Habitat suitability modeling for the eastern hog-nosed snake, *Heterodon platirhinos*, in Ontario

Victor Thomasson

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies of the University the 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 de l'Université d'Ottawa en vue de l'obtention de la maîtrise ès sciences
Institut de biologie d'Ottawa-Carleton

ABSTRACT

With exploding human populations and landscapes that are changing, an increasing number of wildlife species are brought to the brink of extinction. In Canada, the eastern hog-nosed snake, *Heterodon platirhinos,* is found in a limited portion of southern Ontario. Designated as threatened by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC), this reptile has been losing its habitat at an alarming rate. Due to the increase in development of southern Ontario, it is crucial to document what limits the snake's habitat to direct conservation efforts better, for the long-term survival of this species. The goals of this study are: 1) to examine what environmental parameters are linked to the presence of the species at a landscape scale; 2) to predict where the snakes can be found in Ontario through GIS-based habitat suitability models (HSMs); and 3) to assess the role of biotic interactions in HSMs.

Three models with high predictive power were employed: Maxent, Boosted Regression Trees (BRTs), and the Genetic Algorithm for Rule-set Production (GARP). Habitat suitability maps were constructed for the eastern hog-nosed snake for its entire Canadian distribution and models were validated with both threshold dependent and independent metrics. Maxent and BRT performed better than GARP and all models predict fewer areas of high suitability when landscape variables are used with current occurrences. Forest density and maximum temperature during the active season were the two variables that contributed the most to models predicting the current distribution of the species. Biotic variables increased the performance of models not by representing a limiting resource, but by representing the inequality of sampling and areas where forest remains. Although habitat suitability models rely on many assumptions, they remain useful in the fields of

conservation and landscape management. In addition to help identify critical habitat, HSMs may be used as a tool to better manage land to allow for the survival of species at risk.

RÉSUMÉ

Durant les dernières décennies, l’accroissement démographique continu a été à l’origine de grands changements de paysage qui ont mené plusieurs espèces au bord de l'extinction. Au Canada, la couleuvre à nez plat, *Heterodon platirhinos,* est retrouvée dans une portion du sud de l’Ontario. Cette espèce a été désignée comme étant menacée au Canada principalement à cause de la perte d’habitat. En raison du développement rapide du sud de l’Ontario, il est crucial de déterminer ce qui limite l’habitat de ce serpent pour mieux orienter les efforts de conservation. Les buts de ce projet sont: 1) examiner les paramètres environnementaux qui sont liés à la répartition de ce reptile à l’échelle du paysage; 2) prédire la présence de l’espèce en Ontario grâce à des modèles d’indice de qualité de l’habitat (IQH); et 3) évaluer le rôle des interactions biotiques dans les modèles IQH.

Trois modèles IQH de haute performance furent employés: Maxent, Boosted Regression Trees (BRTs), et The Genetic Algorithm for Rule-set Production (GARP). Des cartes d’indice de qualité de l’habitat furent construites pour la distribution canadienne de la couleuvre à nez plat et les modèles furent validés avec différentes métriques de performances. Maxent et BRT performent mieux que GARP et tous les modèles prédisent moins d’habitats favorables lorsque des variables de paysage sont utilisées avec des occurrences récentes. La densité de forêt et la température maximale durant la saison active sont souvent les variables contribuant le plus aux modèles prédisant la distribution actuelle de ce serpent. Les variables biotiques ont permis aux modèles de mieux performer non pas parce qu’elles représentent une limitation de ressources, mais parce qu’elles reflètent les inégalités dans l’échantillonnage et des endroits où il y a encore présence de forêt. Malgré le fait que ces modèles présentent certaines limitations, ils demeurent très utiles en

conservation et en gestion des paysages. Non seulement peuvent-ils nous aider à identifier des habitats essentiels, mais ils peuvent aussi être utilisés comme outil afin de mieux gérer l'utilisation des terres tout en protégeant des espèces en péril.

ACKNOWLEDGEMENTS

A number of people and organizations have made this project possible and helped me throughout my degree. I would like to begin by thanking my supervisor and friend, Dr. Blouin-Demers. I believe it is important to have role models not only as a child but throughout life. During my experience at the University of Ottawa, that person has been you. Your comments and suggestions have always been very useful and your quick replies to my emails have saved me a lot of time. Not only have you been an inspiring figure all the way throughout my degree but you have showed me that it is possible to reconcile academia, hobbies, and family. Thank you for offering me this opportunity to do a Masters in a field that I am very passionate about. Although I am not always sure where I am going in life, this project will have enabled me to leave my own contribution to the fields of conservation and herpetology. Thank you for giving me this opportunity, this means a lot to me.

Thank you to my committee members, Dr. David Curie (University of Ottawa) and Dr. Lenore Fahrig (Carleton University). You have provided me with insightful suggestions that have improved the quality of my project. Thank you for finding time for me; your expertise and knowledge were very useful. Because I had never used a GIS prior to this project, I had to learn it during my first year. Thank you, Dr. Bahram Daneshfar (University of Ottawa), for showing patience when I was bombarding you with emails. You have enabled me to learn ArcGIS as quickly as possible and helped me solve several technical issues. This also applies to Catherine Millar, who I consulted several times to get my models running. As for Nancy Lemay and Pierre Leblanc (University of Ottawa), thank you for helping me acquire ArcGIS 10.0. and countless GIS maps from various databases. Finally, many thanks to all my friends and colleagues: Gabrielle Fortin, Véronique Juneau, Rayan

El-Balaa, Nicola Banger, Laura Robson, Nick Cairns and Julie Châteauvert. I have great memories of times spent with you, whether ice skating, drinking or stuck in a hotel all together for a weekend. At the end of the day, it is the people that you see on a daily basis that make a difference. I understand people come and go but I wouldn't want you to go too far…

This project would have not been possible without the financial support of Parks Canada. I hope my project fulfilled you expectations and that it will help you achieve your goals in the recovery of the eastern hog-nosed snake. In addition to financial support, this project required a lot of data that I could never have gathered by myself. Thank you to : The Natural Heritage Information Center of Ontario (Michael Oldham and colleagues) for giving me access to all the eastern hog-nosed snake occurrences, Ontario Nature (John Urquhart and colleagues) for providing me with amphibian and snake occurrences, and FrogWatch and The Marsh Monitoring Program for sharing data on toad calls in my region of study. Amanda Xuereb who is currently doing a Master at Queen's University has also provided me with additional hog-nosed snake occurrences. All this information represents years and years of field work and came from not only scientist but the general public. Thank you to those who contribute voluntarily to herpetological atlases. This project is an example of what can be done with your data.

I could have never gotten that far if it wasn't for the support of my family. You have always done all you could to help me, especially when I was going through phases of existential crises. You have always been there when I needed you the most. Thank you for guiding me all these years. Finally, I would like to thank all those who came before me. I realize this study will only be a tiny step in the field of conservation and that it could have not been accomplished without the work of the people before me. This is particularly true for those who not only publish a study but also develop computer programs, whether they are models such as Maxent or scripts for R. The work you have accomplished is considerable.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

GENERAL INTRODUCTION

Anthropogenic activity and habitat selection

Understanding how living organisms are distributed around the planet has always been a subject of great interest in ecology. It was one of the topics studied by Charles Darwin on his voyage on the HMS Beagle (Darwin, 1889). Over a century later, this subject is more important than ever. As the human population is increasing and landscapes are changing, numerous wildlife species are becoming at risk of extinction (Pereira et al., 2010b; Stokstad, 2010). Anthropogenic features such as agricultural fields, roads, and cities have reduced habitat availability for many organisms. Continuous natural landscapes are less available and the proximity of wildlife to habitat edges is increasing. This has led to a variety of ecological issues including the reduction of population sizes (Burke and Nol, 2000), reduced genetic diversity (Dixo et al., 2009), lowered dispersal success (Fahrig and Merriam, 1985), and impaired biotic relationships (Tylianakis et al., 2008). Today, habitat loss, degradation, and fragmentation are considered primary threats for countless wildlife species, including snakes (Gibbons et al., 2000; Ernst and Ernst, 2003). This is why it is essential to measure the impact of habitat destruction on the long-term persistence of biodiversity.

Habitat selection, which can be defined as "the disproportionate use of resources or conditions by living things" is often studied to determine how important a habitat is for the conservation of species (Mayor et al., 2009). However, selection is dependent on both spatial and temporal scales (Wiens, 1989). For example, the environmental conditions favored by an organism at the scale of its home range may be different than the ones selected for during hibernation. In relation to spatial scale, Johnson (1980) proposed four orders of selection: species range, home range, feeding site, and food items. At the largest scale, the species range, it may be argued that a species does not select habitat features per se, but that it simply occurs where it can: a species will occupy any areas where it is able to survive and reproduce. The latter three orders of selection have been studied extensively, but there has been less research done at a landscape scale. It is essential to study habitat selection at larger scales to determine what habitat configurations are necessary for the persistence of organisms.

Habitat suitability models (HSMs)

A tool that can help us study habitat selection at the scale of the species range, and that has been particularly useful in the field of conservation, is habitat suitability modeling (Guisan and Thuiller, 2005). Habitat suitability models (HSMs) have a long history and Johnston (1924) was already using them to assess the invasion of cactus species in Australia (quoted in Pearson and Dawson, 2003). Today, HSMs are used for a variety of applications including: defining the ecological requirement of species, evaluating the potential of invasive species, finding unknown populations, predicting the effect of climate change and habitat loss, conservation planning, etc. (Peterson, 2006).

HSMs can be defined as empirical models relating field observations to an array of environmental parameters based on statistically sound algorithms (Guisan and Zimmermann, 2000). HSMs weigh the influence of various environmental variables on the distribution of a given species. The models allow predicting the likelihood of occurrence of a species at a given location. These models are therefore useful in determining what factors are best at predicting the presence of a species at a landscape scale. It is important to remember that a species may not occupy all suitable grounds for various reasons, such as competition with other species or barriers to dispersal, unaccounted for by the model (Pearson, 2007). This is precisely what distinguishes the potential ecological niche from the realized ecological niche. In addition, a species may now occupy a different range, and thus use a different set of environmental conditions, than it did before anthropogenic habitat changes; this is particularly likely for declining species. Finally, a threshold has to be determined to convert the probability of occurrence of a species on a given pixel to a presence or an absence to generate predicted presence/absence maps (Liu et al., 2005).

Two types of data are required to build HSMs: 1) presence/absence points for the species modeled, and 2) environmental characterization data. While some models require presence observations only (more recent models), others require both presence and absence, or presence and pseudo-absence (Miller, 2010). Although real absences have been shown to increase the predictive power of HSMs, this information is seldom available or often unreliable and forces many researchers to use models using presence observations only or presences and pseudo-absences (Brotons et al., 2004). The type of data available will often dictate or influence the decision of which model to use in a given situation (Newbold, 2010).

The second type of data required is environmental characterization data, also known as the environmental predictors. These predictors can be anything from climate to topography and are often collected through remote sensing. In recent years, some researchers have even added variables such as the distribution of other organisms interacting either positively or negatively with the modeled species (Heikkinen et al., 2007; Zielinski et al., 2010; Santos et al., 2006). Predictors should be chosen carefully and should reflect key components of the natural history of the species being modeled. They should be "causal driving forces” shaping the distribution of the species (Guisan and Zimmermann, 2000). Finally, the best predictors are variables having direct physiological impacts on the organism, such as climate or pH. Direct predictors increase the generality of the model and make it more powerful in predicting the distribution of a species at larger special scales (Guisan and Zimmermann, 2000).

Although there have been various HSMs developed in the past 20 years, more recent ones characterizing the background environment tend to outperform more traditional ones (Elith et al. 2006). Similarly, models using pseudo-absences were found to perform better than presence only models (Elith et al., 2006). The predictive abilities of these models were also tested at different scales and sample sizes, and were shown to vary extensively (e.g. Pearson et al., 2007). In a broad study comparing 16 models, boosted regression trees (BRTs) (Elith et al., 2008), maximum entropy (Maxent) (Phillips et al., 2006), and multivariate adaptive regression splines (MARS) (Elith et al., 2006) were found to be the most robust and powerful models (Elith et al., 2006). Similar results were found in a complementary study looking at six models including the ecological niche factor analysis (ENFA) (Ortega-Huerta and Peterson, 2008). Not only do these models weigh the influence of each variable on the distribution of species differently, but they are flexible and can analyze complex situations. The drawback to these qualities is that some models may overfit the training data and have lower predictive power. For more information on habitat suitability modeling, please consult Appendix I.

The potential power of the *Species at Risk Act* of Canada (SARA)

Proclaimed in 2003 by the federal government of Canada, the *Species at Risk Act* (SARA) was put in place to protect wildlife species at risk of extinction and/or extirpation and ensure their recovery in Canada (S.C. 2002, c.29, s.6). The *Act* established the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) as the scientific advisory body of experts responsible for identifying species at risk (S.C. 2002, c.29, s.14,15). This legal document stipulates that a recovery strategy must be prepared by the Minister for any species designated as threatened, endangered, or extirpated by COSEWIC (S.C. 2002, c.29, s.37). This recovery strategy should provide all the necessary tools to protect and ensure the persistence of given species. It also stipulates that critical habitats must be defined "to the extent possible" (S.C. 2002, c.29, s.41). Today, COSEWIC already has 602 species listed in four different "risk" categories. Proportionally, reptiles are the group of vertebrates the most at risk in Canada.

The *Species at Risk Act* has clear provisions concerning the habitat of species at risk. It is prohibited to damage or destroy the residence of threatened, endangered, and extirpated species (S.C. 2002, c.29, s.33). In the same way, it is illegal to destroy the critical habitat of these species (S.C. 2002, c.29, s.58, 61). The Minister may also establish a code of practice as to how a critical habitat will be protected (S.C. 2002, c.29, s.56). However, the *Species at Risk Act* only has legislative power on federal lands, some water surfaces, and migratory birds. In the United-States, the equivalent of the *Species at Risk Act* is the *Endangered Species Act* which was signed in 1973. Provinces and States have complementary legal documents to protect wildlife on provincial and state land. The *Ontario Endangered Species Act* of 2007 protects species at risk in Ontario whether they are endangered, threatened or of special concern. On provincial land, this document forbids the destruction of habitat of endangered or threatened species (2007, c.6, s.10).

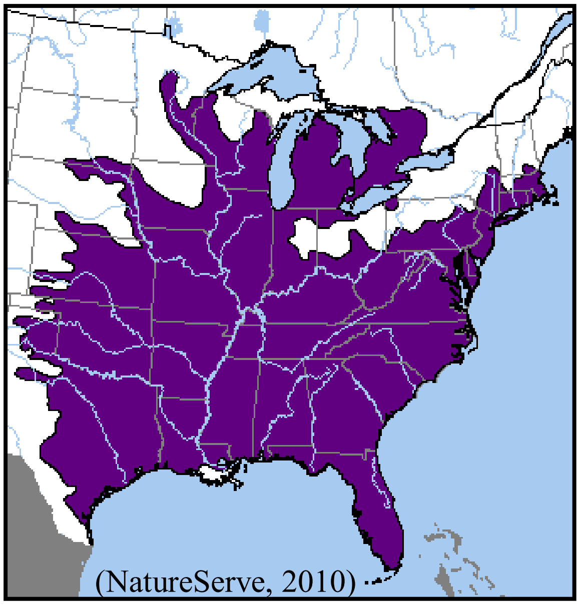
Although the provisions found in these documents provide a legal support for the protection of critical habitats for species at risk, sufficient information on the habitat requirements of a given species is first required. Many terms such as critical habitat and residence must be defined clearly to apply and enforce these legislations. In the same way, important questions remain to be answered (e.g. What land can be considered critical habitat for these species?). It is essential to identify what habitats and conditions are essential to all species at risk to make good use of these legal documents and enforce current legislation. It is also important to know the maximum level of disturbance that can be tolerated by these species to survive to develop useful conservation strategies.

The conservation status of reptiles in Canada

Until 10 000 years ago, most of Canada was covered in ice and was inhospitable for ectothermic animals such as reptiles. Since then, glaciers have receded and 46 species of reptiles are now found in the country (Orchard, 2006). Under the northern climatic conditions of Canada, these reptiles face unique challenges, especially in relation to thermoregulation. For example, many populations of turtles are known to select areas where temperatures are the most favorable (Picard et al., 2011; Dubois et al., 2009). Some snakes were found to use edges for thermoregulation (Blouin-Demers and Weatherhead, 2001). At these latitudes, the active season is short and leaves little time to accumulate the resources that are essential for survival. Among lizards, this has favored certain life-history characteristics including small body size, viviparity, tolerance to freezing temperatures, and long-lived adults (Powell and Russell, 2006). Snakes often use communal hibernation sites that are deep enough to avoid frost (Gregory, 2006). These dens are in limited supply and can be responsible for long-distance migrations of up to several kilometers (Gregory, 2006). All these characteristics are some of the reasons why, in Canada, reptiles are particularly vulnerable to human activity.

Today, almost 70% of the reptile species found in Canada are considered nationally at risk (Seburn and Bishop, 2006). This includes 17 out of the 25 species of snakes found in Canada. Many species now also have scattered populations that make natural recolonization almost impossible (Seburn and Bishop, 2006). Canadian reptiles face many threats including environmental pollution, disease and parasitism, unsustainable use, invasive species, and climate change, but the most important one remains habitat loss and degradation (Gibbons, 2006). This is amplified in Canada because people live where it is the warmest, and hence where reptile diversity is the richest.

The eastern hog-nosed snake



(NatureServe, 2010)

The eastern hog-nosed snake, *Heterodon platirhinos*, is a thick bodied burrowing snake of 50 to 85 cm in length (Seburn, 2009; Davis, 1946) found throughout most of the eastern United States, from Texas and Florida to New England and Ontario (Figure 1.) (NatureServe, 2010). In Canada they are only found in two regions: in the Carolinian Life Zone of southwestern Ontario and in central Ontario, south of the French River and Lake Nipissing (Seburn, 2009). Whether these two populations are genetically isolated remains unknown, but a significant distance now separates them. It is also important to note that less than 10% of the eastern hog-nosed snake total distribution is found in Canada (Seburn, 2009).

Although they have been observed in a range of habitats, eastern hog-nosed snakes are often found in upland deciduous forests in close proximity to open vegetative cover and edges (Lagory et al., 2009; Platt, 1969). In New England, they are found in pine barrens and white pine forest ecosystems (Michener and Lazell Jr, 1989). This reptile also has a preference for well drained sandy and loose soil (Michener and Lazell Jr, 1989; Plummer and Mills, 2000). This type of soil is used by this snake in various ways such as for oviposition (Cunnington and Cebek, 2005; Robson, 2011). These environmental conditions are all believed to help hog-nosed snakes keep relatively high body temperatures (Plummer and Mills, 2010; Lagory et al., 2009; Platt, 1969). Among reptiles, thermoregulation is known to be a driving force for habitat selection, especially in northern populations (Seigel and Collins, 1993; Row and Blouin-Demers, 2006; Dubois et al., 2009). The eastern hog-nosed snake is considered a specialist as a large portion of its diet is composed of amphibians, especially American Toads (*Anaxyrus americanus*) and Fowler's Toads (*Anaxyrus fowleri*) (Platt, 1969; Uhler et al., 1939; Edgren, 1955).

Once common throughout most of its range, the eastern hog-nosed has declined in many areas, including in Canada; it was extirpated from many localities such as Point Pelee National Park and Pelee Island (Seburn, 2009). It was estimated that at least 8% of known element occurrences, which can be considered as small pockets of individuals, have been extirpated in Canada (Oldham, 2000). In the United-States, it is still considered very common although there are no studies that have ever looked at its total distribution or at population densities (NatureServe, 2010). Because the eastern hog-nosed snakes found in Canada are at the most northern limit of their distribution, some populations may be at higher risk of extirpation. For example, recruitment rates could be low and overwintering mortality rates high. This species is also considered of regional concern in some of the northeastern United-States, such as in New Hampshire where it is listed as threatened (Taylor and Marchand, 2005).

Although many factors are responsible for the decline of this species in Canada, habitat loss, degradation, and fragmentation are the most important ones. As many municipalities of Ontario are growing and landscapes changing, fewer habitats remain available for wildlife such as the eastern hog-nosed (Seburn, 2009). This is particularly true in the Carolinian Life Zone that has been used extensively for agriculture. The community of Wasaga Beach has rapidly grown leaving fewer suitable areas for hog-nosed snakes (Seburn, 2009). Several populations are now isolated, particularly in Southwestern Ontario, and may quickly be extirpated due to stochastic events. These small populations probably have small gene pools and might not be viable in the future. Although these snakes have been shown to avoid paved roads (Robson, 2011; Rouse et al., 2011), road mortality was found to be a problem in several areas of Ontario (Rouse et al., 2011; Seburn, 2009). Compared to other snakes such as the Eastern Massasauga, *Sistrurus catenatus*, the hog-nosed is vagile and capable of traveling up to 120 m in a single day (Rouse et al., 2011; Plummer and Mills, 2000). The active foraging nature of this snake was shown to decrease survivorship and render it more vulnerable to roads (Rouse et al., 2011; Plummer and Mills, 2000).

Snakes are often feared by people and sometimes destroyed whether they are venomous or not. Eastern hog-nosed snakes are known for their defensive display. They will bluff by flattening their heads like cobras and hissing, but they are harmless. Because of this unusual behaviour, the snake is often killed by landowners (Seburn, 2009). Finally, the decline of this species may also be related to its specialized diet. In Hog Island, Virginia, the decline of a small population was hypothesized to be related to a decline in toad abundance (Conant et al., 1990). Because of its decline in Canada, COSEWIC has designated the eastern hog-nosed as “Special Concern” in 1997 and as “Threatened” in 2001. This species is also considered as “Threatened” in Ontario, the only province in Canada where it is found.

General objectives

The aim of this project is to define the potential habitat of the eastern hog-nosed snake, *Heterodon platirhinos*, at the scale of the province of Ontario and to provide governmental agencies and the scientific community with useful information that will help design effective conservation strategies for the species. In Chapter 1, I use habitat suitability models to recreate both the historical and current distribution of the snake in the province of Ontario. I also identify the conditions in which the species is found at a landscape scale and discuss where the species could be declining in Canada. In Chapter 2, I assess the role of biotic variables in habitat suitability models. Because the eastern hog-nosed snake is considered a specialist in terms of diet, it was proposed that some populations are restricted to areas where food items are plentiful. I build models using prey distribution variables and assess whether these models perform better or not. It is important to know if biotic variables increase the performance of habitat suitability models because they may help recreate realized distributions from potential distributions, which is often needed in various fields, including conservation.

CHAPTER I

Using habitat suitability models to delineate critical habitat and estimate the historical and current distributions of the eastern hog-nosed snake in Ontario.

Abstract

Habitat suitability models have been used in a variety of fields, such as conservation, and are now considered a powerful tool to reconstruct the potential niche of species. Presence-only models have been particularly useful to monitor and study rare species at a landscape scale. These species are often more difficult to study, and in some cases, the only information available are occurrences in atlases. In Canada, the *Species at Risk Act* not only protects any species at risk, but also their residences and critical habitats. It is thus essential to study species at risk at a landscape scale and identify what geographic areas these species depend on and which habitat can be considered suitable. In this study, I identify areas of high suitability and estimate habitat loss through presence-only habitat suitability models (HSMs).

I employ three models, Maxent, BRTs and GARP, to recreate both the historical and current distributions of a threatened species of snake in Ontario, the eastern hog-nosed snake, *Heterodon platirhinos*. All models are validated with a number of performance metrics, some of which are threshold-dependent and others threshold-independent. The strongest models are then combined using a consensus approach and categorical maps showing four conservation scenarios are built using three thresholds, one of which is the same to calculate the minimum predicting area (MPA). Although these models should be interpreted with care, Maxent and BRT performed better than GARP, with AUCs ranging between 0.85 and 0.95. Habitat suitability is lower in the south of the province when land cover variables are used in models in conjunction with recent occurrences. While forest density increases habitat suitability, cropland density limits the distribution of this reptile. Habitat suitability models can therefore be used to assess habitat loss.

Introduction

Studying species at risk using habitat suitability models

An increasing number of wildlife species are brought to the brink of extinction. Recent rates of extinction were shown to be 10 to 1000 times what they were before human presence and will continue to rise as the human population increases (Pimm et al., 1995; Butchart et al., 2010; Pereira et al., 2010a). It was even proposed that we have already entered earth's sixth mass extinction (Barnosky et al., 2011). Although extinction rates vary from one taxonomic group to another, biodiversity loss is global and affects reptile species in various parts of the world, including Canada (Seburn and Seburn, 2000). A number of causes are responsible for current loss in biodiversity including invasive species, climate change, pathogens and over-harvesting, but the main culprit remains habitat loss and fragmentation. Approximately one million square kilometers of rainforest is cut down every 5 to 10 years (Pimm and Raven, 2000). Although earth is currently undergoing significant changes in landscape, the distributions of most species is still poorly documented and rarely monitored, particularly in developing countries (Gaston and Rodrigues, 2003).

In response to this loss in natural heritage, governments have been slowly putting together pieces of legislations to regulate anthropogenic activity in relation to species considered at risk. For example, in Canada the *Species at Risk Act* proclaimed in 2003 protects any species at risk on federal land (S.C. 2002, c. 29). These pieces of legislation regulate not only the harvest of species at risk, but also the destruction of their habitats. In Canada, both the residence and the critical habitat of species at risk are protected and the Minister has the obligation to produce a recovery strategy for each of these species. It is also stated that critical habitat should be defined "to the extent possible" (S.C. 2002, c.29, s.41).

Unfortunately, this is hard to achieve and there has been no agreement on how to identify critical areas in a standardized approach. In many cases, insufficient information is known on a given species to develop proper recovery strategies and protect its critical habitat. It is essential to find techniques to identify the parcels of lands that are essential to species at risk to be able to apply current legislation and to secure the long-term survival of biodiversity.

In the past decades, geographic information systems (GIS) have been used increasingly to study habitat selection at various scales, including at the scale of the home range via radio telemetry (Row and Blouin-Demers, 2006). This GIS data available, whether high-resolution spatial environmental data or species occurrences in atlases, render possible the study of many species at risk with already existing data. Although species occurrences kept by museums and governmental databases present a number of limitations, including taxonomic and spatial biases, they remain important data that took years to collect (Newbold, 2010). The accessibility of satellite images, or maps derived from them, and species occurrences has led to the rise of a tool which has grown significantly in the past 20 years: habitat suitability models (HSMs). These empirical models relate field observations to an array of environmental parameters based on statistically sound algorithms (Guisan and Zimmermann, 2000). Although some of these models require both presences and absences, presence-only habitat suitability models characterizing habitat availability using pseudo-absences were shown to be effective for modeling species distributions (Elith et al., 2006). In the past decade, HSMs have been successfully used to: find unknown populations, predict the effect of climate change and habitat loss on biodiversity, and to assess the potential of invasive species (Peterson, 2006). With respect to species at risk, these models were shown to be a powerful tool that can help answer questions such as where is habitat suitability the highest for a given species or what conditions it prefers (Hu and Jiang, 2010). Considering all the advances that were made on habitat suitability models, would it be possible to use these models to identify critical habitat of species at risk to make good use of current legislative documents protecting these species?

The eastern hog-nosed snake, *Heterodon platirhinos*, is a small snake ranging between 50 and 85 cm in length (Seburn, 2009; Davis, 1946) found in the United-States and Canada. In Canada, its distribution is limited to the south of Ontario, where it is divided is two populations: one near Long Point and the other in the area of Georgian Bay (Seburn, 2009). This reptile has a specialized diet consisting mostly of toads and is often found in sandy areas including pine forests (Michener and Lazell Jr, 1989). It is considered semi-fossorial and will often burry, not only to avoid predation but to thermoregulate and find suitable areas for oviposition (Cunnington and Cebek, 2005). Due to the increasing development of southern Ontario, the species has been losing its habitat at an alarming rate and is now considered threatened in Canada (Seburn, 2009). Although other studies have looked at the ecology of this species at the scale of its home range in a number of location in the northern part of its range (Robson, 2011; Rouse et al., 2011; Lagory et al., 2009; Plummer and Mills, 2000), it is crucial to determine the habitat requirements at a larger scale to identify areas that should be protected for the long-term survival of this species.

Objectives

In this study, I identify areas of high suitability for the eastern hog-nosed snake and propose three thresholds that may be used with habitat suitability models, for conservation applications, to protect the habitats of species at risk. I attempt to recreate both the historical and current distributions of this reptile in the province of Ontario. I start by building HSMs in a hierarchical approach using tree algorithms, and combine the best models following a consensus method. I evaluate all models using various metrics of performance. I identify the environmental conditions used by the species at a landscape scale and discuss areas where it could have lost habitat in Canada. Finally, I discuss both the conservation implications for the species, and the potential use of habitat suitability models to identify critical habitat.

Methods

Eastern hog-nosed snake occurrences

Two types of data are required to run presence-only habitat suitability models: species occurrences, and environmental predictors that consists in GIS layers in raster format for the entire area of study. Most *Heterodon platirhinos* sightings were retrieved from the Natural Heritage Information Centre of Ontario (NHIC) (Oldham, 1998). These sightings were collected opportunistically from many sources including the public, and are classified according to their level of accuracy. All sightings with 100 m of accuracy or better were kept and were divided in two categories: all sightings (1899-2011) and recent sightings (1990-2011). To complement these records, I also obtained sightings found in one published study (Cunnington and Cebek, 2005), one M.Sc. thesis at the University of Ottawa (Robson, 2011), and one ongoing study conducted by A. Xuereb at Queen’s University. Because this snake is semi-fossorial, threatened and inconspicuous, I had relatively few points to work with and many points were clustered in provincial parks or other conservation areas. Ideally, all occurrences should have been collected with a consistent sampling effort following a sampling design to cover all regions in the area of study. This is hard to achieve for a species as elusive as the eastern hog-nosed snake and would require years of fieldwork. To reduce spatial autocorrelation and overrepresentation of highly sampled regions, I only used a subset of points separated by at least 1 km. Points were selected using Focus Tool (Holland et al., 2004). This radius was judged appropriate because the home range of this species is usually less than a 1 km² (Plummer and Mills, 2000). Many occurrences were therefore excluded in highly sampled regions. Finally, this dataset was divided randomly in two: 75% to build the models and 25% for validation. The final dataset consisted of 254 occurrences, 126 of which have been collected since 1990

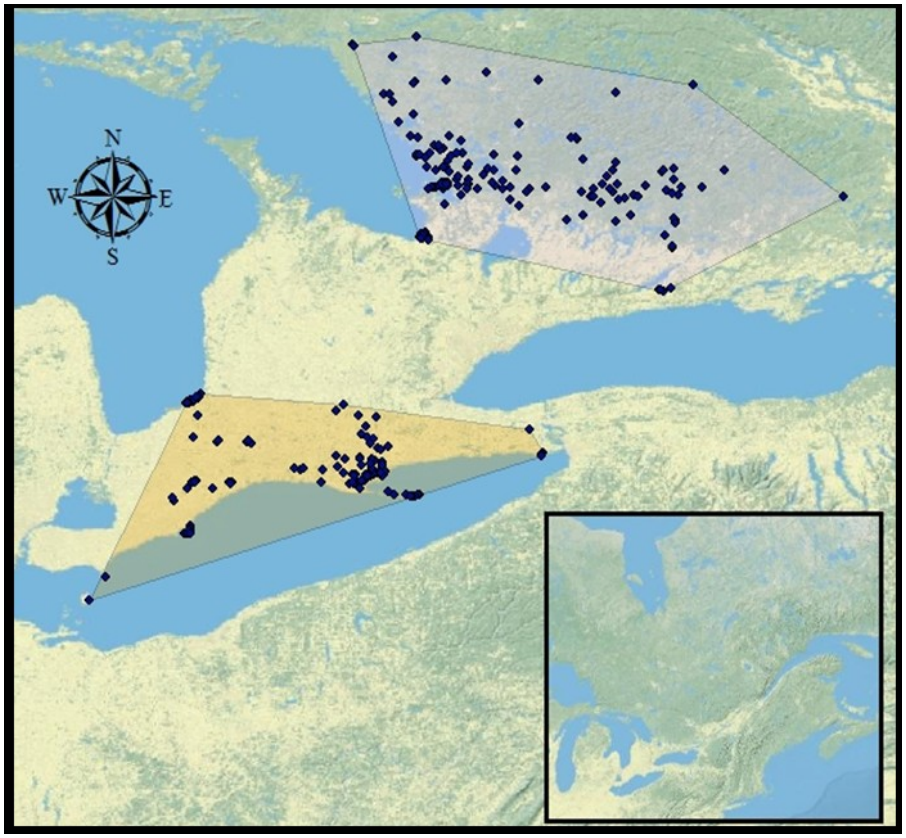


Figure 1-1. Final hog-nosed snake dataset including both historical and current occurrences.

Explanatory variables

The variables chosen to predict the distribution of the eastern hog-nosed were all thought to have clear implications for the distribution of this snake. These variables were chosen in relation to the natural history of the species and the area studied. They were also found to be relevant predictors in other studies looking at the distributions of reptiles including snakes (Santos et al., 2006; Penman et al., 2010; Bombi et al., 2009; Santos et al., 2009). They include both direct predictors (e.g. mean maximum temperature for the active season) and indirect predictors (e.g. elevation), and can be grouped in 6 categories: topographical, geological, ecological, climatic, land cover, and anthropogenic variables (Table 1-1.). They were acquired from different sources and were all built or resampled in ArcGIS 10.0.

Both the slope (SLOP) and the aspect (ASPE) were calculated using the Ontario Provincial Digital Elevation Model (DEM) v.2.0.0 at a resolution of 10 or 20 m (OMNR, 2005). They were calculated at fine resolution (~15 m) with the slope and aspects tools in the Spatial Analyst Tools of ArcGIS 10.0, and were then resampled at the resolution of the model. The only categorical variable was GEOG, representing the surficial geology of Ontario (OGS, 2010), reclassified for practical reasons (Table 1-2.). From this variable, I calculated both sand density (SAND) and Precambrian bedrock density (PRBE) in each pixel. These three variables were added to models because hog-nosed snakes are known to occupy sandy areas and because the northern population of hog-nosed snake in Ontario is found mostly on Precambrian bedrock.

Land cover variables were built using the Ontario Land Cover Dataset (OMNR, 1998) made up of 28 classes at a resolution of 25 m, and represented the proportion of each class within a pixel (Table 1-3.). It is important to note that the Ontario Land Cover Dataset was derived from LANDSAT images obtained between 1986 and 1997, but that most images are from the early 1990s. To build a variable representing population density (PEOP), I divided the number of people found in each dissemination area of the 2006 Canadian Census (StatisticsCanada, 2006) and divided it by its area in km². To build road density (ROAD), I used the Ontario Road Network (OMNR, 2006) and the kernel density tool in the Spatial Analyst Tools of ArcGIS 10.0. The radius of the buffer was always half the resolution of the model and the number of times a road was counted was equal to the number of lanes it had.

Finally all climatic variables were retrieved from the WorldClim database (Hijmans et al., 2005) at a resolution of 30 arc s (~1 km). The GIS data offered by Worldclim is a compilation of interpolated historical data from weather stations all over the world, collected between 1950 and 2000. To calculate the maximum temperature and mean monthly precipitation for the active season, I averaged values for the six months of active season at this latitude for that particular species (April - September). Isothermality, which is a measure of temperature fluctuation, was another variable considered in these models because it was found to limit the distribution of species in other studies (e.g. Bombi et al., 2009).

This total combination of 21 predictors was deemed appropriate as I used various direct and indirect variables and because the extent of this study is only a portion of Ontario. In some cases, indirect predictors can replace direct predictors or resources (Guisan et al., 1999), but will reduce the generality of HSMs (Guisan and Zimmermann, 2000). Because habitat suitability models are often affected by multicollinearity (Miller, 2010), I made sure none of these variables were highly correlated (all $|r| < 0.8$, mean $|r| = 0.08$). The Pearson correlation coefficients among all of these variables, at a resolution of 1 km, are found in Appendix V.

Until 10,000 years ago, Canada was covered in ice and was inhospitable for any ecothermic animals, including reptiles (Orchard, 2006). Because the historical distribution of the eastern hog-nosed snake could have been highly influenced by climate, I also built models using strictly climatic variables (Table 1-4.). These variables comprised 10 WorldClim bioclimatic variables at a resolution of 30 arc s (~1 km), all thought to limit species distributions by representing extreme conditions (Hijmans et al., 2005). These variables were the largest subset of bioclimatic variables that were not highly correlated (all $|r| < 0.8$, mean $|r| = 0.20$) and included both variables of precipitation and temperature at different periods of the year. Again, the Pearson correlation coefficients among all climatic variables, at a resolution of 1 km, are found in Appendix V.

Table 1-1. Variables predicting the historical and current distribution of the hog-nosed snake in Ontario.

Type	**Code**	**Description and Units**	**Used for model**	**Original resolution**	**Source of original data**
Topographical	**ALTI**	Elevation (m)	both	10-20m	(OMNR, 2005)
	SLOP	Slope (°)	both	"	n/a
	ASPE	Aspect (°)	both	"	n/a
Geological	**GEOG**	Surficial geology (8 categories)	both	n/a	(OGS, 2010)
	SAND	Sand density (%)	[1]	"	"
	PRBE	Precambrian bedrock density (%)	[1]	"	"
Land Cover	**ALVA**	Alvar density (%)	current only	25m	(OMNR, 1998)
	ROCK	Bedrock density (%)	"	"	"
	CROP	Cropland density (%)	"	"	"
	CUBU	Cuts & burns density (%)	"	"	"
	FORE	Forest density (%)	"	"	"
	MUDF	Mudflat density (%)	"	"	"
	PAST	Pasture density (%)	"	"	"
	SETT	Settlement density (%)	"	"	"
	WATR	Water density (%)	"	"	"
	WETL	Wetland density (%)	"	"	"
Climatic	**TMAX**	Mean maximum temp. for active season (°C*10)	both	1km	(Hijmans *et al.* , 2005)
	PREC	Mean monthly precip. for active season (mm)	both	"	"
	ISOT	Isothermality (%)	both	"	"
Anthopogenic	**PEOP**	Total population density (people/km^2)	both	n/a	(StatisticsCanada, 2006)
	ROAD	Total road density (kernel, m/km^2)	both	n/a	(OMNR, 2006)

[1] SAND and PRBE was used with all GARP model because it does not take categorical variables.

Table 1-2. Categories encompassed in the variable "Surficial Geology".

Categories	**Descriptions**
(1) Bedrock	Incudes both peloozoic and precambrian bedrock.
(2) Gravel	Coarse sand and gravel.
(3) Diamicton	Includes a mix of different size sediments (coarse to fine).
(4) Organic Deposits	Muck, peat and marl.
(5) Sand	Sand dunes, sand, silty sand.
(6) Silt	Silt, clay.
(7) Clay	Fine clay, silty clay, silt.
(8) Others	Includes fills and non classified areas.

Table 1-3. Descriptions of the land cover variables used to build models.

Land Cover Classes	**Descriptions**
Bedrock density	Mine tailings, quarries and bedrock outcrops.
Cropland density	Croplands and open soil for agricultural activity.
Cuts & burns density	Includes recent cutovers, recent burns and old cuts and burns.
Forest density	Includes any type of forests including plantations.
Mudflat density	Costal mudflats near the Great Lakes.
Pasture density	Pastures and abandoned fields with open grasslands.
Settlement density	Settlement and developped land (e.g. cities).
Water density	All waters bodies not included in wetland density.
Wetland density	Includes all marshes, bogs, and swamps.

Table 1-4. Descriptions all 10 bioclimatic variable (Hijmans *et al.*, 2005).

Code	**Descriptions**
bio2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
bio3	Isothermality (Mean Diurnal Range/Temperature Annual Range)*100
bio5	Max Temperature of Warmest Month
bio6	Min Temperature of Coldest Month
bio8	Mean Temperature of Wettest Quarter
bio9	Mean Temperature of Driest Quarter
bio13	Precipitation of Wettest Month
bio14	Precipitation of Driest Month
bio15	Precipitation Seasonality (Coefficient of Variation)
bio18	Precipitation of Warmest Quarter

A multi-scale hierarchical approach

In Ontario, eastern hog-nosed snakes are at the most northern part of their distribution. Thus, their distribution is not only affected by human activity, but also by cold climatic conditions. To discern the effects of climate and human activity, and estimate habitat loss, I built three series of habitat suitability models in a multi-scale hierarchical approach. To predict the potential historical distribution of the snake (prior to 1990), all snake occurrences including recent ones were used with all variables except land cover because satellite images were not retrieved at the same time as some of the old snake sightings. HSMs make better predictions when predictors match temporally with the species occurrences (Newbold, 2010). Models predicting the current distribution of the snake only used snake occurrences collected since 1990, with all variables including land cover. Finally, I built two climatic models, one using all snake occurrences and the other using only recent occurrences, to assess whether the historical distribution of the snake could have been defined by climatic conditions. It is important to note that these models are not directly comparable because they all use different variables and occurrences collected at different periods. Thus, some caution is warranted in the interpretation of the model comparisons.

Considering I had 126 recent snake occurrences, I had too many variables in models predicting the current distribution of the snake: up to 20. To reduce the number of variables in current models, I built full models and removed the variables that contributed the least to the models. I used a technique similar to backward stepwise selection, but instead of optimizing an AIC value I optimized the AUC (Doetsch et al., 2009). Always taking out the variable that contributed the least to the model, I pruned the models until I had the highest

AUCs. Similar approaches were used in various fields including medicine (Abeel et al., 2010; Wang et al., 2007) and ecology (Millar and Blouin-Demers, 2011).

Models predicting the historical and current distribution of the species were built at 3 resolutions (scales): 1 km, 750 m, and 500 m. These resolutions were judged appropriate considering that 1 km^2 corresponds to the maximum home range size of the eastern hog-nosed in Ontario and New England and that it is suitable for conservation applications (Cunnington, 2004; Lagory et al., 2009). Climatic models were only built at the original resolution of the variables, 1 km, because climatic conditions rarely vary significantly across short distances. These resolutions are also appropriate because some of the predictors are only found at these resolutions (Table 1-1) and the species modeled is relatively vagile. As for the extent of this study, it extends 1 km in all directions beyond the historical distribution of the eastern hog-nosed snake.

Models Employed

While climatic models were built with a Maxent only, two other machine learning algorithms with high predictive power were also employed to predict the current and historical distribution of the eastern hog-nosed snake: Boosted Regression Trees (BRTs) and the Genetic Algorithm for Rule-set Production (GARP). Maxent and BRTs were shown to be amongst the most accurate HSMs in an extensive study comparing 16 different models across 226 species in 6 regions of the World (Elith et al., 2006). As for GARP, it was developed more than 10 years ago (Stockwell and Peters, 1999), is still used today (e.g. (Vega et al., 2010), and is capable of making good predictions in broad unsampled regions (Terribile et al., 2010).

Maxent relies on the principle of maximum entropy and “estimates a target probability distribution by finding the probability distribution of maximum entropy (i.e., that is most spread out, or closest to uniform), subject to a set of constraints that represent our incomplete information about the target distribution” (Phillips et al., 2006). In other words, the algorithm assigns the highest probability possible to each pixel in the area of study (the sum of which must equal to one) based on the species’ occurrences and a number of variables. I used version 3.3.3e of the Maxent software with default settings because they were found to give good results with various datasets (Phillips and Dudík, 2008). Each model was built using 10 random replicates always setting aside 25% of the occurrences and pseudo-absences for validation. Maxent retains its high predictive capacities even when the sample size is small (Hernandez et al., 2006; Wisz et al., 2008). All maps produced with Maxent are in logistic format, with probabilities of suitability ranging between 0 and 1, and can be interpreted as an estimate of the probability of presence (Phillips et al., 2006).

Boosted Regression Trees (BRTs) combine the strengths of classification trees (also known as regression tree) and boosting to build a final model in a forward stage-wise fashion (Elith et al., 2008). In the context of habitat suitability models, classification trees are intuitive, easy to visualize, and can model interactions. As for boosting, it combines a number of simple trees together and aims at increasing the accuracy of a final model by finding an average of rough rules (Elith et al., 2008). All BRTs models were fitted in R version 2.12.0 (R, Development Core Team. 2011) using both the gmb package (Ridgeway, 2010) and other codes for R specific to BRTs (Elith et al., 2008). For all models, I used a tree complexity of 5, a bag fraction of 0.5, and the fastest learning rate (*lr*) that reached a minimum of 1000 trees. These settings were chosen based on both our sample size and past studies (Elith et al., 2008; Millar and Blouin-Demers, 2011; Young et al., 2011).

The Genetic Algortihm for Rule Set Production (GARP) is another machine-learning algorithm that develops a set of conditional rules to relate species occurrences to a number of environmental predictors (Stockwell and Peters, 1999). It builds a set of rules in an iterative process (evaluation, testing, and incorporation or rejection rules) to predict a species' distribution. It uses envelope (e.g. if temperature is between 23 and 29ºC = present), atomic (e.g. if geology is sand = present), and logistic regression rules (regression equation where the output is a probability) in a progressive approach (Stockwell and Peters, 1999). I implemented all GARP models in openModeller (Muñoz et al., 2011) following a procedure proposed to identify a subset of runs that performed the best (Anderson et al., 2003). All models were built using 100 runs, a convergence limit of 0.01, and a maximum number of iterations of 999. Out of the initial 100 runs, only the 10 best models were kept selecting 20% of the models with the least omissions, of which 50% with a commission rate the closest to the median. Because GARP does not perform well with categorical variables (Elith and Graham, 2009), I substituted GEOG by sand density (SAND) and precambrian bedrock density (PRBE) in all GARP models (Table 1.2.).

In total, 10 000 pseudo-absences were generated randomly in the area of study to run all models. Seventy-five percent of these points were used to build the models and twenty-five percent to validate them. For a better comparison of the models, I always used the same subsets to build and validate all models. To reduce the uncertainty associated with each algorithm, I combined the best models using a consensus methods proposed by Marmion *et al.* (2009): arithmetic mean (Mean all). I therefore have a final model composed of one, two or three algorithms depending on their individual scores. This is particularly important because presence-only models are difficult to evaluate and the performance of HSMs is influenced by a number of factors including the resolution (Guisan et al., 2007a), the number of occurrences (Wisz et al., 2008), the sampling design (Dennis et al., 1999), and the prevalence of the species (Hernandez et al., 2006).

Validating models

Validating habitat suitability models is considered a challenge (Araújo and Guisan, 2006; Terribile et al., 2010) and although many studies only use Receiver Operating Characteristics (ROC) analyses to do so, it remains hard to validate HSMs, particularly when real absences are missing. In recent years, the use of ROC analyses to validate HSMs has been criticized (Lobo et al., 2008; Peterson et al., 2008). For example, it ignores the goodness-of-fit of the predictions and treats omission and commission errors equally while this should not always be the case (Lobo et al., 2008). To better assess the performance of our models and avoid relying strictly on ROC analyses, I used several techniques including novel threshold-dependent metrics. I present sensitivity that represents the proportion of correctly classified presences and specificity, the proportion of correctly classified absences. A sensitivity and specificity of one could be interpreted as the perfect prediction of the test dataset. I also calculated Cohen's kappa (Cohen, 1960) and true skill statistic (TSS) (Allouche et al., 2006) which both correct the overall accuracy of model predictions by the accuracy expected to occur by chance. However, TSS is independent of prevalence, uses both commission and omission errors, and was shown to be a strong method to evaluate presence absence models (Allouche et al., 2006). For both of these metrics, +1 indicates perfect agreement between the model and the test dataset while 0 or less, no better than random. For all threshold dependent metrics, the probabilities of occurrence were converted to presences and absences using the threshold that maximized the sum of the specificity and sensitivity. This threshold was not only shown to perform well for validation (Liu et al., 2005), but is appropriate for this study considering I am aiming at reconstructing the distributions of a species. As for evaluating models with ordinal scores, the receiver operating characteristic (ROC) analysis (Fielding and Bell, 1997) is an effective technique used in the majority of studies evaluating HSMs. The area under the ROC curve (AUC) must be determined and should range from 0 to 1. The scores can be interpreted as following: 0.5–0.6, insufficient; 0.6–0.7, poor; 0.7–0.8, average; 0.8–0.9, good and 0.9–1, excellent (Araújo and Guisan, 2006). Here I present two AUC values, one which was calculated with the software of each model (Int. AUC) and the other calculated independently in R (Ext. AUC). This second AUC value was calculated the same way for all models treating pseudo absences as true absences. Because the species studied here is rare and because most pseudo-absences are most likely true absences, the AUC is an appropriate measure of performance for this study. The AUC was found to evaluate models predicting realized distributions more accurately than modes predicting potential distributions (Jiménez-Valverde, 2012). Finally, I used a new metric called the minimal predicted area (MPA) (Engler et al., 2004). To calculate the MPA, continuous maps are converted to binary maps by applying the minimum threshold required to consider 90% of the species occurrences as present. The MPA is the proportion of the map considered as presences and a low score suggests a low level of commission error. According to the principle of parsimony, a good model should predict an area as small as possible while comprising a maximum number of occurrences (Engler et al., 2004). All validation metrics were calculated using each model's software and a package called PresenceAbsence (Freeman and Moisen, 2008) in R version 2.12.0. (R, Development Core Team. 2011). The MPA was calculated using ArcGIS 10.0.

Delineating critical habitats for the eastern hog-nosed snake

No consensus has been reached regarding how to delineate critical or essential habitat with habitat suitability models but these areas of interest may be identified somewhat arbitrarily through various techniques. All habitat suitability models produce maps showing continuous probabilities of occurrence ranging between 0 and 1, which can then be converted to a presence or an absence using a threshold (Liu et al., 2005). The question remains: at what probability can a pixel be considered as critical habitat? This not only depends on the prevalence of the species but on the model used, the quality of the data, etc. One way to address this question is to consider all probabilities above a certain threshold as critical. Although they might not always be critical, we know these areas are highly suitable for the species and worthy of conservation efforts. For example, this threshold could be the one used to calculate the MPA (Rupprecht et al., 2011; Engler et al., 2004). The area comprised in the MPA not only includes most known occurrences, but areas of high suitability. The choice of threshold should also depend on financial resources; with a limited budget it should be increased to comprise only areas that are the most suitable. For species highly at risk of extinction, the threshold could be lowered to include all occurrences. In addition to the MPA, I propose using two additional thresholds to delineate critical habitat: one allowing the predictions of 50% of the occurrences and the other 70%. Maps produce by these thresholds will therefore show four suitability categories: poor, moderate, good, and excellent. The main advantage behind these thresholds is that they are determined only with the information that is known: occurrences. They are also more accurate than simple cut-off thresholds because they are calculated according to probability distributions (Rupprecht et al., 2011). Finally they are easy to visualize, indicating the proportion of the species' distribution considered for conservation.

Results

The performance of historical and current models

Maxent and BRTs usually performed better than GARP in both historical and current models. Maxent and BRTs had higher AUCs, specificities, sensitivities and TSSs, and lower MPAs (Tables 1-5 and 1-6). For this reason, consensus models were built using only Maxent and BRTs (Figures 1-8, 1-9 and 1-10). For most models, the difference in performance between Maxent and BRTs was very small; making it impossible to say which algorithm made the best predictions under these conditions. In current models, Maxent usually required fewer variables (8 to 9) than BRTs and GARP to perform well and achieve high AUCs. Although the differences in scales did not seem to influence the performance of the models, Maxent, BRTs and the consensus models always performed well with AUCs ranging between 0.85 and 0.95. With values always above zero, the kappa statistics and TSS both indicate that all models performed better than what could be expected by chance (Tables 1-5 and 1-6). The models that had the highest kappa statistics were always BRTs with values between 0.15 and 0.25 in historical models (Table 1-6). It is important to note that the prevalence varied a lot from one model to another and that the threshold that maximized the sum of specificity and sensitivity was highly dependent on the algorithm. BRT models always had a low MPA which reduced chances of having false presences (Figure 4-4, 4-5 and 4-6). In the same way, maps produced by this model had few areas of high suitability. GARP does the opposite, having high MPAs and assigning high probabilities to most of the area of study (Figure 4-7, 4-8 and 4-9). Maxent had fewer areas of high suitability than GARP, but more than BRTs (Figure 4-1, 4-2 and 4-3). Finally, according to most metrics, current models performed better than historical ones, but kappa statistics usually dropped to remain slightly above zero (Table 1-6). Current models also had fewer areas of high suitability, particularly in the south of the province near Long Point (e.g. Figures 1-8, 1-9 and 1-10). Habitat suitability maps for each model separately are found in appendix II (Figures 4-1 to 4-9). Consensus habitat suitability maps presenting four scenarios are found in appendix III (Figures 5-1 to 5-3). To convert these continuous maps to categorical ones, a probability of suitability ranging between 0.0 and 0.5 was defined as poor, 0.5 to 0.7 moderate, 0.7 to 0.9 good, and 0.9 to 1.0 excellent.

Contributions of variables in historical and current models

The contribution of each variable in the models varied a lot from one algorithm to another, but usually remained similar between resolutions. I only present the contributions of the variables for Maxent and BRT models because these were the models used to build consensus models. In order of importance, the five variables with the highest contributions in Maxent models predicting the historical distribution of the eastern hog-nosed snake were: (1) surficial geology, (2) altitude, (3) istothermality, (4) mean precipitation during the active season, and (5) road density (Figure 1-2). In historical models produced with BRTs, these variables were: (1) altitude, (2) mean precipitation during the active season, (3) total population density, (4) road density, and (5) slope (Figure 1-2). While surficial geology was the most important variable in historical models produced with Maxent, it was the least important one in historical models built with BRTs. In all historical models, the probability of suitability increased as isothermality and mean precipitation during the active season increased, and decreased when population density increased and when altitude increased beyond ~180 m (Figure 1-4 and 1-5).

In models predicting the current distribution of the snake, the variables used in historical models acted in the same way on the probabilities of suitability, but two additional land cover variables became more influential: cropland density and forest density. In order of importance, the five variables with the highest contributions in Maxent models predicting the current distribution of the eastern hog-nosed snake were: (1) isothermality, (2) cropland density, (3) mean maximum temperature during the active season, (4) forest density, and (5) surficial geology (Figure 1-3). In current models produced with BRTs these variables were: (1) forest density, (2) mean maximum temperature during the active season, (3) altitude, (4) road density, and (5) mean precipitation during the active season (Figure 1-3). These models show that eastern hog-nosed snake sightings that have been reported since 1990 were in forested areas where maximum summer temperatures are relatively high (Figure 1-6 and 1-7). Finally, the probability of high suitability increased with increasing forest density, and decreased as cropland density increased and when road density exceeded a certain threshold (~20 m/km^2).

Table 1-5. Metrics of performance for models predicting the historical distribution of the eastern hog-nosed snake.

Model	Resolution	No. var	Int. AUC	Ext. AUC	Threshold	Sensisivity	Specificity	Kappa	TSS	MPA
Maxent	1000m	9	0.867	0.86	0.298	0.79	0.85	0.17	0.64	0.314
Maxent	750m	9	0.866	0.87	0.32	0.81	0.86	0.18	0.67	0.339
Maxent	500m	9	0.858	0.86	0.256	0.83	0.82	0.14	0.65	0.391
BRT	1000m	9	0.863	0.86	0.031	0.75	0.91	0.25	0.66	0.112
BRT	750m	9	0.885	0.89	0.018	0.84	0.82	0.15	0.66	0.127
BRT	500m	9	0.892	0.89	0.022	0.79	0.87	0.19	0.66	0.097
GARP	1000m	10	0.74	0.78	0.949	0.78	0.74	0.088	0.52	0.526
GARP	750m	10	0.78	0.79	0.949	0.76	0.77	0.1	0.53	0.52
GARP	500m	10	0.74	0.74	0.949	0.81	0.65	0.059	0.46	0.494
Consensus	1000m	n/a	n/a	0.86	0.184	0.76	0.88	0.2	0.64	0.259
Consensus	750m	n/a	n/a	0.88	0.17	0.81	0.85	0.18	0.66	0.249
Consensus	500m	n/a	n/a	0.87	0.135	0.83	0.81	0.14	0.64	0.262

Table 1-6. Metrics of performance for current models.

Model	Resolution	No. var	Int. AUC	Ext. AUC	Threshold	Sensisivity	Specificity	Kappa	TSS	MPA
Maxent	1000m	9	0.896	0.93	0.192	0.94	0.83	0.1	0.77	0.16
Maxent	750m	8	0.886	0.93	0.206	0.91	0.85	0.11	0.76	0.195
Maxent	500m	8	0.881	0.94	0.238	0.91	0.88	0.14	0.79	0.212
BRT	1000m	14	0.934	0.93	0.01	0.91	0.87	0.13	0.78	0.052
BRT	750m	11	0.945	0.95	0.007	0.94	0.82	0.096	0.76	0.068
BRT	500m	10	0.944	0.94	0.007	0.94	0.84	0.11	0.78	0.052
GARP	1000m	9	0.88	0.88	0.649	0.84	0.78	0.068	0.62	0.443
GARP	750m	12	0.88	0.88	0.849	0.88	0.8	0.08	0.68	0.382
GARP	500m	13	0.82	0.82	0.949	0.78	0.74	0.049	0.52	0.416
Consensus	1000m	n/a	n/a	0.93	0.104	0.94	0.83	0.11	0.77	0.144
Consensus	750m	n/a	n/a	0.94	0.108	0.91	0.85	0.11	0.76	0.182
Consensus	500m	n/a	n/a	0.94	0.123	0.91	0.88	0.14	0.79	0.153

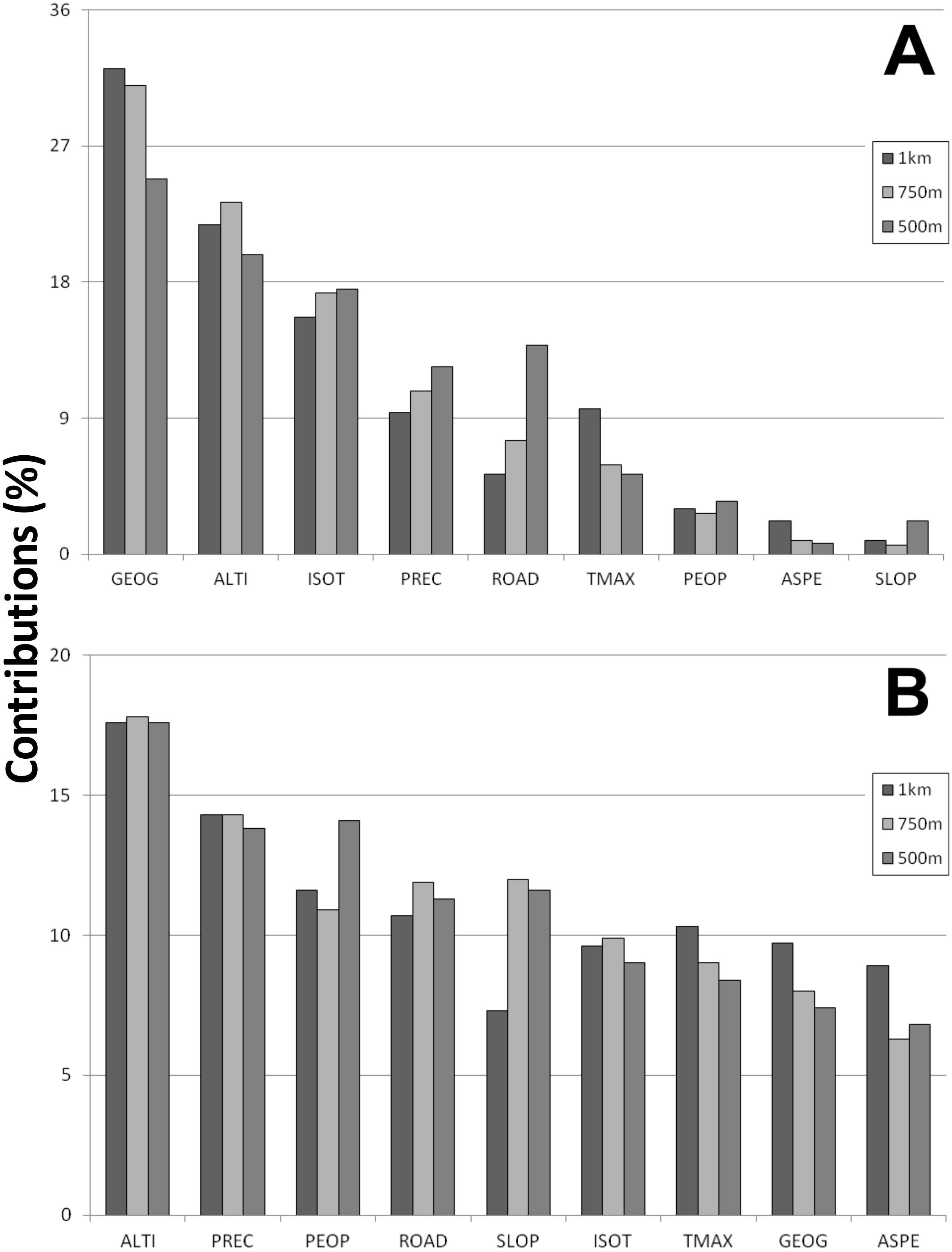


Figure 1-2. Contributions of all variables in habitat suitability models used to estimate the historical distribution of the eastern hog-nosed snake: (**A**) Maxent and (**B**) BRT.

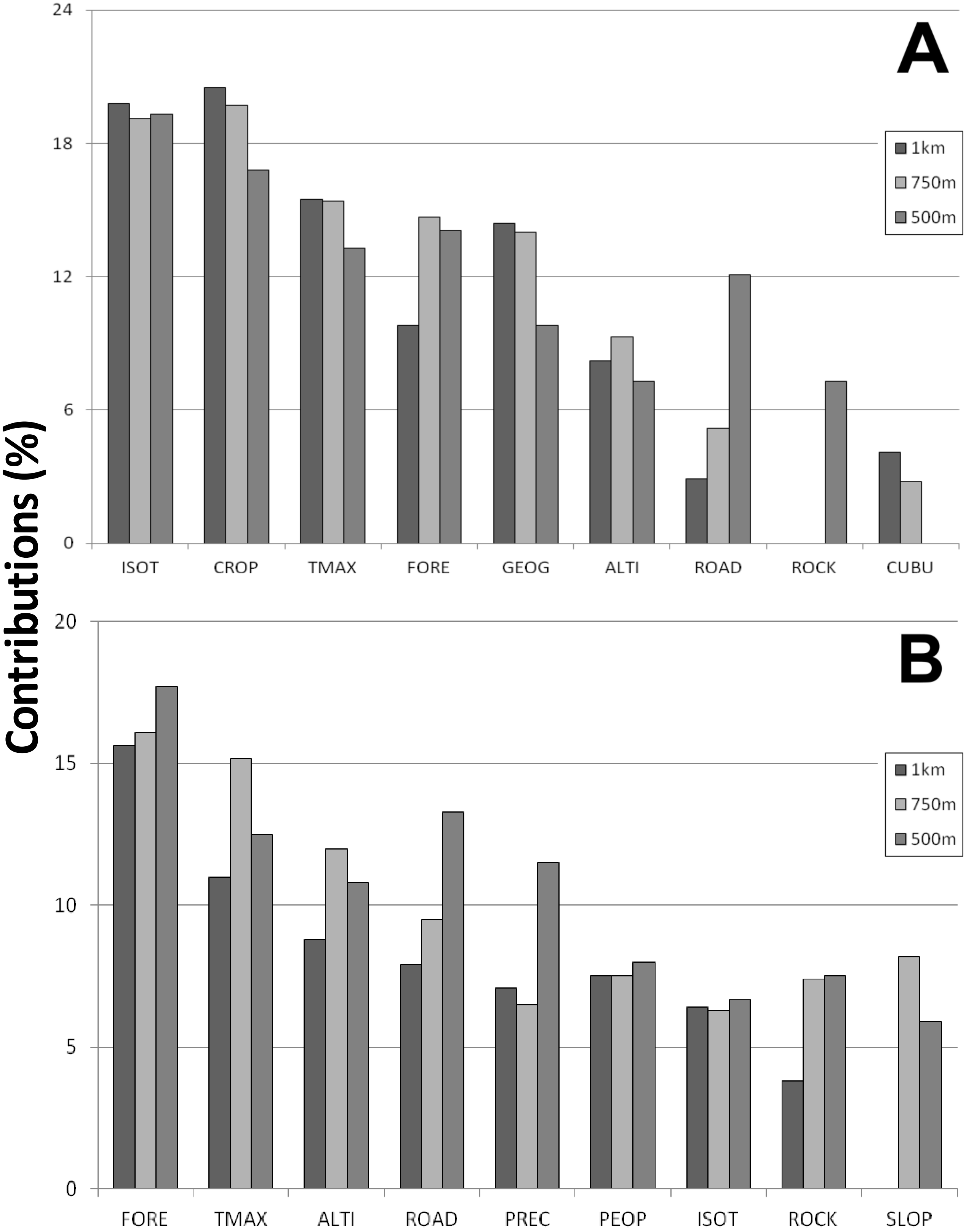


Figure 1-3. Contributions of the nine most influential variables in habitat suitability models used to estimate the current distribution of the eastern hog-nosed snake: (**A**) Maxent and (**B**) BRT.

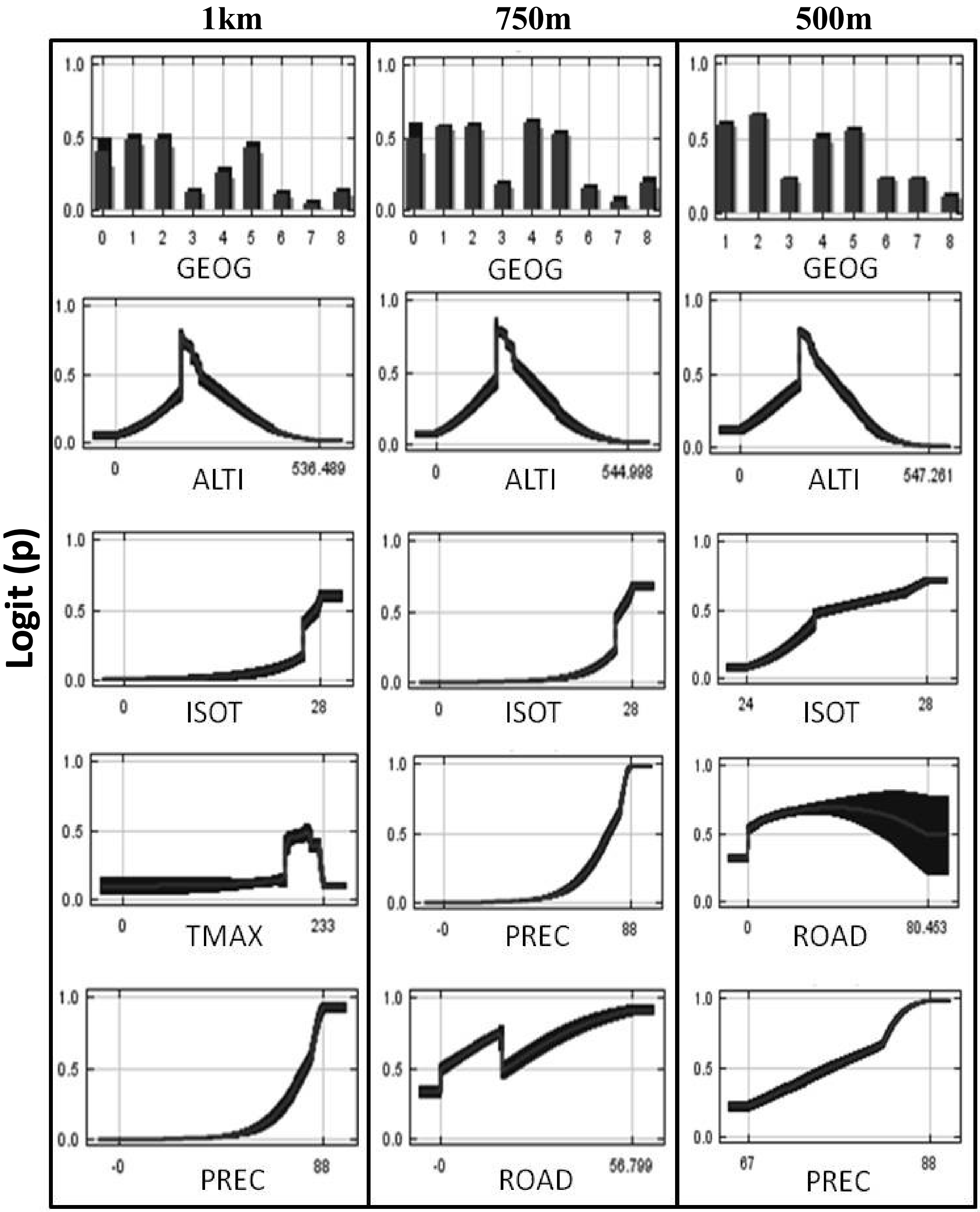


Figure 1-4. Marginal response curves ± SD, for the five most influential variables in the models built with Maxent, estimating the historical distribution of the eastern hog-nosed snake. The curves show how the logistic prediction (p) varies, when one variable is gradually changed, while all others are kept at their average value.

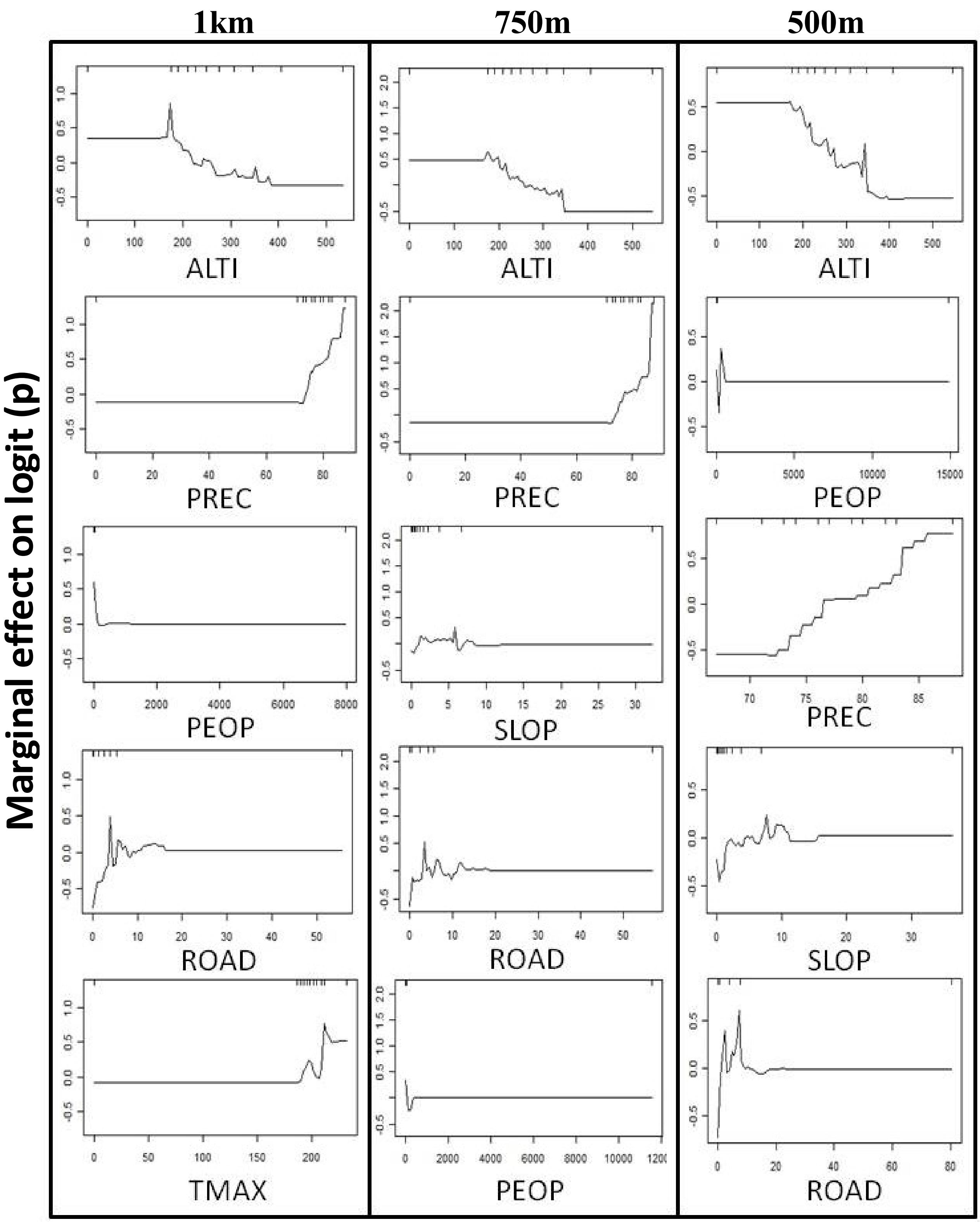


Figure 1-5. Partial dependence plots for the five most influential variables in the models built with BRTs, estimating the historical distribution of the eastern hog-nosed snake. These curves show the influence of each variable on predicted probability of suitability, after accounting for the average effects of all other variables in the model.

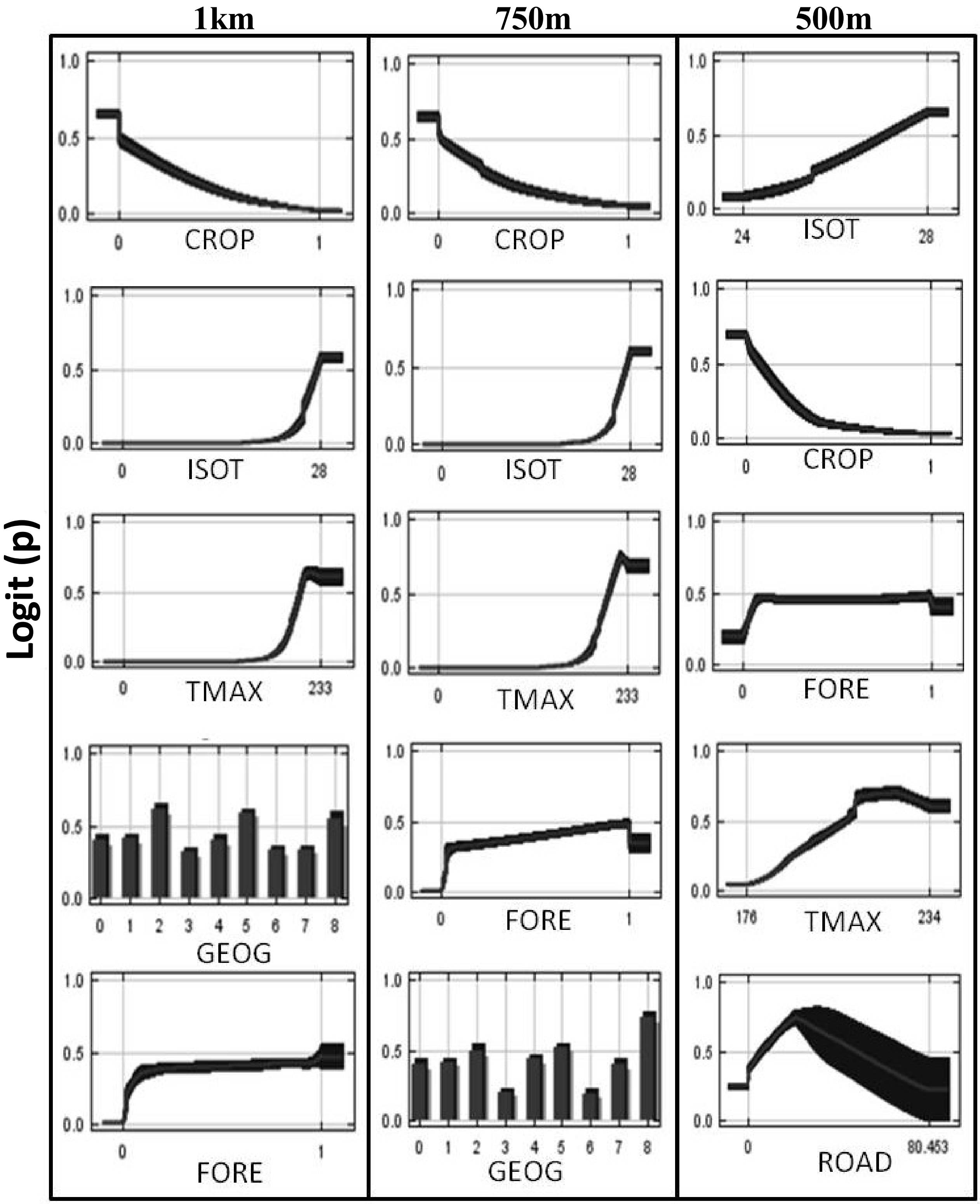


Figure 1-6. Marginal response curves ± SD, for the five most influential variables in the models built with Maxent, estimating the current distribution of the eastern hog-nosed snake. The curves show how the logistic prediction (p) varies, when one variable is gradually changed, while all others are kept at their average value.

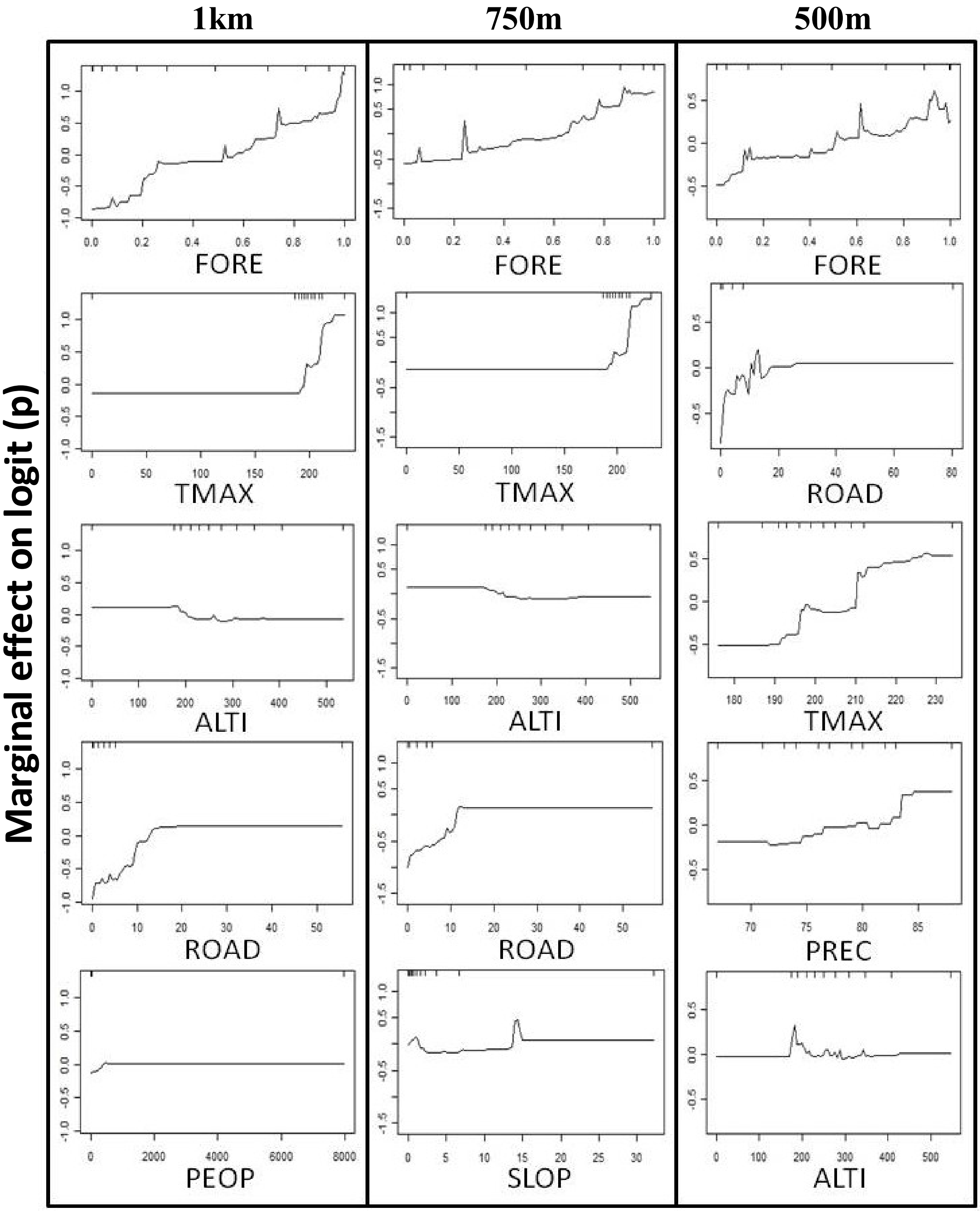


Figure 1-7. Partial dependence plots for the five most influential variables in the models built with BRTs, estimating the current distribution of the eastern hog-nosed snake. These curves show the influence of each variable on predicted probability of suitability, after accounting for the average effects of all other variables in the model.

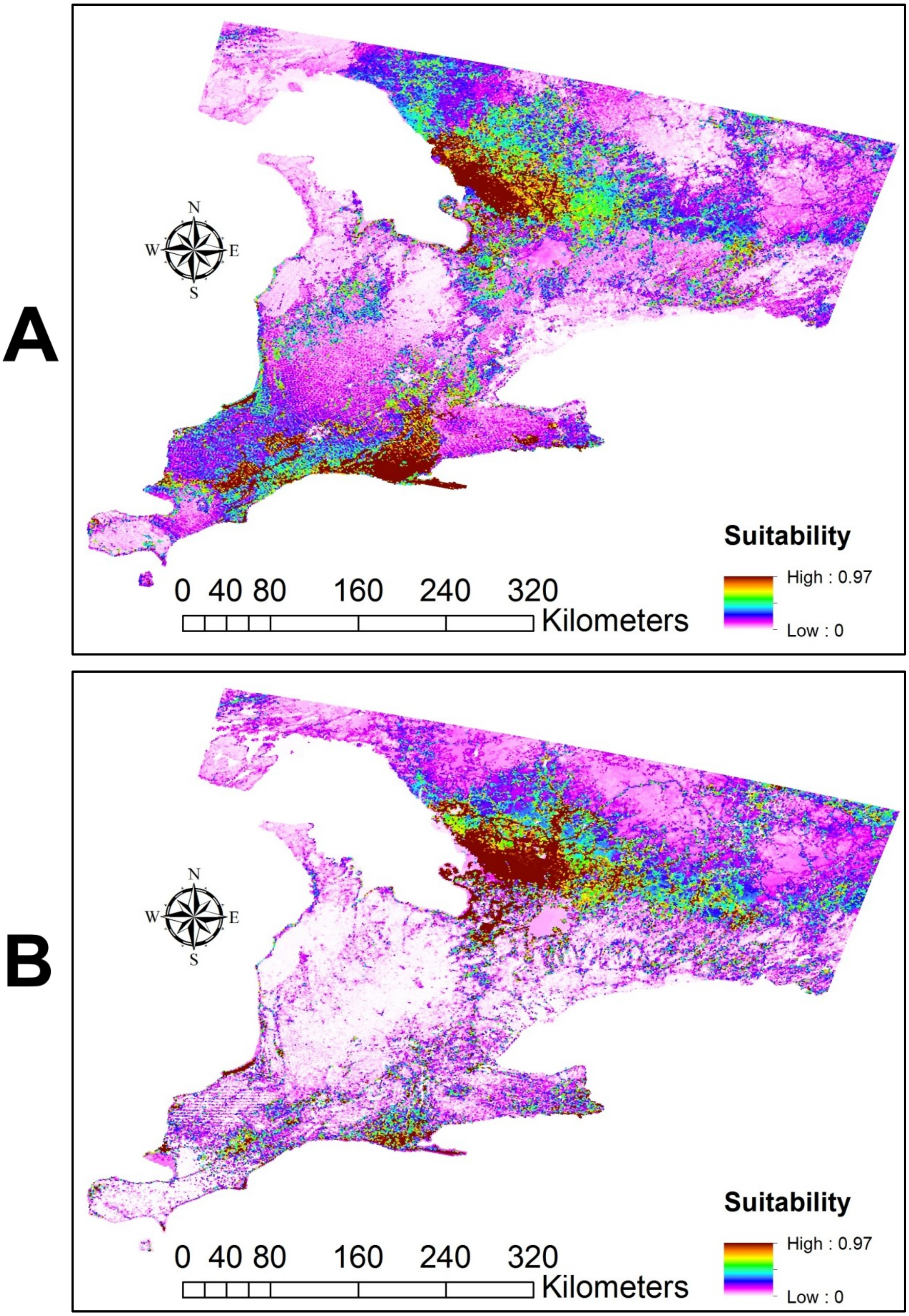


Figure 1-8. Estimated distribution of the eastern hog-nosed snake in Canada, using the consensus approach at a resolution of 1km: (**A**) historically and (**B**) currently.

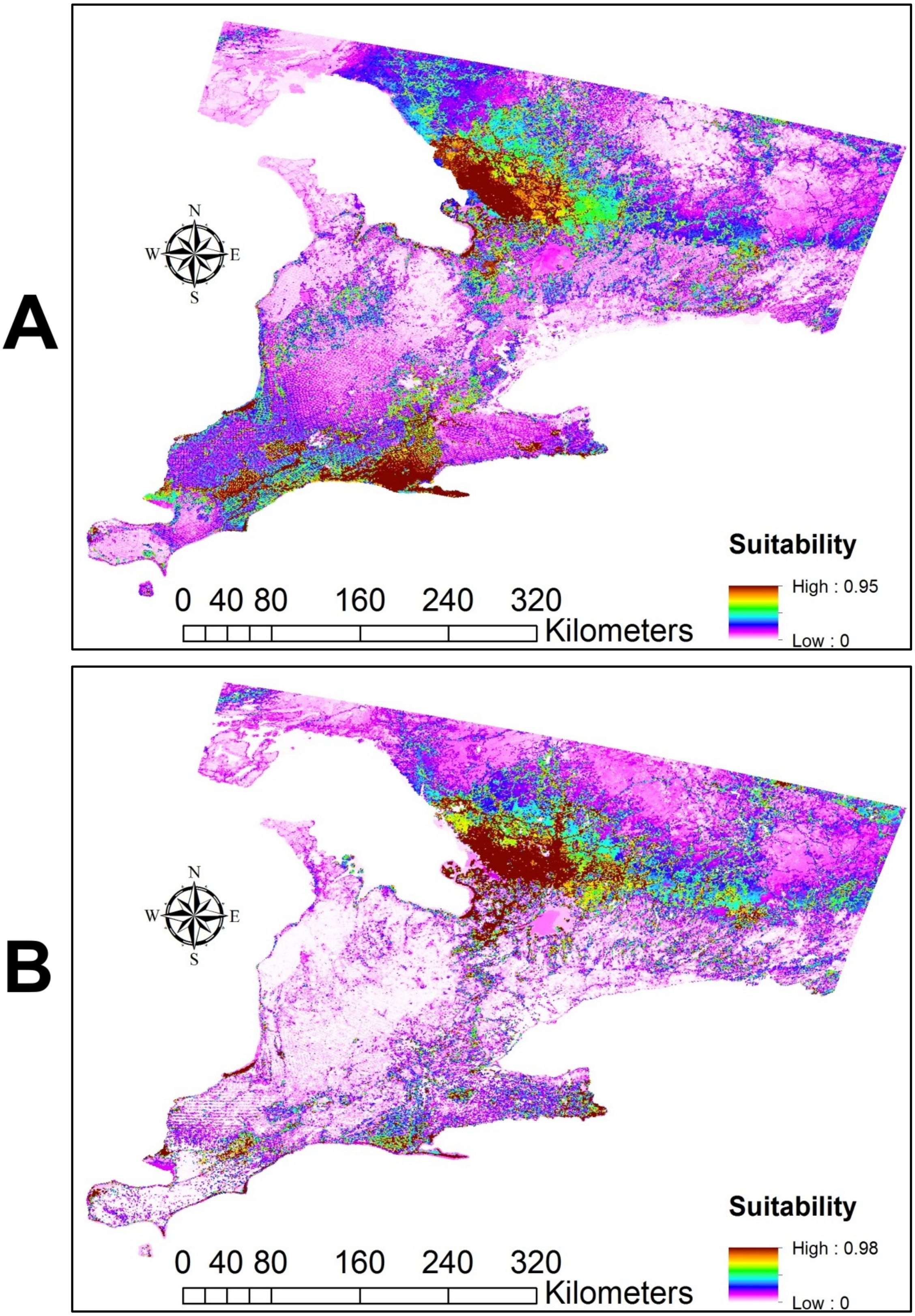


Figure 1-9. Estimated distribution of the eastern hog-nosed snake in Canada, using the consensus approach at a resolution of 750m: (**A**) historically and (**B**) currently.

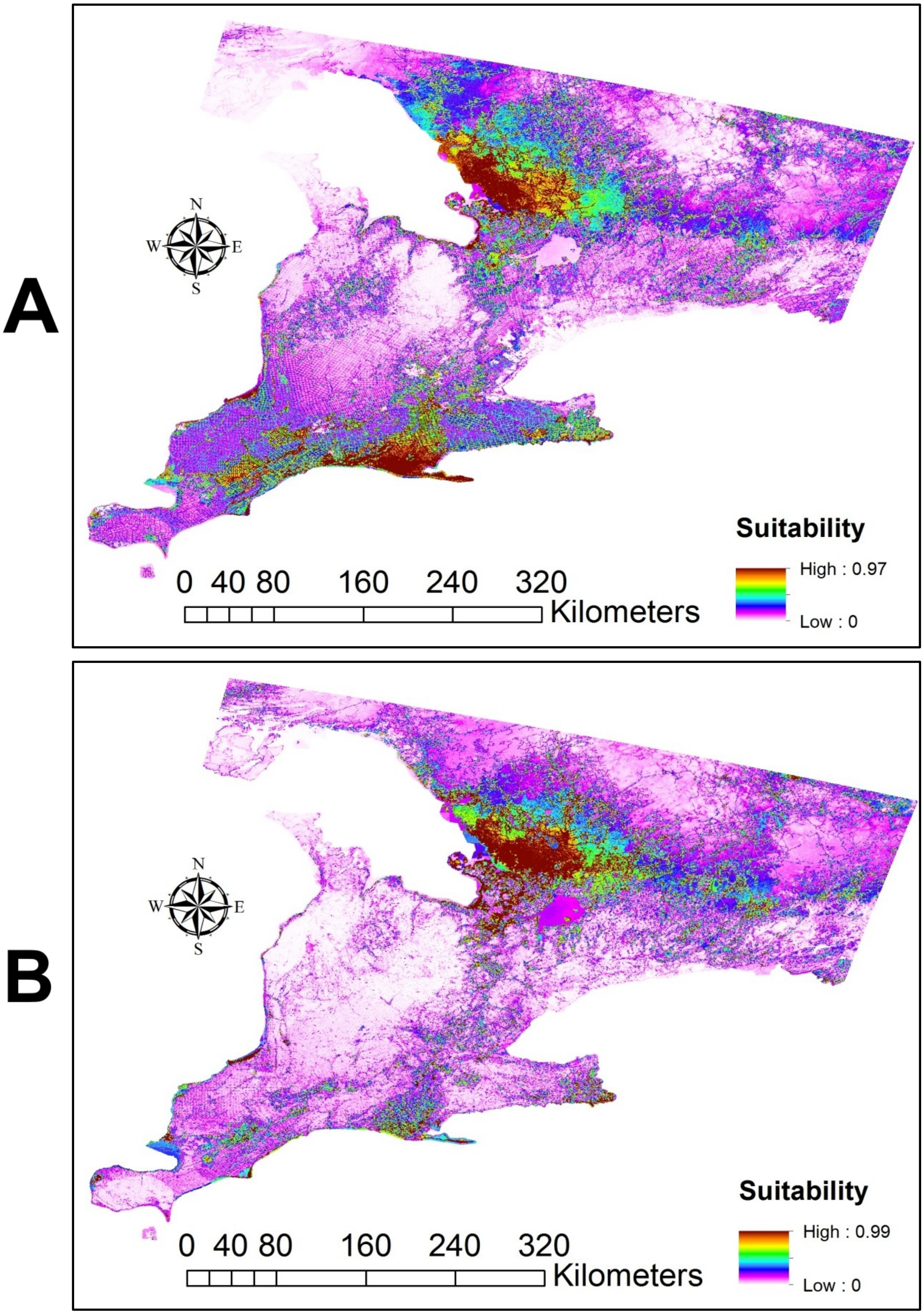


Figure 1-10. Estimated distribution of the eastern hog-nosed snake in Canada, using the consensus approach at a resolution of 500m: (**A**) historically and (**B**) currently.

Results for climatic models

I built models using climatic variables only to assess whether the distribution of this species could have been shaped partly by climate. Models performed well with AUCs ranging between 0.85 and 0.91 (Table 1-7). The predictions of both models (one using all hognose occurrences and the other using only recent ones) were similar which was expected considering that climatic suitability should have remained mostly the same throughout the past century and that the climatic variables used were identical in both models (Figure 1-13). The areas of high suitability were also similar to those found in historical models (Figure 1-8, 1-9 and 1-10). On the two maps built with climatic models, eastern hog-nosed snakes can be found in most high suitability regions, which indicate that climate could be in part responsible for the distribution of this reptile in Ontario. In order of importance, the five variables with the highest contributions in Maxent models built with only climatic variables were: (1) precipitation of the driest month, (2) minimum temperature of the coldest month, (3) isothermality, (4) mean temperature of the wettest quarter, and (5) maximum temperature of the warmest month (Figure 1-2). However, the contributions of each variable vary a lot from one model to the other, suggesting that the algorithm is flexible enough to overfit the dataset and always appear to make good predictions. Overfitting is also supported by the fact that the only variables presenting nice marginal response curves are bio6, the minimum temperature of the coldest month, and bio3, isothermality (Figure 1-11). The response curves of most other variables are usually broken, showing no specific trend in the climatic conditions required by the species.

Table 1-7. Metrics of performance for both climatic models

Model	Occurrences	Resolution	No. var	Int. AUC	Ext. AUC	Threshold	Sensisivity	Specificity	Kappa	TSS	MPA
Maxent	All	1000m	10	0.85	0.88	0.359	0.81	0.81	0.13	0.62	0.172
Maxent	Recent only	1000m	10	0.85	0.91	0.328	0.91	0.79	0.076	0.7	0.212

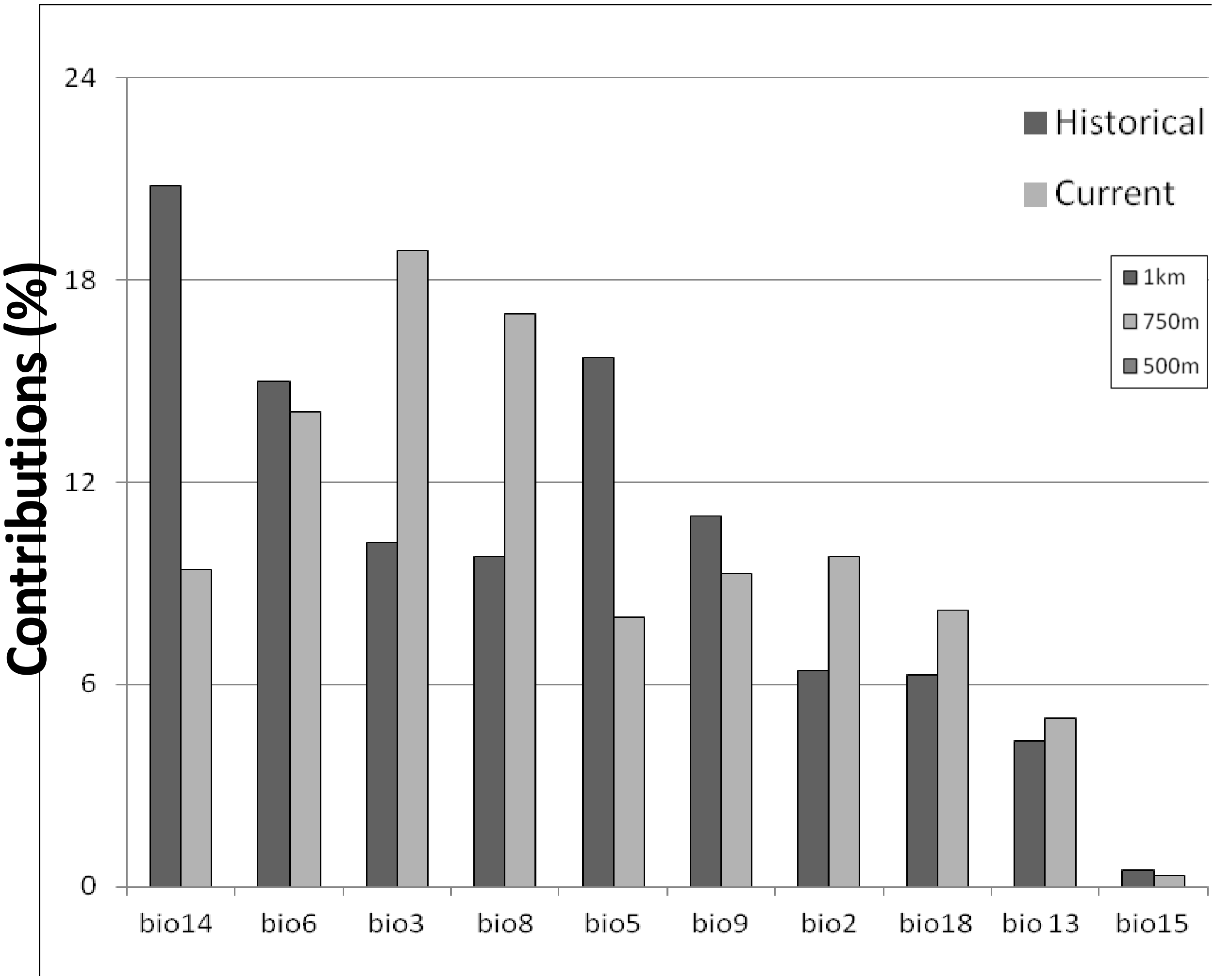


Figure 1-11. Contributions of all variables in habitat suitability models built with Maxent, with strictly climatic variables, for the eastern hog-nosed snake in Ontario. The historical model was built with all 254 occurrences and the current model was built with 126 occurrences collected since 1990.

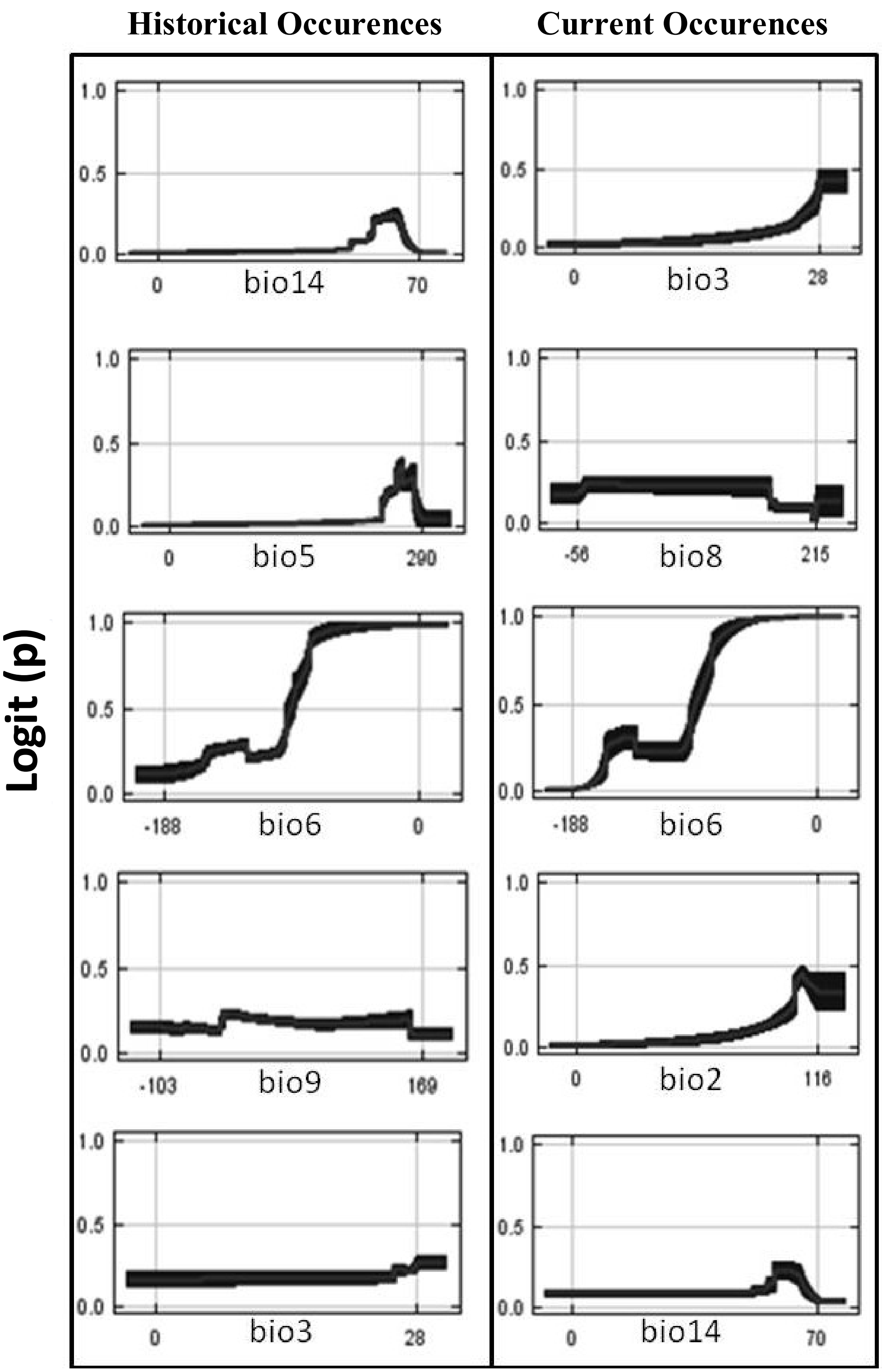


Figure 1-12. Marginal response curves ± SD, for the five most influential variables in the two climatic models built with Maxent, for the eastern hog-nosed snake in Ontario. The curves show how the logistic prediction (p) varies, when one variable is gradually changed, while all others are kept at their average value. The historical model was built with all 254 occurrences and the current model was built with 126 occurrences collected since 1990.

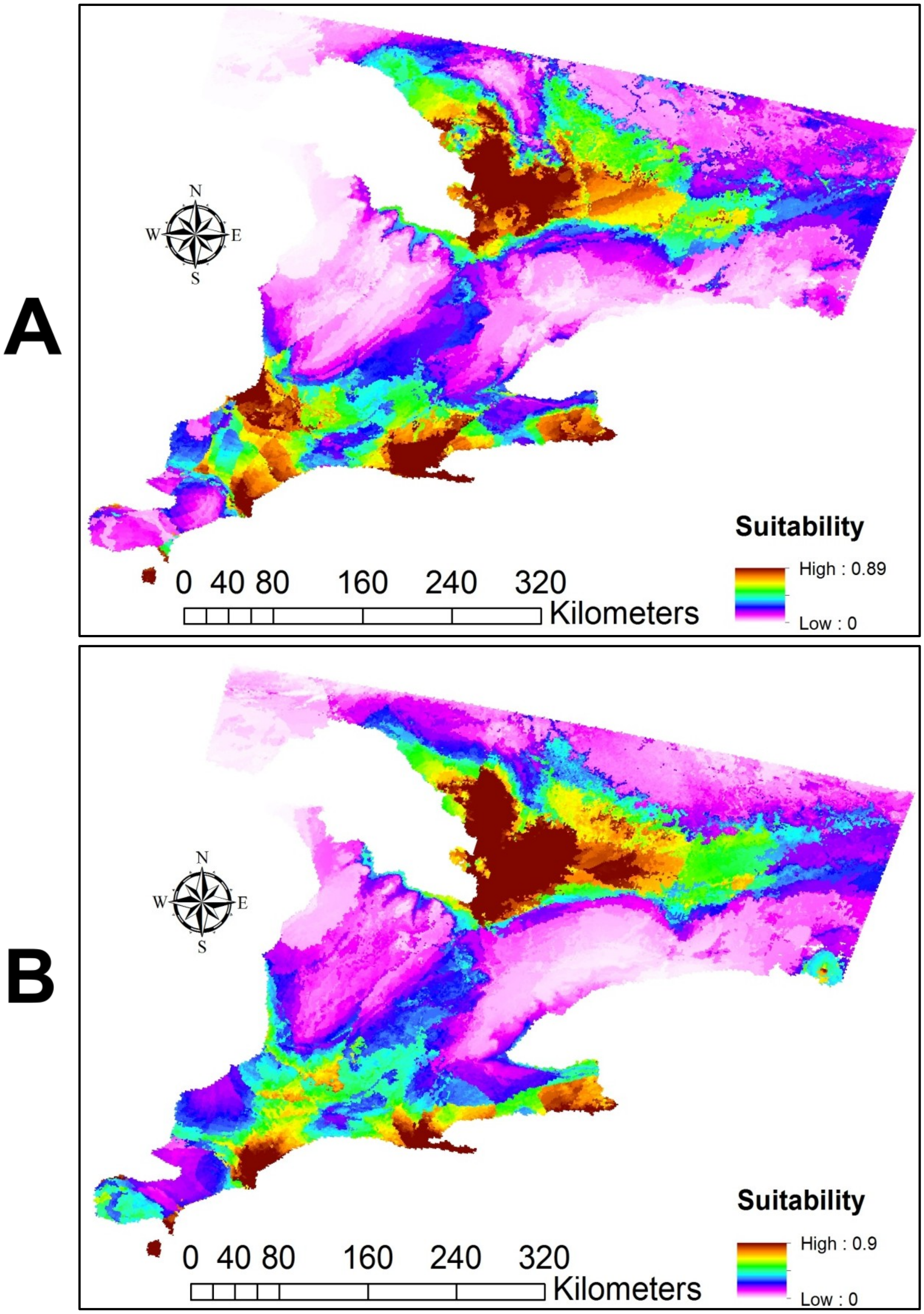


Figure 1-13. Habitat suitability maps created with Maxent, with strictly climatic variables, for the eastern hog-nosed snake in Ontario. The historical model (**A**) was built with all 254 occurrences and the current model (**B**) was built with 126 occurrences collected since 1990.

Discussion

A hierarchical framework to study the decline of species

A number of factors are responsible for shaping the distribution of species including historical events, climate, biotic interactions, and anthropogenic activity. Because these variables are numerous and act on species at different spatial scales, a single model could only approximate habitat suitability. In fact, some factors such as historical events and colonization are almost never considered in habitat suitability models. In this study, I used a multi-scale hierarchical framework to reconstruct the historical and current distribution of the eastern hog-nosed snake. I also identified which parameters are responsible for shaping the distribution of this species in the northern part of its distribution. Hierarchical techniques have been used to describe ecological systems for more than 20 years (e.g. Urban et al., 1987), but have been rarely applied to habitat suitability models (Anadón et al., 2007; Pearson et al., 2004). Yet they present two important advantages: the variables are incorporated to the models in a biologically meaningful manner and the issue of scale is solved because each model can be built at different resolutions (Anadón et al., 2007). A hierarchical approach was applied in this study for two additional reasons. First, the hog-nosed snakes in Ontario are at the most northern part of their distribution. Climate has probably played a significant role in shaping their historical distribution. Also, land cover variables were unavailable to build historical models. I circumvented this issue by approximating how habitat suitability could have looked like historically without considering land cover. However, it is important to note that these models are not directly comparable as they each use different sets of variables.

The use of categorical variables in HSMs

While most habitat suitability models were shown to accept categorical variables except GARP (Elith and Graham, 2009), our results suggest that Maxent and BRTs differed significantly in the way they used the geological categorical variable in our models. While that categorical variable was used successfully by Maxent, it is difficult to say if it was with BRTs. Was the algorithm unable to use that variable because it was categorical or were the contributions of that variable low in BRTs because other variables were more important? In the original study explaining how to use BRTs in habitat suitability models, the authors included a categorical variable composed of five classes in their models (Elith et al., 2008). One of the reasons why the geological variable had low contributions in our BRT models could have been because it had too many classes (eight). This high number of classes also made it difficult to interpret response curves and added noise to our results. Because it is difficult to evaluate presence-only habitat suitability models, studies comparing these models should stay away from categorical variables to avoid discrediting algorithms unable to process categorical variables properly. In the future, I would suggest using categorical variables only when needed, limiting the number of classes as much as possible.

The number of explanatory variables

The number of variables required by models to make good predictions, or at least to obtain good performance scores, was found to be highly dependent on the algorithm. In models predicting the current distribution of the eastern hog-nosed snake, Maxent models required fewer variables to achieve high AUC values than BRTs and GARP. These differences can be explained by the fact that these three models differ in the type of formulas they use to make predictions. Although eliminating variables by optimizing AUC values is novel in habitat suitability model, similar results have been found in the past (Millar and Blouin-Demers, 2011). Although studies discuss how explanatory variables should be chosen to increase the performance and generality of HSMs (Guisan and Zimmermann, 2000; Austin, 2007), I was unable to find any information on how the number of explanatory variable employed affects the performance of models. Some habitat suitability models, such as GLM, were shown to easily overfit their training datasets (Randin et al., 2006) and I believe this may also be the case with BRTs considering it performed very well, but predicted few areas of high suitability. It would be interesting to use these models on separate datasets to test the cross-validation abilities of BRTs. Because some models are prone to overfitting, I suggest assessing how the number of explanatory variables affects the performance and the generality of habitat suitability models. A recursive feature elimination approach similar to what has been used in this study could not only allow answering this question, but help select only the most relevant predictors.

Low kappa scores

While TSS scores were high, kappa scores were low, only slightly above zero. Compared to both historical and climatic models, I also noted lower kappa scores in current models. Cohen's kappa is often used to evaluate presence-absence predictions because it is simple, accounts for both commission and omission errors, and is tolerant to zero values in the confusion matrix (Manel et al., 2001). However, it was found to be affected by prevalence (Allouche et al., 2006; Ridenour and Heath, 1999), which is a serious limitation when dealing with rare species or any small dataset. Kappa responds to prevalence in a unimodal fashion, being at its highest (1) when prevalence is at mid point (Allouche et al., 2006). The low kappa scores in most of our models could be attributed to the small size of our dataset, particularly in current models (126 occurrences), and to the low prevalence of this species. As for climatic models, they simply did not perform as well as historical models, probably because the distribution of eastern hog-nosed snakes is in part shaped by variables other than climate (e.g. anthropogenic activity). Cohen's kappa was not originally formulated to deal with prevalence; it was used to assess the reliability in agreement between two observers in a medical context (Cohen, 1960). I encourage the use of TSSs to evaluate habitat suitability models because it is not affected by low prevalence and designed specifically to assess the performance of HSMs (Allouche et al., 2006).

Resolution

Different conditions or resources are selected by living organisms at different scales (Johnson, 1980). It is important to try to identify these scales to be able to use these variables and detect a response in habitat selection. In all habitat suitability models currently available, all predictors must be at the same resolution. This assumes the species responds to all variables at the same scale, which is not realistic and can prevent us from detecting differences from one resolution to another. I built models at three resolutions hoping to identify at what resolution models performed better, but the outcome was equivocal. Some of the reasons why the differences in resolution did not seem to influence the performance of models could be because the original resolutions of the variables and the occurrences were always the same, or because the metrics of performance employed are not significantly affected by these changes. The effect of grain size on habitat suitability models has been studied before, but with few conclusive results. Depending on conditions, coarsening grain size can either increase or reduce the performance of HSMs (Guisan et al., 2007b). Yet, aggregating variables will eventually lead to a loss of information that will always reduce the performance of models passed a given threshold. Changes in resolutions were found to have the highest effect on models built with highly accurate datasets (Guisan et al., 2007a). Another possibility why I was unable to detect differences could be because the resolutions chosen did not differ enough. For conservation application, I suggest referring to maps at all three resolutions as I was unable to detect differences in performances between resolutions.

Model performance

According to all validation metrics employed, habitat suitability models built with Maxent and BRTs outperformed GARP which is consistent with some studies (Elith et al., 2006), but not with others (Terribile et al., 2010). Most studies employing habitat suitability models only use ROC analyses for validation, a metric which has been heavily criticized (Lobo et al., 2008; Peterson et al., 2008). In a recent study comparing Maxent and GARP, both models were found to achieve similar AUCs, but the first one avoiding commission errors and the second avoiding omission errors (Peterson et al., 2007). That same study shows that GARP can be better than Maxent in making predictions in unsampled regions, something I was unable to verify with our methodology. While GARP may be more transferable, Maxent is good at reconstructing a species' distribution in specific regions (Peterson et al., 2007). For our purpose, the main disadvantage with GARP was that it predicted high probabilities in most of the area of study, making these maps less useful for conservation. If all the region of study was highly suitable, one would expect more occurrences in remaining forested areas such as the Bruce Peninsula (see Figure 4-9.). I also noticed great variability in the predictions made by this algorithm between runs and resolutions. In addition to achieving good performance scores, Maxent was also favoured because it retains its high predictive power even when the sample size is small (Pearson et al., 2007), a good quality when dealing with rare species. Finally, there has been only one study comparing BRTs to other habitat suitability models (Elith et al., 2006), which I believe is not enough to assess the performance of a model, especially considering that models were evaluated using mostly ROC analyses. Because each habitat suitability model has its own advantages and disadvantages, I suggest carefully choosing which one to use taking into consideration the goals of the study. I also encourage other studies to compare habitat suitability models using metrics of validation other than ROC analyses. Because HSMs are difficult to evaluate, validation should always be done with several techniques and models that are judged "equivalent" should be combined using a consensus approach (Marmion et al., 2009). Because of the nature of habitat suitability models, it very difficult to determine to what extent they overfit data. HSMs are useful if they are robust and allow making predictions in other regions or at different time periods (Guisan and Thuiller, 2005). Yet, there are no absolute measures of robustness and it is rarely assessed in recent HSMs studies. It would be important to test the transferability of these models by trying to make predictions in other regions (Araújo and Guisan, 2006). In my case, for instance, it would be worthy to try predicting the distribution of the eastern hog-nosed snake in New England with models based on data from Ontario.

Limitations

Our species occurrences were mostly collected by the public with no specific sampling design. Differences in sampling design can influence the performance of models by disproportionally accounting for environmental conditions found in sampled areas only (Zimmermann et al., 2010). A study shows that museum records for dung beetles in Madrid do not account for all conditions tolerated by the species due to spatial bias (Hortal et al., 2008). Ideally, the sampling effort should have been the same for the entire area of study to cover all habitats and landscapes equally (Newbold, 2010). Yet, sampling an area the size of the distribution of a species is almost impossible, particularly for a species as elusive as the eastern hog-nosed snake. I advise governmental agencies to sample suitable areas that are more remote and less accessible. This is probably the best way to find unknown populations and add valuable occurrences to atlases. Because true absences were unavailable, I used 10 000 pseudo-absences to build and validate all models. These pseudo-absences may include areas where the species is present, which will have a direct impact on both the building and the validation of the models. As described by Jiménez-Valverde (2012), when using the AUC to validate HSMs built with peudo-absences, the maximum achievable score is no longer 1 and is inversely related to the area of distribution. In the same way, the number of pseudo-absences used will influence the AUC because the more pseudo-absences are predicted as absences, the higher the AUC. The interpretation of the AUC is therefore no longer simple, and depends on the modeling design.

Another limitation of this project is the extent of the study. Here I focus on the eastern hog-nosed snakes in Ontario while this species ranges all the way to Texas. Our models are built with occurrences accounting for the conditions found in Ontario only. For example, this species is definitely able to cope with warmer temperatures, but it is unaccounted for by our models. The occurrences used in this study do not represent the entire gradient of conditions that can be tolerated by the species. The factors limiting a species' distributions are also different at edges as conditions are sometimes suboptimal (Braunisch et al., 2008). A future study looking at the range of the eastern hog-nosed snake in the United-States might help identifying more accurately the factors limiting habitat suitability and increase the transferability of habitat suitability models. Within Canada, the northern and southern populations of hognose may also experience different environmental conditions and utilize different habitats. Depending on how and when the species invaded Canada, these two populations may be genetically distinct. To better understand the conditions required by these two populations, it may be necessary to study habitat suitability and selection at a finer scale using other modelling techniques (e.g. Wu and Smeins, 2000).

Studying rare or threatened species is also difficult because we tend to study the habitat a species is left with and not all habitats that were historically available and used. My models were built with occurrences that are currently available and take into consideration recent environmental conditions; they do not account for the possible historical distribution of the species. It is possible that the species was more widespread two hundred years ago and that other important factors shaped its distribution. The variables that seem important in shaping the distribution of a species today could be very different from the variables that dictated the distribution historically. I also assume the species has reached equilibrium and do not consider population source-sink dynamics. Some of the occurrences I used might represent conditions that are too poor, under which the species will eventually be extirpated. Conservation efforts should be prioritized in areas representing population sources.

To better assess habitat loss for this species, remote sensing data spanning a longer period of time would have been required, but are currently unavailable. The models predicting the historical and current distribution of the eastern hog-nosed snake were not built using the same variables because historical land cover data are not currently available in readily available digital format. As a result, habitat loss cannot be determined precisely because models are not directly comparable. I circumvented this limitation by applying a hierarchical scheme using different variables to predict the climatic, historical, and current distribution of the eastern hog-nosed snake. Similar hierarchical methods were used before (Anadón et al., 2007; Pearson et al., 2004) and may be the only way to use habitat suitability models to assess habitat loss until sufficient historical GIS data become available. Finally, I used pseudo-absences instead of real absences, which could include a number of false absences that will affect the performance of the models. The impact of these false-absences on the metrics of accuracy will depend on how many there are in relation to the number of true presences and true absences.

Conservation implications

The models built with Maxent and BRTs are useful for conservation purposes not only because they produce maps that are easily interpretable, but also response curves indicating how the probability of suitability varies along with the variables in each model. These response curves can help determine what limits habitat suitability and what actions should be taken to manage or protect the habitat of species at risk. Although the geological categorical variable I used had too many classes making it difficult to interpret, hog-nosed snakes that were found in the south of the province were often on sand while individuals found near Georgian Bay were on precambrian bedrock (APPENDIX IV). Under a temperate climate, one important factor shaping the distribution of reptiles is winter conditions. For example, the availability of hibernacula was suggested to limit the distribution of ratsnakes in Ontario (Gregory, 1984). It is possible that eastern hog-nosed snakes can survive in the area of Georgian Bay by hibernating in crevasses in bedrock that are deep enough to avoid frost. In the south of the province, these snakes do not need to go as deep and can simply dig their own hibernacula in sand (Robson, 2011). Alternatively, sand could be selected at a finer scale and not be noticeable when looking at GIS maps of surficial geology. Although bedrock is the primary material forming surficial geology around Georgian Bay, these snakes may be able to find enough sand within their home range to fulfill all biological needs. I suggest protecting areas where sand is abundant as this species is often found in sandy areas (Robson, 2011; Lagory et al., 2009; Cunnington and Cebek, 2005), particularly in the south of the province.

Based on historical models, the eastern hog-nosed snakes found in Ontario were often near roads or in populated rural areas. The response curves of ROAD and PEOP could be explained by the fact that most occurrences used in this study were retrieved opportunistically by the public, often near roads, small towns, and parks. These findings are the result of poor sampling with a spatial bias for areas where people are most likely going to reports sightings. This would also explain why probabilities of suitability usually drop down beyond a given threshold of either road density or population density. These results should not be interpreted as a preference for roads. In fact, road mortality was found to be a serious threat for eastern hog-nosed snakes as they are vagile and active foragers (Rouse et al., 2011). To better identify the conditions where this species is found in Ontario and have access to better data, I encourage building a strategic sampling design that will aim at sampling different habitats, both disturbed and undisturbed, equally. The habitat suitability maps produced in this study could help develop this sampling design by indicating which regions and habitats should targeted. This sampling design would produce a stronger dataset that could identify more precisely where this species is declining and could lead in the development of stronger habitat suitability models. Habitat suitability models were found to perform better with systematic sampling and larger sample size (Hirzel and Guisan, 2002).

Cropland density reduced habitat suitability in Maxent models predicting the current distribution of the eastern hog-nosed snake, which could explain why current models are always less suitable in the area of Long Point and Point Pelee. Most of southern Ontario is heavily used for agriculture, which could suggest habitat loss in these regions. Agricultural activity is also destructive for amphibians which may further affect this snake by limiting food availability. Agriculture is known to be a driving cause of species endangerment in Canada, particularly in southern Ontario (Kerr and Cihlar, 2004). In BRT models predicting the current distribution of the eastern hog-nosed snake, forest density increased habitat suitability which makes sense considering that most hog-nosed were found in forested areas, both in the south and the north of the Province. The reason why it was either cropland density or forest density that contributed the most to current models is because these two variables were the most correlated ($r = -0.71$). The conclusion that should be drawn from current models is that reduced forest availability limits habitat suitability for this species. Today, the few pockets of individuals found in southern Ontario are isolated in protected areas where forest remains.

Average precipitation and maximum temperature during the active season both seem to limit the distribution of this species. Being at the most northern extreme of their distribution, the eastern hog-nosed snakes found in Ontario could be limited by heat availability. At similar latitudes, this snake was rarely found to take cover under rocks or logs, which could also suggest that thermoregulation is an important factor limiting the distribution of this species (Plummer and Mills, 2010). In the north, the season is short and weather conditions must be good enough for the species to survive and propagate. As for increased precipitations, it may favour amphibian populations which can in turn favour their predators including the eastern hog-nosed snake. According to both climatic models, the south and the north populations could be experiencing similar climatic conditions, enough for Maxent to make relatively good predictions. However, these results should be interpreted with care as the contributions of each variable were different in the two models. Yet, there seems to be a gap between the two populations that would be unsuitable climatically for this species. The climatic conditions found along the Great Lakes could be different from the ones found more inland. Minimum temperature during winter could limit hibernation opportunities. Although land cover definitely plays a role in habitat suitability for this species, our models suggest that its distribution has been shaped at least in part by climate.

Using habitat suitability models to protect species at risk

One goal of this study is to use habitat suitability models as a tool for conservation to help identify areas worth protecting for hog-nosed snakes. Ideally, these areas should correspond to critical habitat as defined by the *Species at Risk Act* or the *Endangered Species Act,* to easily apply current legislations, but there is no consensus on how to do this.

I propose using three thresholds to identify the areas most suitable for a species. The categorical maps produced with these thresholds are tools that can be easily used by governmental agencies to know what land to prioritize in conservation efforts. I built models in a hierarchical process to better understand the complexity behind the distribution of a species at risk. Models were built using three algorithms and various explanatory variables to assess where the species could have declined. To identify areas worth protecting, I suggest applying the following three-step approach in a hierarchical process: (1) Use habitat suitability maps to identify areas that are climatically suitable in regions where conservation is possible, (2) use the consensus maps of models predicting the current distribution of the species to identify areas of high suitability within the regions that are climatically suitable, and (3) ground truth any of these previous areas to make sure the species is present. This approach allows identifying areas that are not only occupied by the species, but climatically suitable and still suitable today after considering other variables such as human activity. Similar hierarchical approaches have been used successfully in the past (Pearson et al., 2004; Anadón et al., 2007) and may offer a simple and efficient comprise on how to use habitat suitability models in conservation.

CHAPTER II

Assessing the importance biotic predictors in habitat suitability models

Abstract

Habitat suitability models have proven to be a useful tool in many fields of biology, including conservation, where predictions allow identifying conditions required by species at risk. Yet, an important aspect that remains to be studied more in depth, and that can affect the predictions of these models, is the use of biotic variables. The presence of a species at a given location depends on a number of factors, including the right combination of biotic and abiotic conditions. Most studies currently published employing habitat suitability models use mostly abiotic explanatory variables such as elevation and climate. While this may be enough to predict the fundamental niche of some species, the use of biotic variables may be essential to predict the realized distribution of species having specialized ecological niches.

In this study, I assess the importance of biotic variables in habitat suitability models using a species considered a specialist in terms of diet, the eastern hog-nosed snake, *Heterodon platirhinos*. I use the Maxent models predicting the historical and current distribution of this species in Ontario developed in Chapter I, and add extra biotic variables representing food resources: toad occurrences and toad calls. All models are evaluated with the same metrics as in Chapter I, calculating the change in each one of them. To confirm that the differences in performance are the result of a limitation in resources, I repeat the analyses substituting toad occurrences by snake occurrences. Both the toad and snake variables were found to have little effect on the metrics of accuracy, but in some cases had high contributions in the models. Similar results between the two analyses suggest that these variables helped building the models by representing areas that were sampled, in accessible regions near roads and small towns. These variables are therefore accounting for the sampling bias and not a limitation in food resources.

Introduction

The importance of biotic variables in habitat suitability models

Habitat suitability models have been used extensively in the past decade to answer a number of questions in various fields of biology. They have been used to: study the effects of climate change on species distribution, find unknown population, identify the conditions required by species at risk, and study invasive species. (Peterson, 2006). Several models have been developed using both different algorithms and statistical techniques. Their gain in popularity can be attributed to the fact that some use presences only, data that are often available in atlases. These models have been evaluated and compared in many studies, but they remain difficult to classify in order of performance (Elith and Graham, 2009). To help make better predictions and increase the accuracy of these models, several aspects related to habitat suitability modeling have been studied such as sample size and resolution. For example, some models, such as Maxent, were found to retain their predictive power even when the sample size is small (Wisz et al., 2008). One important limitation that has been identified more than a decade ago (Guisan and Zimmermann, 2000), but that has yet to be investigated in depth is the omission of biotic interactions. Many variables contribute in shaping the distribution of species, including biotic and abiotic factors (Cunningham et al., 2009). With respect to abiotic factors, climate is considered a major driver of species distributions (MacArthur, 1958) influencing species both directly and indirectly. The mean fitness of a species may decrease at the extremities of its distribution because it is physiologically restrained by climatic conditions. Under these conditions, competitor species may be more abundant which will further affect the fitness and distribution of this initial species (Gross and Price, 2000). The variables most often used in habitat suitability models are climatic because data are available and they affect species directly. However, other important factors such as biotic interactions and historical events are also responsible for shaping realized distributions, acting independently on species.

HSMs are a direct operational application of the ecological niche theory (Hirzel and Le Lay, 2008). For this reason, it is important to define this concept to know how these models should be built. Habitat suitability models have taken a variety of names such as "predictive habitat distribution modeling" (Guisan and Zimmermann, 2000), "bioclimatic envelope modeling" (Pearson and Dawson, 2003), and "niche modeling" (Stockwell, 2007). Most of these models could be argued to be associated with Hutchinson's (1957) definition of multidimensional niche (Guisan et al., 2006). However, this niche is further divided into the fundamental niche and realized niche (Hutchinson, 1957). While the fundamental niche can be defined by environmental conditions only, the realized niche also incorporates biotic processes such as dispersal and competition (Miller, 2010). Yet, defining precisely each of these niches has been shown to be difficult and has lead to contradictory interpretations (Araújo and Guisan, 2006). In relation to HSMs, these two notions of niche incorporate biotic and abiotic interactions that are rarely taken into consideration in distribution modeling. This is why most models looking at the distribution of species do not model niches per se, but the suitable environments through space (Miller, 2010; Pearson, 2007).

For most applications, including conservation, it could be argued that the realized ecological niche is the one that matters because what is sought after is where a species is found, ideally after considering all variables affecting its distribution. Most current validation techniques also require some form of absences, hence, defining the realized distribution. For those reasons, any biotic interaction significantly affecting the distribution of a species should be included in habitat suitability models. Along an environmental gradient, species can be argued to be stressed biologically in one direction and physically in the other (Brown et al., 1996; Guisan et al., 1998; MacArthur, 1972). The BAM diagram (Soberon, 2007) is useful to visualize why incorporating biotic interactions in habitat suitability can be important (Figure 2-1). “A” represents regions where scenopoetic conditions are adequate and “B” regions where biological conditions allow the presence of viable populations. Solid circles represent source populations, open triangles sink populations due to competitive exclusion, and open squares sink populations due to negative intrinsic growth rates. Open circles are sink populations due to both of the above. Incorporating biotic variables in habitat suitability models could help predict potential areas

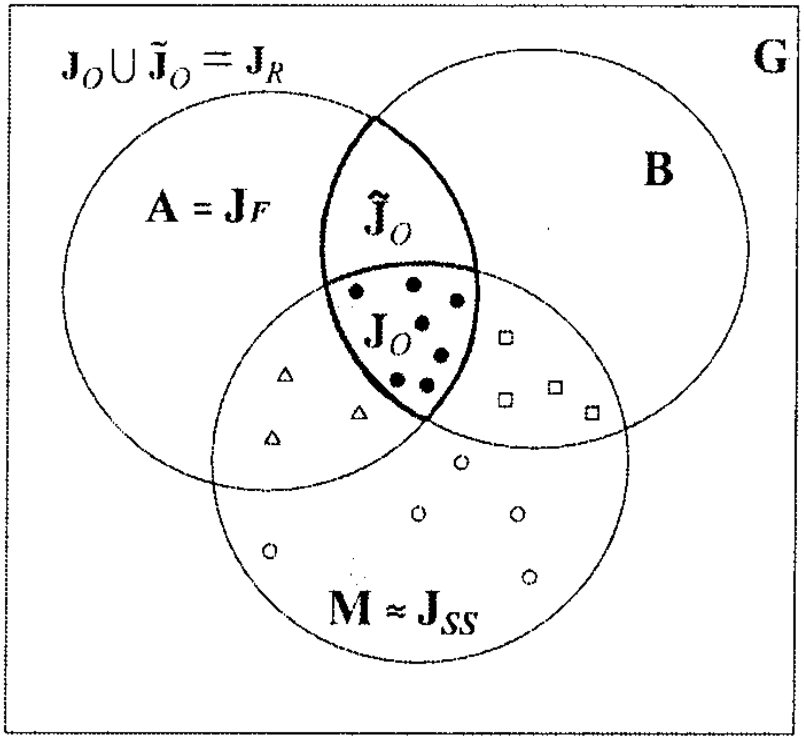


Figure 2-1. The BAM diagram (Soberon, 2007).

that could be occupied by the species because both biological and scenopoetic conditions are suitable (Soberon, 2007).

Variables reflecting biotic interactions are sometimes available, but rarely used in HSMs (Guisan and Thuiller, 2005). The biotic variables most often used in habitat suitability models are land cover variables derived from satellite images such as the density of agricultural areas (Lahoz-Monfort et al., 2010; Loiselle et al., 2010). Occurrences of other competing species have also been used successfully (Santos et al., 2006; Cunningham et al., 2009). The few studies using these kinds of variables suggest that realized distributions are affected by biotic interactions even at a coarse resolution, suggesting that this subject should be studied more in depth (Guisan and Thuiller, 2005). Most habitat suitability models currently rely on the assumption that biotic interactions do not influence the range of species, which is clearly incorrect. It is crucial to study the role of biotic variables in habitat suitability models to help make reliable predictions that will be useful in the context of conservation.

The eastern hog-nosed snake, *Heterodon platirhinos,* is a semi-fossorial reptile native to both the United-States and Canada, known to occupy a variety of habitats, particularly open upland areas. This species is particularly well suited to study the effect of biotic interactions in habitat suitability models because it has a specialized diet, feeding mostly on toads (Platt, 1969; Uhler et al., 1939; Edgren, 1955). Toads were found to constitute between 40% (Uhler et al., 1939) and 75% (Surface, 1906) of their diet. The distribution of this snake could therefore be limited in part to areas where toad populations are high enough to support viable snake populations. In fact, it was proposed that a decline in eastern hog-nosed snakes in Hog Island, Virginia, could have been related to a decline in toad abundance (Conant et al., 1990).

Objectives

The goal of this study is to answer the following question: (1) Can biotic interaction variables improve the predictive power of habitat suitability models? I assess the role of biotic interactions in habitat suitability models predicting the historical and current distribution of the eastern hog-nosed snake in Ontario. I hypothesize that prey availability restrains the distribution of the eastern hog-nosed snake to some areas only, preventing it from occupying its entire potential distribution. I use the Maxent models built in Chapter I and add additional variables representing food resources, toad occurrences and toad calls, to determine whether prey abundance improve model fit. If prey abundance limits the distribution of hognose snake, then HSMs incorporating prey distribution should perform better than those that do not. All models are evaluated with the same metrics as in Chapter I, calculating the changes in each one of them. To confirm that model fit improvements are due to a limitation in food resources and not the potential effect of unequal sampling on HSMs, I then substitute toad occurrences by snake occurrences. For all models, changes in performance are calculated and the contributions of all biotic variables are determined. Finally, I discuss why these variables may have helped building models and why it may be difficult to study biotic interactions in the context of habitat suitability models.

Methodology

Additional biotic variables

The first step was to build variables that would represent as much as possible a limiting resource for the eastern hog-nosed snake. In Ontario, this snake consumes both american toads, *Anaxyrus americanus*, and fowler's toads, *Anaxyrus fowleri* (Platt, 1969). I built two types of variables with theses two species: the presence of toads and the relative abundance of toads (Table 2-1). I used toad occurrences from both the Ontario Reptile and Amphibian Atlas (OntarioNature, 2011) and the Ontario Herpetofaunal Summary Atlas (Oldham, 2000) to build two variables representing toad presence. The variable TOAD was built with all existing toad occurrences in Ontario having an accuracy of 100m or less, and represents the historical distribution of toads. T1990 was built with toad occurrences collected since 1990 only, and represents the current distribution of toads. For these two categorical variables, presences were assigned the number 1 while the rest of the territory was assigned 0. Toad calls from the Marsh Monitoring Program were used to build the variable CALL representing the relative abundance of toads (BirdStudiesCanada, 2008). That variable represents the average number of toads that were heard at each survey station since the beginning of the program in 1995. This variable was therefore continuous and a full chorus was assigned the number 25. It is important to note that all of these variables represent points in space that were converted into rasters to be compatible with habitat suitability models.

To confirm that the eastern hog-nosed snake responds to an actual limitation in resources and not differences between these models and those presented in Chapter 1, I also built variables with the occurrences of four snake species: *Nerodia sipedon, Opheodrys* *vernalis, Storeria dekayi,* and *Thamnophis sirtalis* (OntarioNature, 2011). I made sure to select species that exploit different habitats to represent various areas in the region of study. The variable SNAK included all occurrences of these four species with an accuracy of 100 m or less while S1990 included occurrences collected since 1990 only (Table 2-1). Finally, all variables were built at three resolutions (1 km, 750 m, and 500 m) to be compatible with previous models.

Table 2-1. Additional variables used to predict the historical and current distribution of the eastern hog-nosed snake in Ontario.

Organisms	Code	Description and Units	Used for model	Source of original data
Toads	**TOAD**	All toad occurrences	historical	(OntarioNature, 2011) (Oldham, 2000)
	T1990	Toad occurrences since 1990	current	"
	CALL	Toad calls since 1995	current	(BirdStudiesCanada, 2008)
Snakes	**SNAK**	All snake occurrences	historical	(OntarioNature, 2011)
	S1990	Snake occurrences since 1990	current	"

Modeling approach

I added each of these variables one at the time in the pruned Maxent models developed in Chapter I. Five Maxent models were therefore built, each one containing an additional biotic variable. Two models predicting the historical distribution of the eastern hog-nosed snake were built: one using all toad occurrences (TOAD) and the other all snake occurrences (SNAK). Three models predicting the current distribution of the snake were built: one using recent toad occurrences (T1990), one toad calls (CALL), and the last recent snake occurrences (S1990). I always added these biotic variables to the pruned models developed in Chapter I that optimized the AUC and used the same sets of occurrences and pseudo-absences. All models were built with Maxent version 3.3.3e (Phillips et al., 2006) with the same settings and variable datasets as in Chapter I, at three spatial scales (1 km, 750 m, and 500 m). Each model consisted of 10 random replicates always setting aside 25% of the occurrences and pseudo-absences for validation. Finally, all maps built with Maxent were produced in logistic format that represents an estimate of the probability of presence (Phillips et al., 2006).

Validating Models

To compare the models produced in Chapter I with the models produced here containing biotic variables, all models were validated with the same metrics of performance as in Chapter I, recording the changes in each one of them. Presence-only habitat suitability models are difficult to evaluate (Elith and Graham, 2009) and it was proposed that no single validation technique should be used alone (Miller, 2010). This is why I used a number of different techniques, some of which are threshold-dependent and others which are threshold-independent.

Results

The changes in the metrics of performance

All changes in the metrics of performance were very subtle (Table 2-2). Values only rarely varied more than two or three hundredths of a point and many remained unchanged. Out of the three series of models built with toad variables, the only one that seemed to make better predictions is the one including the variable built with all toad occurrences (TOAD). For those historical models, all internal AUCs were increased by 0.009 on average. For those same models, the MPA was reduced by 1.3% on average. Some of these models have also increased Kappa and TSS scores. It is important to note that the model that has improved the most is the one built at the smallest resolution, 500 m. The two other series of models built with toad variables (T1990 and CALL) seemed to perform more poorly as their internal AUCs, Kappas, and TSSs decreased and their MPAs increased. Finally, the changes in sensitivity and specificity varied in both directions making it impossible to notice any trends.

For the models built with snake variables, only the historical models built with all snake occurrences (SNAK) seemed to perform better (Table 2-3). Their internal AUCs increased by 0.011 on average and their MPA decreased by 2.5% on average. Again, it is the model built at a resolution of 500 m that has improved the most. The models predicting the current distribution of the eastern hog-nosed snake built with the variables representing recent snake occurrences (S1990) were hardly improved as their internal AUC increased by only 0.001 point and their MPA decreased by 0.6%. All other metrics did not change considerably or remained unchanged.

The contributions of each biotic variable in the models

The contributions of the biotic variables in each model were highly variable, but were usually higher in models predicting the historical distribution of the eastern hog-nosed snake. Across resolutions, the contribution of the variable TOAD in historical models was on average 8.3%. It was the highest at a resolution of 1 km where it reached 13.1% (Figure 2-2). In models predicting the current distribution of the eastern hog-nosed snake, the contributions of both T1990 and CALL were both low at all resolutions. Across resolutions, the contribution of T1990 was on average 1.5% (3.5% at a resolution of 1 km). Finally the lowest contributions were given to the variables CALL, with only 1.1% on average across resolutions. In all models, the order of importance of the variables remained very similar between the models built in Chapter I and the models built here comprising an additional biotic variable (Figures 2-2, 2-3 and 2-4).

Of all models, the biotic variable that reached the highest contribution was SNAK in models predicting the historical distribution of the eastern hog-nosed snake. Across resolution, this variable had a contribution of 21.3% on average (Figure 2-4). In current models, the variable S1990 had an average contribution of 12.6% across resolutions. At a resolution of 1 km, it was the most important variable in the models with a contribution of 19% (Figure 2-4). These results show that biotic variables representing the distribution of other snake species could contribute more to models predicting the distribution of the eastern hog-nosed snake than variables representing toad distribution and relative abundance.

All biotic variables increased the probability of occurrence of the eastern hog-nosed snake in all models where they had significant contributions (Figure 2-5, 2-6 and 2-7). The response curves of the snake variables were in fact very similar to those of the toad variables, but the probabilities increased more in models using snake occurrences. Finally, the maps produced by Maxent for models using biotic variables are almost the same as for the models without them (Figure 2-8).

Table 2-2. Changes in the metrics of performance for models using an additional toad variable.

Time Period	Add. Variable	Resolution	Δ Int. AUC	Δ Ext. AUC	Threshold	Δ Sensisivity	Δ Specificity	Δ Kappa	Δ TSS	Δ MPA
Historical	**TOAD**	1000m	0.009	0.01	0.321	-0.03	0.03	0.03	0	-0.007
Historical	**TOAD**	750m	0.006	0.02	0.294	0.02	-0.01	0	0.01	-0.003
Historical	**TOAD**	500m	0.011	0	0.251	0.01	0	0.01	0.01	-0.021
Current	**T1990**	1000m	-0.004	0	0.161	0.03	-0.04	-0.015	-0.01	0.007
Current	**T1990**	750m	0.004	0	0.202	0	0	0	0	0.002
Current	**T1990**	500m	-0.006	0	0.238	0	0	0	0	0.004
Current	**CALL**	1000m	-0.01	0	0.165	0.03	-0.03	-0.012	0	-0.001
Current	**CALL**	750m	-0.004	-0.03	0.205	-0.03	0	0	0	0.037
Current	**CALL**	500m	-0.003	0	0.195	0.03	-0.04	-0.03	-0.01	0.004

Table 2-3. Changes in the metrics of performance for models using an additional snake variable

Time Period	Add. Variable	Resolution	Δ Int. AUC	Δ Ext. AUC	Threshold	Δ Sensisivity	Δ Specificity	Δ Kappa	Δ TSS	Δ MPA
Historical	**SNAK**	1000m	0.008	0.01	0.302	0	0.03	0.04	0.03	0.001
Historical	**SNAK**	750m	0.005	0.01	0.297	-0.02	0	0	-0.02	-0.028
Historical	**SNAK**	500m	0.019	0	0.201	0.01	-0.02	0	-0.01	-0.049
Current	**S1990**	1000m	0.001	0	0.195	0	0.02	0.01	0.02	-0.009
Current	**S1990**	750m	0.001	0	0.203	0	0.01	0.01	0.01	-0.007
Current	**S1990**	500m	-0.001	0	0.232	0	0.01	0.01	0.01	-0.003

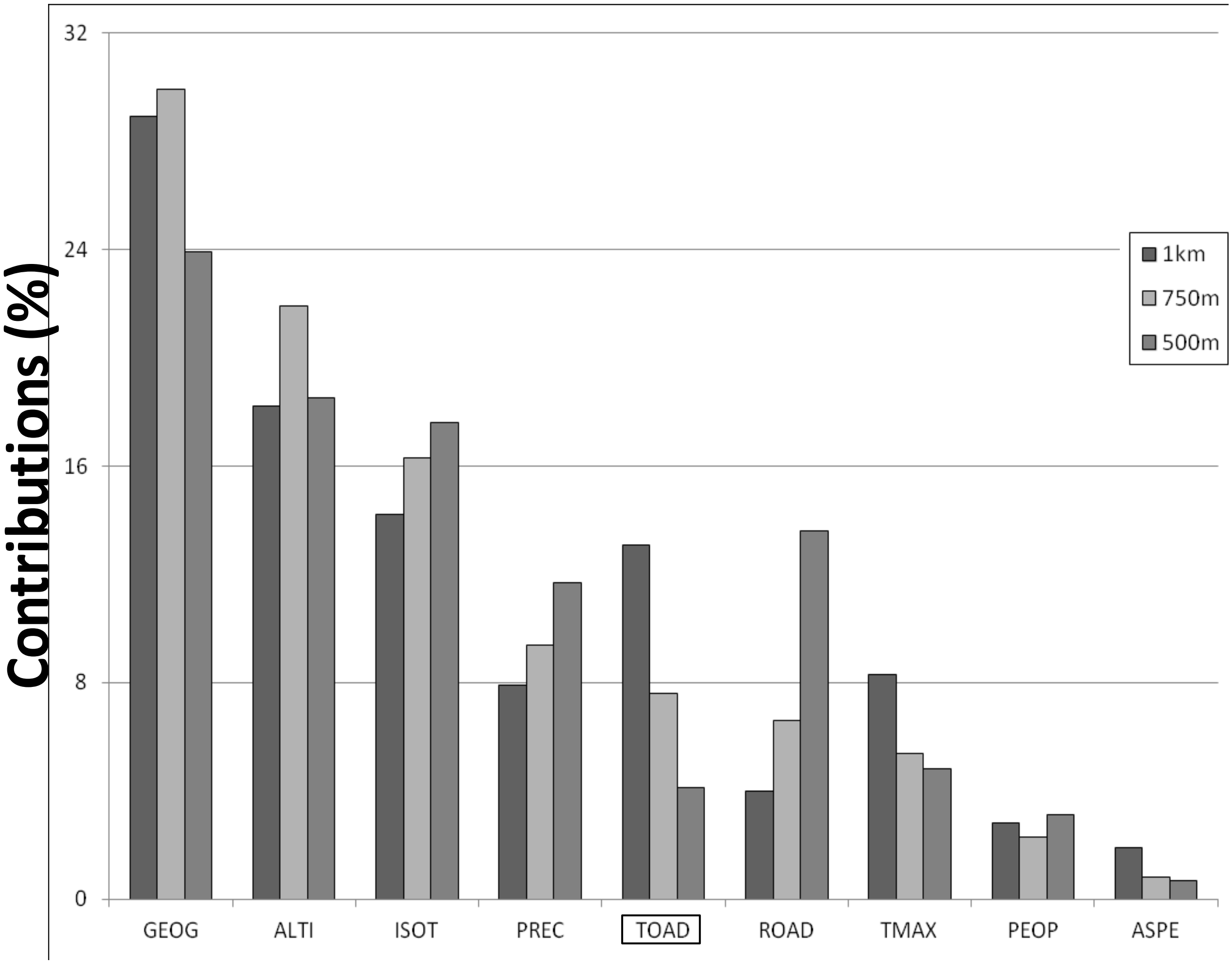


Figure 2-2. Contributions of the nine most influential variables in the Maxent habitat suitability model estimating the historical distribution of the eastern hog-nosed snake. An additional variable, TOAD, was added to the model.

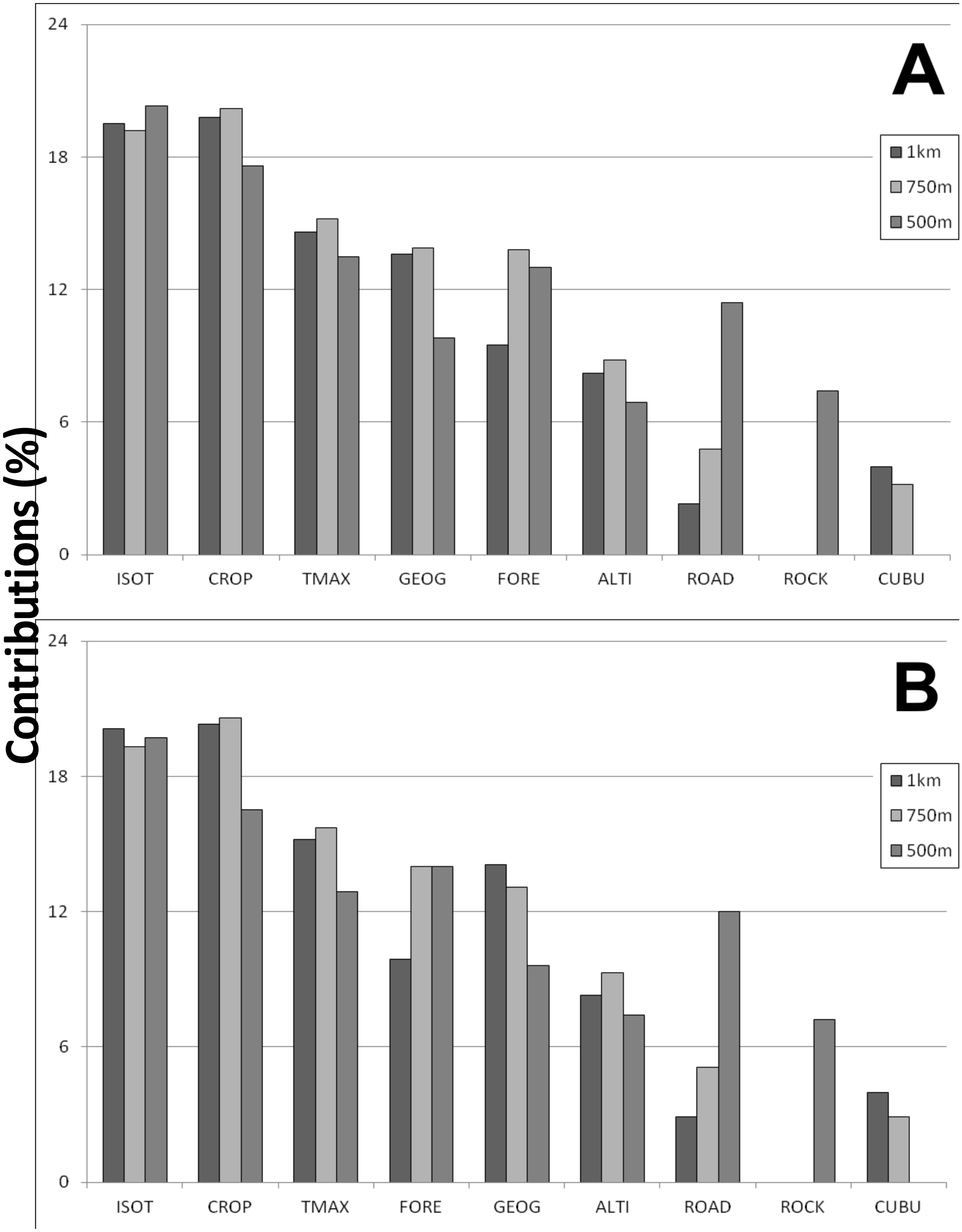


Figure 2-3. Contributions of the nine most influential variables in the Maxent habitat suitability models estimating the current distribution of the eastern hog-nosed snake. An additional toad variable was used in each model: (**A**) T1990 (**B**) CALL.

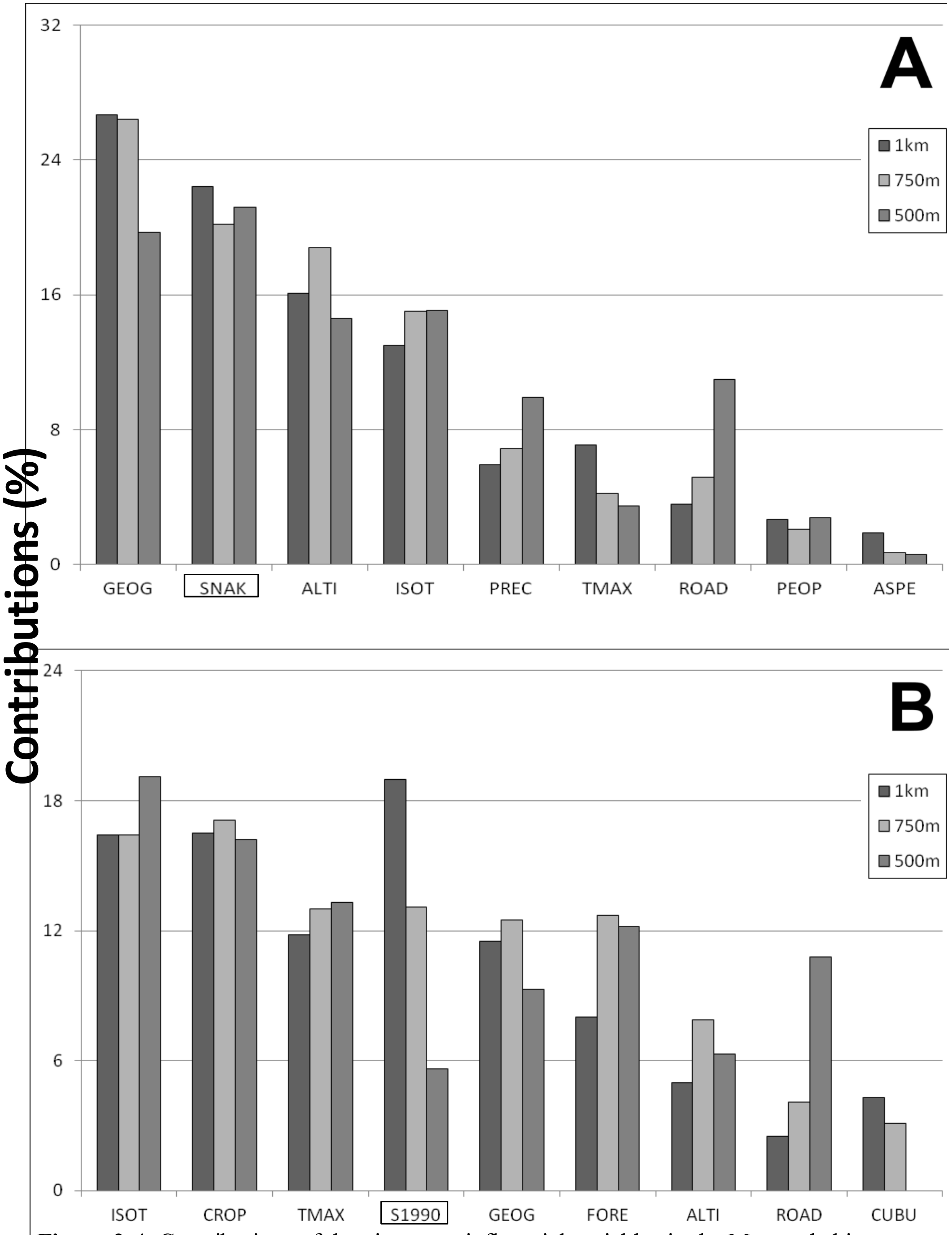


Figure 2-4. Contributions of the nine most influential variables in the Maxent habitat suitability models estimating the historical (A) and current (B) distribution of the eastern hog-nosed snake. An additional snake variable was used in each model: (**A**) SNAK (**B**) S1990.

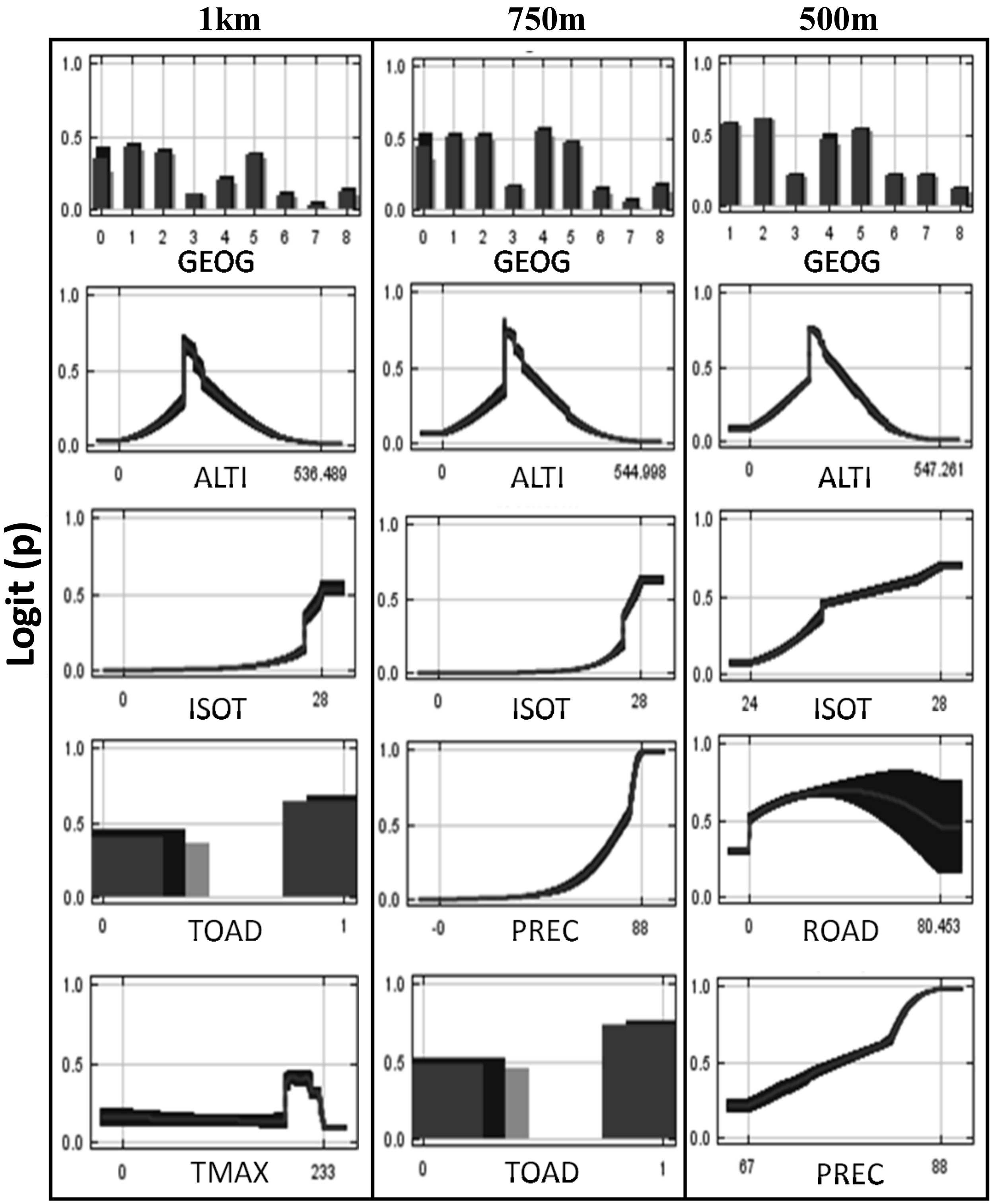


Figure 2-5. Marginal response curves ± SD, for the five most influential variables in the models built with Maxent, estimating the historical distribution of the eastern hog-nosed snake. The variable **TOAD** was added to the model. The curves show how the logistic prediction (p) varies, when one variable is gradually changed, while all others are kept at their average value.

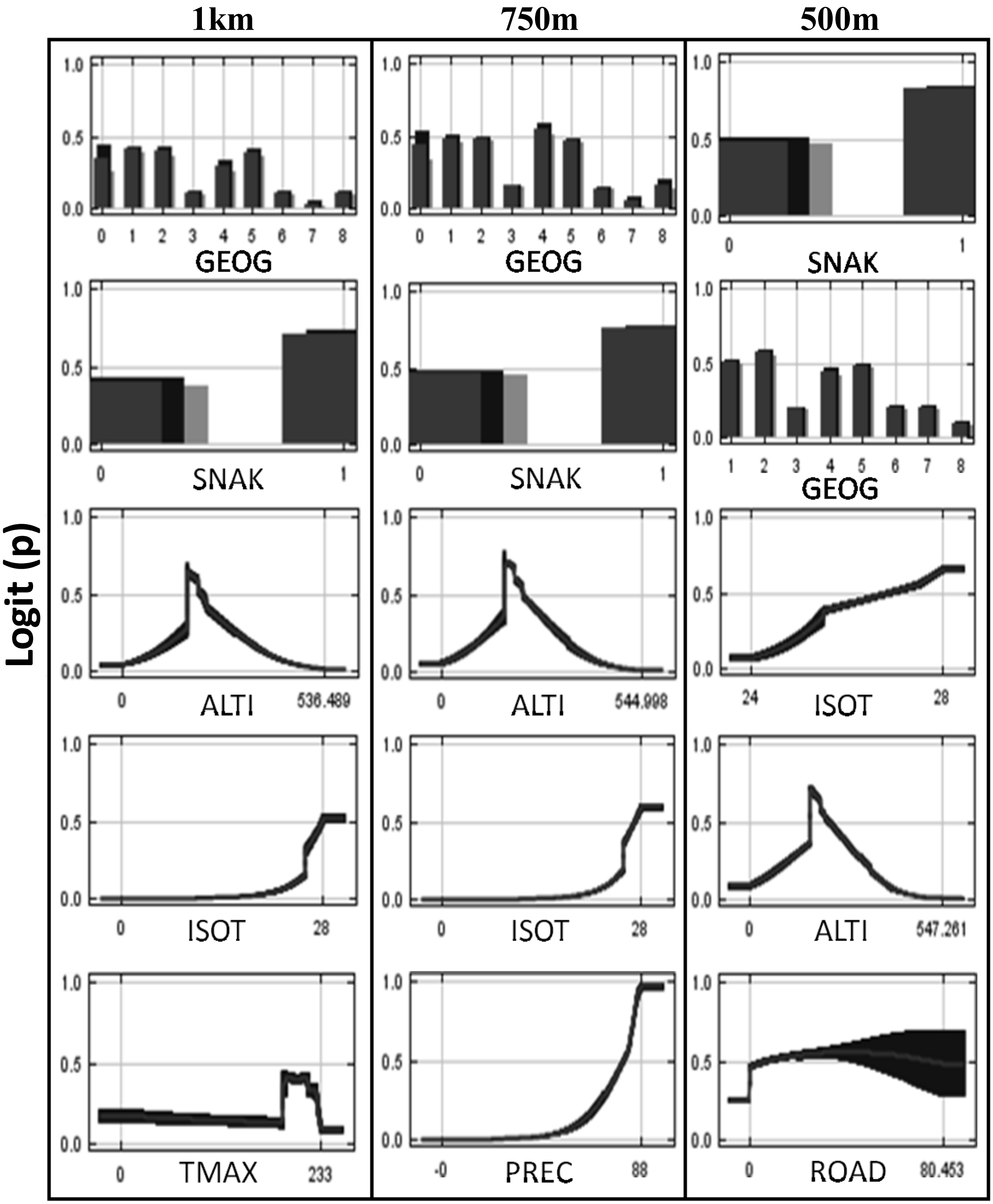


Figure 2-6. Marginal response curves ± SD, for the five most influential variables in the models built with Maxent, estimating the historical distribution of the eastern hog-nosed snake. The variable **SNAK** was added to the model. The curves show how the logistic prediction (p) varies, when one variable is gradually changed, while all others are kept at their average value.

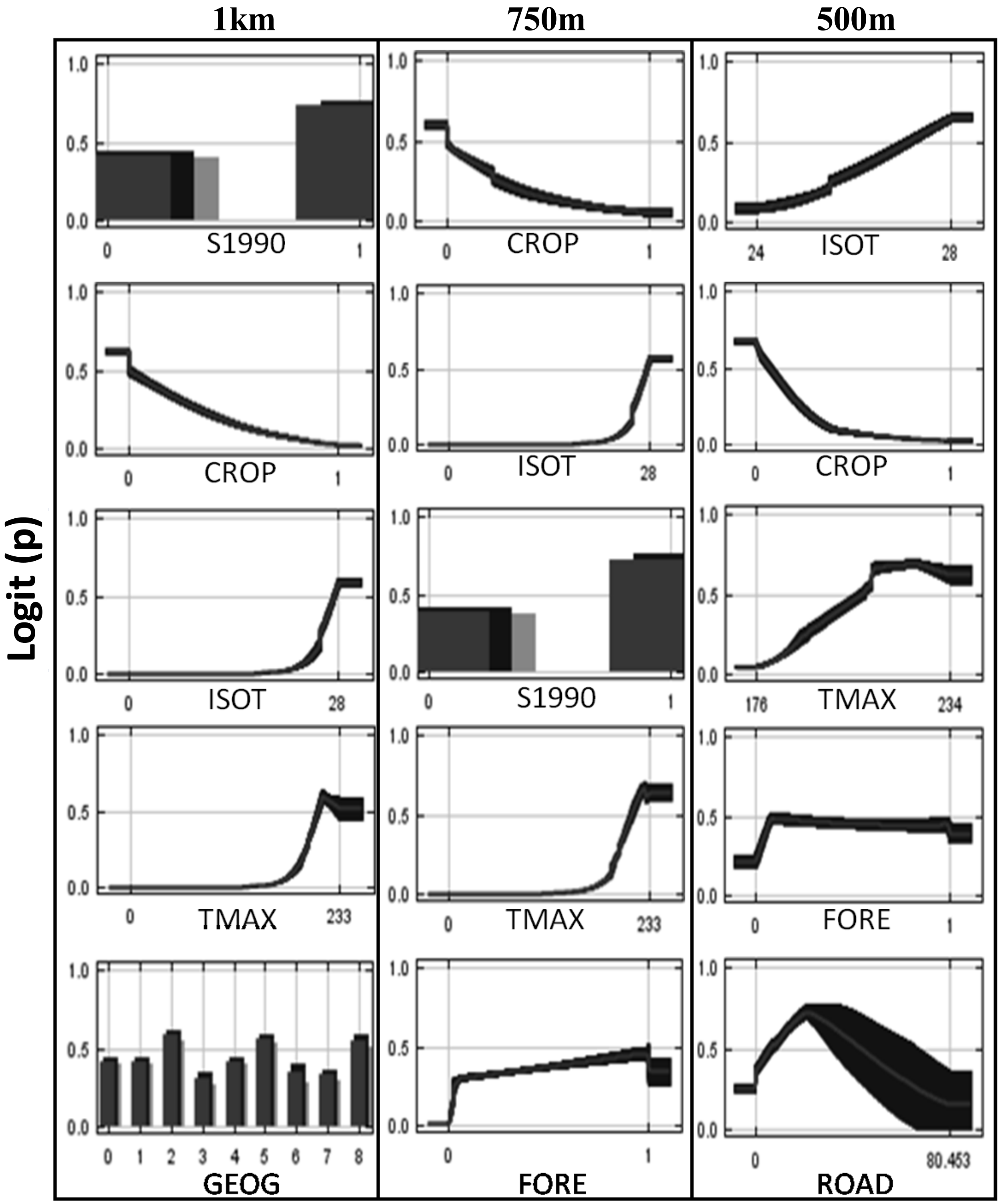


Figure 2-7. Marginal response curves ± SD, for the five most influential variables in the models built with Maxent, estimating the current distribution of the eastern hog-nosed snake. The variable **S1990** was added to the model. The curves show how the logistic prediction (p) varies, when one variable is gradually changed, while all others are kept at their average value.

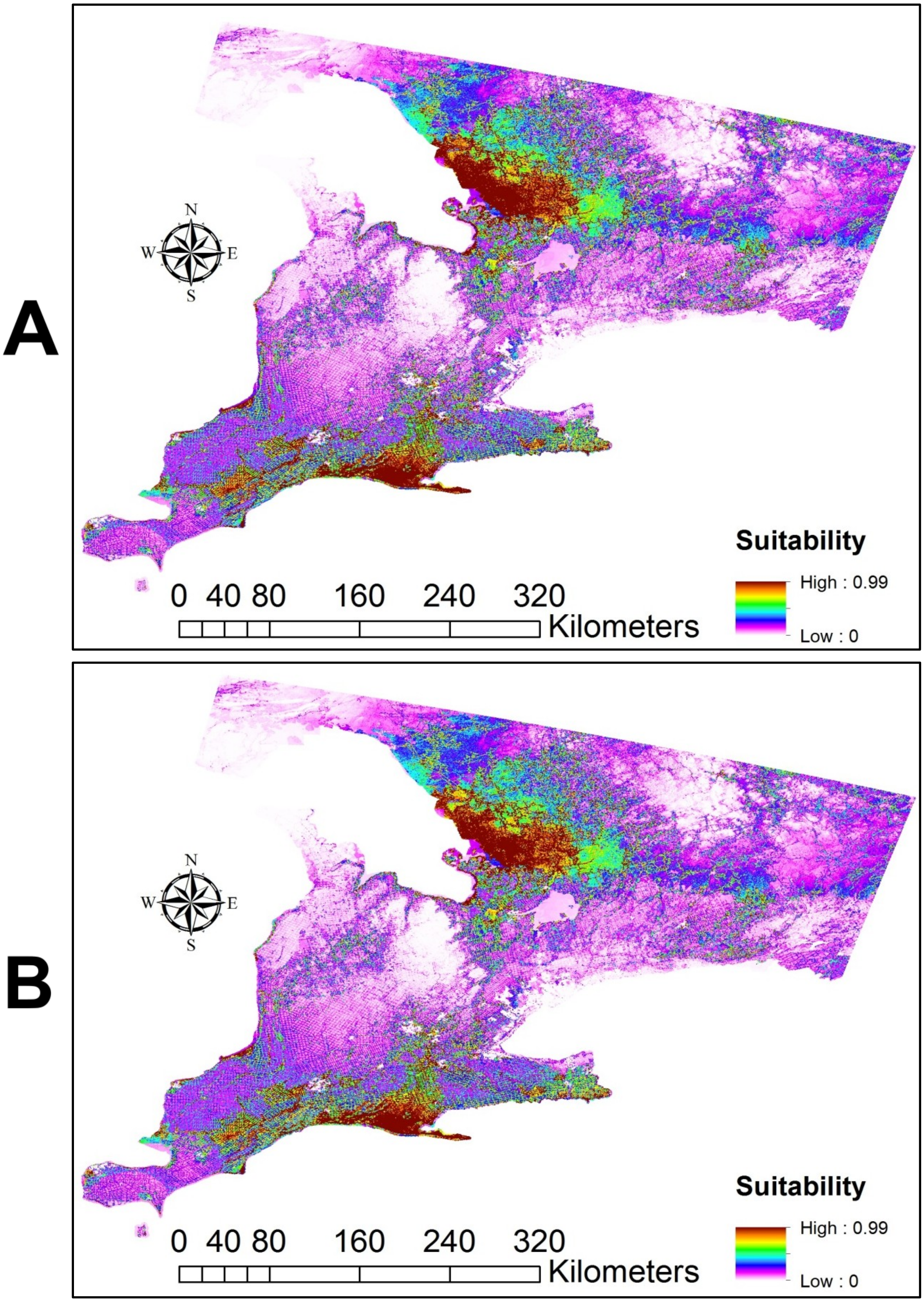


Figure 2-8. Estimated distribution of the eastern hog-nosed snake in Canada at a resolution of 500m. Each model was built with a biotic variable: (**A**) TOAD and (**B**) SNAK.

Discussion

Small changes in the metrics of performance

There are several reasons why the metrics of performance were unaffected by the use of biotic variables. Toads are found almost everywhere in Southern Ontario, which could indicate that their presence is not a factor limiting the distribution of the eastern hog-nosed snakes. In fact, toads are probably found everywhere throughout the range of the eastern hog-nosed snake in Ontario. This snake is also capable of eating other species of frogs and salamanders (Platt, 1969). Thus, it is possible that the distribution of the eastern hog-nosed snake is not influenced by toad distribution and that variables representing toad presence are truly unnecessary to predict where eastern hog-nosed snake are found. For these reasons, it is possible that the models were not improved and that my hypothesis was incorrect.

Because the distribution of toads is so vast, I also used toad calls (CALL) as a measure of toad relative abundance. It is possible that hog-nosed snakes are unable to subsist when toad populations are too low and that they are only found where toads are plentiful. However, that variable decreased the performance of current models and only had very low contributions. Toad calls may have not been an important variable in current models because of insufficient quality. I used all the data available from the Marsh Monitoring Program (BirdStudiesCanada, 2008), which consisted of only 3121 entries, most of which were clustered in some regions. These entries also comprise various surveys that were done at the same location at different times. The goal of the Marsh Monitoring Program is to survey the same locations at different times to notice trends in amphibian decline and provide data for conservation. Similarly to data found in many atlases, it lacks a sampling design. Finally, few points were available for the area of Georgian Bay where the distribution of the eastern hog-nosed snake remains extensive. It is therefore possible that there were simply not enough surveys done across the area studied to build a good variable representing toad relative abundance. Obtaining reliable data to study biotic interactions at a landscape scale will be challenging. This is one of the reasons why few studies use biotic variables to build habitat suitability models.

Alternatively, it is possible that an additional variable adds noise to the model and that it is not an optimal set of variables for the algorithm to make better predictions. In Chapter I, all original models were pruned selecting the minimum number of variables required to optimize the AUCs during validation. This is why the Maxent models predicting the current distribution of the eastern hog-nosed snake do not use all 19 variables available, but only 8 or 9. By adding this additional biotic variable, it is possible that AUCs are not at their highest. Unfortunately, the effect of the number of variables on habitat suitability models has not been studied extensively and will depend on the statistics used by each algorithm. For example, in multivariable regression analyses, too many variables will reduce the probability of detecting associations and will give outcomes similar to what could be expected by chance (Lewis, 2007). This subject should be assessed more in depth in future studies and may help identify the level at which some models overfit their training dataset.

Finally, the two models that seemed to make better predictions with biotic variables (TOAD and SNAK) were built at the finest spatial scale, 500 m. Two reasons could explain why biotic interactions influenced these models at a finer grain. First, biotic interactions are known to affect species distributions at both coarse and fine scales (Heikkinen et al., 2007; Cunningham et al., 2009). In this study, it is possible that the eastern hog-nosed snake only responded to these variables at 500 m. The second reason comes from the GIS manipulations required to build these variables. When converting points to a raster, fewer points are lost at a resolution of 500 m, creating a variable that resembles more the original data. At a larger resolution, the data are coarsened, and it is possible that too much information was lost.

High contributions

It is surprising that although the biotic variables did not seem to help the models to make better predictions, some of these variables (both toads and snakes) had high contributions in the models. We would expect that these highly modified models would make better predictions because the algorithm favoured the biotic variables over other variables for that specific reason. Yet, the validation techniques employed here are unable to pick-up these changes. The directions of the response curves are also very informative because both the toad and snake variables increased the probability of occurrence of the eastern hog-nosed snake. This was expected with toads, but not with snakes. The presence of other snake species should not favour the presence of hog-nosed snakes, but do the opposite by accounting for competitor species. The only reason that can explain why the variable SNAK was one of the variables with the highest contributions is that it represents the areas that have been sampled for the presence of reptiles: areas that have been sampled for other snake species have probably been sampled for eastern hog-nosed snakes as well.

Sampling bias

Both toad and snake variables had high contributions in some of the models. At a resolution of 500 m, SNAK was the most important variable used to predict the historical distribution of the eastern hog-nosed snake. Similarities between the models built with toad and snake variables indicate that these variables probably represent sampling bias. Various datasets used to build habitat suitability models exhibit strong sampling bias (Dennis and Thomas, 2000). However, most studies using these data to build habitat suitability models do not take this into consideration. Although funding resources are often limited and data of better quality not always accessible, various ways have been proposed to deal with sampling biases including better selection of background points (pseudo-absences). Some have proposed selecting background points that have a sampling bias similar to the occurrences (Phillips et al., 2009). The aim behind this technique is that the models will focus on differences between the occurrences and the background points and not differences inherent to the sampling bias. To generate background points reflecting the same sampling bias as the occurrences, it is possible to use other species occurrences sampled in the same way; target groups (Anderson, 2003). Because the issue of sampling bias is rarely assessed in studies using habitat suitability models and that it may be a serious problem with some datasets, I suggest not only verifying the impact of this problem on the models but also controlling for it by better selecting background points.

Implications

Species occurrences found in atlases is an important source of data that have allowed studying various phenomena in different fields of ecology (Newbold, 2010). Yet, this study clearly shows that the spatial bias and lack of sampling design often associated with these data is a serious limitation that deserves more attention. It may be difficult to evaluate habitat suitability models and make good predictions when the quality of the data is deficient. In the same way, it may not be possible to study fine-scale ecological associations such as biotic interactions with poor quality data. Presence-only models are more often used than presence-absence models, and are also more affected by sampling biases (Phillips et al., 2009). Because the quality of sampling designs is known to influence the reliability of habitat suitability models (Edwards et al., 2006), I encourage building models with quality data taking the time to develop strategic sampling designs. Alternatively, choosing a right set of background points exhibiting a similar sampling bias can decrease the magnitude of this problem (Phillips et al., 2009). As for conservation, we should use habitat suitability maps to help better sample regions of importance. This data would not only allow developing more accurate habitat suitability models but also help assess the true status of these species.

APPENDIX I

Sources of bias in habitat suitability modelling

The performance of models is influenced by many factors, but the quality of the observations is particularly important. Although researchers often use preexisting presences or absences to build HSMs, these data may not reflect the environmental conditions where a species is found because they were not collected to achieve this goal (Pearson, 2007). These datasets rarely cover the entire known distribution of the species or the region that is being modeled (Wisz et al., 2008). In the same way, some observation records kept by governments are gathered with the help of the public without any specific sampling design. Some location points are sometimes determined only through "georeferencing" using location descriptions only, especially for historic data (Newbold, 2010). These sampling procedures always lead to biases that are usually unaccounted for. For elusive and rare species, non-detectability does not always translate to absences. The species may not be detected with the sampling design (Thuiller and Münkemüller, 2010). These absences are known as false absences and are considered a major source of bias in presence/absence models (Hirzel et al., 2002). There is also a degree of spatial precision associated with the observations. If the observations are spatially imprecise, the environmental layer/predictors may not correspond to the conditions where the species is found. Finally, models assume that distributions are at equilibrium with the environment (Pearson, 2007). This is difficult to verify as models do not generally take into consideration biotic interactions and dispersal, and because models are not verified through time.

APPENDIX II

Historical and current distributions estimated with each algorithm separately

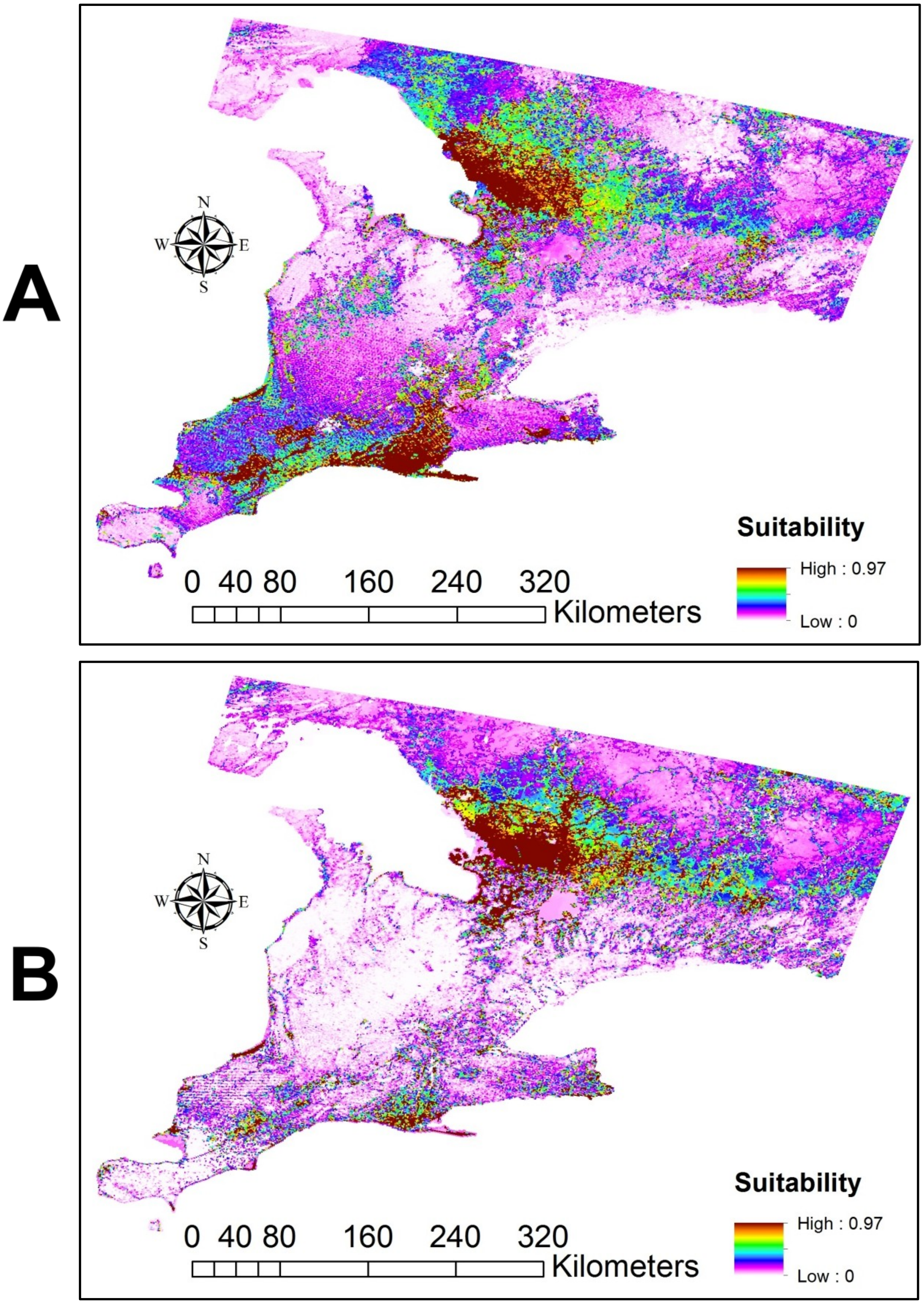


Figure 4-1. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm Maxent at a resolution of 1km: (**A**) historically and (**B**) currently.

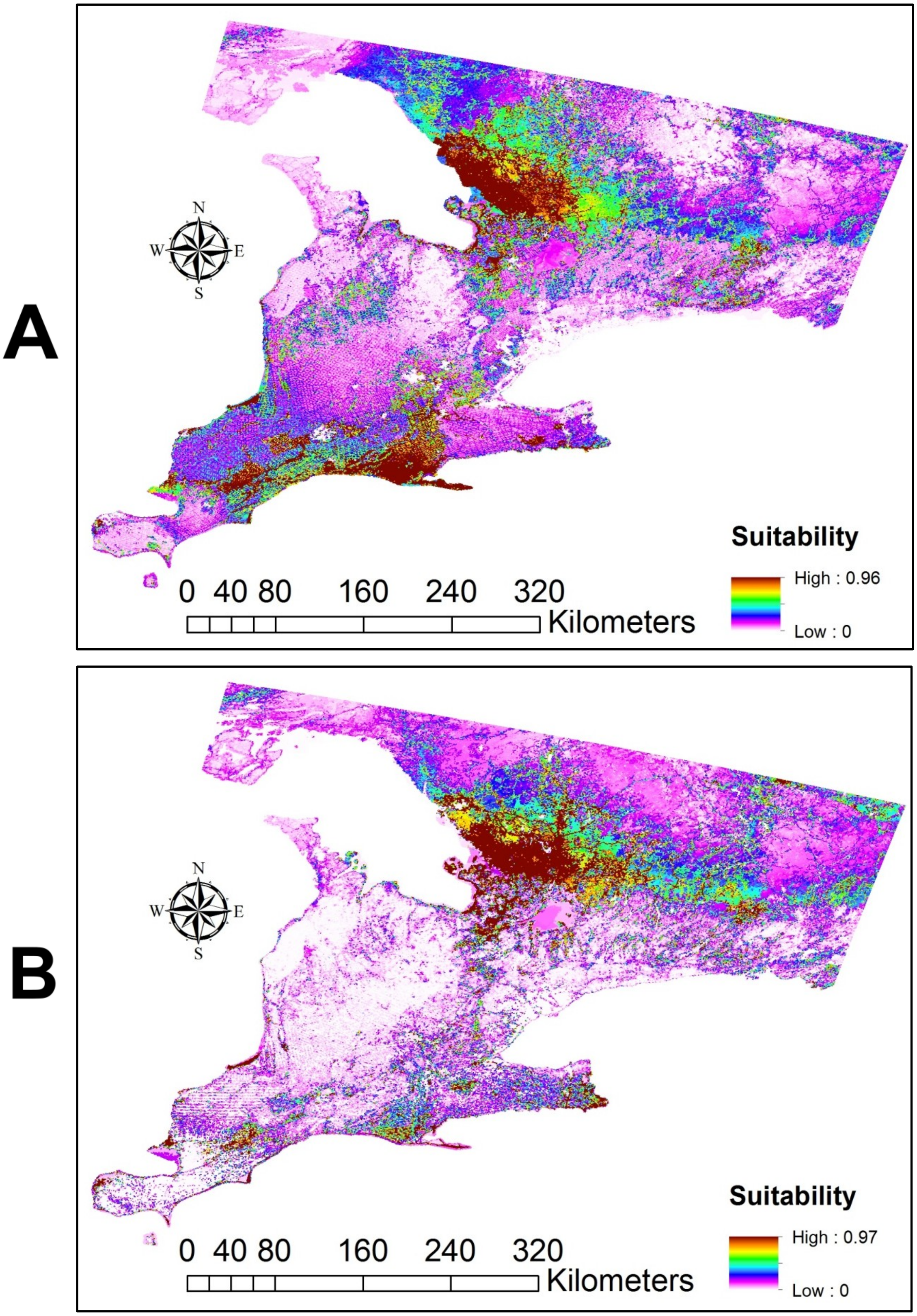


Figure 4-2. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm Maxent at a resolution of 750m: (**A**) historically and (**B**) currently.

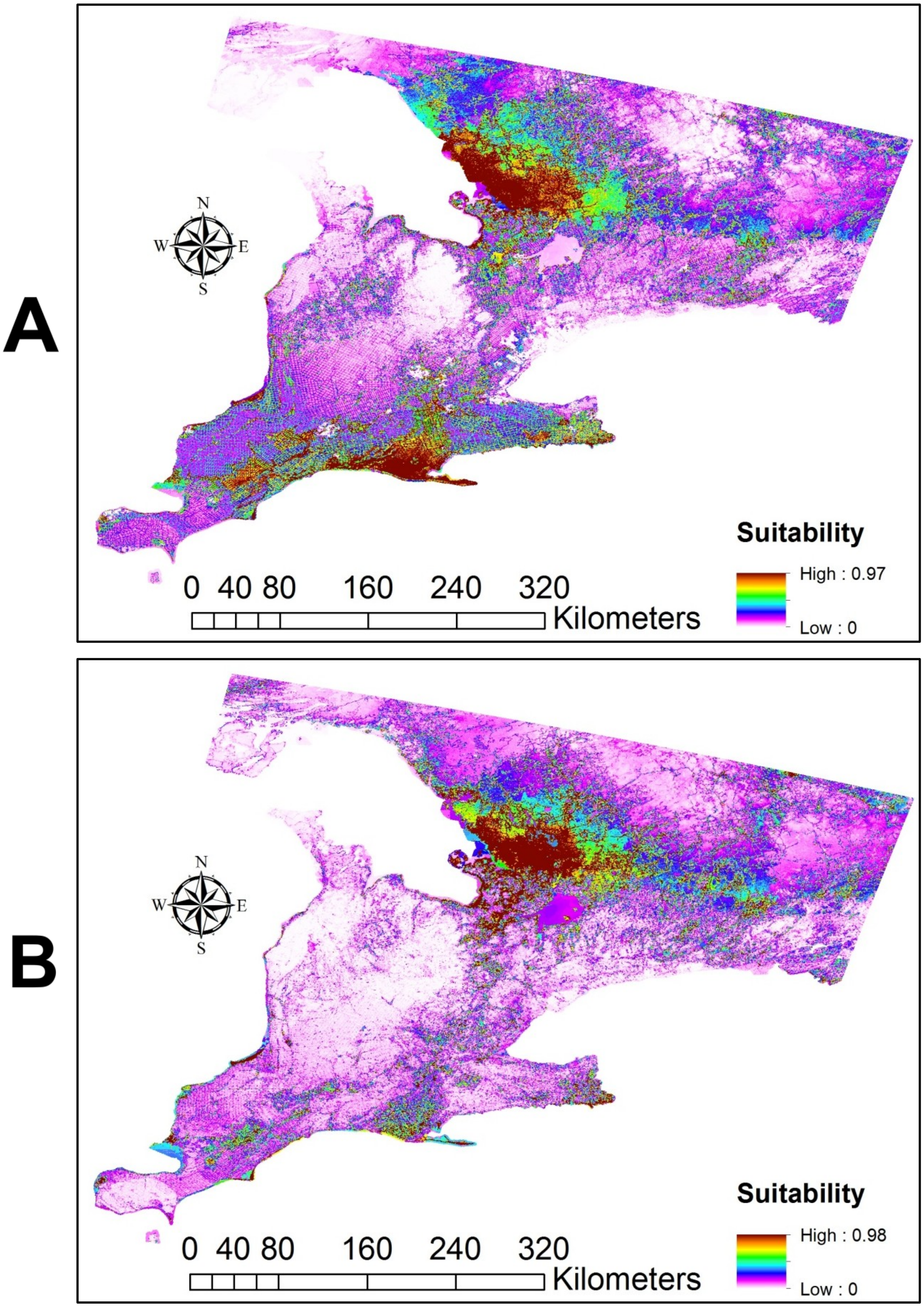


Figure 4-3. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm Maxent at a resolution of 500m: (**A**) historically and (**B**) currently.

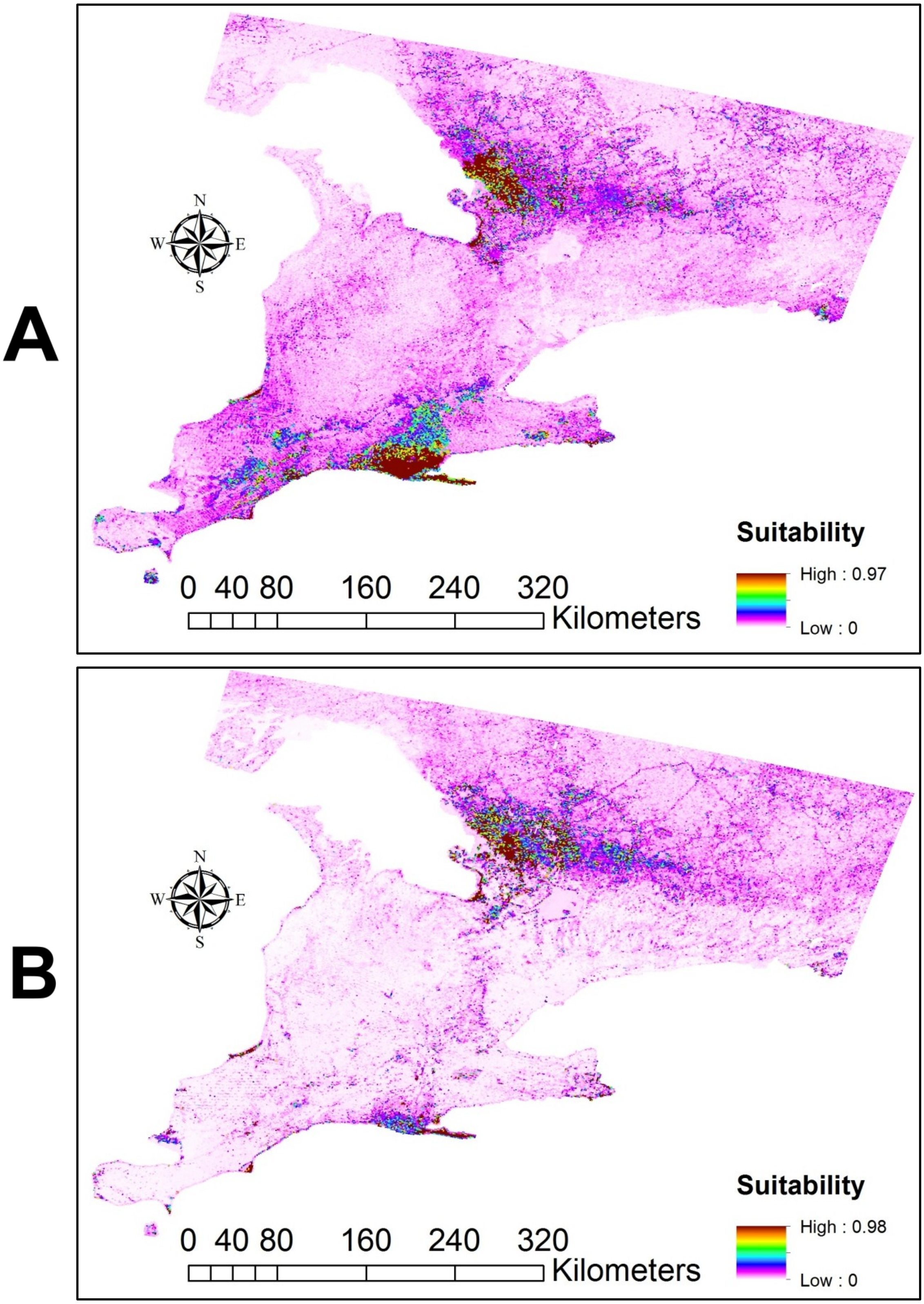


Figure 4-4. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm BRT at a resolution of 1km: (**A**) historically and (**B**) currently.

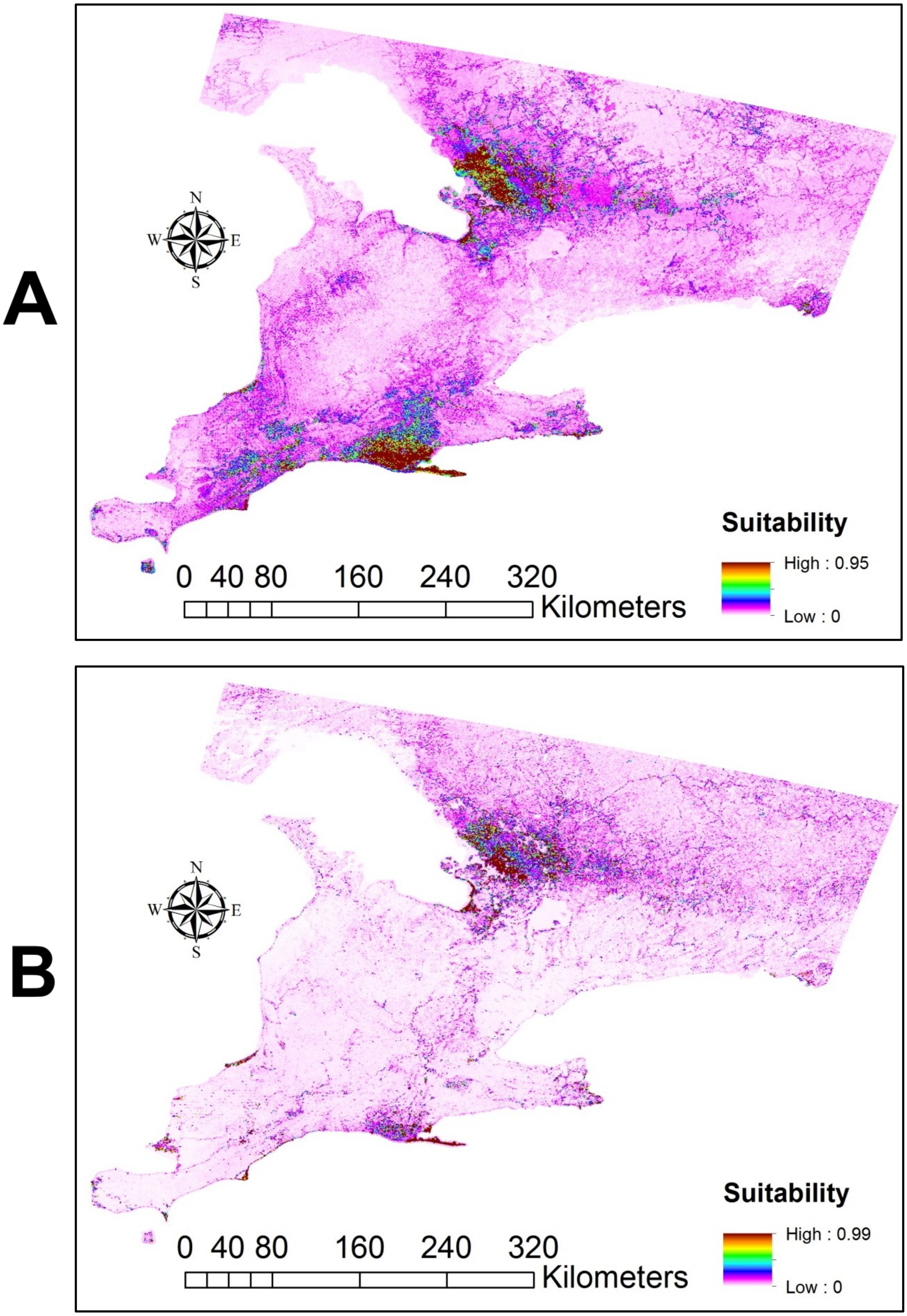


Figure 4-5. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm BRT at a resolution of 750m: (**A**) historically and (**B**) currently.

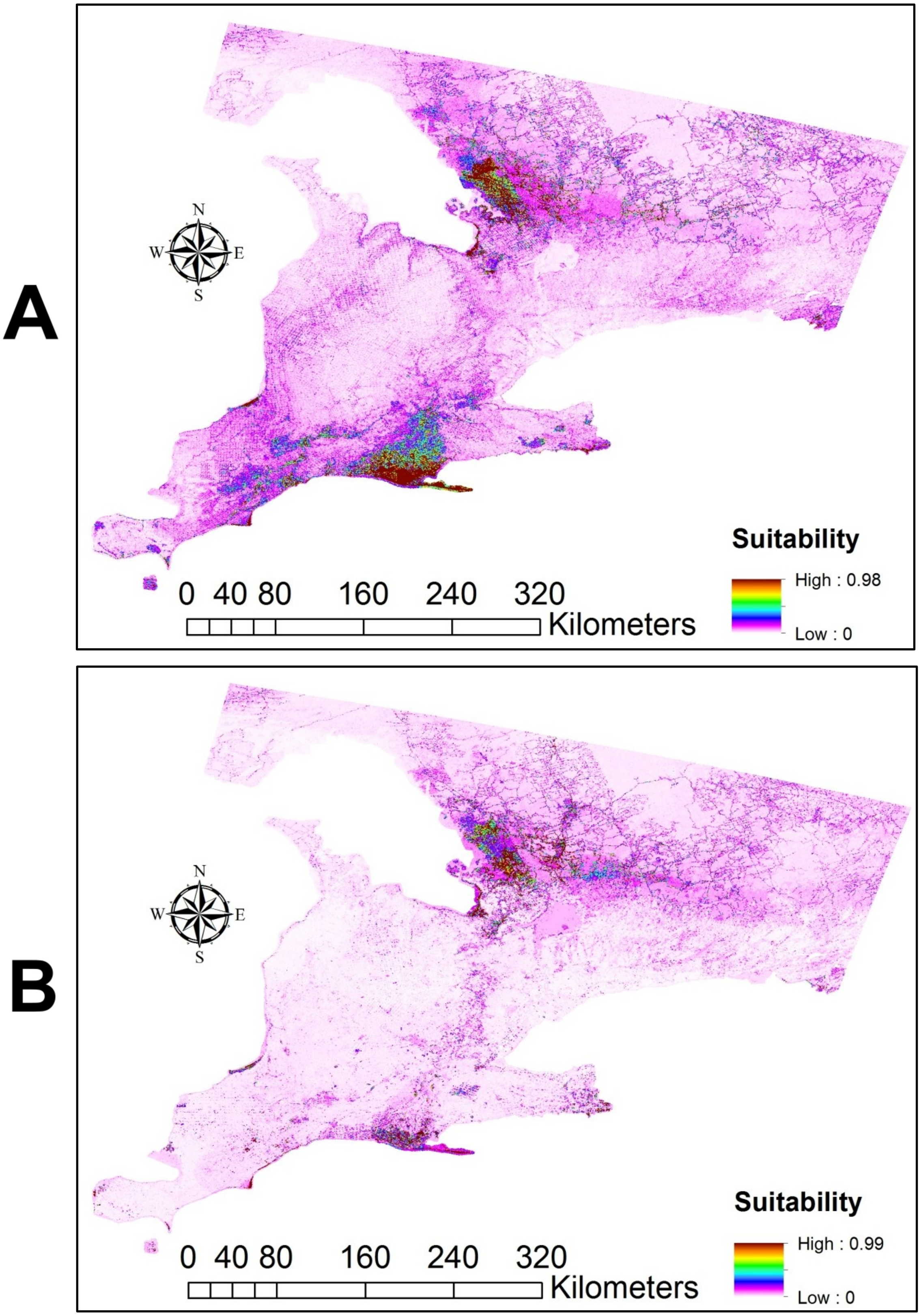


Figure 4-6. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm BRT at a resolution of 500m: (**A**) historically and (**B**) currently.

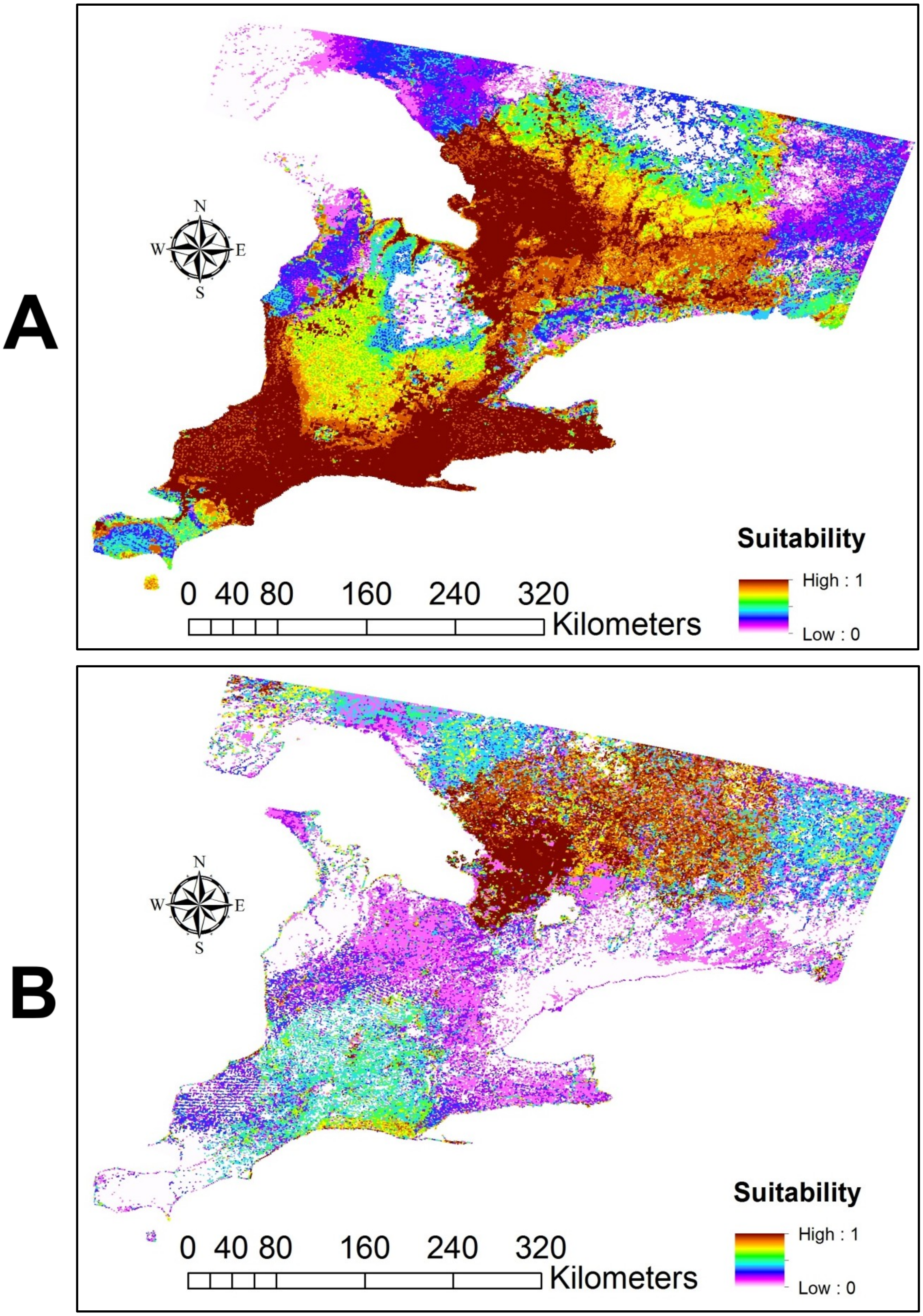


Figure 4-7. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm GARP at a resolution of 1km: (**A**) historically and (**B**) currently.

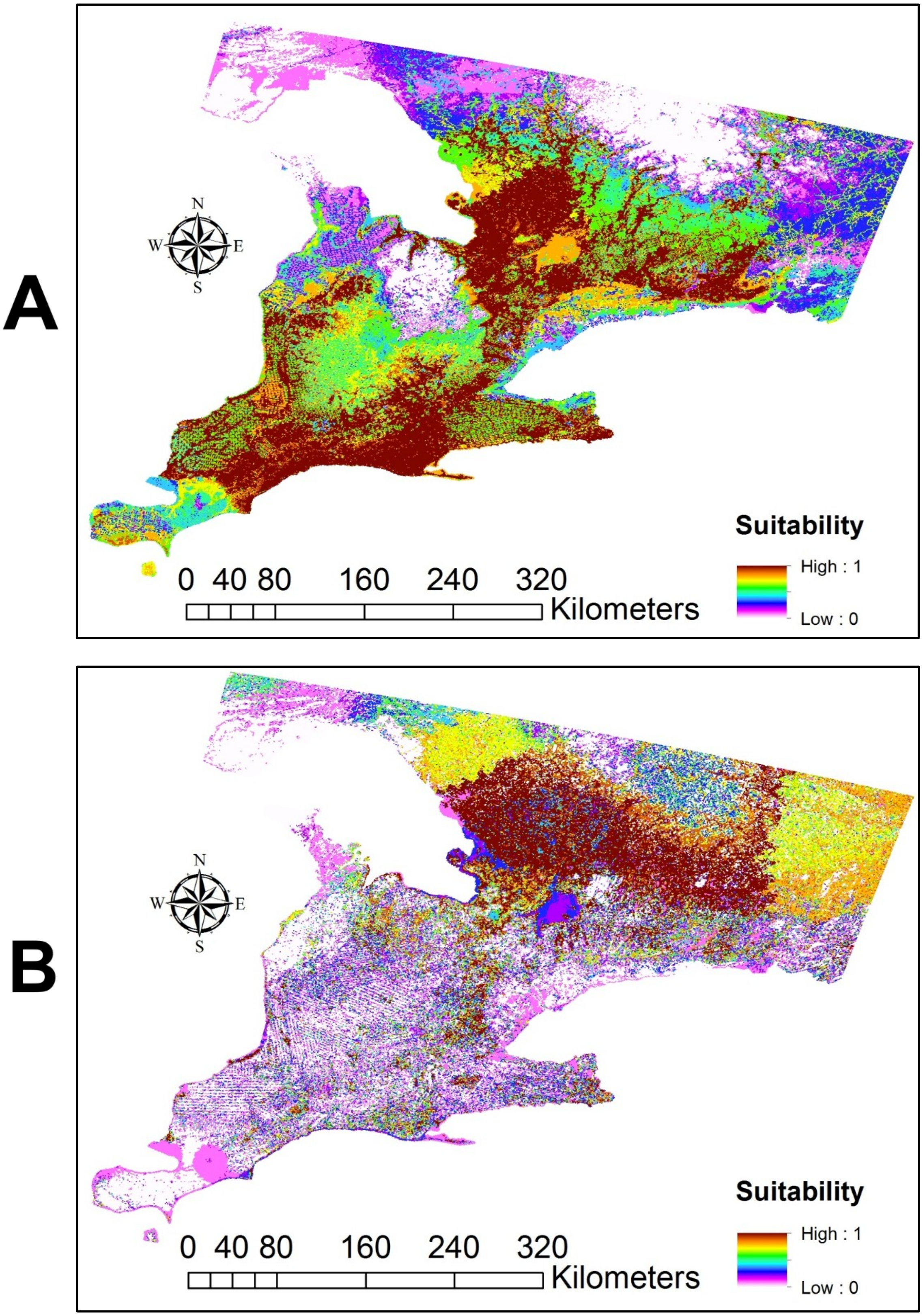


Figure 4-8. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm GARP at a resolution of 750m: (**A**) historically and (**B**) currently.

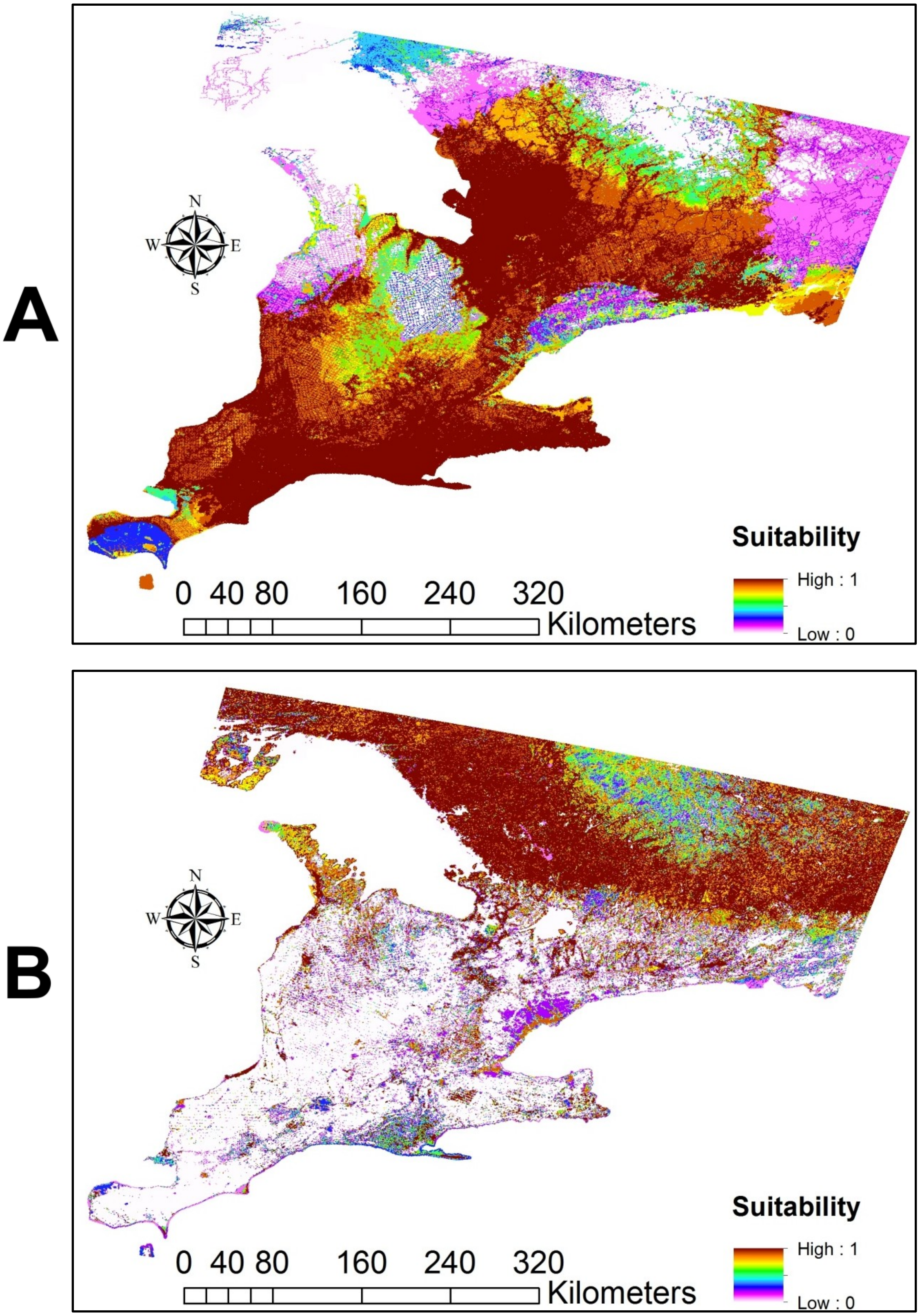


Figure 4-9. Estimated distribution of the eastern hog-nosed snake in Canada, using the algorithm GARP at a resolution of 500m: (**A**) historically and (**B**) currently.

APPENDIX III

Habitat suitability maps using a consensus approach and three thresholds

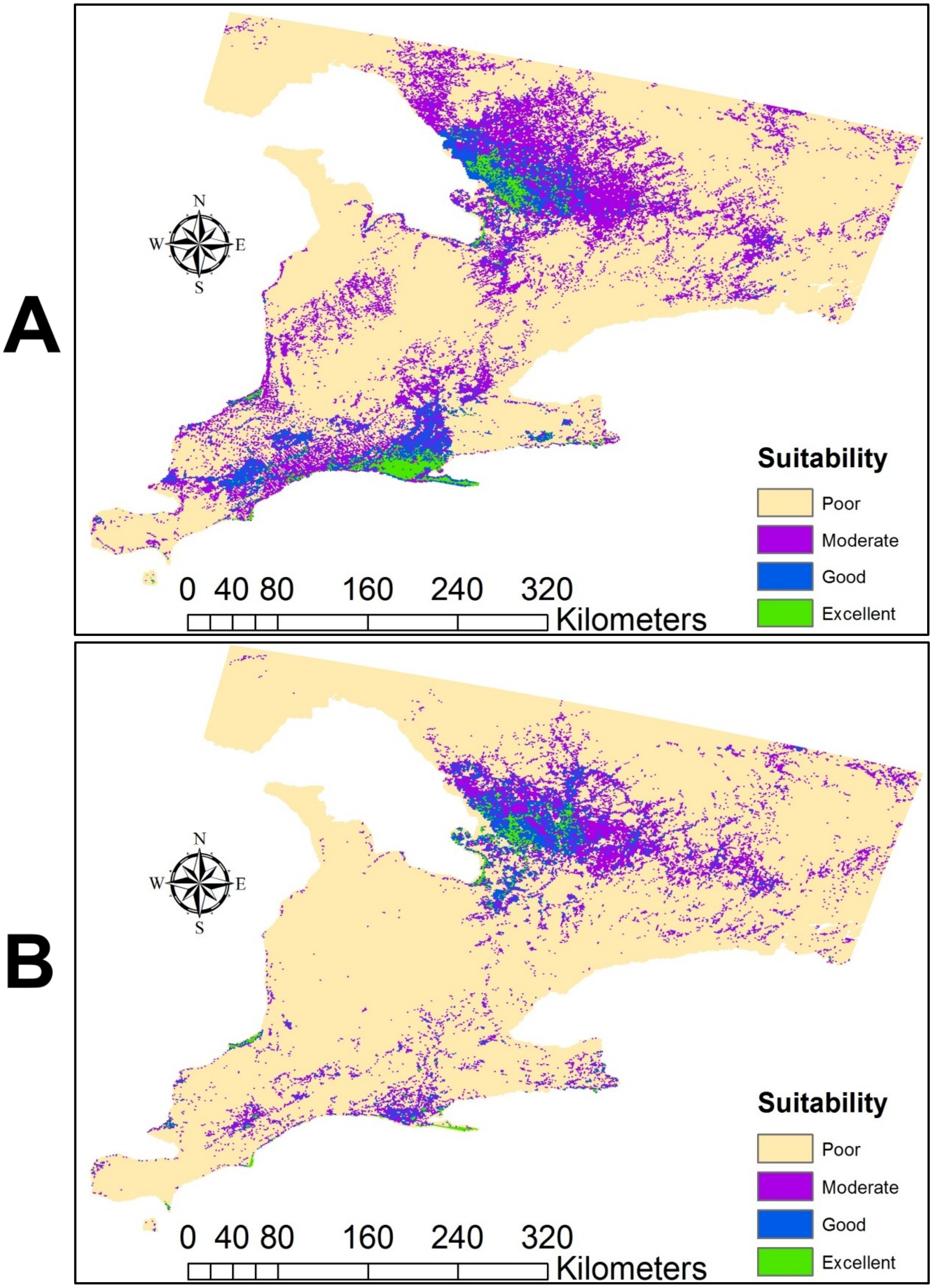


Figure 5-1. Habitat suitability maps for the eastern hog-nosed snake in Ontario, using consensus models at a resolution of 1km: (**A**) historically and (**B**) currently.

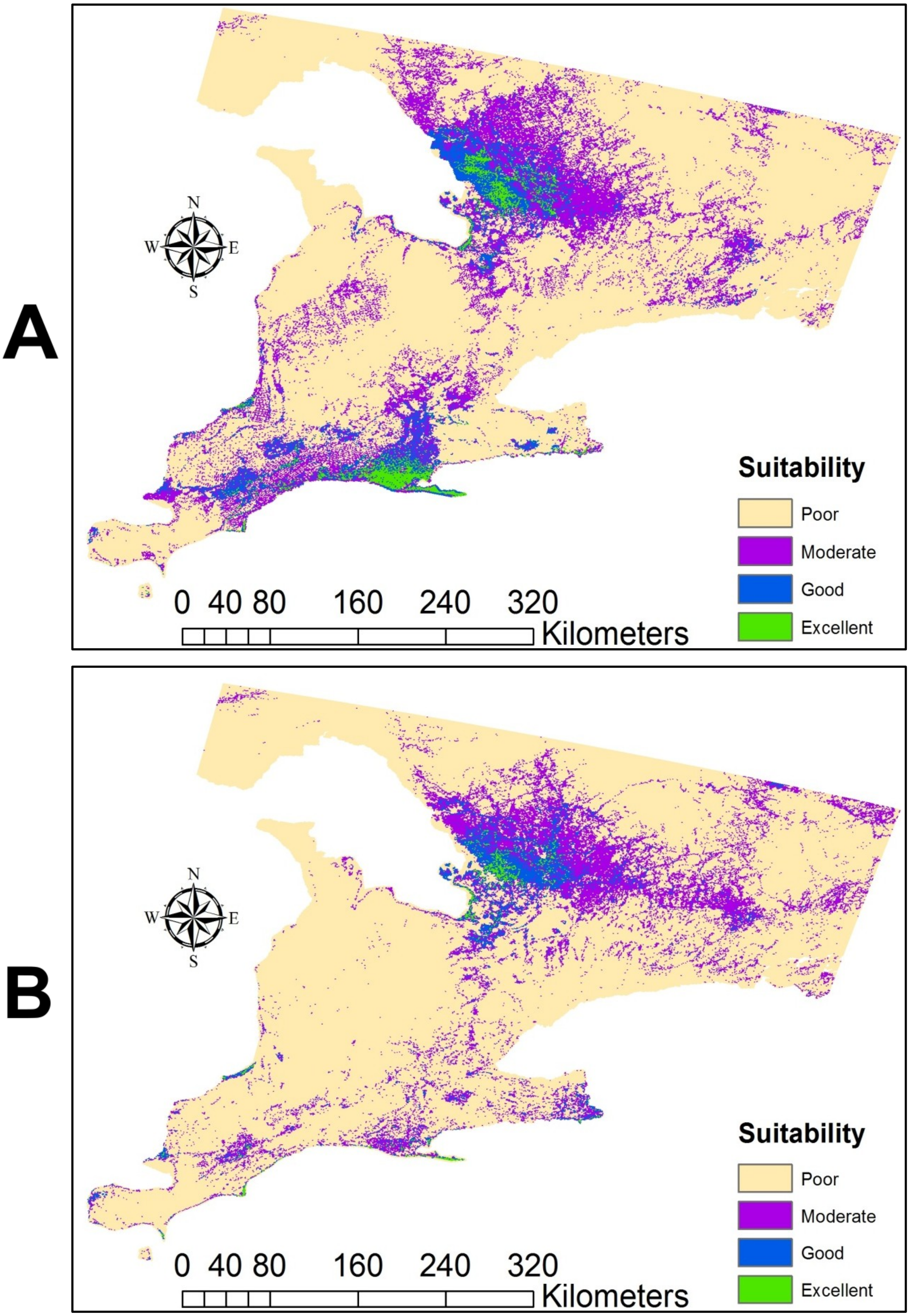


Figure 5-2. Habitat suitability maps for the eastern hog-nosed snake in Ontario, using consensus models at a resolution of 750m: (**A**) historically and (**B**) currently.

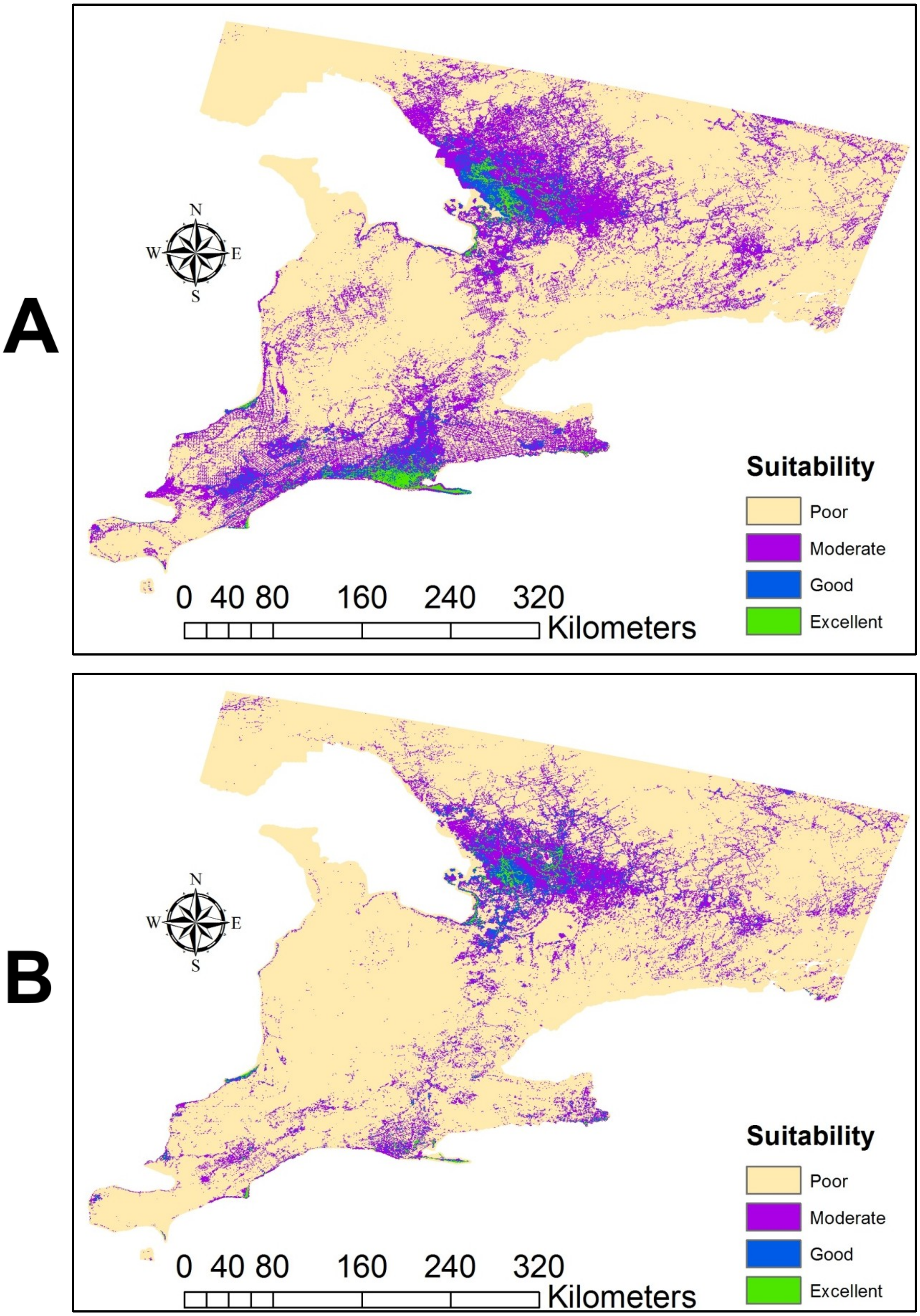


Figure 5-3. Habitat suitability maps for the eastern hog-nosed snake in Ontario, using consensus models at a resolution of 500m: (**A**) historically and (**B**) currently.

APPENDIX IV

Surficial geology conditions on which eastern hog-nosed are found in Ontario.

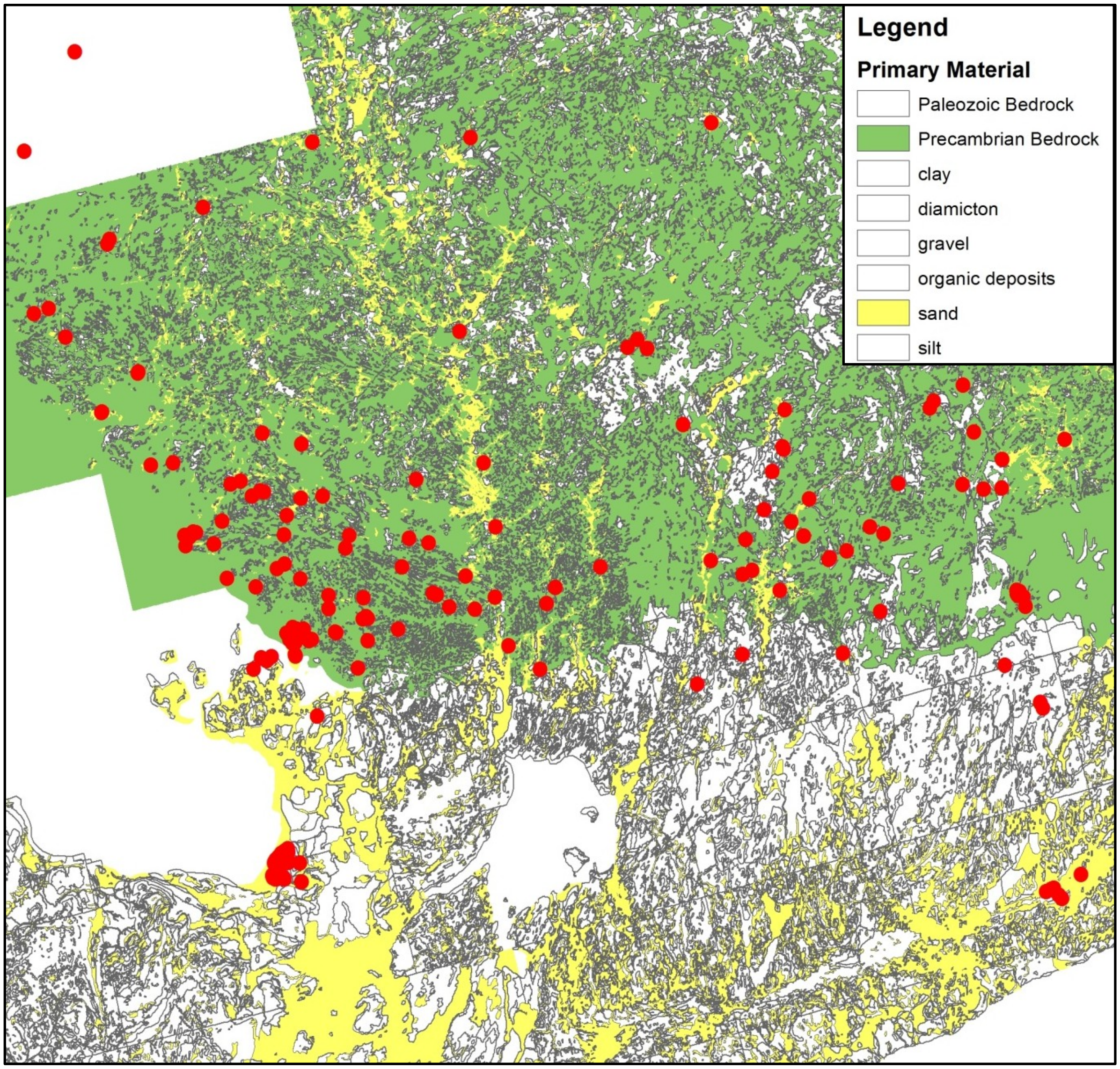


Figure 6-1. Single primary material of the surficial geology found in the region occupied by the northern populations of eastern hog-nosed snakes in Ontario.

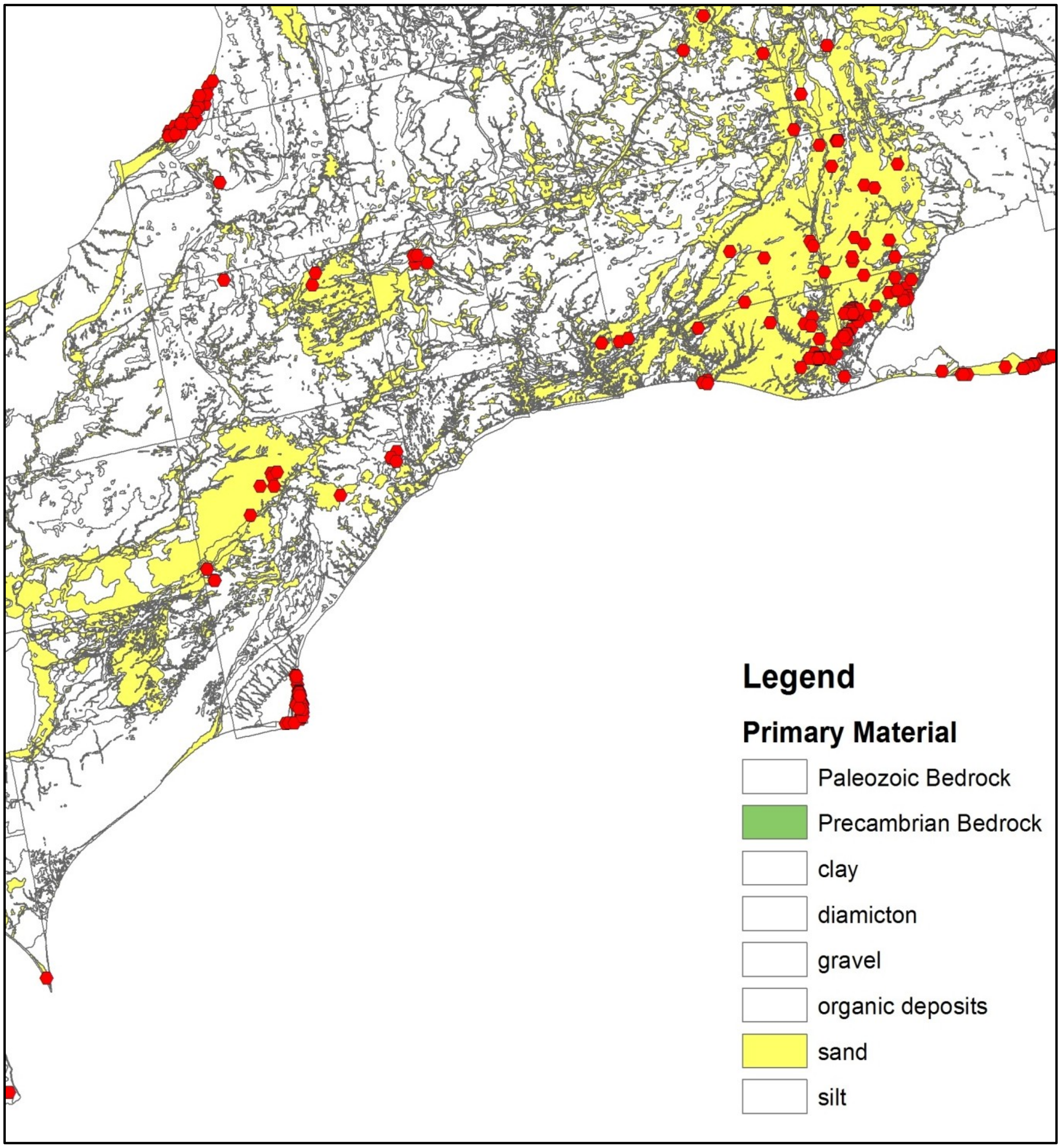


Figure 6-2. Single primary material of the surficial geology found in the region occupied by the southern populations of eastern hog-nosed snakes in Ontario.

APPENDIX V

Correlation matrix of all variables included in the habitat suitability models predicting the distribution of the eastern hog-nose snake.

Variables	**ALTI**	**SLOP**	**ASPE**	**GEOG**	**SAND**	**PRBE**	**ALVA**	**ROCK**	**CROP**	**CUBU**	**FORE**	**MUDF**	**PAST**
ALTI	**1.000**	0.273	0.002	-0.295	-0.118	0.133	-0.061	-0.097	-0.071	-0.039	0.345	0.034	-0.074
SLOP	0.273	**1.000**	-0.008	-0.279	-0.071	0.291	-0.037	0.023	-0.281	0.035	0.393	0.149	-0.068
ASPE	0.002	-0.008	**1.000**	0.006	0.002	-0.015	0.003	0.000	0.018	-0.003	-0.018	0.011	-0.009
GEOG	-0.295	-0.279	0.006	**1.000**	0.342	-0.649	-0.058	-0.120	0.468	-0.042	-0.548	-0.300	0.145
SAND	-0.118	-0.071	0.002	0.342	**1.000**	-0.334	-0.037	-0.048	0.213	0.031	-0.180	-0.158	0.099
PRBE	0.133	0.291	-0.015	-0.649	-0.334	**1.000**	-0.034	0.183	-0.647	0.024	0.630	0.416	-0.290
ALVA	-0.061	-0.037	0.003	-0.058	-0.037	-0.034	**1.000**	-0.004	-0.052	0.036	-0.027	-0.027	0.042
ROCK	-0.097	0.023	0.000	-0.120	-0.048	0.183	-0.004	**1.000**	-0.120	0.292	-0.009	0.059	-0.025
CROP	-0.071	-0.281	0.018	0.468	0.213	-0.647	-0.052	-0.120	**1.000**	-0.083	-0.712	-0.404	0.106
CUBU	-0.039	0.035	-0.003	-0.042	0.031	0.024	0.036	0.292	-0.083	**1.000**	0.013	-0.074	0.192
FORE	0.345	0.393	-0.018	-0.548	-0.180	0.630	-0.027	-0.009	-0.712	0.013	**1.000**	0.253	-0.233
MUDF	0.034	0.149	0.011	-0.300	-0.158	0.416	-0.027	0.059	-0.404	-0.074	0.253	**1.000**	-0.210
PAST	-0.074	-0.068	-0.009	0.145	0.099	-0.290	0.042	-0.025	0.106	0.192	-0.233	-0.210	**1.000**
SETT	-0.155	-0.048	-0.034	0.096	0.089	-0.131	-0.011	-0.031	-0.089	-0.036	-0.175	-0.080	-0.011
WATR	-0.222	-0.097	0.014	0.059	-0.117	0.178	-0.023	0.003	-0.339	-0.152	-0.227	0.246	-0.187
WETL	-0.102	-0.021	0.018	-0.079	-0.040	0.054	0.051	-0.002	-0.271	-0.029	0.094	0.314	-0.040
TMAX	-0.489	-0.285	-0.002	0.575	0.250	-0.546	0.000	-0.113	0.567	-0.089	-0.544	-0.208	0.025
PREC	0.662	0.095	0.007	-0.022	-0.025	-0.027	-0.072	-0.123	0.194	-0.128	0.110	0.012	-0.220
ISOT	0.115	-0.070	0.001	0.242	0.203	-0.341	0.028	-0.150	0.296	-0.071	-0.212	-0.115	-0.002
PEOP	-0.121	-0.042	-0.028	0.078	0.068	-0.108	-0.012	-0.025	-0.016	-0.037	-0.139	-0.064	-0.002
ROAD	-0.164	-0.082	-0.034	0.176	0.144	-0.249	-0.020	-0.058	0.147	-0.004	-0.274	-0.140	0.086
TOAD	-0.056	-0.032	-0.011	0.106	0.062	-0.111	-0.014	-0.014	0.077	0.022	-0.074	-0.024	0.036
T1990	-0.052	-0.020	0.002	0.065	0.040	-0.068	-0.009	-0.009	0.047	0.003	-0.050	-0.010	0.016
CALL	-0.029	-0.009	0.003	0.018	0.007	-0.020	-0.002	0.000	-0.002	0.000	-0.019	0.041	0.006
SNAK	-0.093	-0.021	-0.003	0.059	0.032	-0.068	0.004	0.004	0.001	0.023	-0.027	0.016	0.030
S1990	-0.057	-0.012	0.003	0.003	0.007	-0.028	0.014	0.009	-0.028	0.011	-0.002	0.028	0.012

Figure 7-1. The first section of a correlation matrix including the Pearson correlation coefficients among variables used in models predicting both the historical and current distribution of the eastern hog-nosed snake. All variables were at a resolution of 1 km.

Variables	**SETT**	**WATR**	**WETL**	**TMAX**	**PREC**	**ISOT**	**PEOP**	**ROAD**	**TOAD**	**T1990**	**CALL**	**SNAK**	**S1990**
ALTI	-0.155	-0.222	-0.102	-0.489	0.662	0.115	-0.121	-0.164	-0.056	-0.052	-0.029	-0.093	-0.057
SLOP	-0.048	-0.097	-0.021	-0.285	0.095	-0.070	-0.042	-0.082	-0.032	-0.020	-0.009	-0.021	-0.012
ASPE	-0.034	0.014	0.018	-0.002	0.007	0.001	-0.028	-0.034	-0.011	0.002	0.003	-0.003	0.003
GEOG	0.096	0.059	-0.079	0.575	-0.022	0.242	0.078	0.176	0.106	0.065	0.018	0.059	0.003
SAND	0.089	-0.117	-0.040	0.250	-0.025	0.203	0.068	0.144	0.062	0.040	0.007	0.032	0.007
PRBE	-0.131	0.178	0.054	-0.546	-0.027	-0.341	-0.108	-0.249	-0.111	-0.068	-0.020	-0.068	-0.028
ALVA	-0.011	-0.023	0.051	0.000	-0.072	0.028	-0.012	-0.020	-0.014	-0.009	-0.002	0.004	0.014
ROCK	-0.031	0.003	-0.002	-0.113	-0.123	-0.150	-0.025	-0.058	-0.014	-0.009	0.000	0.004	0.009
CROP	-0.089	-0.339	-0.271	0.567	0.194	0.296	-0.016	0.147	0.077	0.047	-0.002	0.001	-0.028
CUBU	-0.036	-0.152	-0.029	-0.089	-0.128	-0.071	-0.037	-0.004	0.022	0.003	0.000	0.023	0.011
FORE	-0.175	-0.227	0.094	-0.544	0.110	-0.212	-0.139	-0.274	-0.074	-0.050	-0.019	-0.027	-0.002
MUDF	-0.080	0.246	0.314	-0.208	0.012	-0.115	-0.064	-0.140	-0.024	-0.010	0.041	0.016	0.028
PAST	-0.011	-0.187	-0.040	0.025	-0.220	-0.002	-0.002	0.086	0.036	0.016	0.006	0.030	0.012
SETT	**1.000**	-0.069	-0.066	0.157	-0.131	-0.037	0.624	0.698	0.036	0.022	0.005	0.036	0.025
WATR	-0.069	**1.000**	-0.020	-0.103	-0.199	-0.056	-0.054	-0.152	-0.043	-0.020	0.009	-0.014	0.007
WETL	-0.066	-0.020	**1.000**	-0.056	-0.169	-0.081	-0.050	-0.109	0.009	0.017	0.050	0.047	0.044
TMAX	0.157	-0.103	-0.056	**1.000**	-0.032	0.356	0.136	0.262	0.121	0.074	0.030	0.088	0.021
PREC	-0.131	-0.199	-0.169	-0.032	**1.000**	0.318	-0.108	-0.106	0.026	-0.004	-0.006	-0.029	-0.038
ISOT	-0.037	-0.056	-0.081	0.356	0.318	**1.000**	-0.043	0.040	0.045	0.013	0.004	0.011	-0.011
PEOP	0.624	-0.054	-0.050	0.136	-0.108	-0.043	**1.000**	0.566	0.034	0.022	0.006	0.037	0.023
ROAD	0.698	-0.152	-0.109	0.262	-0.106	0.040	0.566	**1.000**	0.097	0.052	0.010	0.073	0.037
TOAD	0.036	-0.043	0.009	0.121	0.026	0.045	0.034	0.097	**1.000**	0.329	0.044	0.269	0.155
T1990	0.022	-0.020	0.017	0.074	-0.004	0.013	0.022	0.052	0.329	**1.000**	0.030	0.116	0.133
CALL	0.005	0.009	0.050	0.030	-0.006	0.004	0.006	0.010	0.044	0.030	**1.000**	0.027	0.034
SNAK	0.036	-0.014	0.047	0.088	-0.029	0.011	0.037	0.073	0.269	0.116	0.027	**1.000**	0.578
S1990	0.025	0.007	0.044	0.021	-0.038	-0.011	0.023	0.037	0.155	0.133	0.034	0.578	**1.000**

Figure 7-2. The second section of a correlation matrix including the Pearson correlation coefficients among variables used in models predicting both the historical and current distribution of the eastern hog-nosed snake. All variables were at a resolution of 1 km.

variables	bio2	bio3	bio5	bio6	bio8	bio9	bio13	bio14	bio15	bio18
bio2	1.000	0.298	-0.239	-0.783	-0.125	-0.233	0.113	0.152	0.063	0.130
bio3	0.298	1.000	0.277	0.254	0.036	0.097	0.053	0.143	-0.179	0.341
bio5	-0.239	0.277	1.000	0.668	0.521	0.026	-0.461	-0.440	-0.270	-0.006
bio6	-0.783	0.254	0.668	1.000	0.307	0.250	-0.232	-0.216	-0.227	0.051
bio8	-0.125	0.036	0.521	0.307	1.000	-0.368	-0.596	-0.560	-0.278	0.330
bio9	-0.233	0.097	0.026	0.250	-0.368	1.000	0.512	0.126	0.375	-0.185
bio13	0.113	0.053	-0.461	-0.232	-0.596	0.512	1.000	0.567	0.647	0.159
bio14	0.152	0.143	-0.440	-0.216	-0.560	0.126	0.567	1.000	-0.131	0.346
bio15	0.063	-0.179	-0.270	-0.227	-0.278	0.375	0.647	-0.131	1.000	-0.089
bio18	0.130	0.341	-0.006	0.051	0.330	-0.185	0.159	0.346	-0.089	1.000

Figure 7-3. Correlation matrix including the Pearson correlation coefficients among all variables used in models predicting both the climatic distribution of the eastern hog-nosed snake. All variables were at a resolution of 1 km.

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.

ALLOUCHE, O., A. TSOAR, and R. KADMON. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology. 43:1223-1232.

ANADÓN, J., A. GIMÉNEZ, M. MARTÍNEZ, J. PALAZÓN, and M. ESTEVE. 2007. Assessing changes in habitat quality due to land use changes in the spur-thighed tortoise *Testudo graeca* using hierarchical predictive habitat models. Diversity and Distributions. 13:324-331.

ANDERSON, R. P. 2003. Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. Journal of Biogeography. 30:591-605.

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.

ARAÚJO, M., and A. GUISAN. 2006. Five (or so) challenges for species distribution modelling. Journal of Biogeography. 33:1677-1688.

AUSTIN, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecological modelling. 200:1-19.

BARNOSKY, A. D., N. MATZKE, S. TOMIYA, G. O. U. WOGAN, B. SWARTZ, T. B. QUENTAL, C. MARSHALL, J. L. MCGUIRE, E. L. LINDSEY, K. C. MAGUIRE, B. MERSEY, and E. A. FERRER. 2011. Has the Earth's sixth mass extinction already arrived? Nature. 471:51-57.

BIRDSTUDIESCANADA. 2008. Marsh Monitoring Program. Amphibian Surveys. Data accessed from NatureCounts, a node of the Avian Knowledge Network, Bird Studies Canada. Available: http://www.naturecounts.ca/. Accessed: 7 February 2011.

BLOUIN-DEMERS, G., and P. J. WEATHERHEAD. 2001. Habitat use by black rat snakes (*Elaphe obsoleta obsoleta*) in fragmented forests. Ecology. 82:2882-2896.

BOMBI, P., D. SALVI, L. VIGNOLI, and M. BOLOGNA. 2009. Modelling Bedriaga's rock lizard distribution in Sardinia: An ensemble approach. Amphibia-Reptilia. 30:413-424.

BRAUNISCH, V., K. BOLLMANN, R. F. GRAF, and A. H. HIRZEL. 2008. Living on the edge--Modelling habitat suitability for species at the edge of their fundamental niche. Ecological modelling. 214:153-167.

BROTONS, L., W. THUILLER, M. ARAÚJO, and A. HIRZEL. 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography. 27:437-448.

BROWN, J., G. STEVENS, and D. KAUFMAN. 1996. The geographic range: size, shape, boundaries, and internal structure. Annual review of ecology and systematics. 27:597-623.

BURKE, D., and E. NOL. 2000. Landscape and fragment size effects on reproductive success of forest-breeding birds in Ontario. Ecological Applications. 10:1749-1761.

Butchart, S. H. M., M. Walpole, B. Collen, A. Van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, and J. Bruno. 2010. Global biodiversity: indicators of recent declines. Science. 328:1164.

Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement. 20:37-46.

Conant, R., J. Mitchell, and C. Pague. 1990. Herpetofauna of the Virginia barrier islands. Virginia Journal of Science. 41:364-380.

Cunningham, H., L. Rissler, and J. Apodaca. 2009. Competition at the range boundary in the slimy salamander: using reciprocal transplants for studies on the role of biotic interactions in spatial distributions. Journal of Animal Ecology. 78:52-62.

Cunnington, G. 2004. Eastern Hognose Snake research program – Provincial species at risk year-end report 2003. *In:* Unpublished report to Ontario Ministry of Natural Resources, Ministry of Natural Resources. 17 pages.

Cunnington, G., and J. Cebek. 2005. Mating and nesting behavior of the eastern hognose snake (*Heterodon platirhinos*) in the northern portion of its range. The American Midland Naturalist. 154:474-478.

Darwin, C. 1889. Journal of researches into the natural history and geology of the countries visited during the voyage of HMS "Beagle" round the world. Ward, Lock and co.

Davis, D. D. 1946. Observations on the Burrowing Behavior of the Hog-Nosed Snake. Copeia:75-78.

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.

Dennis, R. L. H., T. H. Sparks, and P. B. Hardy. 1999. Bias in butterfly distribution maps: the effects of sampling effort. Journal of Insect Conservation. 3:33-42.

Dixo, M., J. Metzger, J. Morgante, and K. Zamudio. 2009. Habitat fragmentation reduces genetic diversity and connectivity among toad populations in the Brazilian Atlantic Coastal Forest. Biological Conservation. 142:1560-1569.

Doetsch, P., C. Buck, P. Golik, N. Hoppe, M. Kramp, J. Laudenberg, C. Oberdörfer, 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.

Dubois, Y., G. Blouin-Demers, B. Shipley, and D. Thomas. 2009. Thermoregulation and habitat selection in wood turtles *Glyptemys insculpta*: chasing the sun slowly. Journal of Animal Ecology. 78:1023-1032.

Edgren, R. A. 1955. The natural history of the hog-nosed snakes, genus *Heterodon*: a review. Herpetologica. 11:105-117.

Edwards, T. C., D. R. Cutler, N. E. Zimmermann, L. Geiser, and G. G. Moisen. 2006. Effects of sample survey design on the accuracy of classification tree models in species distribution models. Ecological modelling. 199:132-141.

Elith, J., and C. Graham. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography. 32:66-77.

Elith, J., C. Graham, R. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. Hijmans, F. Huettmann, J. Leathwick, and A. Lehmann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography. 29:129-151.

ELITH, J., J. LEATHWICK, and T. HASTIE. 2008. A working guide to boosted regression trees. Journal of Animal Ecology. 77:802-813.

ENGLER, R., A. GUISAN, and L. RECHSTEINER. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology. 41:263-274.

ERNST, C., and E. ERNST. 2003. Snakes of the United States and Canada. Smithsonian Books Washington, DC.

FAHRIG, L., and G. MERRIAM. 1985. Habitat patch connectivity and population survival. Ecology. 66:1762-1768.

FIELDING, A., and J. BELL. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental conservation. 24:38-49.

FREEMAN, E. A., and G. MOISEN. 2008. PresenceAbsence: An R Package for Presence Absence Analysis. Journal of Statistical Software. 23:1-31.

GASTON, K. J., and A. S. L. RODRIGUES. 2003. Reserve selection in regions with poor biological data. Conservation Biology. 17:188-195.

GIBBONS, J., D. SCOTT, T. RYAN, K. BUHLMANN, T. TUBERVILLE, B. METTS, J. GREENE, T. MILLS, Y. LEIDEN, and S. POPPY. 2000. The global decline of reptiles, déjà vu amphibians. BioScience. 50:653-666.

GIBBONS, J. W. 2006. Forward. *In:* Ecology, Conservation, and Status of Reptiles in Canada. C. N. Seburn and C. A. Bishop (eds.). The Society for the Study of Amphibians and Reptiles, Herpetological Conservation 2., United-States of America.

GREGORY, P. T. 1984. Communal denning in snakes., p. 57-75. *In:* Vertebrate Ecology and Systematics: A Tribute to Henry S. Fitch. A. Seigel, L. E. Hunt, J. L. Knight, L. Malaret, and N. L. Zuschlag (eds.). Museum of Natural History, University of Kansas, Lawrence.

GREGORY, P. T. 2006. Biology and Conservation of a Cold-Climate Snake Fauna. *In:* Ecology, Conservation, and Status of Reptiles in Canada. C. N. Seburn and C. A. Bishop (eds.). The Society for the Study of Amphibians and Reptiles, Herpetological Conservation 2., United-States of America.

GROSS, S., and T. PRICE. 2000. Determinants of the northern and southern range limits of a warbler. Journal of Biogeography. 27:869-878.

GUISAN, A., C. H. GRAHAM, J. ELITH, and F. HUETTMANN. 2007a. Sensitivity of predictive species distribution models to change in grain size. Diversity and Distributions. 13:332-340.

GUISAN, A., A. LEHMANN, S. FERRIER, M. AUSTIN, J. OVERTON, R. ASPINALL, and T. HASTIE. 2006. Making better biogeographical predictions of species' distributions. Journal of Applied Ecology. 43:386-392.

GUISAN, A., J. THEURILLAT, and F. KIENAST. 1998. Predicting the potential distribution of plant species in an alpine environment. Journal of Vegetation Science. 9:65-74.

GUISAN, A., and W. THUILLER. 2005. Predicting species distribution: offering more than simple habitat models. Ecology letters. 8:993-1009.

GUISAN, A., S. WEISS, and A. WEISS. 1999. GLM versus CCA spatial modeling of plant species distribution. Plant Ecology. 143:107-122.

GUISAN, A., and N. ZIMMERMANN. 2000. Predictive habitat distribution models in ecology. Ecological modelling. 135:147-186.

GUISAN, A., N. ZIMMERMANN, J. ELITH, C. GRAHAM, S. PHILLIPS, and A. PETERSON. 2007b. What matters for predicting the occurences of trees: techniques, data, or species' characteristics? Ecological Monographs. 77:615-630.

HEIKKINEN, R., M. LUOTO, R. VIRKKALA, R. PEARSON, and J. KÖRBER. 2007. Biotic interactions improve prediction of boreal bird distributions at macro scales. Global Ecology and Biogeography. 16:754-763.

HERNANDEZ, P., C. GRAHAM, L. MASTER, and D. ALBERT. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography. 29:773-785.

HIJMANS, R., S. CAMERON, J. PARRA, P. JONES, and A. JARVIS. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology. 25:1965-1978.

HIRZEL, A., and A. GUISAN. 2002. Which is the optimal sampling strategy for habitat suitability modelling. Ecological modelling. 157:331-341.

HIRZEL, A., J. HAUSSER, D. CHESSEL, and N. PERRIN. 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? Ecology. 83:2027-2036.

HIRZEL, A., and G. LE LAY. 2008. Habitat suitability modelling and niche theory. Journal of Applied Ecology. 45:1372-1381.

HOLLAND, J., D. BERT, and L. FAHRIG. 2004. Determining the spatial scale of species' response to habitat. BioScience. 54:227-233.

HORTAL, J., A. JIMÉNEZ-VALVERDE, J. F. GÓMEZ, J. M. LOBO, and A. BASELGA. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos. 117:847-858.

HU, J., and Z. JIANG. 2010. Predicting the potential distribution of the endangered Przewalski's gazelle. Journal of Zoology. 282:54-63.

HUTCHINSON, G. E. 1957. Population Studies - Animal Ecology and Demography - Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology. 22:415-427.

JIMÉNEZ-VALVERDE, A. 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecology and Biogeography. 21:498-507.

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

KERR, J. T., and J. CIHLAR. 2004. Patterns and causes of species endangerment in Canada. Ecological Applications. 14:743-753.

LAGORY, K., L. WALSTON, C. GOULET, R. VAN LONKHUYZEN, S. NAJJAR, and C. ANDREWS. 2009. An Examination of Scale-Dependent Resource Use by Eastern Hognose Snakes in Southcentral New Hampshire. Journal of Wildlife Management. 73:1387-1393.

LAHOZ-MONFORT, J. J., G. GUILLERA-ARROITA, E. MILNER-GULLAND, R. P. YOUNG, and E. NICHOLSON. 2010. Satellite imagery as a single source of predictor variables for habitat suitability modelling: How Landsat can inform the conservation of a critically endangered lemur. Journal of Applied Ecology. 47:1094-1102.

LEWIS, S. 2007. Regression analysis. Practical Neurology. 7:259-264.

LIU, C., P. BERRY, T. DAWSON, and R. PEARSON. 2005. Selecting thresholds of occurrence in the prediction of species distributions. Ecography. 28:385-393.

LOBO, J. M., V. JIMÉNEZ, A., and R. REAL. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography. 17:145-151.

LOISELLE, B. A., C. H. GRAHAM, J. M. GOERCK, and M. C. RIBEIRO. 2010. Assessing the impact of deforestation and climate change on the range size and environmental niche of bird species in the Atlantic forests, Brazil. Journal of Biogeography. 37:1288-1301.

MACARTHUR, R. H. 1958. Population ecology of some warblers of northeastern coniferous forests. Ecology. 39:599-619.

MACARTHUR, R. H. 1972. Geographical Ecology: Patterns in the Distribution of Species. Harper and Row, New York.

MANEL, S., H. C. WILLIAMS, and S. J. ORMEROD. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology. 38:921-931.

MARMION, M., M. PARVIAINEN, M. LUOTO, R. K. HEIKKINEN, and W. THUILLER. 2009. Evaluation of consensus methods in predictive species distribution modelling. Diversity and Distributions. 15:59-69.

MAYOR, S. J., D. C. SCHNEIDER, J. A. SCHAEFER, and S. P. MAHONEY. 2009. Habitat Selection at Multiple Scales. Ecoscience. 16:238-247.

MICHENER, M., and J. LAZELL JR. 1989. Distribution and relative abundance of the hognose snake, *Heterodon platirhinos*, in eastern New England. Journal of Herpetology. 23:35-40.

MILLAR, C. S., and G. BLOUIN-DEMERS. 2011. Habitat suitability modelling for species at risk is sensitive to algorithm and scale: A case study of Blanding's turtle, *Emydoidea blandingii*, in Ontario, Canada. Journal for Nature Conservation. In Press.

MILLER, J. 2010. Species Distribution Modeling. Geography Compass. 4:490-509.

MUÑOZ, M. E. S., R. GIOVANNI, M. F. SIQUEIRA, T. SUTTON, P. BREWER, R. S. PEREIRA, D. A. L. CANHOS, and V. P. CANHOS. 2011. OpenModeller: A generic approach to species' potential distribution modelling. GeoInformatica. 15:111-135.

NATURESERVE. 2010. NatureServe Explorer: An online encyclopedia of life [web application]. Version 7.1. NatureServe, Arlington, Virginia.

NEWBOLD, T. 2010. Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography. 34:3.

OGS. 2010. Surficial geology of southern Ontario; Ontario Geological Survey. Data Set 128.

OLDHAM, M., AND M. AUSTEN. 1998. *Heterodon platirhinon:* NHIC element report. Natural Heritage Information Centre, Ontario Ministry of Natural Resources, Peterborough, Ontario.

OLDHAM, M., AND W. WELLER. 2000. Ontario Herpetofaunal Atlas. Natural Heritage Information Centre, Ontario Ministry of Natural Resources, Peterborough, Ontario.

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

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

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

ONTARIONATURE. 2011. Ontario Nature. The Ontario Reptile and Amphibian Atlas. Available: http://www.ontarionature.org Accessed: 7 February 2011.

ORCHARD, S. A. 2006. Introduction: The Conservation of Reptiles in Canada. *In:* Ecology, Conservation and Status of Reptiles in Canada. C. N. Seburn and C. A. Bishop (eds.). The Society for the Study of Amphibians and Reptiles, Herpetological Conservation 2., United-States of America.

ORTEGA-HUERTA, M., and A. PETERSON. 2008. Modeling ecological niches and predicting geographic distributions: a test of six presence-only methods. Revista Mexicana de Biodiversidad. 79:205-216.

PEARSON, R. 2007. Species' distribution modeling for conservation educators and practitioners. Synthesis. American Museum of Natural History. 50.

PEARSON, R., and T. DAWSON. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography. 12:361-371.

PEARSON, R., T. DAWSON, and C. LIU. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. Ecography. 27:285-298.

PEARSON, R., C. RAXWORTHY, M. NAKAMURA, and A. PETERSON. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography. 34:102-117.

PENMAN, T., D. PIKE, J. WEBB, and R. SHINE. 2010. Predicting the impact of climate change on Australia's most endangered snake, *Hoplocephalus bungaroides*. Diversity and Distributions. 16:109-118.

PEREIRA, H. M., P. W. LEADLEY, V. PROENÇA, R. ALKEMADE, J. P. W. SCHARLEMANN, J. F. FERNANDEZ-MANJARRÉS, M. B. ARAÚJO, P. BALVANERA, R. BIGGS, and W. W. L. CHEUNG. 2010a. Scenarios for global biodiversity in the 21st century. Science. 330:1496.

PEREIRA, H. M., P. W. LEADLEY, V. PROENÇA, R. ALKEMADE, J. P. W. SCHARLEMANN, J. F. FERNANDEZ-MANJARRÉS, M. B. ARAÚJO, P. BALVANERA, R. BIGGS, W. W. L. CHEUNG, L. CHINI, H. D. COOPER, E. L. GILMAN, S. GUÉNETTE, G. C. HURTT, H. P. HUNTINGTON, G. M. MACE, T. OBERDORFF, C. REVENGA, P. RODRIGUES, R. J. SCHOLES, U. R. SUMAILA, and M. WALPOLE. 2010b. Scenarios for Global Biodiversity in the 21st Century. Science. 330:1496-1501.

PETERSON, A. 2006. Uses and requirements of ecological niche models and related distributional models. Biodiversity Informatics. 3:59-72.

PETERSON, A. T., M. PAPEŞ, and M. EATON. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. Ecography. 30:550-560.

PETERSON, A. T., M. PAPEŞ, and J. SOBERÓN. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological modelling. 213:63-72.

PHILLIPS, S., R. ANDERSON, and R. SCHAPIRE. 2006. Maximum entropy modeling of species geographic distributions. Ecological modelling. 190:231-259.

PHILLIPS, S., and M. DUDÍK. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography. 31:161-175.

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.

PICARD, G., M. A. CARRIÈRE, and G. BLOUIN-DEMERS. 2011. Common Musk Turtles (*Sternotherus odoratus*) select habitats of high thermal quality at the northern extreme of their range. Amphibia-Reptilia. 32:83-92.

PIMM, S., G. RUSSELL, J. GITTLEMAN, and T. BROOKS. 1995. The future of biodiversity. Science. 269:347.

PIMM, S. L., and P. RAVEN. 2000. Biodiversity: extinction by numbers. Nature. 403:843-845.

PLATT, D. 1969. Natural history of the hognose snakes, *Heterodon platyrhinos* and *Heterodon nasicus*. University of Kansas, Museum of Natural History. 18:253-420.

PLUMMER, M., and N. MILLS. 2000. Spatial ecology and survivorship of resident and translocated hognose snakes (*Heterodon platirhinos*). Journal of Herpetology. 34:565-575.

PLUMMER, M., and N. MILLS. 2010. Body Temperature Variation in Free-Ranging Hognose Snakes (*Heterodon platirhinos*). Journal of Herpetology. 44:471-474.

POWELL, G. L., and A. P. RUSSELL. 2006. Life History Implications for Conservation and Monitoring of Lizards in Canada. *In:* Ecology, Conservation and Status of Reptiles in Canada. C. N. Seburn and C. A. Bishop (eds.). The Society for the Study of Amphibians and Reptiles, Herpetological Conservation 2., United-States of America.

R. Development Core Team. 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RANDIN, C., T. DIRNBÖCK, S. DULLINGER, N. ZIMMERMANN, M. ZAPPA, and A. GUISAN. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography. 33:1689-1703.

RIDENOUR, T. A., and A. C. HEATH. 1999. A note on issues in meta-analysis for behavioral genetic studies using categorical phenotypes. Behavior genetics. 29:155-162.

RIDGEWAY, G. 2010. Generalized boosted regression models. Documentation on the R Package 'gbm', version 1.6-3.1. Available: http://cran.r-project.org/web/packages/gbm/, Accessed June 2011.

ROBSON, L. E. 2011. The spatial ecology of Eastern Hognose Snakes (*Heterodon platirhinos*): habitat selection, home range size, and the effect of roads on movement patterns. *In:* Faculty of Graduate and Postdoctoral Studies. M.Sc. University of Ottawa. Ottawa.

ROUSE, J. D., R. J. WILLSON, R. BLACK, and R. J. BROOKS. 2011. Movement and Spatial Dispersion of *Sistrurus catenatus* and *Heterodon platirhinos*: Implications for Interactions with Roads. Copeia:443-456.

ROW, J. R., and G. BLOUIN-DEMERS. 2006. Thermal quality influences habitat selection at multiple spatial scales in milksnakes. Ecoscience. 13:443-450.

RUPPRECHT, F., J. OLDELAND, and M. FINCKH. 2011. Modelling potential distribution of the threatened tree species *Juniperus oxycedrus*: how to evaluate the predictions of different modelling approaches? Journal of Vegetation Science. 22:647-659.

SANTOS, X., J. BRITO, J. CARO, A. ABRIL, M. LORENZO, N. SILLERO, and J. PLEGUEZUELOS. 2009. Habitat suitability, threats and conservation of isolated populations of the smooth snake (*Coronella austriaca*) in the southern Iberian Peninsula. Biological Conservation. 142:344-352.

SANTOS, X., J. C. BRITO, N. SILLERO, J. M. PLEGUEZUELOS, G. A. LLORENTE, S. FAHD, and X. PARELLADA. 2006. Inferring habitat-suitability areas with ecological modelling techniques and GIS: A contribution to assess the conservation status of *Vipera latastei*. Biological Conservation. 130:416-425.

SEBURN, C. N., and C. A. BISHOP. 2006. Ecology, Conservation and Status of Reptiles in Canada. The Society for the Study of Amphibians and Reptiles, Herpetological Conservation 2., United-States of America.

SEBURN, D. 2009. Recovery Strategy for the Eastern Hog-nosed Snake (*Heterodon platirhinos*) in Canada, p. 24. *In:* Species at Risk Act Recovery Strategy Series. Parks Canada Agency, Ottawa.

SEBURN, D., and C. SEBURN. 2000. Conservation priorities for the amphibians and reptiles of Canada. World Wildlife Fund Canada, Toronto, ON, and the Canadian Amphibian and Reptile Conservation Network, Delta, BC.

SEIGEL, R., and J. COLLINS. 1993. Snakes: ecology and behavior. McGraw-Hill Companies.

SOBERON, J. 2007. Grinnellian and Eltonian niches and geographic distributions of species. Ecology letters. 10:1115-1123.

STATISTICSCANADA. 2006. 2006 Census of Population. Statistics Canada, Catalogue no. 92-160-GIE, Ottawa, Canada.

STOCKWELL, D. 2007. Niche modeling: predictions from statistical distributions. Chapman & Hall/CRC, Boca Raton, Florida.

STOCKWELL, D., and D. PETERS. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of Geographical Information Science. 13:143-158.

STOKSTAD, E. 2010. Despite Progress, Biodiversity Declines. Science. 329:1272-1273.

SURFACE, H. A. 1906. The serpents of Pennsylvania. *In:* Monthly bulletin of the Division of Zoology. Vol. 4. Pennsylvania. Dept. of Agriculture, Harrisburg, Pennsylvania.

TAYLOR, J., and M. MARCHAND. 2005. Species Profile, Hognose Snake, *Heterodon platirhinos.*, p. 184-189. Vol. New Hampshire Wildlife Action Plan. New Hampshire Heritage Bureau.

TERRIBILE, L. C., J. A. F. DINIZ, and P. DE MARCO. 2010. How many studies are necessary to compare niche-based models for geographic distributions? Inductive reasoning may fail at the end. Brazilian Journal of Biology. 70:263-269.

THUILLER, W., and T. MÜNKEMÜLLER. 2010. Habitat suitability modelling. Effects of Climate Change on Birds:77.

TYLIANAKIS, J., R. DIDHAM, J. BASCOMPTE, and D. WARDLE. 2008. Global change and species interactions in terrestrial ecosystems. Ecology letters. 11:1351-1363.

UHLER, F., C. COTTAM, and T. CLARKE. 1939. Food of snakes of the George Washington National Forest, Virginia, p. 605-622. Vol. 4.

URBAN, D., R. O'NEILL, and H. SHUGART. 1987. A hierarchical perspective can help scientists understand spatial patterns. Landscape Ecology. 37:119-127.

VEGA, R., C. FLOJGAARD, A. LIRA-NORIEGA, Y. NAKAZAWA, J.-C. SVENNING, and J. B. SEARLE. 2010. Northern glacial refugia for the pygmy shrew *Sorex minutus* in

Europe revealed by phylogeographic analyses and species distribution modelling. Ecography. 33:260-271.

WANG, Z., Y. I. CHANG, Z. YING, L. ZHU, and Y. YANG. 2007. A parsimonious threshold-independent protein feature selection method through the area under receiver operating characteristic curve. Bioinformatics. 23:2788.

WIENS, J. 1989. Spatial scaling in ecology. Functional ecology. 3:385-397.

WISZ, M., R. HIJMANS, J. LI, A. PETERSON, C. GRAHAM, and A. GUISAN. 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions. 14:763-773.

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

YOUNG, N. E., T. J. STOHLGREN, P. H. EVANGELISTA, S. KUMAR, J. GRAHAM, and G. NEWMAN. 2011. Regional data refine local predictions: modeling the distribution of plant species abundance on a portion of the central plains. Environmental Monitoring and Assessment:1-13.

ZIELINSKI, W., J. DUNK, J. YAEGER, and D. LAPLANTE. 2010. Developing and testing a landscape-scale habitat suitability model for fisher (*Martes pennanti*) in forests of interior northern California. Forest Ecology and Management. 260:1579–1591.

ZIMMERMANN, N., T. EDWARDS JR, C. GRAHAM, P. PEARMAN, and J. SVENNING. 2010. New trends in species distribution modelling. Ecography. 33:985-989.