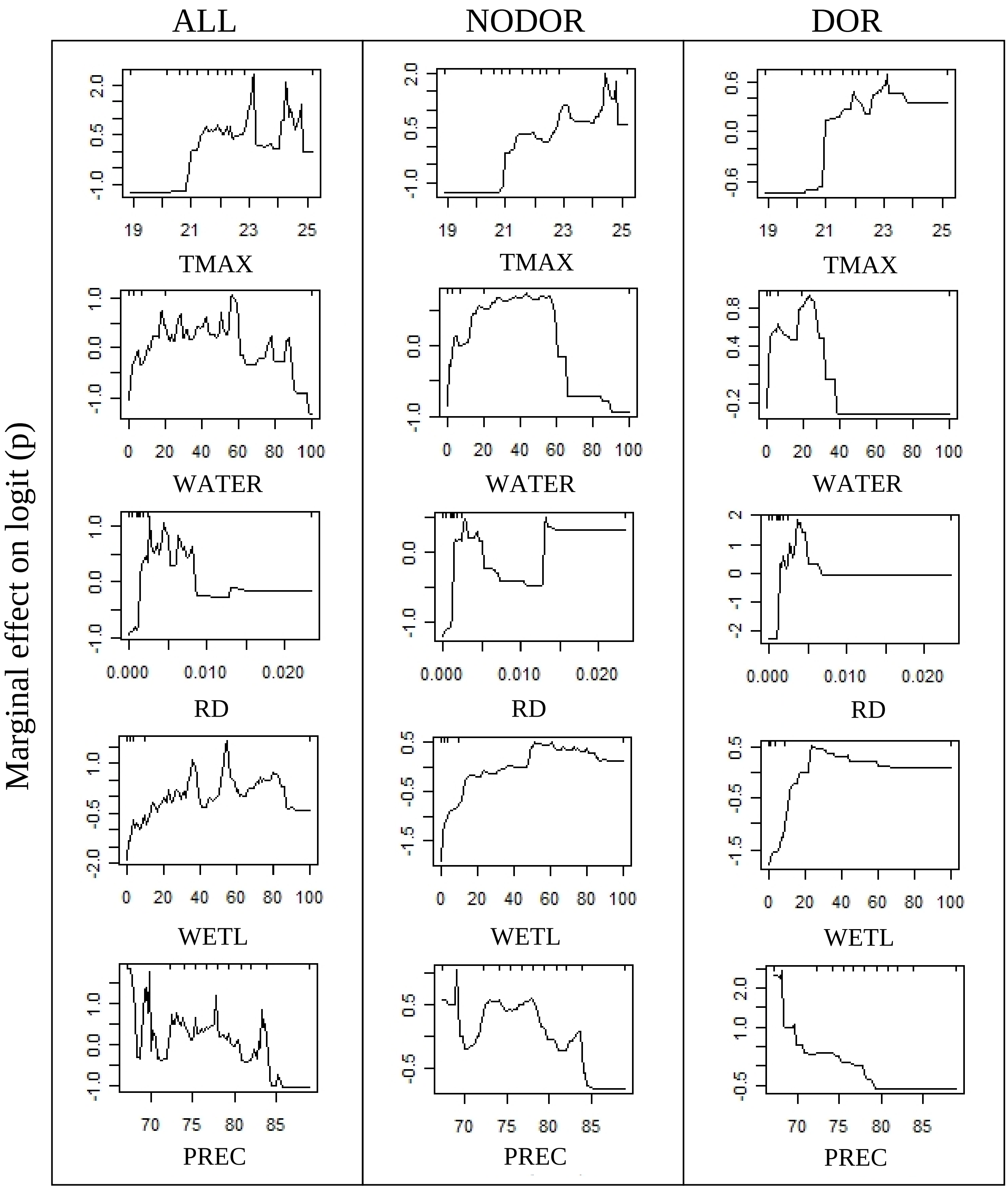

Figure 6-7. Response curves produced by univariate models of the five most important predictor variables in habitat suitability models built using BRTs and a 500 m buffer. Y axes are on the logit scale and are centered to have zero mean over the data distribution and represent the predicted probability of suitable conditions based on each predictor variable independently. The top most rug plots on the inside top of each plot shows the distribution of sites, in deciles, across each variable.

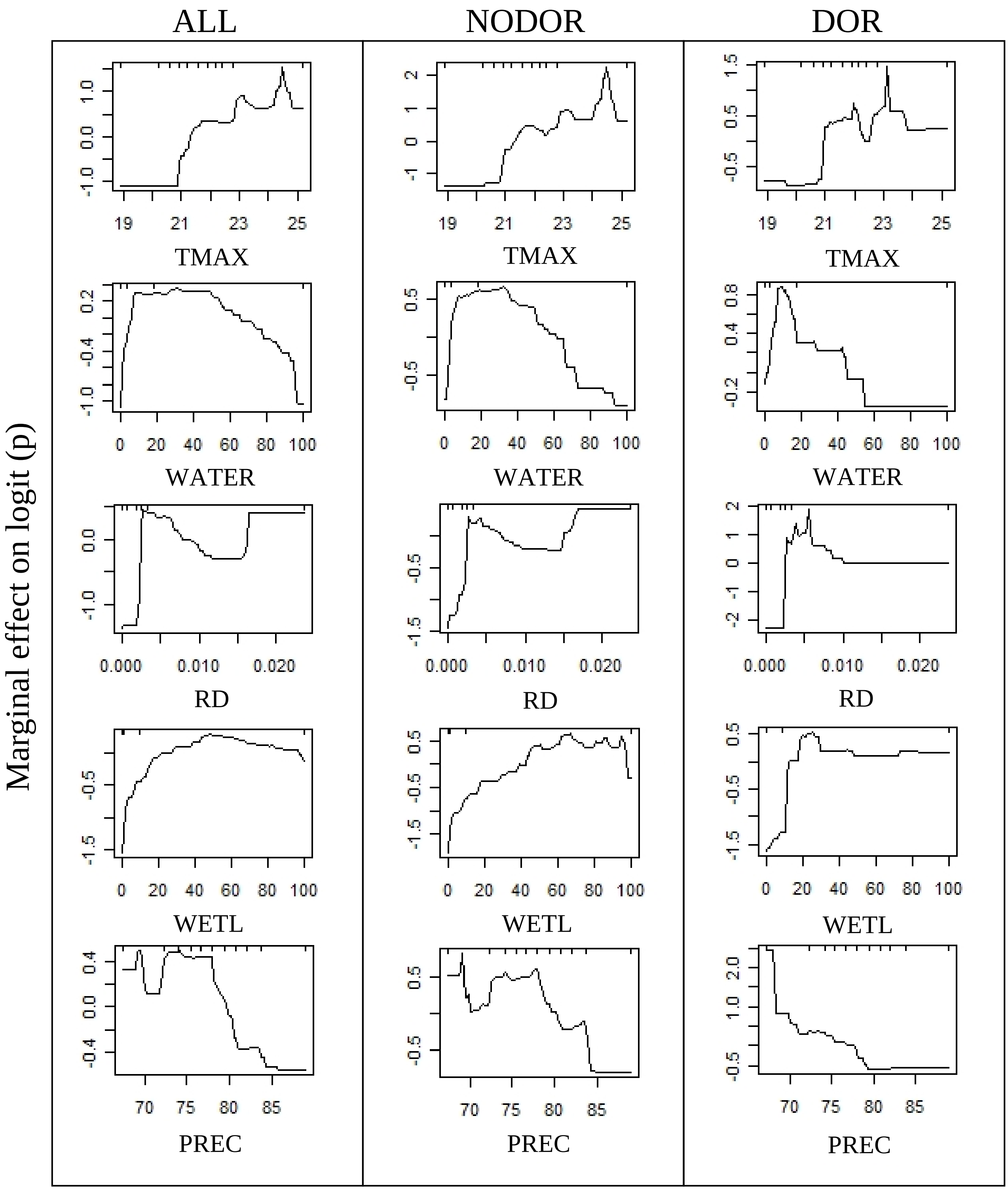

Figure 6-8. Response curves produced by univariate models of the five most important predictor variables in habitat suitability models built using BRTs and a 250 m buffer. Y axes are on the logit scale and are centered to have zero mean over the data distribution and represent the predicted probability of suitable conditions based on each predictor variable independently. The top most rug plots on the inside top of each plot shows the distribution of sites, in deciles, across each variable.

LITERATURE CITED

Abeel, T., T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys. 2010. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* **26**: 392-398.

Aebischer, N. J., P.A. Robertson, and R. E. Kenward. 1993. Compositional analysis of habitat use from animal radio-tracking data. *Ecology* **74**:1313-1325.

Aitken, GM. 1998. Extinction. *Biology and Philosophy* **13**: 393-411.

Akçakaya, H.R., V.C. Radeloff, D.J. Mladenoff, and H.S. He. 2004. Integration landscape and metapopulation modeling approaches: viability of the sharp-tailed grouse in a dynamic landscape. *Conservation Biology* **18**: 526-537.

Anderson, R.P., A.T. Peterson, and M. Gómez-Laverde. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos* **98**: 3-16.

Anderson, R.P., D. Lew, and A.T. Peterson. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling* **162**: 211-232.

Anderson, R.P. and E. Martínez-Meyer. 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (*Heteromys*) of Ecuador. *Biological Conservation* **116**: 167-179.

Araújo, M.B. and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* **33**:1677–1688.

Araújo, M. B., M. Cabeza, W. Thuiller, L. Hannah, and P.H Williams. 2004. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. *Global Change Biology* **10**: 1618-1626.

Araújo, M. B. and P.H. Williams. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* **96**: 331-345.

Aresco, M. J. 2005. The effect of sex-specific terrestrial movements and roads on the sex ratio of freshwater turtles. *Biological Conservation* **123**:37-44.

Ashton, K. G., and C. R. Feldman. 2003. Bergmann’s rule in nonavian reptiles: turtles follow it, lizards and snakes reverse it. *Evolution* **57**:1151–1163.

Attum, O., Y.M. Lee, J.H. Roe, and B.A. Kingsbury. 2008. Wetland complexes and upland–wetland linkages: landscape effects on the distribution of rare and common wetland reptiles. *Journal of Zoology* **275**:245-251.

Barry, J.F., and R. Shine. 1980. Sexual size dimorphism and sexual selection in turtles (order *Testudines*). *Oecologia (Berl.)* **44**:185-191.

Belleau, P. 2008. Habitat selection, movement patterns, and demography of common musk turtles (*Sternotherus odoratus*) in southwestern Québec. Unpublished PhD thesis. McGill University.

Bernstein, N.P., R.J. Richtsmeier, R.W., Black, and B.R. Montgomery. 2007. Home range and philopatry in the ornate box turtle, *Terrapene ornata ornata,* in Iowa. *American Midland Naturalist* **157**:162-174.

Beyer, H.L. 2004. Hawth's Analysis Tools for ARCGIS. *In*: www.spatialecology.com/htools.

Blancher, P.J. 1984. Natural Resource Description and Management Considerations, Cataraqui Marsh – Rideau Canal. Prepared for Parks Canada, Regional Office, Cornwall, Ontario. 70 pp., plus appendices.

Blouin-Demers, G., and P.J. Weatherhead. 2001. Thermal ecology of black rat snakes (*Elaphe obsoleta*) in a thermally challenging environment. *Ecology* **82**: 3025-3043.

Boersma, P.D., P. Kareiva, W.F. Fagan, J.A. Clark, and J.M. Hoekstra. 2001. How good are endangered species recovery plans? *BioScience* **51**: 643-650.

Boyce, M.S., J.S. Meyer, and L.L. Irwin. 1994. Habitat-based PVA for the northern spotted owl. *In*: Fletcher, D.J., Manly, B.F.J. (Eds.). Statistics in Ecology and Environmental Monitoring, Otago Conference Series No. 2. University Otago Press, Dunedin, New Zealand, pp. 63-85.

Boyce, M.S., P.R. Vernier, S.E. Nielsen, and F.K.A. Schmiegelow. 2002. Evaluating resource selection functions. *Ecological Modelling* **157**: 281-300.

Breiman, L., J.H. Friedman, R. A.Olshen, and C.G. Stone. 1984. Classification and regression trees. Wadsworth International Group, Belmont, California, USA.

Browne, C.L., and S.J. Hecnar. 2007. Species loss and shifting population structure of freshwater turtles despite habitat protection. *Biological Conservation* **138**: 421-429.

Breslow, N. E., and N. E. Day. 1980. Statistical Methods in Cancer Research: Vol. 1 – The Analysis of Case-Control Studies. Lyon, France, IARC Scientific Publications.

Burt, W.H. 1943. Territoriality and home range concepts as applied to mammals. *Journal of Mammalogy* **24**: 346-352.

Bury, R.B.1979. Population ecology of freshwater turtles. *In*: M. Harless and H. Morelock (eds). Turtles: perspectives and research. John Wiley and Sons. New York, N.Y. pp 571-602.

Cagle, F.R. 1939. A system for marking turtles for future identification. *Copeia* **1939**: 170-173.

Cagle, F. R. 1944. Home range, homing behavior and migration in turtles. *Miscellaneous Publications of the Museum of Zoology, University of Michigan* **61**: 1-34.

Carr, A. 1952. Handbook of turtles: The turtles of the United States, Canada, and Baja California. Cornell University Press, Ithaca, New York, USA.

Carrière, M.A. 2007. Movement patterns and habitat selection of common map turtles (*Graptemys geographica*) in St. Lawrence Islands National Park, Ontario, Canada. Unpublished M.S. Thesis, University of Ottawa, Ottawa. 120 pp.

Carrière, M.A., G. Bulté, and G. Blouin-Demers. 2009. Spatial ecology of northern map turtles (*Graptemys geographica*) in a lotic and a lentic habitat. *Journal of Herpetology* **43**: 597-604.

Carroll, T.E., and D.W. Ehrenfeld. 1978. Intermediate range homing in the wood turtle, *Clemmys insculpta*. *Copeia* **1978**: 117-126.

Carter, S.L., C.A. Haas, and J.C. Mitchell. 1999. Home range and habitat selection of bog turtles in southwestern Virginia. *Journal of Wildlife Management* **63**: 853–860.

Christian, K.A., and C.R. Tracy. 1981. The effect of the thermal environment on the ability of hatchling Galapagos land iguanas to avoid predation during dispersal. *Oecologia* **49**: 218–223.

Claussen, D.L., R. Lim, M. Kurz, and K. Wren. 2002. Effects of Slope, Substrate, and

Temperature on the Locomotion of the Ornate Box Turtle, *Terrapene ornate. Copeia* **2**: 411-418.

Collinge, S.K. 2001. Spatial ecology and biological conservation. *Biological Conservation* **100**: 1-2.

Compton, B.W., J.M. Rhymer, and M. McCollough. 2002. Habitat selection by wood turtles (*Clemmys insculpta*): an application of paired logistic regression. *Ecology* **83**: 833-843.

Congdon, J. D., A.E. Dunham, and R. C. van Loben Sels. 1993. Delayed sexual maturity and demographics of Blanding's turtles (*Emydoidea blandingii*): Implications for conservation and management of long-lived organisms. *Conservation Biology* **7**: 826-833.

Congdon, J.D., T.E. Graham, T.B. Herman, J.W. Lang, M.J. Pappas, and B.J. Brecke. 2008. *Emydiodea blandingii* (Holbrook 1838) - Blanding's turtle. In: Rhodin, A.G.J., P.C.H Pritchard, P.P. van Dijk, R.A. Saumure, K.A. Buhmann, and J.B. Iverson (Eds.) Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group. Chelonian Research Monographs No.5, pp. 015.1-015.12, doi: 10:3854/crm.5.015.blandingii.v1.2008, http:www.iucn-tftsg.org/cbftt/.

Congdon, J.D., R.D. Nagle, O.M. Kinney, M. Ostentoski, H.W. Avery, R.C. van Loben Sels, and D.W. Tinkle. 2000. Nesting ecology and embryo mortality: Implications for hatchling success and demography of Blanding's Turtles (*Emydoidea blandingii*). *Chelonian Conservation Biology* **3**: 569–579.

Congdon, J. D. and P. C. Rosen. 1983. Egg components and utilization during development in aquatic turtles. *Copeia* **1983**: 264–268.

Congdon, J.D., D.W. Tinkle, G.L. Breitenbach, and R.C. van Loben Sels. 1983. Nesting ecology and hatching success in the turtle *Emydoidea blandingii*. *Herpetologica* **39**: 417-429.

Corsi, F., E. Duprè, and L. Boitani. 1999. A large scale model of wolf distribution in Italy for conservation planning. *Conservation Biology* **13**: 150-159.

COSEWIC. 2005. COSEWIC assessment and update status report on the Blanding's Turtle Emydoidea blandingii in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. viii + 40 pp. (www.sararegistry.gc.ca/status/status_e.cfm).

Dawson, W.R. 1975. On the physiological significance of the preferred body temperatures of Reptiles. In: Perspectives of Biophysical Ecology (D.M. Gates and R.B. Schmerl, eds.) pp.443-473. Springer-Verlag. New York.

Dennis, R., and C. Thomas. 2000. Bias in butterfly distribution maps: the influence of hot spots and recorder’s home range. *Journal of Insect Conservation* **4**: 73-77.

DeSolla, S. R., R. Bonduriansky, and R. J. Brooks. 1999. Eliminating autocorrelation reduces biological relevance of home range estimates. *Journal of Animal Ecology* **68**: 221-234.

De’ath, G. 2002. Multivariate regression trees: a new technique for constrained classification analysis. *Ecology* **83**: 1103–1117.

De’ath, G. 2007. Boosted Trees for Ecological Modeling and Prediction. *Ecology* **88**: 243-251.

De’ath, G. and K.E. Fabricius. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* **81**: 3178-3192.

Doetsch, P., C. Buck, P. Golik, N. Hoppe, M. Kramp, J. Laudenberg, C. Oberd, P. Steingrube, J. Forster, and A. Mauser. 2009. Logistic Model Trees with AUC Split Criterion for the KDD Cup 2009 Small Challenge. *JMLR: Workshop and Conference Proceedings* **7**: 77-88.

Dreiseitl, S., and M. Osl. Feature Selection Based on Pairwise Classification Performance. In: Moreno-Díaz, R., F. Pishler, and A. Quesada-Arencibia. Eds. *Computer Aided Systems Theory - EUROCAST 2009*, *LNCS* 5717, Springer-Verlag Berlin Heidelberg, Germany, pp. 769-776.

Dudík, M., S.J. Phillips, R.E. Schapire. 2004. Performance guarantees for regularized maximum entropy density estimation. In: Proceedings of the 17th Annual Conference on Computational Learning Theory, ACM Press, New York, pp. 655-662.

Edge, C.B., B.D. Steinberg, R.J. Brooks, J.D. Litzgus. 2010. Habitat selection by Blanding's turtles (*Emydoidea blandingii*) in a relatively pristine landscape. *Ecoscience* **17**: 90-99.

Edmonds, J.H. 1998. Population ecology of the stinkpot turtle (*Sternotherus odoratus*) in Georgian Bay, Ontario. Unpublished M.Sc. Thesis. University of Guelph.

Ehrlich, P.R. 1995. The scale of the human enterprise and biodiversity loss, p.214-226. *In* J.H. Lawton and R.M. May (ed.), *Extinction Rates*. Oxford University Press, Oxford, United Kingdom.

Elith, J. 2002. Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Ferson, S., Burgman, M. (Eds.), Quantitative Methods for Conservation Biology. Springer-Verlag, New York, 39–58.

Elith, J., C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R.J. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, J. Li, L.G. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J.McC. Overton, A.T. Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R.E. Schapire, J. Soberón, S. Williams, M.S. Wisz, and N.E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129–151.

Elith, J. and J.R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* **40**: 677-697.

Elith, J., J.R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* **77**: 802–13.

Emrich, M.E. 1991. Blanding's turtle (*Emydoidea blandingii*) nesting behaviour and response to an artificial nest habitat. Unpublished M.Sc. thesis. Bard College, Annandale, NY.

Ernst, C. H. 1986. Ecology of the turtle, *Sternotherus odoratus*, in southeastern Pennsylvania. *Journal of Herpetology* **20**: 341-352.

Ernst, C.H., R.W. Barbour, and J.E. Lovich. 1994. Turtles of the United States and Canada. Smithsonian Institution Press. Washington, DC. 578. pp.

ESRI 1982-1996. ArcInfo 8.1. Environmental Systems Research Institute, Redlands, CA.

ESRI. 1995-2010. ArcGIS Desktop Evaluation Edition 9.3. Environmental Systems Research Institute, Redlands, California.

Evans, D. L., W. J. Streever, and T. L. Crisman. 1999. Natural flatwoods marshes and created freshwater marshes of Florida: factors influencing aquatic invertebrate distribution and comparisons between natural and created marsh communities. Pages 81–104 in D. P. Batzer, R. B. Rader, and S. A. Wissinger, editors. Invertebrates in freshwater wetlands of North America. John Wiley & Sons, New York, New York, USA.

Evink, G. L. 1980. Studies of causeways in the Indian River, Florida. Report FL-ER-7-80, Florida Department of Transportation, Tallahassee, USA.

Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? *Systematic Biology* **51**: 331-363.

Fielding, A. H. and J.F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**: 38-49.

Findlay, C.S., and J. Bourdages. 2000. Response time of wetland biodiversity to road construction on adjacent lands. *Conservation Biology* **14**: 86-91.

Fisher, J., D.B. Lindenmayer, and A. Cowling. 2004. The challenge of managing multiple species at multiple scales: reptiles in an Australian grazing landscape. *Journal of Applied Ecology* **41**: 32-44.

Fleishman, E., R. MacNally, J.P. Fay, and D.D. Murphy. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation Biology* **15**: 1674-1685.

Forman, R.T.T., and L.E. Alexander.1998. Roads and their major ecological effects. *Annual Review of Ecology and Systematics* **29**: 207–231.

Funk, V. and K. Richardson. 2002. Systematic data in biodiversity studies: use it or lose it. *Systematic Biology* **51**: 303-316.

Frazer, N.B., J.W. Gibbons, and J.L. Green. 1991. Growth, survivorship and longevity of painted turtles *Chrysemys picta* in a southwestern Michigan marsh. *American Midland Naturalist* **125**: 245-258.

Friedman, J.H. 2002. Stochastic gradient boosting. *Computational Statistics and Data Analysis* **38**: 367–378.

Friedman, J.H., and J.J. Meulman. 2003. Multiple additive regression trees with application in epidemiology. *Statistics in Medicine* **22**: 1365–1381.

Gaston, G. R. 1999. Bayous of the northern Gulf of Mexico: distribution and trophic ecology of invertebrates. *In*: D. P. Batzer, R. B. Rader, and S. A. Wissinger (eds). Invertebrates in freshwater wetlands of North America. John Wiley & Sons, New York, New York, USA. pp 919–947

Germano, D. J., R.B. Bury, and M. Jennings. 2000. Growth and population structure of *Emydoidea blandingii* from western Nebraska. *Chelonian Conservation and Biology* **3**: 618–625.

Gibbons, J. W. 1970. Terrestrial activity and the population dynamics of aquatic turtles. *The American Midland Naturalist* **83**: 404-414.

Gibbons, J.W. 2003. Terrestrial habitat: a vital component for herpetofauna of isolated wetlands. *Wetlands* **23**: 630-635.

Gibbons, W.J., J.L. Greene, and J.D. Congdon. 1990. Temporal and spatial movement patterns of sliders and other turtles. Pages 201-215 in W. J. Gibbons, editor. Life History and Ecology of the Slider Turtle. Smithsonian Institution Press, Washington D.C

Gibbons, J.W., D.E Scott, T.J. Ryan, K.A. Buhlmann, T.D. Tuberville, B.S. Metts, T.M. Greene, Y. Leiden, S. Poppy, and C.T. Winne. 2000. The global decline of reptiles, déjà vu amphibians. *BioScience* **50**: 653-666.

Gibbs, J.P., and W.G. Shriver. 2002. Estimating the effects of road mortality on turtle populations. *Conservation Biology* **16**: 1647-1652.

Gibbs, J.P. and D.A. Steen. 2005. Trends in sex ratios of turtles in the United States: implications of road mortality. *Conservation Biology* **19**: 552-556.

Gibson, L.A., B.A. Wilson, D.M. Cahill, and J. Hill. 2004. Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology* **41:** 213-223.

Goolsby, J. A. 2004. Potential distribution of the invasive old world climbing fern, *Lygodium microphyllum*, in North and South America. *Natural Areas Journal* **24**: 351-353.

Graham, T.E., and T.S. Doyle. 1979. Dimorphism, courtship, eggs, and hatchlings of the Blanding's turtle, *Emydoidea blandingii* (Reptilia, Testudines, Emydidae) in Massachusetts. *Journal of Herpetology* **13**: 125-127.

Graham, C.H., S. Ferrier, F. Huettman, C. Moritz, and A.T. Peterson. 2004. New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution* **19**: 497–503.

Grgurovic, M. and P.R. Sievert. 2005. Movement patterns of Blanding’s turtles

(*Emydoidea Blangingii*) in the suburban landscape of eastern Massachusetts. *Urban Ecosystems* **8**: 203-213.

Guisan, A. and U. Hofer. 2003. Predicting reptile distributions at the mesoscale: relation to climate and topography. *Journal of Biogeography* **30**: 1233-1243.

Guisan, A., A. Lehmann, S. Ferrier, M. Austin, J.McC. Overton, and R. Aspinall. 2006. Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* **43**: 386-392.

Guisan, A. and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**: 993–1009.

Guyon, I., and A. Elisseeff. 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research* **3:** 1157-1182.

Hamernick, M.G. 2000. Home ranges and habitat selection of Blanding's turtles (*Emydoidea blandingii*) at the Weaver Dunes, Minnesota. Final report submitted to the Nongame Wildlife Program, Minnesota Department of Natural Resources. 18 pp.

Hanley, J.A., and B.J. McNeil. 1983. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* **148**: 839-843.

Hansteen, T.L., H.P. Andreassen, and R.A. Ims. 1997. Effects of spatiotemporal scale on autocorrelation and home range estimators. *Journal of Wildlife Management* **61**: 280-290.

Harding, J.H. 1997. Amphibians and Reptiles of the Great Lakes Region. U. of Mich. Press, Ann Arbor, MI. 378pp.

Hartwig, T.S., and E. Kiviat. 2007. Microhabitat Association of Blanding's turtles in natural and constructed wetlands in Southeastern New York. *Journal of Wildlife Management* **71**: 576-582.

Hastie, T., R. Tibshirani, and J.H. Friedman. 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.

Haxton, T. 2000. Road mortality of snapping turtles, *Chelydra serpentina,* in central Ontario during their nesting period. *Canadian Field-Naturalist* **114**: 106-110.

Herman, T.B., J.S. Boates, C. Drysdale, S. Eaton, J. McNeil, S. Mockford, E. Alcorn, S. Bleakney, M. Elderkin, J. Gilhen, C. Jones, J. Kierstead, J. Mills, I. Morrison, S. O'Grady, D. Smith. 2003. National Recovery Plan for the Blanding's Turtle (*Emydoidea blandingii*) Nova Scotia Population. January 2003. 63 pp.

Herman, T.B., T.D. Power, and B.R. Eaton. 1995. Status of Blanding's Turtles, *Emydoidea blandingii*, in Nova Scotia, Canada. *Canadian Field-Naturalist* **109**: 182–191.

Hernandez, P.A., C.H. Graham, L.L. Master, and D.L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* **29**: 773-785.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965-1978.

Hoekstra, J.M. 2005. Confronting a biome crisis: global disparities of habitat loss and protection. *Ecology Letters* **8**: 23-29

Holland, J.D., D.G. Bert, and L. Fahrig. 2004.Determining the spatial scale of species' response to habitat. *Bioscience* **54**: 227-233.

Hooge, P.N., and B. Eichenlaub. 2000. Animal movement extension to ArcView 2.0. In: Alaska Science Center - Biological Science Office, U.S. Geological Survey, Anchorage, AK, USA.

Hosmer, D.W., and S. Lemeshow. 2000. Applied Logistic Regression: Second edition. John Wiley and Sons, NY, USA.

Huey, R.B. 1982. Temperature, physiology, and the ecology of reptiles. In: Biology of the Reptilia, vol. 12 (C.Gans and F.H. Pough, eds.) pp. 25-91. Academic Press. New York.

Huey, R.B. and J.G. Kingsolver. 1989. Evolution of thermal sensitivity of ectotherm performance. *Trends in Ecology and Evolution* **4**: 131–135.

Huey, R.B., C.R. Peterson, S.J. Arnold, and W.P. Porter. 1989. Hot rocks and not-so-hot rocks: retreat-site selection by garter snakes and its thermal consequences. *Ecology* **70**: 931–944.

Innes, R.J., K.J. Babbit, and J.J. Kanter. 2008. Home Range and Movement of Blanding's Turtles (*Emydoidea blandingii*) in New Hampshire. *Northeastern Naturalist* **15:** 431-444.

Jaynes, E. T. 1980. The Minimum Entropy Production Principle. *Annual Review of Physical Chemistry* **31**: 579-601.

Jenkins, R. J., and K. J. Babbitt. 2003. Life-history notes. *Emydoidea blandingii* mating behavior. *Herpetological Review* **34**: 361-362.

Johnston, C.A. 1994. Cumulative impacts to wetlands. *Wetlands* **14**: 49-55.

Johnson, D.H. 1980. The comparison of usage and availability measurements for evaluating resource preference. *Ecology* **61**: 65-71.

Johnson, C.J., D.R. Seip, and M.S. Boyce. 2004. A quantitative approach to conservation planning: Using resource selection functions to identify important habitats for mountain caribou. *Journal of Applied Ecology* **41**: 238-251.

Jones, R.L. 1996. Home range and seasonal movements of the turtle *Graptemys flavimaculata*. *Journal of Herpetology* **30**: 376-385.

Joyal, L.A., M. McCollough, and M.L. Hunter Jr. 2000. Population structure and reproductive ecology of Blanding's Turtle (*Emydoidea blandingii*) in Maine, near the Northeastern edge of its range. *Chelonian Conservation Biology* **3**: 580–588.

Joyal, L.A, M. McCollough, and M.L. Hunter Jr. 2001. Landscape ecology approaches to wetland species conservation: a case study of two turtle species in southern Maine. *Conservation Biology* **15**: 1755-1762.

Kaliontzopoulou, A., J.C. Brito, M.A. Carretero, S. Larbes, and D.J. Harris. 2008. Modelling the partially unknown distribution of wall lizards (*Podarcis*) in North Africa: ecological affinities, potential areas of occurrence, and methodological constraints. *Canadian Journal of Zoology* **86**: 992-1001.

Kofron, C.P., and A.A. Schreiber. 1985. Ecology of two endangered aquatic turtles in Missouri: *Kinosternon flavescens* and *Emydoidea blandingi*. *Journal of Herpetology* **19**: 27-40.

Kramer-Schadt, S., E. Revila, T. Wiegand, and W. Breitenmoser. 2004. Fragmented landscapes, road mortality and patch connectivity: modeling influences on the dispersal of Eurasian lynx. *Journal of Applied Ecology* **41**: 711–723.

Krishnapuram, B., L. Carin, and A. Hartemink. 2004. Gene expression analysis: joint feature selection and classifier design. In: Schölkopf, B., K. Studa, and J-P. Vert. Eds. *Kernel Methods in Computational Biology*. MIT Press, Cambridge, MA, pp. 299–317.

Lagler, K.F. 1943. Food habits and economic relations of the turtles of Michigan with special reference to fish management. *American Midland Naturalist* **29**: 257–312.

Laver, P. 2005. ABODE Ecological Tools for ARCGIS. *In:* http://filebox.vt.edu/users/plaver/abode/home.html.

Linck, M.H. and J.J. Moriarty. 1997. The effects of habitat fragmentation on Blanding's Turtles in Minnesota. pp 30–37. In: Moriarty, J. J. and D. Jones, editors. (Eds.). Minnesota's Amphibians and Reptiles: Their Conservation and Status. Serpent's Tales Books. Lanesboro, MN.

Linck , M.H., J.A. DePari, B.O. Butler, and T.E. Graha. Nesting behaviour of the turtle, *Emydoidea blandingii*, in Massachusetts. *Journal of herpetology* **23**: 442-444.

Litzgus, J.D. and T.A. Mousseau. 2004. Home range and seasonal activity of southern spotted turtles (*Clemmys guttata*): implications for management. *Copeia* **2004**: 804-817.

Litzgus, J. D., S.E. Durant, and T. A. Mousseau. 2004. Clinal variation in body size and cell size in a widely distributed vertebrate ectotherm. *Oecologia* **140**: 551–558.

Litvaitis, J.A., and J.P. Tash. 2008. An approach toward understanding wildlife-vehicle collisions. *Environmental Assessment* **42**: 688-697.

Luck, G.W. 2002. The habitat requirements of the rufous treecreeper (*Climacteris rufa*). Preferential habitat use demonstrated at multiple spatial scales. *Biological Conservation* **105**: 383-394.

Lütolf, M., F. Kienast, and A. Guissan. 2006. The ghost of past species occurrence: improving species distribution models for presence-only data. *Journal of Applied Ecology* **43**: 802-815.

Marchand, M.N., and J.A. Litvaitis. 2004. Effects of Habitat Features and Landscape Composition on the Population Structure of a Common Aquatic Turtle in a Region Undergoing Rapid Development. *Conservation biology* **18**: 758-767.

Mayor, S.J., D.C. Schneider, J.A. Schaefer, S.P. Mahoney. 2009. Habitat selection at multiple scales. **16**: 238-247.

McMaster, N.L., and T.B. Herman. 2000. Occurrence, habitat selection, and movement patterns of juvenile Blanding's Turtles (*Emydoidea blandingii*) in Kejimkujik National Park, Nova Scotia. *Chelonian Conservation and Biology* **3**: 602-610.

Metcalf, E., and A.L. Metcalf. 1970. Observations on ornate box turtles (*Terrapene ornata ornata* Agassiz). *Transactions of the Kansas Academy of Science* **73**: 96-117.

Miller, A.J. 1990. Subset Selection in Regression. Chapman & Hall, London.

Milner-Gulland, E.J., O.M. Bukreeva, T. Coulson, A.A. Lushchekina, M.V. Kholodova, A.B. Bekenov, and I.A. Grachev. 2003. Conservation: Reproductive collapse in saiga antelope harems. *Nature* **422**: 135.

Mitsch, W.J., and J.G. Gosselink. 2000. *Wetlands*. Third edition. J. C. Wiley and Sons, New York, USA.

Morin, P., D. Berteaux, and I. Klvana. 2005. Hierarchical habitat selection by North American porcupines in south boreal forest. *Canadian Journal of Zoology* **83**: 1333-1342.

Morreale, S.J., J.W. Gibbons, and J.D. Congdon. 1984. Significance of activity and movement in the Yellow-bellied Slider Turtle (*Pseudemys scripta*). *Canadian Journal of Zoology* **62**: 1038–1042.

Muegel, L.A., and D. L. Claussen. 1994. Effects of slope on voluntary locomotor performance in the turtle, *Terrapene carolina carolina*. *Journal of Herpetology* **28**: 6-11.

Murphy, A. H., and R. L. Winkler. 1992. Diagnostic verification of probability forecasts. *International Journal of Forecasting* **7:** 435-455.

NatureServe. 2009. NatureServe Explorer: An online encyclopedia of life [web application]. Version 7.1. NatureServe, Arlington, Virginia. http://www.natureserve.org/explorer. (Accessed: May 20, 2010).

North, M.P., and J.H. Reynolds. 1996. Microhabitat analysis using radiotelemetry locations and polytomous logistic regression. *Journal of Wildlife Management* **60**: 639-653.

Oldham, M.J., and W.F. Weller. 2010. Ontario Herpetofaunal Atlas. Natural Heritage Information Centre, Ontario Ministry of Natural Resources. http://nhic.mnr.gov.on.ca/MNR/nhic/herps/ohs.html.

OMNR. 1998. Ontario 28-Class Land Cover Database. 1998. Ontario Ministry of Natural Resources. Peterborough, Ontario, Canada.

OMNR. 2005. Provincial Digital Elevation Model, Version 2.0.0. Ontario Ministry of Natural Resources. Peterborough, Ontario, Canada.

OMNR. 2006. Ontario Road Network. Ontario Ministry of Natural Resources. Peterborough, Ontario, Canada.

Orians, G. H., and J. F. Wittenberger. 1991. Spatial and temporal scales in habitat selection. *The American Naturalist* **137**: 29-49.

Pappas, M.J. and B.J. Brecke. 1992. Habitat selection of juvenile Blanding's turtles, *Emydoidea blandingii*. *Journal of Herpetology* **26**: 233-234.

Pappas, M.J., B.J. Brecke, and J. D. Congdon. 2000. The Blanding's Turtle (*Emydoidea blandingii*) of Weaver Dunes, Minnesota. *Chelonian Conservation and Biology* **3**: 557-568.

Parker, W.S. 1984. Immigration and dispersal of the slider turtles *Pseudemys scripta* in Mississippi farm ponds. *The American Midland Naturalist* **112**: 280-293.

Pearce, J.L., and S. Ferrier. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* **128**: 127-147.

Pearson, R.G., and T.P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are climate envelope models useful? *Global Ecology and Biogeography* **12**: 361-371.

Pearson, R.G., T.P. Dawson, and C. Lin. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography* **27:** 285-298.

Peterson, A.T. 2001. Predicting species' geographic distributions based on ecological niche modeling. *The Condor* **103**: 599-605.

Peterson, A.T. 2003.Predictability of the geography of species' invasions via ecological niche modeling. *Quarterly Review of Biology* **78**: 419-433.

Peterson, A. T., E. Martínez -Meyer, and C. Gonzalez-Salazar. 2004. Reconstructing the pleistocene geography of the Aphelocoma jays (Corvidae). *Diversity and Distribution* **10**: 237-246.

Peterson, A.T., and J.J. Shaw. 2003. *Lutzomyia* vectors for cutaneous leishmaniasis in southern Brazil: ecological niche models, predicted geographic distributions, and climate change effects. *International Journal of Parasitology* **33**: 919-931.

Peterson, A.T., J. Soberon, and V. Sanchez-Cordero. 1999. Conservatism of ecological niches in evolutionary time. *Science* **285**: 1265-1267.

Peterson, C.R., A.R. Gibson, and M.E. Dorcas. 1993. Snake thermal ecology: the causes and consequences of body temperature variation. Pages 241–314 in R. A. Seigel, and J. T. Collins, editors. Snakes: ecology and behavior. Mc-Graw-Hill, New York, New York, USA.

Petokas, P. J. 1986. Patterns of reproduction and growth in the freshwater turtle, *Emydoidea blandingii*. Unpublished Ph.D. thesis. University of New York. Binghamton, NY. 185. pp.

Phillips, S.J. 2008. Transferability, sample selection bias and background data in presence-only modeling: a response to Peterson et al. *Ecography* **31**: 272-278.

Phillips, S.J., R.P. Anderson, and R.E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231-259.

Phillips, S.J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**: 161-75.

Phillips, S.J., M. Dudík, J. Elith, C.H. Graham, A. Lehmann, J. Leathwick, and S.

Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**: 181-197.

Phillips, S. J., M. Dudík, R.E. Schapire. 2004. A maximum entropy approach to species distribution modeling. *In*: Proceedings of the Twenty-First International Conference on Machine Learning. ACM Press, New York. pp. 472-486

Piepgras, S.A. 1998. Summer and seasonal movements and habitats, home ranges, and buffer zones of a central Minnesota population of Blanding's Turtles. Unpublished M.Sc. Thesis. University of North Dakota. Grand Forks, ND.

Piepgras, S.A., and J.W. Lang. 2000. Spatial ecology of Blanding's Turtle in central Minnesota.*Chelonian Conservation and Biology* **3**: 589-601.

Powell, R.A. 2000. Animal home ranges and territories and home range estimators. *In*: L. Boitani and T. Fuller (Eds). Research in Animal Ecology. Controversies and Consequences. Columbia University Press, New York. pp 65-110

Pluto, T.G., and E.D. Bellis. 1988. Seasonal and annual movements of riverine map turtles, *Graptemys geographica*. *Journal of Herpetology* **22**: 152-158.

Raup, D.M. and J.J. Sepkoski Jr. 1984. Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences USA* **81**: 801-805.

R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reddy, S., and L.M. Dávalos. 2003. Geographical sampling bias and its implications for conservation priorities in Africa. *Journal of Biogeography* **30**: 1719-1727.

Regional Municipality of Halton. 2009. Halton Region Greenland Securement Strategy. pp 48. http://www.halton.ca/common/pages/UserFile.aspx?fileId=33577

Reunanen, J. 2003. Overfitting in Making Comparisons Between Variable Selection Methods. *Journal of Machine Learning Research* **3**: 1371-1382.

Rickets, T.H. 2001. The matrix matters: effective isolation in fragmented landscapes. *American Naturalist* **158**: 87-99.

Ridgeway, G. 2006. Generalized boosted regression models. Documentation on the R Package ‘gbm’, version 1·5–7. http://www.i-pensieri.com/gregr/gbm.shtml, accessed March 2008.

Riley, S.J., S.D. DeGloria, and R. Elliot. 1999. A terrain ruggedness index that quantifies topographic heterogeneity. *Intermountain Journal of Sciences* **5**: 23-27.

Roe, J.H., J. Gibson, and B.A. Kingsbury. 2006. Beyond the wetland border: estimating the impact of roads for two species of water snakes. *Biological Conservervation* **130**: 161-168.

Ross, D.A. and R.K. Anderson. 1990. Habitat use, movements, and nesting of *Emydoidea blandingii* in Central Wisconsin. *Journal of Herpetology* **24**: 6-12.

Ross, D.A., and J.E. Lovich. 1992. Does the color pattern of two species of turtles imitate duckweed? *Journal of the Pennsylvania Academy of Science* **66**: 39-42.

Row, J.R., and G. Blouin-Demers. 2006. Kernels are not accurate estimators of home range size for herpetofauna. *Copeia* **2006**: 797-802.

Rowe, J.W. 1987. Seasonal and daily activity in a population of Blanding's Turtle

(*Emydoidea blandingii*) in northern Illinois. Unpublished M.Sc. Thesis. Eastern Illinois University. Charleston, IL.

Rowe, J.W. 1992. Dietary habits of the Blanding's turtle (*Emydoidea blandingi*) in northeastern Illinois. *Journal of Herpetology* **26**: 111–114.

Rowe, J.W. and E.O. Moll. 1991. A radiotelemetric study of activity and movements of the Blanding's turtle (*Emydoidea blandingii*) in Northeastern Illinois. *Journal of Herpetology* **25**: 178-185.

Rutherford, P.L. and P.T. Gregory. 2003. Habitat Use and Movement Patterns of Northern Alligator Lizards (*Elgaria coerulea*) and Western Skinks (*Eumeces skiltonianus*) in Southeastern British Columbia. *Journal of Herpetology* **37**: 98-106.

Rubin, C.S., R.E. Warner, and D.R. Ludwig. 2001. Habitat use and movements of radiotagged Blanding's turtles (*Emydoidea blandingii*) in a suburban landscape. *Chelonian Conservation and Biology* **4**: 136–141.

Rushton, S. P., S.J. Ormerid, and G. Kerby. 2004. New paradigms for modeling species distributions? *Journal of Applied Ecology* **41**: 193-200.

Saeys,Y., I. Inza, and P. Larrañaga. 2007. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23:** 2507-2517.

Sajwaj, T.D., and J.W. Lang. 2000. Thermal ecology of Blanding's Turtle in central Minnesota. *Chelonian Conservation and Biology* **3**: 626-636.

Sajwaj, T.D., S.A. Piepgras, and J.W. Lang. 1998. Blanding's turtles (*Emydoidea* *blandingii*) at Camp Ridley: critical habitats, population status, and management guidelines. Report submitted to the Nongame Wildlife Program, Minnesota Department of Natural Resources. 185 pp.

Sappington, J.M., K.M. Longshore, and D.B. Thompson. 2007. Quantifying landscape ruggedness for animal habitat analysis: A Case Study Using Bighorn Sheep in the Mojave Desert. *Journal of Wildlife Management* **71**: 1419–1426.

SAS Institute Inc. 2005-2010. JMP®6.0.2. Statistical Discovery. Cary, NC, USA.

Schapire, R. 2003. The boosting approach to machine learning – an overview. In: Denison, D.D., M. H. Hansen, C. Holmes, B. Mallick, and B. Yu (Eds.). MSRI Workshop on Nonlinear Estimation and Classification, Springer, New York.

Schuler, M., and R.P. Thiel. 2008. Annual vs. Multiple-Year Home Range Sizes of Individual Blanding's Turtles, *Emydoidea blandingii*, in Central Wisconsin. *Canadian Field-Naturalist* **122**: 61-64.

Schulman, L., T. Toivonen, and K. Ruokolainen. 2007. Analysing botanical collecting effort in Amazonia and correcting for it in species range estimation. *Journal of Biogeography* **34**: 1388-1399.

Scott, J.M., P.J. Heglund, M.L. Morrison, J.B. Haufler, M.G. Raphael, W.A. Wall, F.B. Samson (Eds.). 2002. Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington, DC.

Seaman, D.E., B. Griffith, and R.A. Powell. 1999. Effects of sample size on kernel home range estimates. *Journal of Wildlife Management* **63**: 739-747.

Segurado, P., M.B. Araújo, and W.E. Kunin. 2006. Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology* **43**: 433-444.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**: 379-423, 623-656.

Shine, R. 1991. Australian Snakes: A Natural History. Ithaca (NY): Cornell University Press.

Species at Risk Act S.C. 2003, c. 29

Standing, K.L., T.B. Herman, and I. P. Morrison. 1999. Nesting ecology of Blanding's Turtle (*Emydoidea blandingii*) in Nova Scotia, the northeastern limit of the species' range. *Canadian Journal of Zoology* **77**: 1609–1614.

Steen, D.A., and J.P. Gibbs. 2004. Effects of Roads on the Structure of freshwater Turtle Populations. *Conservation Biology* **18**: 1143-1148.

Stickel, L.F. 1950. Populations and home range relationships of the box turtle, *Terrapene c. carolina* (Linnaeus). *Ecological Monographs* **20**: 353-378.

Stockwell, D.R.B., and A.T. Peterson. 2002. Effect of sample size on accuracy of species distribution models. *Ecological Modelling* **148**: 1-13.

Strang, C.A. 1983. Spatial and temporal activity patterns in two terrestrial turtles. *Journal of Herpetology* **17**: 43-47.

Swets, J.A. 1988. Measuring the accuracy of diagnostic systems. *Science* **240**: 1285-1293.

Swihart, R. K., and N. K. Slade. 1985. Influence of sampling interval on estimates of home range size. *Journal of Wildlife Management* **49:** 1019-1025.

Tingley, R., and T.B. Herman. 2009. Land-cover data improve bioclimatic models for anurans and turtles at a regional scale. *Journal of Biogeography* **36**: 1656-1672.

Thuiller,W., S. Lavorel, M.B. Araújo, M.T. Sykes, and I.C. Prentice. 2005. Climate change threats to plant diversity in Europe. *Proceedings of the National Academy of Science USA* **102**: 8245-8250.

Tibshirani, R. 1996. Bias, variance, and prediction error for classification rules. Technical report, University of Toronto.

Van Dam, B. 1993. Element stewardship abstract for *Emydoidea blandingii* (Holbrock) Blanding's turtle. Michigan Natural Features Inventory, Lansing, M.I. 25pp.

Vayssières, M.P., R.E. Plant, and B.H. Allen-Diaz. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science* **11**: 679-694.

Walker, P. A., and K.D. Cocks. 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Global Ecology and Biogeography Letters* **1**: 108-118.

Wang, Z, U-C. I. Chang, Z. Ying, L. Zhu, Y. Yang. 2007. A parsimonious threshold-independent protein feature selection method through the area under receiver operating characteristic curve. *Bioinformatics* **23**: 2788-2794.

Ward, G., T. Hastie, S. Barry, J. Elith, and J.R. Leathwick. 2009. Presence-Only Data and the EM Algorithm. *Biometrics* **65**: 554-563.

Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* **3**: 385-397.

Wiley, E.O., K.M. McNyset, A.T. Peterson, C.R. Robins, and A.M. Stewart. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography* **16**: 120 -127.

Whittaker, R.H. 1967. Gradient analysis of vegetation. *Biological Reviews* **42**: 207–264.

Whitaker, P.B. and R. Shine. 2003. A radiotelemetric study of movements and sheltersite selection by free-ranging Brownsnakes (*Pseudonaja textilis,* Elapidae). *Herpetological Monographs* **17**: 130-144.

Wood, R.C., and R. Herlands, R. 1997. Turtles and tires: the impact of roadkills on the northern diamondback terrapin, *Malaclemys terrapin*, populations on the Cape May peninsula, southern New Jersey, USA. Pages 46-53 *in* J. Van Abbema, technical coordinator. Proceedings: conservation, restoration, and management of tortoises and turtles an international conference, 11-16 July 1993, State University of New York, Purchase. New York Turtle and Tortoise Society, New York, USA.

Welk, E., K. Schubert, and M.H. Hoffmann. 2002. Present and potential distribution of invasive mustard (*Alliara petiolata*) in North America. *Diversity and Distributions* **8**: 219-233.

Wu, X.B., and F.E. Smeins. 2000. Multiple-scale habitat modeling approach for rare plant conservation. *Landscape and Urban Planning* **51**: 11-28.

Yang, G-Z., and X-P. Hu. 2009. U.S. Patent Application Publication No. US 2009/0157584 A1. Washington, DC: U.S. Patent and Trademark Office.

Yom-Tov, Y., and R. Kadmon. 1998. Analysis of the distribution of insectivorous bats in Israel. *Diversity and Distributions* **4**: 63-70.

Yozzo, D. J., and R. J. Diaz. 1999. Tidal freshwater wetlands: invertebrate diversity, ecology, and functional significance. pp. 889–918. In: D.P. Batzer, R.B. Rader, and S.A. Wissinger (Eds). Invertebrates in freshwater wetlands of North America. John Wiley & Sons, New York, New York, USA.

Zaniewski, A.E., A. Lehmann, and J.McC. Overton . 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* **157**: 261-280.

Zheng, B., and A. Agresti. 2000. Summarizing the predictive power of a generalized linear model. *Statistics in Medicine* **19**: 1771-1781.