

---

## Original Research

---

# Using Habitat Suitability Models Considering Biotic Interactions to Inform Critical Habitat Delineation: An Example with the Eastern Hog-nosed Snake (*Heterodon platirhinos*) in Ontario, Canada

Victor THOMASSON<sup>1</sup> and Gabriel BLOUIN-DEMERS<sup>1</sup>

<sup>1</sup> Department of Biology, University of Ottawa, 30 Marie Curie, Ottawa, ON, K1N 6N5, Canada

## Abstract

Habitat suitability models have been used in a variety of fields, including conservation, and are considered a powerful tool to model the potential niche of species. Presence-only models have been particularly useful to define suitable habitat for rare species at a landscape scale. In Canada, the *Species at Risk Act* not only protects species at risk, but also their residences and critical habitats. It is thus necessary to identify which geographic areas species at risk depend on and which habitats can be considered suitable. In this study, we identify areas of high suitability for the eastern hog-nosed snake, *Heterodon platirhinos*, in Ontario, Canada. We employ three models - Maxent, Boosted Regression Trees, and the Genetic Algorithm for Rule Set Production (GARP) - to model the current distribution of the species. Because the eastern hog-nosed snake is a diet specialist, we also assess the importance of biotic interactions in habitat suitability models by including variables representing prey availability. The best models were combined using a consensus approach and categorical maps showing 4 conservation scenarios were built. Maxent and Boosted Regression Trees performed better than GARP. While forest density is positively related to habitat suitability, cropland density limits the distribution of this snake. Climate also played an important role in shaping the distribution of this species. Biotic variables allowed better interpretation of the predictions made by the models by reflecting spatial bias in sampling. We discuss how habitat suitability models can help delineate the critical habitat of species at risk and whether variables representing biotic interactions should be included.

**Key Words:** Boosted Regression Trees, Conservation, Maxent Habitat Model.

## INTRODUCTION

Current rates of extinction are 10 to 1000 times what they were before human presence and are predicted to continue to rise as the human population increases (Pimm *et al.* 1995; Butchart *et al.* 2010; Pereira *et al.* 2010). The primary factors causing extinctions are habitat loss and habitat fragmentation; for instance, an estimated one million square kilometers of rainforest are cut every 5 to 10 years (Pimm and Raven 2000). In Canada, most species at risk are found in the southern portions of the country where biodiversity is the richest, but where human activity is also the highest (Kerr and Cihlar 2004). In response to the biodiversity crisis, many governments have enacted legislations to regulate anthropogenic activity in relation to species at risk. For instance, in Canada, the *Species at Risk Act* proclaimed in 2003 protects species at risk on federal lands (SARA 2002); both the residence (an area that is occupied by one or more individuals during all or part of their life cycle) and the critical habitat (habitat that is necessary for the survival or recovery of a wildlife species) of species at risk are protected. According to the Act, critical habitat should be defined “to the extent possible” (S.C. 2002, c.29, s.33, 41, 58, 61). Sufficient information on the habitat requirements of species at risk is required to enforce habitat-related legislation. In Ontario, the Endangered Species Act of 2007 is an equivalent piece of legislation that aims at protecting both threatened and endangered species as well as their habitat on provincial lands (ESA 2007).

The eastern hog-nosed snake is found in the USA and Canada, and is known for its large home range (51.7 ha on average, Lagory *et al.* 2009). In Canada, its distribution is limited to 2 populations in southern Ontario: one near Long Point and the other around Georgian Bay (Seburn 2009). Both of these regions are heavily impacted by human activity. In southern Ontario, the proximity to the Great Lakes, the long growing season, and the fertile soil have resulted in intensive agricultural activity. Many cities have also expanded, resulting in major habitat loss and fragmentation. Georgian Bay, on the other hand, is a popular destination for recreational activities. Due to increasing development in southern Ontario, the species has been losing its habitat at an alarming rate and is now considered threatened by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) and The Committee on the Status of Species at Risk in Ontario (COSSARO). Eastern hog-nosed snakes are also known for their extensive movements, making the species particularly susceptible to road mortality (Rouse *et al.* 2011; Robson and Blouin-Demers 2013). Although other studies have looked at habitat selection of this species at the scale of its home range (Plummer and Mills 2000; Lagory *et al.* 2009; Rouse *et al.* 2011; Robson and Blouin-Demers 2013), it is crucial to determine the habitat requirements at the scale of the province to identify areas that should be protected for its long-term survival.

Geographic information systems (GIS) have been used to study habitat selection at various scales (e.g., Row and Blouin-Demers 2006). High-resolution spatial environmental data coupled with species occurrence records found in atlases render possible the study of habitat associations at large spatial scales at low cost (e.g., Millar and Blouin-Demers 2012). Although species occurrences kept by museums and governmental databases present a number of limitations, including taxonomic and spatial biases, they are sometimes the best we have (Newbold 2010). Atlases containing occurrences that cover many decades may also be the only source of information used to report changes in species distribution. The increasing availability of spatial data has encouraged the development of new habitat modelling approaches (Elith *et al.* 2006; Guisan *et al.* 2006). Habitat suitability models (HSMs) relate species occurrences to an array of environmental parameters based on statistical algorithms (Guisan and Zimmermann 2000). Although some HSMs require both presence and absence data, presence-only HSMs characterizing habitat availability using pseudo-absences are effective for modelling species distributions (Elith *et al.* 2006). HSMs have been successfully used to find unknown populations (Araújo and Williams 2000), to predict the effect of climate change and habitat loss on biodiversity (Hu and Jiang 2010), and to assess the potential risk of invasive species (Pyron *et al.* 2008). With respect to species at risk, these models are a powerful tool that can help determine where habitat suitability is the highest (Hu and Jiang 2010). HSMs have been used extensively to study endangered and rare species (Cianfrani *et al.* 2010; Marino *et al.* 2011; Mbatudde *et al.* 2012), but few studies integrate these models with conservation legislation. Could HSMs help delineate critical habitat of species at risk as mandated by legislations such as the Species at Risk Act of Canada and the Endangered Species Act of Ontario? Our first objective is to use HSMs to delineate areas of high habitat suitability for the eastern hog-nosed snake, *Heterodon platirhinos*, in Ontario, Canada, and to identify the factors responsible for the current distribution of this species. We also propose 3 thresholds that may be used for conservation applications, and produce maps that highlight the areas that are the most important for the species.

One important limitation of HSMs that was 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. One of the reasons biotic variables are often omitted from HSMs is that, with the exception of satellite imagery, appropriate and reliable biological data are rarely available. Biological data, such as species occurrences, collected according to an established sampling design, are not readily available for most taxa. Many variables contribute to shaping the distribution of species, including biotic and abiotic factors (Cunningham *et al.* 2009). 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 or the

presence of hardwood forest (Lahoz-Monfort *et al.* 2010; Loisel *et al.* 2010). Occurrences of other competing species have also been used (Santos *et al.* 2006; Cunningham *et al.* 2009). The few studies that have included biotic variables in HSMs suggest that realized distributions are indeed affected by biotic interactions (Heikkinen *et al.* 2007; Pellissier *et al.* 2010), even at larger spatial scales (Wis *et al.* 2013). The eastern hog-nosed snake is well suited to study the effect of biotic interactions in HSMs because it has a specialized diet, feeding mostly on toads (Uhler *et al.* 1939; Edgren 1955; Platt 1969). Toads constitute between 40% (Uhler *et al.* 1939) and 75% (Surface 1906) of the diet of eastern hog-nosed snakes. Our second objective is to assess the role of biotic interactions in HSMs by including explanatory variables representing prey availability. We hypothesize that prey availability constrains the distribution of the eastern hog-nosed snake to some areas only, preventing it from occupying its entire potential distribution.

## METHODOLOGY

### Species occurrences and explanatory variables

All eastern hog-nosed occurrences were retrieved from the Natural Heritage Information Centre of Ontario (Oldham and Austen 1998) and 2 studies conducted in Ontario (Cunnington *et al.* 2005; Robson Blouin-Demers 2013). To reduce spatial autocorrelation and overrepresentation of highly sampled regions, only a subset of presences separated by at least 1 km were identified and retained, using Focus Tool (Holland *et al.* 2004). The final dataset consisted of 126 presences (between 1990 and 2011) with an accuracy  $\leq 100$  m. The variables used in the models can be grouped in 5 categories: topographical, geological, climatic, land cover, and anthropogenic variables, as well as biotic variables (Table 1). The only categorical variable was GEOG, representing the surficial geology of Ontario, and it was used to calculate sand

Table 1. Variables used to predict the distribution of the eastern hog-nosed snake in Ontario, Canada.

Type	Code	Description and units	Data resolution	Source of original data
Topographical	ALTI	Elevation (m)	10-20 m	(OMNR 2005)
	SLOP	Slope (°)	10-20 m	(OMNR 2005)
	ASPE	Aspect (°)	10-20 m	(OMNR 2005)
Geological	GEOG	Surficial geology (8 categories)	n/a	(OGS 2010)
	SAND <sup>a</sup>	Sand density (%)	n/a	(OGS 2010)
	PRBE <sup>a</sup>	Precambrian bedrock density (%)	n/a	(OGS 2010)
Land Cover	ALVA	Alvar density (%)	25 m	(OMNR 1998)
	ROCK	Bedrock density (%)	25 m	(OMNR 1998)
	CROP	Cropland Density (%)	25 m	(OMNR 1998)
	CUBU	Cuts and burns density (%)	25 m	(OMNR 1998)
	FORE	Forest density (%)	25 m	(OMNR 1998)
	MUDF	Mudflat density (%)	25 m	(OMNR 1998)
	PAST	Pasture density (%)	25 m	(OMNR 1998)
	SETT	Settlement density (%)	25 m	(OMNR 1998)
	WATR	Water density (%)	25 m	(OMNR 1998)
	WETL	Wetland density (%)	25 m	(OMNR 1998)
Climatic	TMAX	Mean maximum temperature for the active season (°C*10)	1 km	(Hijmans <i>et al.</i> 2005)
	PREC	Mean monthly precipitation for the active season (mm)	1 km	(Hijmans <i>et al.</i> 2005)
	ISOT	Isothermality (%)	1 km	(Hijmans <i>et al.</i> 2005)
Anthropogenic	PEOP	Total population density (people/km <sup>2</sup> )	n/a	(Statistics Canada 2006)
	ROAD	Total road density (kernel, m/km <sup>2</sup> )	n/a	(OMNR 2006)
Biological	T1990	Toad occurrences since 1990	n/a	(Oldham and Weller 2000; Nature 2011)
	CALL	Toad calls since 1995	n/a	(Bird Studies Canada 2008)
	S1990	Snake occurrences since 1990	n/a	(Ontario Nature 2011)

<sup>a</sup> SAND and PRBE were only used with GARP models because it does not handle categorical variables.

density (SAND) and Precambrian bedrock density (PRBE). To build a variable representing human population density (PEOP), the number of people found in each dissemination area (areas occupied by 400 to 700 individuals) of the 2006 Canadian Census was divided by its area in km<sup>2</sup> (Statistics Canada 2006). All climatic variables were retrieved from the WorldClim database (Hijmans *et al.* 2005). For the maximum temperature and the mean monthly precipitation for the active season, the values for the 6 months of active season were averaged (April - September). None of these variables were highly correlated (all  $|r| < 0.8$ , mean  $|r| = 0.08$ ). In Canada, the eastern hog-nosed snake consumes both American toads, *Anaxyrus americanus*, and Fowler's toads, *Anaxyrus fowleri* (Platt 1969). Occurrences of these 2 species were used to build a categorical variable representing the incidence of prey; 1 = presence, 0 = absence (Table 1). Data from toad call surveys were used to build a continuous variable representing relative toad abundance. For that variable, the number on each pixel represented the number of toads calling; a full chorus was assigned the number 25.

Finally, a control variable was built with the occurrences of 4 snake species: northern watersnake (*Nerodia sipedon*), smooth greensnake (*Opheodrys vernalis*), Dekay's brownsnake (*Storeria dekayi*), and eastern gartersnake (*Thamnophis sirtalis*) (Ontario Nature 2011) (Table 1). This variable was employed to assess sampling bias: models should not respond to the presence of other snake species in the same way they respond to prey availability (toads). If this snake occurrence variable increases the predictive power of the models, it would suggest that areas sampled for eastern hog-nosed snakes were also sampled for other snake species.

#### Models employed

Three machine learning algorithms with high predictive power were employed to predict the distribution of the eastern hog-nosed snake: Maxent, 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 models for 226 species in 6 regions of the World (Elith *et al.* 2006). GARP 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 one) based on the species' occurrences and the explanatory variables employed. The version 3.3.3e of the Maxent software was used 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. 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 combine the strengths of classification trees (also known as regression trees) 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 complex 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 BRT models were fitted in R version 2.12.0 (R Development Core Team 2011) using both the gbm package (Ridgeway 2010) and other codes for R specific to BRTs (Elith *et al.* 2008). All models were built with 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; Young *et al.* 2011; Millar *et al.* 2012).

The Genetic Algorithm for Rule Set Production (GARP) is a machine-learning method that 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 29°C = 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). All GARP models were implemented in openModeller (Muñoz *et al.* 2011) following a procedure that allows identifying 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: the 20% of models with the least omissions from which 50% with a commission rate closest to the median were kept. Because GARP does not perform well with categorical variables (Elith and Leathwick 2009), the variable representing surficial geology (GEOG) was substituted with sand density (SAND) and Precambrian bedrock density (PRBE) in all GARP models (Table 1).

#### Modelling approach

A first series of models was built employing all 3 algorithms, without considering biotic interactions; 75% of the occurrences were used to build the models, and 25%, chosen randomly, to validate them. All models were built at 3 resolutions (1 km, 750 m, and 500 m), but results are only presented at a resolution of 1 km because all models achieved similar performance and gave qualitatively similar results (Thomasson 2012). The study area extended 1 km in all directions beyond the historical distribution

of the eastern hog-nosed snake in Ontario (all occurrences ever reported to the government). This area excludes all of the Great Lakes. Considering the number of snake sightings available, there were too many variables (up to 20) to build full models. To reduce their number, the variables that contributed the least to the models were omitted. A technique similar to backward stepwise selection was employed aiming at optimizing the Area Under the Curve (AUC) (Doetsch *et al.* 2009). Taking out the variables with the lowest contributions, the models were pruned until the highest AUCs were reached. Similar variable reduction approaches have been used in various fields including medicine (Wang *et al.* 2007; Abeel *et al.* 2010) and ecology (Millar and Blouin-Demers 2012). In addition, 10,000 pseudo-absences were generated randomly in the area of study to run all models. Again, 75% of these points were used to build the models and 25% to validate them. For a better comparison of the algorithms, the same subsets were always used to build and validate models across algorithms. To reduce the uncertainty associated with each algorithm, the predictions given by the best models were combined using a consensus method proposed by Marmion *et al.* (2009): the arithmetic mean of the matrices produced by the models. The model could therefore be composed of 1, 2 or 3 algorithms depending on their individual scores. This approach was chosen because HSMs are difficult to evaluate and their performances are influenced by a number of factors including the resolution (Guisan *et al.* 2007), the sample size (Wisz *et al.* 2008), the sampling design (Dennis *et al.* 1999), and the prevalence of the species (Hernandez *et al.* 2006).

A second series of models was built to assess the role of biotic interactions in habitat suitability models. These models were built by adding the biotic variables to the pruned Maxent model one at a time. Three additional models predicting the distribution of the eastern hog-nosed snake were built: 1 with recent toad occurrences (T1990), 1 with toad calls (CALL), and the last with snake occurrences (S1990). The modelling approach was the same as the one used to build the first series of models.

#### Variable contributions and response curves

To assess the relationship between each variable and the distribution of the species, partial dependence plots were constructed for the models that were retained and the contributions of the variables were determined. With Maxent, the contribution of each variable was determined by looking at the increase in gain of the model when modifying the coefficient of a single feature (Phillips *et al.* 2006). This increase in gain is then assigned to the environmental variable(s) that the feature depends upon. At the end of the training process, these increases in gain are converted to percentages. With BRTs, the contributions of the variables are calculated by looking at the number of times the variables are selected, weighted by the improvement of the model at each split (Elith *et al.* 2008). The contribution of a variable is then estimated by calculating its average importance in all trees. Finally, with

GARP, jackknife analyses were performed to estimate the accuracy of the predictions made by the model when a single variable is removed (Muñoz *et al.* 2011).

#### Validation

Although many studies only use Receiver Operating Characteristics (ROC) analyses to validate habitat suitability models, this technique has been criticized (Lobo *et al.* 2008; Peterson *et al.* 2008). For example, ROC ignores the goodness-of-fit of the predictions and treats omission and commission errors equally (Lobo *et al.* 2008). To better assess the performance of our models and to avoid relying strictly on ROC analyses, we used several validation techniques, including novel threshold-dependent metrics. Both sensitivity (the proportion of correctly classified presences) and specificity (the proportion of correctly classified absences) are presented. A sensitivity and specificity of 1 could be interpreted as the perfect prediction of the test dataset. Cohen's Kappa (Cohen 1960) and true skill statistic (TSS) (Allouche *et al.* 2006) were also calculated, both correcting the overall accuracy of model predictions by the accuracy expected to occur by chance. TSS is independent of prevalence (how common the species is in the area of study), uses both commission and omission errors, and was shown to be a good method to evaluate HSMs (Allouche *et al.* 2006). For both of these metrics, +1 indicates perfect agreement between the model and the test dataset, <1 to >0 better than random, and 0 to -1 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 selection method was shown to perform well with presence only data (Liu *et al.* 2005, 2013), and is suitable for this study considering the goal to predict where current suitable habitats are located. With respect to attributing ordinal scores to models, the receiver operating characteristic (ROC) analysis (Fielding and Bell 1997) is an effective technique used in the majority of studies evaluating HSMs. It looks at the ability of a model to differentiate between true positives and false positives. The area under the ROC curve (AUC) must be determined and should range from 0 to 1. The scores can be interpreted as follows: 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 2 AUC values are presented, one which was calculated with the software of each model (Internal AUC) and the other calculated independently in R (External AUC). This second AUC value was calculated the same way for all algorithms, always treating the same 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. Finally, a metric called the minimal predicted area (MPA) was calculated for all models (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 and minimizing model overfit (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 habitat for the eastern hog-nosed snake

Most 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 depends on the prevalence of the species, the algorithm used, and the quality of the data. More importantly, however, the critical habitat should be delineated based on expert knowledge and through consultations with expert practitioners, but suitable habitat identified via HSMs would be very valuable to inform such discussions. One way, among others, to address this question is to consider all probabilities of presence above a certain threshold as critical for a given model. Although they might not always be critical, 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 (Engler *et al.* 2004; Rupprecht *et al.* 2011). The area comprised in the MPA not only includes most known occurrences, but areas of high suitability. Such a threshold is not subjective, as it depends on the occurrences available. In addition to the threshold used to calculate the MPA, we propose 2 additional examples of thresholds to delineate critical habitat: one allowing the predictions of 50% of the occurrences and the other 70%. Maps produced with these thresholds therefore show 4 suitability categories: excellent (i.e., the range of suitability that captures 50% of presences in the best habitat), good (i.e., the range of suitability that captures 20% of presences in good habitat), moderate (i.e., the range of suitability that captures 20% of presences occupying the next good habitat), and poor (i.e., the range of suitability that captures 10% of presences occupying the

poorest habitat). 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 arbitrary threshold values 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

#### Performance of the HSMs built without biotic variables

Maxent and BRTs performed better than GARP in most models. Maxent and BRTs had higher AUCs, specificities, sensitivities and TSSs, and lower MPAs (Table 2). For this reason, consensus models were built using only Maxent and BRTs (Figure 1). For most models, the differences in performance between Maxent and BRTs were small making it impossible to say which algorithm made the best predictions under these conditions. Maxent usually required fewer variables (on average 9) than BRTs (on average 14) to perform well and achieve high AUCs. Maxent, BRTs and the consensus models always performed well with AUCs ranging between 0.89 and 0.93. With values always above 0, the Kappa statistics and TSS both indicate that all models performed better than what could be expected by chance (Table 2). The models that had the highest Kappa statistics were BRTs with a value of 0.13. It is important to note that the predicted proportion of highly suitable areas varied a lot from one algorithm to another and that the threshold that maximized the sum of specificity and sensitivity was highly dependent on the algorithm. BRT models had a low MPA which reduced chances of having false presences (Figure 2). Accordingly, maps produced by this algorithm had few areas of high suitability. GARP does the opposite, having high MPAs and assigning high probabilities to most of the area of study. Maxent had fewer areas of high suitability than GARP, but more than BRTs (Figure 2). All models predicted fewer areas of high suitability in the south of the province near Long Point. The area where habitat suitability remains high for the eastern hog-nosed snake is the region of Georgian Bay where the northern population is located. Finally, a consensus habitat suitability map presenting 4 conservation scenarios was built by applying the 3 thresholds

Table 2. Metrics of performance for all models used to predict the distribution of the eastern hog-nosed snake in Ontario at a resolution of 1 km.

Model	Number of variables	Internal AUC	External AUC	Threshold <sup>a</sup>	Sensitivity	Specificity	Kappa	TSS	MPA
Maxent	9	0.896	0.93	0.192	0.94	0.84	0.1	0.77	0.16
BRT	14	0.934	0.93	0.01	0.91	0.87	0.13	0.78	0.052
GARP	9	0.88	0.88	0.649	0.84	0.78	0.068	0.62	0.443
Consensus	n/a	n/a	0.93	0.104	0.94	0.83	0.11	0.77	0.144

<sup>a</sup> The threshold employed to calculate threshold dependent metrics maximized the sum of sensitivity and specificity.

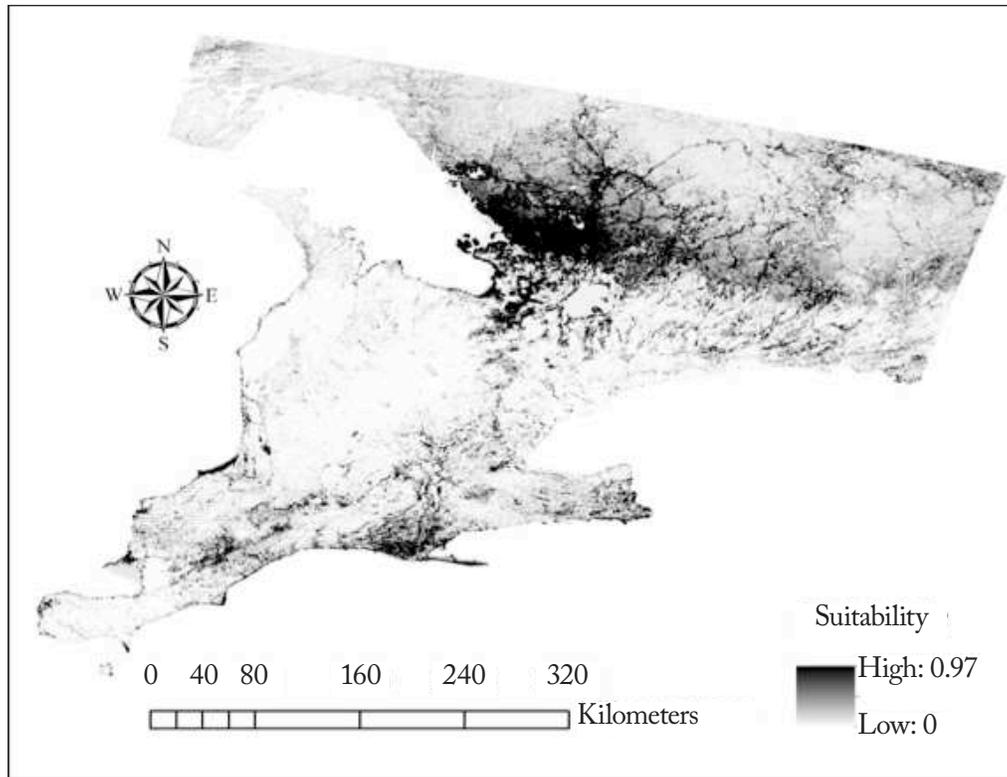


Figure 1. Habitat suitability map for the eastern hog-nosed snake in Ontario using the consensus approach (Maxent and BRT) at a resolution of 1 km.

described above (Figure 3).

#### Contributions of variables in HSMs built without biotic variables

The contribution of each variable in the models varied from one algorithm to another, but usually remained similar between resolutions (Thomasson 2012). We only present the contributions of the variables for Maxent and BRT models because these were the algorithms used to build the consensus model (Figure 4). In order of importance, the 5 variables with the highest contributions in the Maxent model predicting the current distribution of the eastern hog-nosed snake were: cropland density (20.5%), isothermality (19.8%), mean maximum temperature during the active season (15.5%), surficial geology (14.4%), and forest density (9.8%). In the model produced with BRTs the best explanatory variables were forest density (15.6%), mean maximum temperature during the active season (11.0%), altitude (8.8%), road density (7.9%), and total population density (7.5%). The probability of suitability increases with increasing forest density and decreases as cropland density increases. The marginal response to road density is an increase in suitability until road density reaches  $\sim 15$  m/km<sup>2</sup>; this is partially due to a sampling bias (Figure 5). Finally, these models show that eastern hog-nosed snake sightings were generally in forested areas, near roads, and where maximum summer temperatures are relatively high.

#### Inclusion of biotic variables in HSMs

Most changes in the metrics of performance were very small when biotic variables were added to our Maxent models (Table 3). Out of the 3 models built with an additional biotic variable, the only one that performed better than the original models was the one that included the variable representing snake occurrences (S1990). For that model, the internal AUC increased by 0.001 and the MPA decreased by  $\sim 0.01\%$ . The 2 models built with toad variables (T1990 and CALL) performed more poorly as their internal AUCs, Kappas, and TSSs decreased and their MPAs increased. Biotic variables added to the models were therefore inadequate to help better predict habitat suitability for the eastern hog-nosed snake. The contributions of the biotic variables in the models were variable. In the model built with an additional variable representing snake occurrences, the variable S1990 was the variable with the highest contribution (19%). In the 2 other models including an extra biotic variable, the contributions of T1990 and CALL were low (3.5% and 0.1%, respectively). This shows that a biotic variable representing the distribution of other snake species contributes more to models predicting the distribution of the Eastern hog-nosed snake than variables representing prey availability (toad distribution and toad relative abundance). The biotic variables increased the probability of occurrence of the Eastern hog-nosed snake in models where they had significant

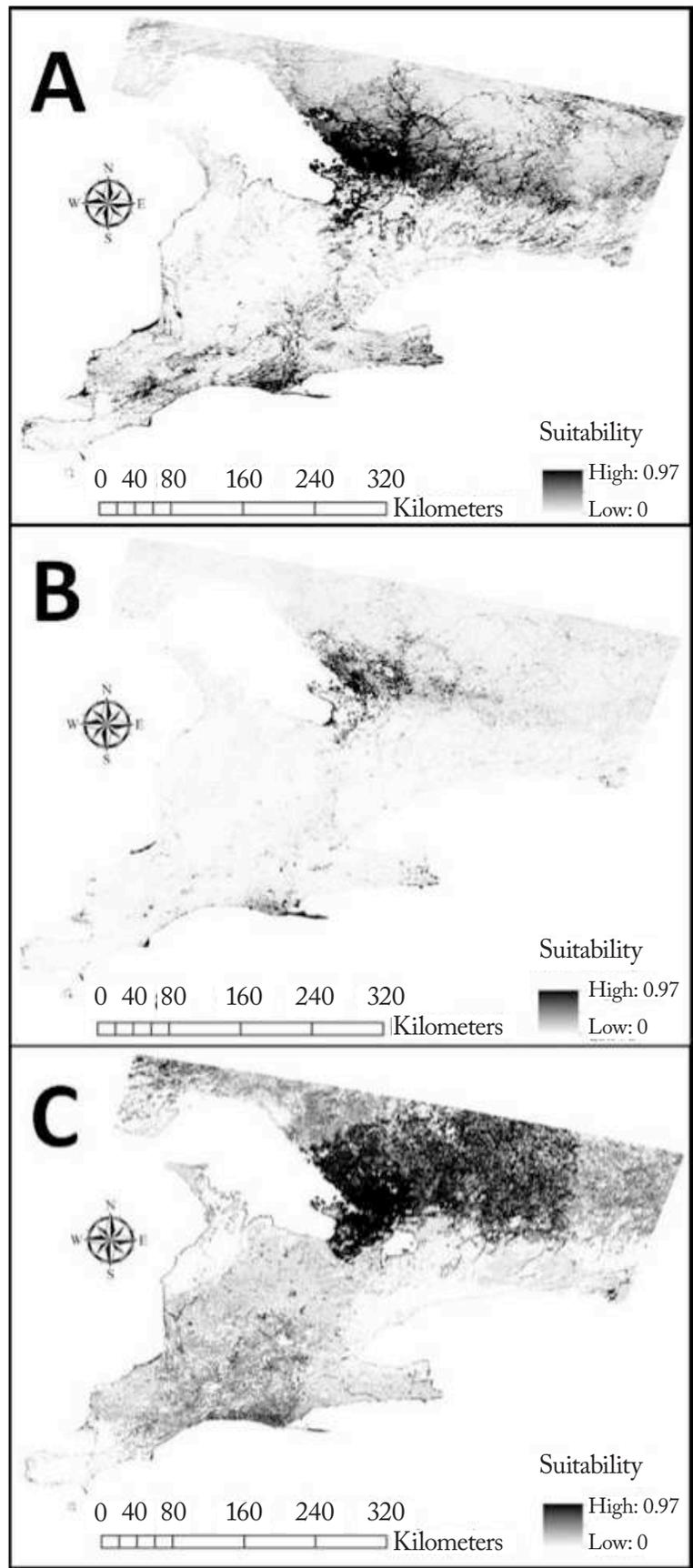


Figure 2. Habitat suitability map for the eastern hog-nosed snake in Ontario using the algorithms Maxent (A), BRT (B), and GARP (C) at a resolution of 1 km.

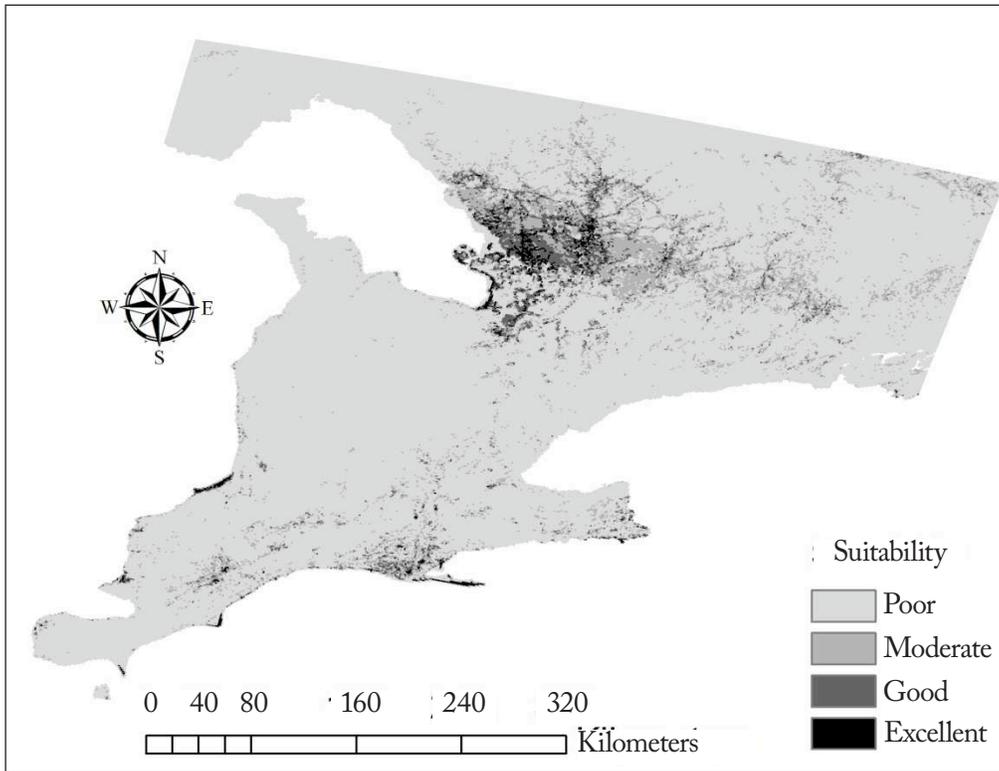


Figure 3. Categorical habitat suitability map, at a resolution of 1 km, for the eastern hog-nosed snake in Ontario. Excellent (i.e., the range of suitability that captures 50% of presences occupying the best habitat), good (i.e., the range of suitability that captures 20% of presences occupying good habitat), moderate (i.e., the range of suitability that captures 20% of presences occupying the next good habitat), and poor (i.e., the range of suitability that captures 10% of presences occupying the poorest habitat).

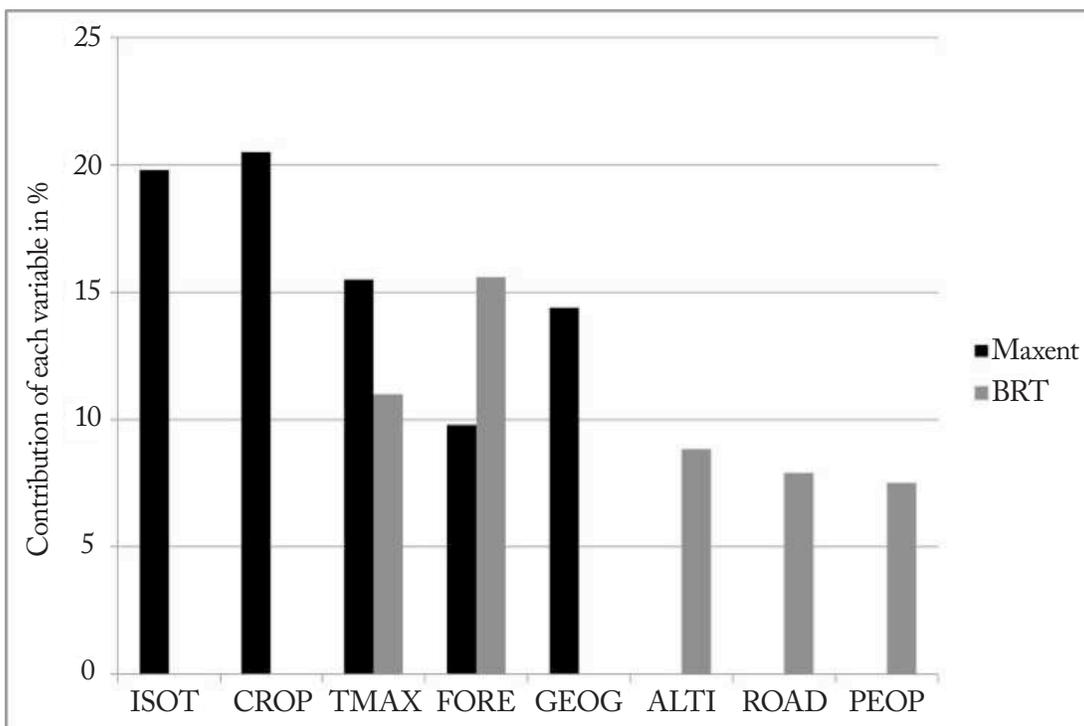


Figure 4. Contributions (%) of the 5 most influential variables in habitat suitability models built to estimate the distribution of the eastern hog-nosed snake (without biotic variables). See Table 1 for explanations regarding the nature of the variables stated above.

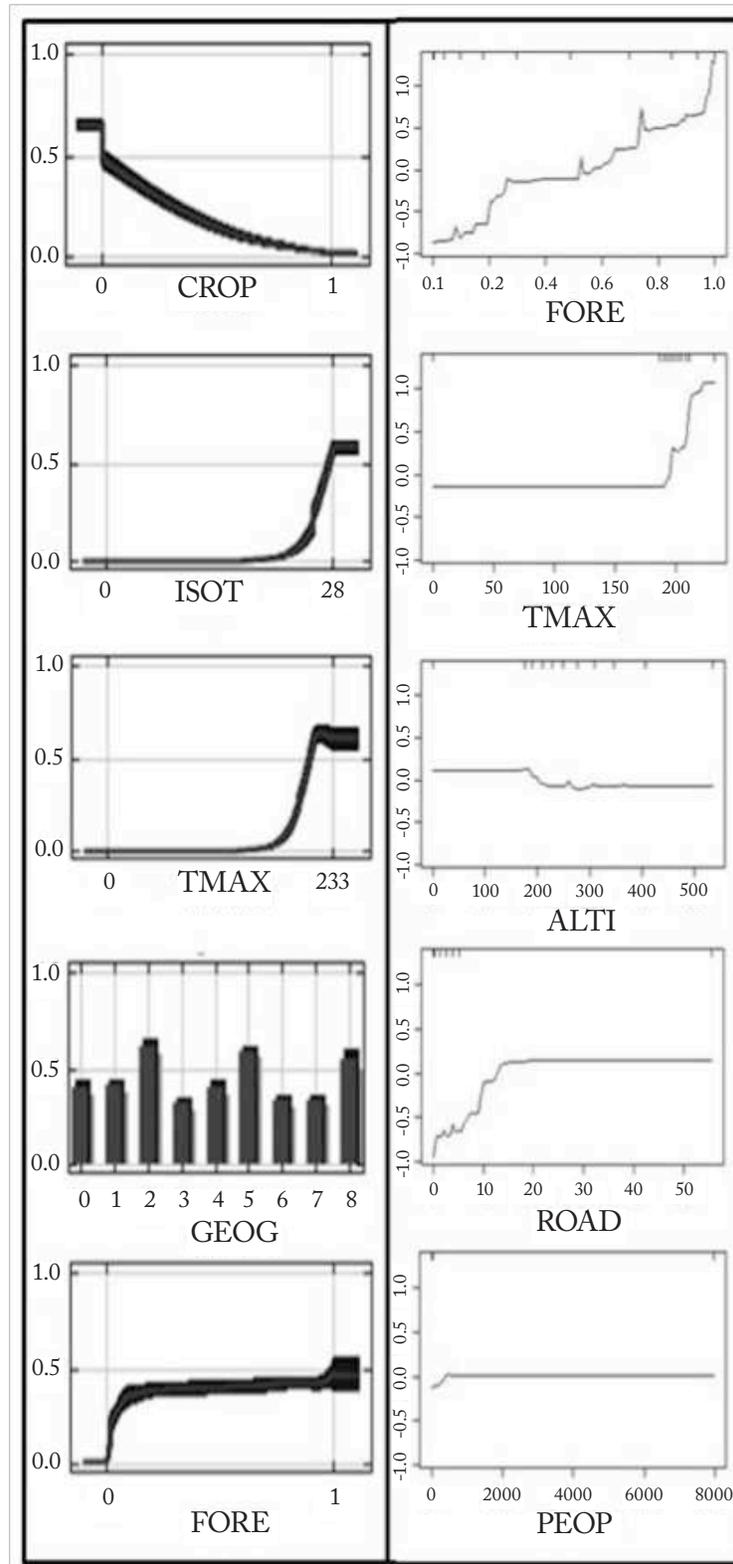


Figure 5. (Left) Marginal response curves  $\pm$  SD for the 5 most influential variables, in the models built with Maxent, estimating the distribution of the eastern hog-nosed snake. The curves show how the logistic prediction varies when one variable is gradually changed while all others are kept at their average value. (Right) Partial dependence plots for the 5 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. See Table 1 for explanations regarding the nature of the variables stated above.

contributions.

## DISCUSSION

### Performance of the models and their application to conservation

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 HSMs only use the AUC, a metric that has been heavily criticized (Lobo *et al.* 2008; Peterson *et al.* 2008). In a recent study comparing Maxent and GARP, both algorithms achieved similar AUCs, but the former avoided commission errors while the latter avoided omission errors (Peterson *et al.* 2007). That same study indicated that GARP can perform better than Maxent in making predictions in unsampled regions, a characteristic that we did not assess in this study. While GARP may be more transferable, Maxent is good at reconstructing a species' distribution in specific regions (Peterson *et al.* 2007). For the purpose of our study, the main disadvantage of GARP was that it predicted high probabilities in most of the area of study, making these maps less useful for prioritization of conservation efforts. We also favoured Maxent because it retains its high predictive power even when the sample size is small (Peterson *et al.* 2007), a good quality when dealing with rare species. Finally, there has been few studies comparing BRTs to other HSMs (e.g., Elith *et al.* 2006); not enough to assess how well it performs. Future studies should aim at comparing the performance of HSMs 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). When evaluating HSMs, some assumptions should also be carefully taken into consideration when interpreting results and making recommendations. For example, in this study, true absences were not available; they were substituted with pseudo-absences chosen randomly in the area of study. This means that false absences calculated to evaluate the models may include unreported presences and affect some of the metrics employed during evaluation. Also, some of the metrics used to evaluate the

models may be high due to model overfitting, which is not assessed in this present study.

Finally, the categorical habitat map (Figure 3) produced by applying 3 thresholds to the continuous habitat suitability map renders possible the application of HSMs in conservation biology. The area included in "excellent" not only comprises 50% of all eastern hog-nosed snake occurrences, but also the most suitable habitat for the species. Considering limitations in funding when protecting species at risk, the categorical habitat map proposed in this study may help conservation authorities better prioritize conservation initiatives (e.g., start by looking at what actions can be taken for land falling under "excellent"). Not only are categorical habitat maps easier to use, but they avoid difficulties related to interpretation. A similar approach could be employed to help inform the process of defining critical habitat.

### The importance of carefully choosing the explanatory variables

The number of variables required by HSMs to make good predictions depended on the algorithm. Maxent models required fewer variables to achieve high AUC values than BRTs and GARP. Although eliminating variables by optimizing AUC values is seldom done in the context of HSMs, similar results have been found in the past (Millar and Blouin-Demers 2012). Several studies discuss how explanatory variables should be chosen to increase the performance and generality of HSMs (Guisan and Zimmermann 2000; Austin 2007), but only a few have attempted to better understand how the number of variables employed affect the performance of HSMs. Yet, some HSMs can easily overfit their training datasets (Randin *et al.* 2006). It would be informative to use the BRT models produced in this study on separate datasets to test their cross-validation abilities. A recursive feature elimination approach similar to what we used in this study can help identify which explanatory variables should be used for modelling. The use of categorical variables in HSMs is problematic when comparing one algorithm to another. While most HSMs allow using categorical variables, except GARP (Elith and Graham 2009), our results suggest that Maxent and BRTs differ significantly in the way they use the geological categorical variable. We believe Maxent may be able to process categorical variables successfully because the

Table 3. Changes in the metrics of performance for Maxent models including an additional biotic variable while predicting the distribution of the eastern hog-nosed snake in Ontario

Add. Variable	$\Delta$ Internal AUC	$\Delta$ External AUC	Threshold	$\Delta$ Sensitivity	$\Delta$ Specificity	$\Delta$ Kappa	$\Delta$ TSS	$\Delta$ MPA
T1990	-0.004	0	0.161	0.03	-0.04	-0.015	-0.01	0.007
CALL	-0.01	0	0.165	0.03	-0.03	-0.012	0	-0.001
S1990	0.001	0	0.195	0	0.02	0.01	0.02	-0.009

contribution of this geological variable was high in Maxent models. It is unclear whether BRTs used this variable as effectively because its contribution was minimal in BRT models. We suggest using categorical variables only when absolutely needed, and limiting the number of classes as much as possible.

#### **Incorporating biotic variables in habitat suitability models**

When biotic variables were added to the HSMs, only small changes were observed in the metrics of performance (Table 3). Several reasons may explain this finding. First, 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 snake. It is probably not necessary to incorporate weak biotic interactions in HSMs, as was the case for *Krameria* plants and *Centris* bees (Giannini *et al.* 2013). Because toads are so widespread, we also used toad calls (CALL) as a measure of toad relative abundance. When toad abundance was added to the models, differences in performance scores were still low and the contributions of that variable in the models were also low. It remains unclear whether toad abundance constrains the distribution of the eastern hog-nosed snake. Our toad abundance variable may have been too coarse. We used data from the Marsh Monitoring Program (Bird Studies Canada 2008) to derive toad abundance. With the exception of data derived from satellite imagery, biotic variables (such as species occurrences) suffer from many shortfalls, including unequal spatial sampling and lack of a robust sampling design. Obtaining reliable data to study biotic interactions at a landscape scale will be challenging. This is probably the main hindrance to using biotic variables to build HSMs.

It is surprising that although the biotic variables did not seem to help the models make better predictions, the variable representing snakes occurrences had high contributions in the models. The directions of the response curves are very informative because an increase in both the toad and snake variables increased the probability of occurrence of the eastern hog-nosed snake. Snake occurrences were better at predicting the presence of eastern hog-nosed snakes than toad occurrences (Thomasson 2012). This was expected with toads, but not with snakes. Similar results were found in a study in which information on unrelated species was included in HSMs (Giannini *et al.* 2013). Perhaps the variable S1990 was one of the variables with the highest contribution because it represents sampling effort. Areas sampled for other snake species are probably more likely to have been sampled for eastern hog-nosed snakes. This would suggest a correlated bias in sampling effort (Newbold 2010). Unless species occurrences are collected following a sampling design specific to the needs of a given study, sampling biases will often limit the predictive capabilities of HSMs (Dennis and Thomas 2000). Due to various limitations such as funding and time, this is rarely taken into consideration in HSMs. Another reason why S1990 contributed highly to HSMs predicting the distribution of eastern hog-nosed snake could be that it reflects

some of the conditions that are generally good for snakes at a coarse resolution, such as forest cover and warm temperatures.

#### **Conservation implications for the eastern hog-nosed snake**

Eastern hog-nosed snakes in southern Ontario were often on sand while individuals found near Georgian Bay were on Precambrian bedrock, or selected sand at a finer spatial scale (Thomasson 2012). Similar results were obtained in other parts of the species' range at smaller spatial scales (Plummer and Mills 2000; Lagory *et al.* 2009; Seburn 2009) and confirm that this species has a clear affinity for sandy areas. Previous studies have also shown that open environments, where temperatures are usually higher, are selected by the eastern hog-nosed snake at the scale of its home range (Lagory *et al.* 2009; Peet-Paré and Blouin-Demers 2012). Our results indicate that maximum temperature during the active season also seems to limit the distribution of this species (Figure 5). Considering the high contribution of the variable TMAX in the models, our result suggest that the species selects areas with higher temperatures even at the scale of its Canadian distribution. Temperature clearly plays a role in shaping the distribution of the species in Ontario. Being at the northern extreme of its distribution in Ontario, the eastern hog-nosed snake appears to be limited by colder temperatures.

Under a temperate climate, one important factor shaping the distribution of reptiles is winter conditions (Gregory 1984). We know that eastern hog-nosed snakes can hibernate in a variety of locations: they may take refuge in mammal burrows or dig their own hibernacula when there is loose soil (Plummer 2002). It is possible that eastern hog-nosed snakes can survive around Georgian Bay by hibernating in bedrock fissures that are deep enough to avoid frost. Alternatively, they may find refuge in isolated sand deposits. In southern Ontario, eastern hog-nosed snakes do not need to go as deep and can simply dig their own hibernacula in sand (Robson 2011). Clearly, habitats with sandy soils are important to conserve as this species is often found in association with these areas (Cunnington and Cebek 2005; Lagory *et al.* 2009; Robson 2011), particularly in southern Ontario (Thomasson 2012).

Eastern hog-nosed snakes were often near roads or in populated rural areas. The positive response curves of ROAD and PEOP could arise because occurrences used in this study were obtained opportunistically by the public, often near roads and small towns, and indicate once again a sampling bias. It is not possible to determine whether eastern hog-nosed snakes exhibit true affinity for roads with the present study. Road mortality is, however, a serious threat for eastern hog-nosed snakes (Rouse *et al.* 2011; Robson and Blouin-Demers 2013). Eastern hog-nosed snakes avoid paved roads within their home ranges, which may have an effect on population persistence (Robson and Blouin-Demers 2013). Another reason many eastern hog-nosed snake occurrences were documented near roads is because this species prefers open

areas and edges (Lagory *et al.* 2009; Peet-Paré and Blouin-Demers 2012).

Cropland density reduced habitat suitability in Maxent models (Figure 5), which could explain why habitat was scored less suitable in Long Point and Point Pelee, areas with intensive agriculture, despite historical presence of eastern hog-nosed snakes. Agriculture is known to be a driving cause of species endangerment in Canada, particularly in southern Ontario (Kerr and Cihlar 2004). Forest density, on the other hand, increased habitat suitability in agreement with the observation that most documented occurrences of eastern hog-nosed snakes were in forest. Cropland density and forest density were negatively correlated ( $r = -0.71$ ). Thus, reduced forest availability limits habitat suitability for this species. While edges seem to be selected by eastern hog-nosed snakes at the scale of their home range, this study indicates that forested areas are required by the species at larger spatial scales.

The categorical map in Figure 3 could be used to help locate land that should be protected to conserve eastern hog-nosed snakes. Prioritizing land of high suitability (category: excellent) could be a starting point, but other important aspects should be considered. For instance, connectivity and parcel size will greatly influence the probability of population persistence and must be considered during the elaboration of conservation plans.

## CONCLUSION

One goal of this study was to use habitat suitability models to illustrate how areas for protection can be identified, using the eastern hog-nosed snake as an example. We proposed 3 thresholds to identify areas most suitable for a species. The categorical habitat maps produced with these thresholds are tools that can be used by conservation agencies to prioritize conservation efforts. To identify areas for protection, we suggest developing HSMs predicting the distribution of the species, and then verifying the species presence in the most suitable areas targeted for protection. Similar hierarchical approaches have been used successfully in the past (Pearson *et al.* 2004; Anadón *et al.* 2007) and may offer an efficient strategy to use HSMs in conservation. Because the quality of the sampling design is known to influence the reliability of HSMs (Edwards *et al.* 2006), implementing strategic sampling for species at risk will improve our ability to predict their habitat requirements and allow developing more robust HSMs.

## ACKNOWLEDGMENTS

This project was funded by Parks Canada and the University of Ottawa. We are grateful to the Natural Heritage Information Center of Ontario for providing us with snake occurrences. Finally, we are indebted to D. Currie and L. Fahrig for providing insightful comments throughout this project.

## LITERATURE CITED

- Abeel, T., T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saey. 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., 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.
- Araújo, M. B., and P. H. Williams. 2000. Selecting areas for species persistence using occurrence data. *Biological Conservation* 96: 331-345.
- Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological modelling* 200: 1-19.
- Bird Studies Canada. 2008. Marsh monitoring program. Amphibian surveys. Data accessed from NatureCounts, a node of the Avian Knowledge Network. Bird Studies Canada, Canada. <http://www.naturecounts.ca/>. Accessed 7 February 2011].
- 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.
- Cianfrani, C., G. Le Lay, A. Hirzel, and A. Loy. 2010. Do habitat suitability models reliably predict the recovery areas of threatened species? *Journal of Applied Ecology* 47: 421-430.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20: 37-46.
- 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., 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.
- 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.
- 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.
- 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. *Journal of Machine Learning Research (JMLR)- Workshop and Conference Proceedings* 7: 77-88.
- 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.
- Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics* 40: 677-697.
- 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.
- ESA. 2007. *Endangered Species Act of Ontario*. Toronto, Ontario, Canada.
- 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.
- Giannini, T. C., D. S. Chapman, A. M. Saraiva, I. Alves-dos-Santos, and J. C. Biesmeijer. 2013. Improving species distribution models using biotic interactions: a case study of parasites, pollinators and plants. *Ecography* 36: 649-656.
- Gregory, P. T. 1984. Communal denning in snakes. Pages 57-75 in A. Seigel, L. E. Hunt, J. L. Knight, L. Malaret, and N. L. Zuschlag, editors. *Vertebrate ecology and systematics: a tribute to Henry S. Fitch*. University of Kansas, Lawrence, Kansas, USA.
- Guisan, A., C. H. Graham, J. Elith, and F. Huettmann. 2007. 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., and N. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling* 135: 147-186.
- 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.
- Holland, J., D. Bert, and L. Fahrig. 2004. Determining the spatial scale of species' response to habitat. *BioScience* 54: 227-233.
- Hu, J., and Z. Jiang. 2010. Predicting the potential distribution of the endangered Przewalski's gazelle. *Journal of Zoology* 282: 54-63.
- 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 Lonkhuizen, 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.
- Liu, C., P. Berry, T. Dawson, and R. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385-393.
- Liu, C., M. White, and G. Newell. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. *Journal of Biogeography* 40: 778-789.
- 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.
- Marino, J., M. Bennett, D. Cossios, A. Iriarte, M. Lucherini, P. Pliscoff, C. Sillero-Zubiri, L. Villalba, and S. Walker. 2011. Bioclimatic constraints to Andean cat distribution: a modelling application for rare species. *Diversity and Distributions* 17: 311-322.
- 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.
- Mbatudde, M., M. Mwanjololo, E. K. Kakudidi, and H. Dalitz. 2012. Modelling the potential distribution of endangered *Prunus africana* (Hook.f.) Kalkm. in East Africa. *African Journal of Ecology* 50: 393-403.
- Millar, C. S., and G. Blouin-Demers. 2012. 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* 20: 18-29.
- 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.
- 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: 3-22.
- OGS. 2010. Surficial geology of southern Ontario; Data Set 128. Ontario Geological Survey, Ontario, Canada.
- Oldham, M., and M. Austen. 1998. *Heterodon platirhinos*: NHIC element report. Updated August 2002 by K. Ramster edition., Natural Heritage Information Centre, Ontario Ministry of Natural Resources, Peterborough, Ontario, Canada.
- Oldham, M., and W. Weller. 2000. Ontario herpetofaunal atlas. Natural Heritage Information Centre, Ontario Ministry of Natural Resources, Peterborough, Ontario, Canada.
- OMNR. 1998. Ontario 28-Class land cover database. Ontario Ministry of Natural Resources, Peterborough, Ontario, Canada.
- OMNR. 2005. Provincial digital elevation model, tiled dataset, Version 2.0.0., Ontario Ministry of Natural Resources, Peterborough, Ontario, Canada.
- OMNR. 2006. Ontario road network. Ontario Ministry of Natural Resources, Peterborough, Ontario, Canada.
- Ontario Nature. 2011. The Ontario reptile and amphibian atlas. Ontario Nature, Ontario, Canada. <http://www.ontarionature.org>. Accessed 7 February 2011.
- Pearson, R. G., T. P. Dawson, and C. Liu. 2004. Modelling species distributions in Britain: A hierarchical integration of climate and land-cover data. *Ecography* 27: 285-298.
- Peet-Paré, C. E., and G. Blouin-Demers. 2012. Female eastern hog-nosed snakes (*Heterodon platirhinos*) choose nest sites that produce offspring with phenotypes likely to improve fitness. *Canadian Journal of Zoology* 90: 1215-1220.
- Pellissier, L., K. A. Brathen, J. Pottier, C. F. Randin, P. Vittoz, A. Dubuis, N. G. Yoccoz, T. Alm, N. E. Zimmermann, and A. Guisan. 2010. Species distribution models reveal apparent competitive and facilitative effects of a dominant species on the distribution of tundra plants. *Ecography* 33: 1004-1014.
- 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. 2010. Scenarios for global biodiversity in the 21st Century. *Science* 330: 1496-1501.
- 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.
- 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. 2002. Observations on hibernacula and overwintering ecology of eastern hog-nosed snakes (*Heterodon platirhinos*). *Herpetological Review* 33: 89-90.
- Plummer, M., and N. Mills. 2000. Spatial ecology and survivorship of resident and translocated hognose snakes (*Heterodon platirhinos*). *Journal of Herpetology* 34: 565-575.
- Pyron, R. A., F. T. Burbrink, and T. J. Guirher. 2008. Claims of potential expansion throughout the US by invasive python species are contradicted by ecological niche models. *PLoS One* 3: e2931.
- R. Development Core Team. 2011. 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.
- Ridgeway, G. 2010. gbm: Generalized Boosted Regression Models. Documentation on the R Package 'gbm', version 1.6-3.1. <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. MSc. thesis, University of Ottawa, Ottawa, Ontario, Canada.
- Robson, L. E., and G. Blouin-Demers. 2013. Eastern hognose snakes (*Heterodon platirhinos*) avoid crossing paved roads, but not unpaved roads. *Copeia* 2013: 507-511.
- 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. 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.
- SARA. 2002. Species at Risk Act. Government of Canada, Ottawa, Ontario, Canada.
- Seburn, D. 2009. Recovery Strategy for the eastern hog-nosed snake (*Heterodon platirhinos*) in Canada. Species at Risk Act Recovery Strategy Series. Parks Canada Agency, Ottawa, Ontario, Canada.
- Statistics Canada. 2006. 2006 Census of population - Catalogue no. 92-160-GIE. Statistics Canada, Ottawa, Ontario, Canada.
- 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.
- Surface, H. A. 1906. The serpents of Pennsylvania - Monthly bulletin of the Division of Zoology. Pennsylvania Department of Agriculture, Harrisburg, Pennsylvania, USA.
- 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.
- Thomasson, V. 2012. Habitat Suitability Modeling for the eastern hog-nosed snake, *Heterodon platirhinos*, in Ontario. MSc thesis, University of Ottawa, Ottawa, Ontario, Canada.
- Uhler, F., C. Cottam, and T. Clarke. 1939. Food of snakes of the George Washington National Forest, Virginia. *Transactions of the North American Wildlife Conference*. 4: 605-622.
- 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.
- 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.
- Wisz, M. S., J. Pottier, W. D. Kissling, L. Pellissier, J. Lenoir, C. F. Damgaard, C. F. Dormann, M. C. Forchhammer, J. A. Grytnes, and A. Guisan. 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews* 88: 15-30.
- 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.

## ABOUT THE AUTHORS

**Victor Thomasson** holds a MSc in biology from the University of Ottawa, a BSc in wildlife Biology from McGill University and a BEd from the University of Ottawa. He always had a general interest in wildlife biology, especially herpetology.

He will always be very happy to talk about about reptiles or amphibians around a cup of coffee. He also enjoys working with the community and educating others about the environment and conservation. Although he has always been extremely passionate about science, education has always been important to him. He is currently a science teacher for the English Montreal School Board in Montreal, Quebec.



**Gabriel Blouin-Demers** holds a BSc in environmental biology from McGill University and a PhD in biology from Carleton University. He held a postdoctoral fellowship at The Ohio State University. He has been a professor in the Department of Biology at the University of Ottawa since 2002. His research program aims at explaining 2 major patterns: 1) why do animals select particular habitats, and 2) why are



some animals polymorphic. He has mostly used reptiles as a model system, but has worked on fish, amphibians, birds, mammals, and insects. Because reptiles are, proportionally, the most threatened vertebrate group in Canada, he has also conducted conservation research on reptiles at risk.

*Submitted 14 August 2014 — Accepted 7 November 2014*