



## Original Articles

## Modelling species habitat suitability from presence-only data using kernel density estimation

Guiming Zhang<sup>a,b</sup>, A-Xing Zhu<sup>c,d,e,f,b,\*</sup>, Steve K. Windels<sup>g</sup>, Cheng-Zhi Qin<sup>f,c</sup><sup>a</sup> Department of Geography & the Environment, University of Denver, Denver, USA<sup>b</sup> Department of Geography, University of Wisconsin-Madison, Madison, USA<sup>c</sup> Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China<sup>d</sup> Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Nanjing, China<sup>e</sup> State Key Laboratory Cultivation Base of Geographical Environment Evolution, Nanjing, China<sup>f</sup> State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China<sup>g</sup> Voyageurs National Park, National Park Service, International Falls, USA

## ARTICLE INFO

## Keywords:

Habitat suitability modelling and mapping  
 Presence-only data  
 Resource availability  
 Kernel density estimation  
 Ecological monitoring

## ABSTRACT

We present a novel approach for modelling and mapping habitat suitability from species presence-only data that is useful for ecosystem and species monitoring. The approach models the relationship between species habitat suitability and environment conditions using probability distributions of species presence over environmental factors. Resource availability is an important issue for modelling habitat suitability from presence-only data, but it is in lack of consideration in many existing methods. Our approach accounts for resource availability by computing habitat suitability based on the ratio of species presence probability over environmental factors to background probability of environmental factors in the study area. A case study of modelling and mapping habitat suitability of the white-tailed deer (*Odocoileus virginianus*) using presence locations recorded in aerial surveys at Voyageurs National Park, Minnesota, USA was conducted to demonstrate the approach. Performance of the approach was evaluated through randomly splitting the presence locations into training data to build the model and test data to evaluate prediction accuracy of the model (repeated 100 times). Results show that the approach fit training data well (average training area under the curve AUC = 0.792, standard deviation SD = 0.029) and achieved better-than-random prediction accuracy (average test AUC = 0.664, SD = 0.025) that is comparable to the state-of-the-art MAXENT method (average training AUC = 0.784, SD = 0.021; average test AUC = 0.673, SD = 0.027). In addition, the suitability-environment responses modelled using our approach are more amenable to ecological interpretation compared to MAXENT. Compared to modelling habitat suitability purely based on species presence probability distribution (average training AUC = 0.743, SD = 0.030; average test AUC = 0.645, SD = 0.023), incorporating background distribution to account for resource availability effectively improved model performance. The proposed approach offers a flexible framework for modelling and mapping species habitat suitability from species presence-only data. The modelled species-environment responses and mapped species habitat suitability can be very useful for ecological monitoring at ecosystem or species level.

## 1. Introduction

Habitat suitability modelling, also referred to as environmental niche modelling or species distribution modelling (Franklin and Miller, 2009), is essential to understanding species habitat requirements and identifying drivers of species distribution (Elith and Leathwick, 2009; Graham et al., 2004b; Leathwick and Austin, 2001; Mac Nally, 2000).

Habitat suitability mapping is achieved by projecting habitat suitability models from environmental space to geographic space to predict spatial variation of species habitat suitability. The resultant habitat suitability models (i.e., species-environment responses) and habitat suitability maps can be used to support a wide range of applications such as ecological monitoring, biodiversity assessment, biological reserve design, habitat restoration, invasive species management, etc. (Ferrier

\* Corresponding author at: Jiangsu Center for Collaborative Innovation in Geographical Information Resource and Application, School of Geography, Nanjing Normal University, 1 Wenyuan Road, Xianlin University District, Nanjing, Jiangsu 210023, China.

E-mail address: [azhu@wisc.edu](mailto:azhu@wisc.edu) (A.-X. Zhu).

<https://doi.org/10.1016/j.ecolind.2018.04.002>

Received 30 September 2017; Received in revised form 1 February 2018; Accepted 1 April 2018

1470-160X/ © 2018 Elsevier Ltd. All rights reserved.

et al., 2002; Lindenmayer and Likens, 2010; Telesco et al., 2007; Thorn et al., 2009; Thuiller et al., 2005).

The key to habitat suitability modelling and mapping is deriving the relationships between species habitat suitability and environmental conditions (i.e., environmental niches) from species data and environmental data (Guisan and Zimmerman, 2000; Hirzel and Lay, 2008; Warren, 2012). With the rapid development of geospatial technologies such as geographic information system (GIS) and remote sensing, environmental data are increasingly available (Gillespie et al., 2008; Kerr and Ostrovsky, 2003; Viña et al., 2008). According to the different species data required for deriving the suitability-environment relationships, methods for habitat suitability modelling and mapping fall into three groups: presence-absence methods, presence-pseudo-absence methods, and presence-only methods (Elith and Leathwick, 2009).

Presence-absence methods require both species presence and absence data to derive suitability-environment relationships. Examples are generalized linear models (GLM), generalized additive models (GAM) (Guisan et al., 2002), and regression trees (De'Ath, 2002). Presence-absence data are often collected through well-designed biological surveys. Thus presence-absence data are usually only available for a target group of species in small geographic areas (Brooks, 2004; Pressey, 2004). Even if absence data are available, accuracy of the data can be problematic. A recorded absence might simply result from the failure to detect the species, or the habitat was suitable but not accessible for the species. In either case, a recorded absence is not a true absence of the species (Gu and Swihart, 2004; Hirzel et al., 2002; Li and Hilbert, 2008).

Presence-pseudo-absence methods use pseudo-absence data (e.g., locations randomly selected in the study area) to replace absence data to train presence-absence models (Elith and Leathwick, 2007; Engler et al., 2004). However, performance of presence-pseudo-absence methods are shown to be very sensitive to strategies used to generate the pseudo-absence data, and there is no consensus of a robust strategy for generating pseudo-absences (Chefaoui and Lobo, 2008; Hanberry et al., 2012; Stokland et al., 2011; Wisz and Guisan, 2009).

Presence-only methods require only species presence data to derive suitability-environment relationships. This group of methods are widely applied in practice as many biological datasets consist of only species presence records, such as historical collections from museums and herbaria (Graham et al., 2004a) and patrol records (Zhang et al., 2017b). Among existing presence-only methods, envelope-based BIOCLIM (Busby, 1991) and HABITAT (Walker and Cocks, 1991) treat habitat suitability as invariant at locations (in environmental space) within the environmental envelopes constrained by the outermost species presences. Thus, they tend to oversimplify the ecological reality that species habitat suitability may vary even within the environmental envelopes.

Environmental similarity-based DOMAIN (Carpenter et al., 1993), EDGM (environmental-distance geometric mean) (Hirzel and Arlettaz, 2003) and LIVES (limiting variable and environmental suitability) (Li and Hilbert, 2008) compute habitat suitability at a location based on either the maxim or the geometric mean of the environmental similarities between the location and all known presence locations. Environmental distance-based SVMs (one-class support vector machines) (Guo et al., 2005) and MDMs (Mahalanobis-distance models) (Farber and Kadmon, 2003) compute habitat suitability at a location based on the environmental distance from the location to the center of all known presence locations. Implicitly, similarity- or distance-based methods model habitat suitability as a linear function of environment similarity or distance. Thus, they tend to oversimplify the ecological reality that species habitat suitability may responds nonlinearly to environmental gradient.

GARP (genetic algorithm for rule-set production) (Stockwell, 1999) and MAXENT (maximum entropy) (Phillips et al., 2006) are machine learning algorithms capable of fitting sophisticated rules or function relations (e.g., nonlinear) based on species presence data and

background data. GARP and MAXENT can often achieve high prediction accuracy in habitat suitability mapping (Elith et al., 2006). But the fitted suitability-environment relationships are often implicit, complex and hard to interpret. GARP and MAXENT are mostly used for predictive mapping of species habitat suitability but have limited power for modelling species environmental niches.

ENFA (ecological niche factor analysis) (Hirzel et al., 2002) models habitat suitability based on frequency distributions of species presence over ecological niche factors (transformed from original environmental predictors using a procedure similar to principal component analysis). ENFA can accommodate nonlinear responses of species habitat suitability to niche factors. But it assumes that the frequency distribution on each factor must be unimodal and symmetrical. It thus oversimplifies the ecological reality that frequency distribution of species presence on niche factors may be multimodal or skewed. Zhu et al. (2015) models species habitat suitability using species presence probability distributions over environmental factors without assuming the unimodality or symmetry of the distributions. However, both ENFA and Zhu et al. (2015) model habitat suitability purely based on the probability distribution of species presence over environmental factors without accounting for background distributions (i.e., probability distribution of environmental factors in the study area). Neither of them accounts for the “availability” of resources, which is an important consideration when modelling habitat suitability from presence-only data (Boyce et al., 2002; Johnson et al., 2006).

This article presents a novel approach for modelling and mapping habitat suitability from species presence-only data. The approach models suitability-environment relationships using probability distributions of species presence over environmental factors. It imposes no assumptions on the shape of species suitability-environment relationships. Moreover, it accounts for resource availability by adjusting the presence probability distributions with background probability distributions. Details of the approach are presented in Section 2. A case study of habitat suitability modelling and mapping for the white-tailed deer (*Odocoileus virginianus*) at the Voyageurs National Park to demonstrate the approach is reported in Section 3. Discussion and conclusions are presented in Section 4 and Section 5, respectively.

## 2. Methodology

### 2.1. Basic idea

#### 2.1.1. Approximating species environmental niche

The theoretical basis of habitat suitability modelling lies on the concept of species environmental niche, which characterizes how species fitness (habitat suitability) responds to environmental conditions (Guisan and Zimmerman, 2000; Hirzel and Lay, 2008; Leibold, 1995). Probability distribution of species presence over environmental gradients is often taken as a natural approximation to species (realized) environmental niche based on which species habitat suitability is modeled. For example, ENFA computes species habitat suitability based on the frequency distribution of species presence over ecological niche factor axes (Hirzel et al., 2002). MAXENT estimates a probability density surface of species occurrence (habitat suitability) over pixels in the study area, with probability density at each pixel related to environmental conditions at that location (Phillips et al., 2006).

Our approach also uses probability distributions (probability density functions) of species presence over environmental factors to approximate species environmental niches and to compute habitat suitability. The approach estimates probability distribution from species presence-only data in a nonparametric fashion. It imposes no assumptions on the shape of the distribution (e.g., unimodality, symmetry), nor on the form of species suitability-environment relationships (e.g., linear, gaussian).

### 2.1.2. Accounting for resource availability

It is important to consider the “availability” of resources when computing habitat suitability from species presence-only data (Boyce et al., 2002; Johnson et al., 2006). With species presence-only data, habitat suitability should be measured by the use of resources relative to resource availability (Johnson et al., 2006). Suppose one species does not differentiate environmental conditions for habitat use in a study area and thus occurs randomly over the study area (i.e., equal habitat suitability everywhere). Variability in the presence probability distribution over environmental factors would still suggest that the species prefers certain environmental conditions over others. But in fact, such variability is an artifact of the probability distribution of available (nonuniform) environmental conditions in the study area (i.e., background distribution). It does not reflect species true habitat preference. However, if the background distribution is accounted for, such artifact can be avoided. For instance, the ratio of species presence probability to background probability can be computed. A ratio greater than one suggests preference of certain environmental condition and a ratio smaller than one suggests avoidance (Johnson et al., 2006).

Our approach accounts for the “availability” of resources when computing habitat suitability from species presence-only data. It computes habitat suitability based on the probability distribution of species presence over environmental factors adjusted by the background probability distribution (representing resource availability). It assumes that the adjusted probability density of species presence is an effective indicator of species habitat suitability: higher adjusted presence probability density indicates higher habitat suitability.

## 2.2. Estimating probability distributions

### 2.2.1. Presence probability distribution over environmental factors

Given species presence locations and GIS data layers characterizing environmental factors, the probability density function (PDF) of species presence with respect to individual environmental factors can be estimated from a sample consisting of values of the corresponding environmental factor at the presence locations. The presence PDF with respect to individual environmental factors was estimated using kernel density estimation (KDE), a nonparametric density estimation method capable of estimating continuous probability density functions (Silverman, 1986). KDE uses the equation:

$$f_{\text{presence}}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_x} K\left(\frac{x-x_i}{h_x}\right), \quad (1)$$

where  $f_{\text{presence}}(x)$  is the estimated PDF of species presence with respect to environmental factor  $x$ ,  $x_i$  is the value of  $x$  at presence location  $i$ ,  $n$  is the total number of presence locations.  $K$  is a kernel density function and here the Gaussian kernel was adopted (Silverman, 1986):

$$K\left(\frac{x-x_i}{h_x}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h_x^2}} \quad (2)$$

$h_x$  is a smoothing parameter for  $x$  called bandwidth and it is a crucial parameter for KDE. A bandwidth that is too large would result in a “flat” PDF that fails to reflect the variability in data, and one that is too small would result in a “spiky” PDF that contains too much noise. The “golden section search optimization procedure” (Brunsdon, 1995) was adopted to find the optimal bandwidth based on maximum likelihood criterion through cross-validation on the sample data. The initial lower and upper search bounds for the optimal bandwidth were set as  $0.01 \times h_{0x}$  and  $2.0 \times h_{0x}$  respectively. Here  $h_{0x}$  is the “rule-of-thumb” bandwidth (Silverman, 1986) for  $x$  determined by:

$$h_{0x} = 1.06 \cdot \sigma_x \cdot n^{-1/5} \quad (3)$$

in which  $\sigma_x$  is the standard deviation of values of  $x$  at presence locations. Details of the “golden section search optimization procedure” are beyond the scope of this article and interested readers should refer

to Brunsdon (1995) for full details.

Presence PDF on a categorical environmental factor (e.g., land cover type) was estimated with the normalized frequency distribution of species presences over the categories of the factor.

### 2.2.2. Background distribution of environmental factors

Probability density function of individual environmental factors over the study area (background distribution) were also estimated using the KDE method. Different from estimating the presence PDF, the sample in this case consists of values of the corresponding environmental factor at all locations (pixels) in the study area. Moreover, the “rule-of-thumb” algorithm (Eq. (3)) was used to determine bandwidth for estimating background PDF due to two reasons. First, the golden section search optimization is computationally intensive when sample size (i.e., number of pixels in the study area) is very large. Second, when sample size is large, the “rule-of-thumb” algorithm can provide a robust estimation of the optimal bandwidth for KDE (Jones et al., 1996; Silverman, 1986).

Background PDF of a categorical environmental factor was estimated with the normalized frequency distribution of the factor in the study area.

## 2.3. Modelling and mapping habitat suitability

### 2.3.1. Accounting for resource availability

Species presence PDF was adjusted using background PDF to account for resource availability (Section 2.1.2) when computing species habitat suitability. Specifically, for each individual environmental factor, the ratio of species presence probability density to the background probability density was computed:

$$f_{\text{ratio}}(x) = \frac{f_{\text{presence}}(x)}{f_{\text{background}}(x)} \quad (4)$$

where  $f_{\text{presence}}(x)$  and  $f_{\text{background}}(x)$  are species presence PDF and background PDF with respect to environmental factor  $x$  respectively,  $f_{\text{ratio}}(x)$  is the ratio function of the two PDFs. This ratio function values in  $[0, +\infty)$  and the value 1 has special meanings. For some value of  $x$ , say  $x'$ ,  $f_{\text{ratio}}(x') = 1$  indicates species habitat use of environmental condition  $x'$  is purely by chance. Accordingly,  $f_{\text{ratio}}(x') > 1$  indicates preference (more likely than random use) and  $f_{\text{ratio}}(x') < 1$  indicates avoidance (less likely than random use).  $f_{\text{ratio}}(x)$  approximates the environment niche or resource selection function (Boyce et al., 2002; Johnson et al., 2006) of the species with respect to environmental factor  $x$ .

### 2.3.2. Modelling habitat suitability

Species habitat suitability with respect to individual environmental factors was modelled based on the ratio of species presence probability density to the background probability density. Habitat suitability often values within  $[0, 1]$  (Guisan et al., 2002; Hirzel et al., 2002; Phillips et al., 2006). The ratio of species presence probability density to background probability density on environmental factors values in  $[0, +\infty)$ . The following equation was used to transform the ratio to compute a habitat suitability value that has an upper limit of 1.0 regarding environmental factor  $x$ :

$$S(x) = \frac{1}{1 + e^{1-f_{\text{ratio}}(x)}} \quad (5)$$

where  $S(x)$  is species habitat suitability with respect to environmental factor  $x$ .  $f_{\text{ratio}}(x)$  is the ratio function computed using Eq. (4).

This transformation (Eq. (5)) has an important property that it maintains the correspondence between ratio value 1 and suitability value 0.5. For some value of  $x$ , say  $x'$ ,  $S(x') = 0.5$  if  $f_{\text{ratio}}(x') = 1$ . Thus, a suitability value 0.5 indicates species habitat use of environmental condition  $x'$  is purely by chance. A suitability value greater than 0.5

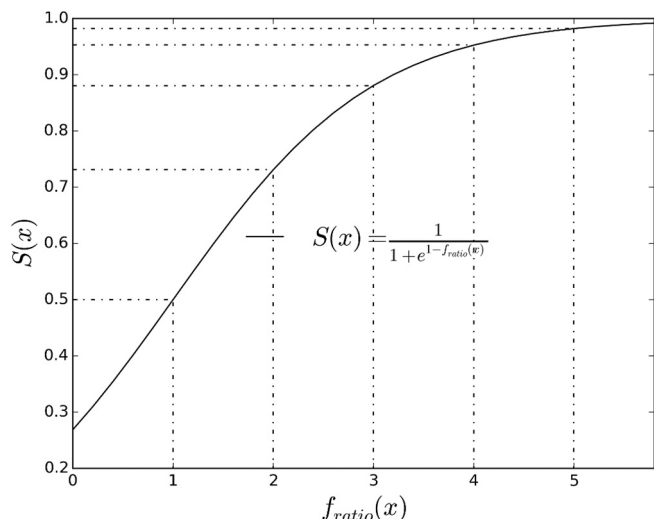


Fig. 1. Modelling habitat suitability based on the ratio of species presence probability density to background probability density on an environmental factor.

indicates habitat preference and a suitability value less than 0.5 indicates avoidance. It should be noted that  $S(x) = 0.269$  when  $f_{ratio}(x) = 0$  (Fig. 1). Thus, the modelled suitability is in fact bounded within  $[0.269, 1)$  instead of  $[0, 1]$ . The authors are aware that an ideal transformation is expected to have a second property that it maintains the correspondence between ratio value 0 and suitability value 0 (i.e., complete avoidance in habitat use). But an elegant transformation simultaneously maintaining the above two properties was not found (although it is mathematically possible to stretch suitability values in  $[0.269, 0.5)$  to  $[0, 0.5)$ , such a manipulation would introduce extra complexity and inconsistency). Nevertheless, the transformation adopted here is an acceptable compromise as maintaining the indicative meaning of suitability value 0.5 (i.e., differentiate habitat preference vs. avoidance) is a more desirable property when modelling habitat suitability from presence-only data.

The overall habitat suitability considering all environmental factors involved was computed by aggregating suitability values with respect to individual environmental factors. Many habitat suitability modelling methods adopt the “weighted average” strategy to determine the

overall effects of environmental factors on species habitat suitability (Guisan et al., 2002; Hirzel et al., 2002; Phillips et al., 2006). Here the “weighted average” strategy was also followed to determine the overall habitat suitability:

$$S(x^1, x^2, x^j, \dots, x^m) = \sum_{j=1}^m w_j S(x^j) \tag{6}$$

where  $S(x^1, x^2, x^j, \dots, x^m)$  is the overall habitat suitability,  $x^j$  is the  $j$ th environmental factor,  $m$  is the total number of environmental factors involved,  $S(x^j)$  is the habitat suitability with respect to  $x^j$ , and  $w_j$  is a weight on  $x^j$ . Here all environmental factors were equally weighted by  $1/m$ . Alternatively, the “limiting factor” strategy can be adopted to determine the overall habitat suitability. Readers are deferred to Section 4.2 for comparisons between the two aggregation strategies. The overall habitat suitability is bounded within  $[0.269, 1)$ .

### 2.3.3. Mapping habitat suitability

A habitat suitability map of the study area was predicted by computing habitat suitability at each location (pixel) in the study area. Values of environmental factors at a location were first extracted from the GIS data layers characterizing environmental factors. Habitat suitability at that location was then computed by following Eqs. (5) and (6).

## 3. Case study

### 3.1. Study area

The study area is Voyageurs National Park (VNP) ( $48^{\circ}18' - 48^{\circ}38'N$ ,  $92^{\circ}27' - 93^{\circ}11.5'W$ ) in northern Minnesota, along the United States-Canada border. VNP has a total area of approximately 828 km<sup>2</sup> of which about forty percent are lakes and other open waters (Fig. 2). The topography is generally flat with maximum relief of 80–90 m (Gogan et al., 1997). In addition to white-tailed deer (*Odocoileus virginianus*), other large mammals occurring within the park are moose (*Alces alces*), gray wolf (*Canis lupus*), black bear (*Ursus americanus*), lynx (*Felis canadensis*), and bobcat (*Felis rufus*). White-tailed deer and moose are the only two extant ungulates in the park. Moose are more limited in their distribution in the park than deer (Cobb et al., 2004; Windels and Olson, 2017). Where they do overlap in space and cover type use, they do not generally compete for food as they have different foraging preferences, especially in winter (Cobb et al., 2004). Gray wolves are the main predators of white-tailed deer year-round, while black bears, lynx,

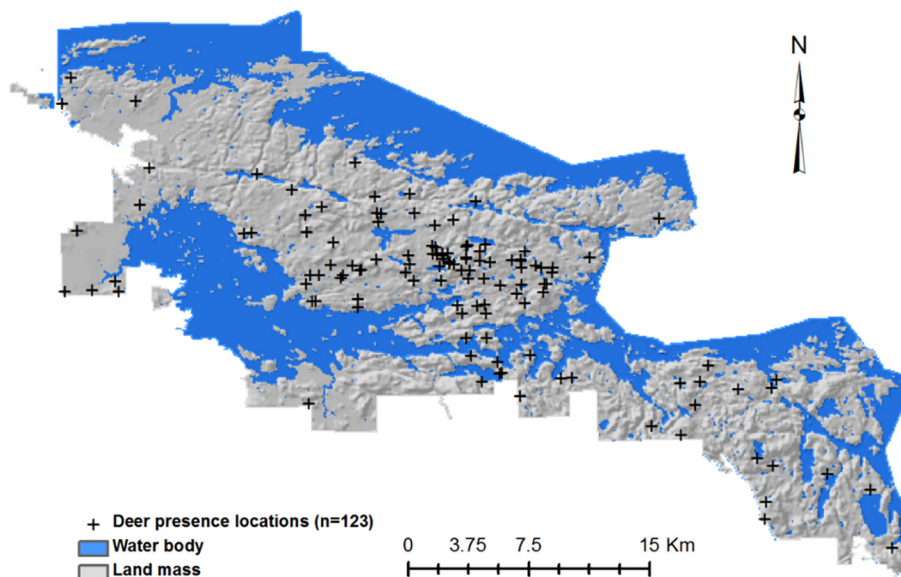


Fig. 2. Deer presence locations recorded during aerial surveys conducted in Voyageurs National Park, MN, USA, January 2009.



**Table 1**  
Pearson's correlation coefficients amongst the five continuous environmental factors.

	Elevation	Tangent slope	Sine aspect	Cosine aspect	Snow depth
Elevation	1.000	0.283	0.018	-0.032	-0.071
Tangent slope		1.000	0.037	-0.001	0.142
Sine aspect			1.000	0.024	-0.014
Cosine aspect				1.000	0.022
Snow depth					1.000

and bobcats generally only prey on deer fawns in the early summer (Gogan et al., 2004). White-tailed deer densities were estimated to be between 3.0 and 6.4 deer/km<sup>2</sup> in 2009 based on winter fecal pellet counts (Gable et al., 2017).

### 3.2. Deer presence locations

Presence locations of the white-tailed deer in VNP were recorded during aerial surveys conducted in January 2009 in conjunction with a moose population survey (Gable et al., 2017). The park was delineated into 33 survey units, ranging in area from 9.8 km<sup>2</sup> to 27.0 km<sup>2</sup> (Gogan et al., 1997; Windels, 2014). All 33 survey units were surveyed in 2009. A pilot and 3 observers searched for white-tailed deer using a de Havilland Beaver aircraft, flying at airspeeds between 125–145 km per hour and an altitude of 150–215 m. Survey units were sampled by flying west-east transects spaced at 0.536 km intervals at an intensity of about 1.5 min per km<sup>2</sup>. Locations of observed deer were plotted on topographic maps with reference to features on the topographic maps and on high-resolution aerial photos (hills, valleys, lakes, wetlands, roads, etc.). Sex composition, age and group size of each moose or deer sighting were recorded on datasheets. Estimated positioning accuracy of the recorded locations was about 50 m.

A total of 140 deer presence locations were recorded during the surveys. 17 of the 140 presence locations were plotted within water bodies close to lake shores because in the winter deer often travel along the edges of frozen lakes where the snow is less deep and travel is easier. These deer presence locations were excluded from analysis. The remaining 123 deer presence locations were used for modelling and mapping deer habitat suitability in the park (Fig. 2).

### 3.3. Environmental data

Environmental data representing topography, snow depth, and vegetation community were obtained to characterize the environmental factors that influence deer habitat suitability in winter. These factors play important roles in determining the mobility, thermal condition, and food availability for deer in winter (Bolvar-Cirín and Gallina, 2012; Moen, 1976; Ozoga and Gysel, 1972; Ozoga and Verme, 1970). For example, in the winter deer tend to prefer areas with less snowpack to improve travel efficiency (Ozoga and Gysel, 1972). Deer also prefer to

stay on leeward sides of slopes to reduce exposure to wind to better maintain body temperature (Moen, 1976). Topography has an impact on the distribution of snow accumulation primarily as it influences the amount of solar radiation reaching the surface. Vegetation communities provide thermal cover, escape cover from predators, and forage for deer in winter (Cobb et al., 2004; Ozoga and Gysel, 1972; Ozoga and Verme, 1970).

A digital elevation model (DEM) at 30 m resolution of VNP was obtained from the National Elevation Dataset (Gesch et al., 2002) provided by the United States Geological Survey (USGS). The DEM was resampled to 90 m resolution with the bilinear method in ArcMap (ESRI, 2013) to accommodate the positioning accuracy of deer presence locations (~50 m). *Elevation*, *tangent of slope* (steepness), *sine of aspect* (eastness), and *cosine of aspect* (northness) were derived from the DEM in ArcMap to represent topographic factors. A *land cover type* map of 2006 at 30 m resolution was obtained from the National Land Cover Database provided by USGS (Fry et al., 2011). The land cover map was resampled to 90 m resolution with the majority method to match the spatial resolution of other environmental data layers. Daily *snow depth* estimation at 1 km resolution of VNP was obtained from the Snow Data Assimilation System Data Products provided by the National Oceanic and Atmospheric Administration (NOAA) (NOHRSC, 2004). The daily snow depth raster data layers in January 2009 were averaged to compute an average snow depth raster. The Ordinary Kriging method in ArcMap was adopted to interpolate the average snow depth raster at 1-km resolution to 90-m resolution.

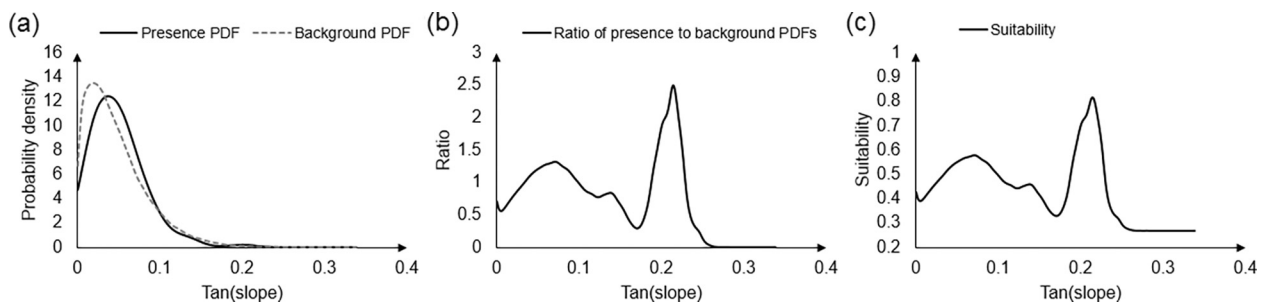
In total, five continuous environmental factors and one categorical environmental factor (i.e., land cover type) were used for modelling and mapping deer habitat suitability (pixels falling into water bodies were excluded). The correlation amongst the five continuous environmental factors were rather weak (Table 1). The highest correlation was found between *elevation* and *tangent of slope* (Pearson's  $r = 0.283$ ) and between *tangent of slope* and *snow depth* (Pearson's  $r = 0.142$ ). The weak correlations across the environmental factors suggests that each selected environmental factor characterizes a relatively unique aspect of the environmental conditions.

### 3.4. Habitat suitability modelling

#### 3.4.1. Suitability-environment relationships

The proposed approach was applied on the deer presence data and environmental data to model and map winter deer habitat suitability in VNP. Fig. 3 shows the relationships between deer habitat suitability and environmental factors using slope factor as an example. Obviously, our approach does not impose any assumptions on the shape of the suitability-environment relationships (e.g., linear, gaussian, symmetrical).

The presence PDF suggests that deer mostly preferred slopes around 2.9° (whose tangent = 0.05) (Fig. 3a). But such an observation is questionable without accounting for availability of the factor itself in the study area. With the background PDF accounted for, the ratio of the presence PDF to the background PDF clearly reveals that the deer



**Fig. 3.** The relationship between deer habitat suitability and environmental factors illustrated with the slope factor. (a) Presence PDF and background PDF on slope. (b) Ratio of presence PDF to background PDF. (c) Suitability-slope relationship.

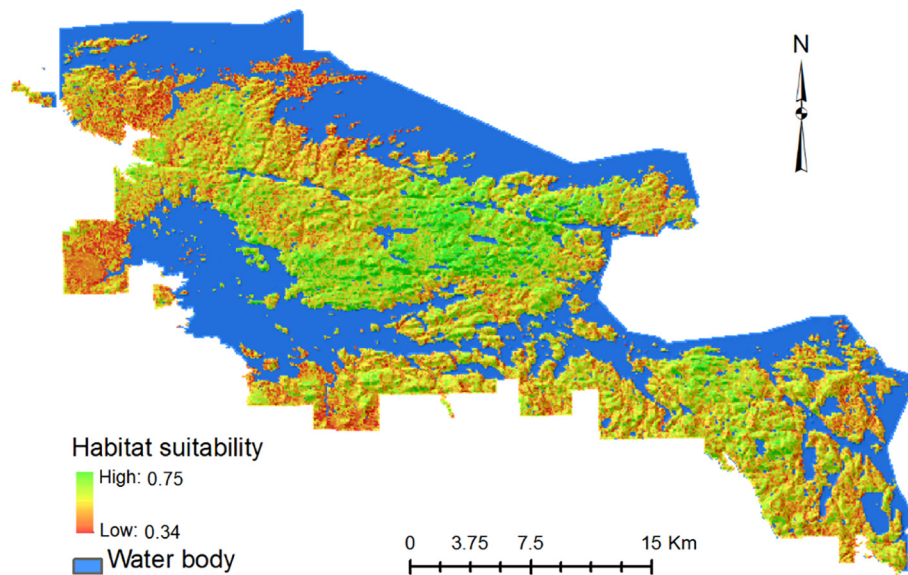


Fig. 4. Habitat suitability map of white-tailed deer in winter in Voyageurs National Park, MN, USA, predicted using our approach.

preferred slopes around  $12.4^\circ$  (whose tangent = 0.22) more than slopes around  $2.9^\circ$  (Fig. 3b). The suitability-slope relationship, characterizing how deer habitat suitability responds to slope conditions (Fig. 3c), can be readily interpreted as the realized environmental niche or resource selection function of the deer in winter.

### 3.4.2. Habitat suitability map

Fig. 4 shows the predicted winter deer habitat suitability map in VNP. Large patches of high suitability areas were in the central parts of the park. Smaller patches of relatively high suitability areas were also in the northwestern and southeastern parts of the park. Results generally conform to known distribution of white-tailed deer in the park in winter (Gable et al., 2017; Vanderwaal et al., 2015).

## 3.5. Evaluation

### 3.5.1. Performance of the approach

Performance of the proposed approach was evaluated by evenly splitting the 123 presence locations at random into a training set and a test set. Presence locations in the training set (61 locations) were used to predict a habitat suitability map and compute its *training accuracy*. Presence locations in the test set (62 locations) was used to compute the *test accuracy* of the suitability map. The above procedures (random split, prediction, evaluation) were repeated 100 times.

The area under the ROC (receiver operating characteristic) curve (AUC) was adopted as an accuracy measure of the predicted suitability map. AUC can be computed for a suitability map given species presence locations (positive) and background locations chosen uniformly at random from the study area (negative) (Phillips and Dudík, 2008). 5000 locations were randomly sampled from the VNP study area and were used as background locations for computing AUC. A ROC curve is obtained by plotting all true positive fraction values on the y axis against their equivalent false positive fraction for all available suitability thresholds on the x axis (Fielding and Bell, 1997). The AUC is the probability that the predicted suitability at a randomly chosen presence location will be higher than that at a randomly chosen background location (Phillips et al., 2006). The AUC ranges from 0.5 to 1.0. A value of 0.5 indicates that the prediction is no better than random predictions. A value of 1.0 indicates perfect model performance. But with presence-only test data the maximum achievable AUC is less than 1.0 (Wiley et al., 2003). AUC provides a single accuracy measure that is independent of any choice of threshold.

Running the proposed approach on the 100 random splits, the average training AUC was 0.792 (standard deviation, SD = 0.029), indicating that the approach can fit training data fairly well. The average test AUC was 0.664 (SD = 0.025), suggesting that the approach achieved better-than-random prediction performance.

### 3.5.2. Comparison against MAXENT

Performance of the proposed approach was compared against the MAXENT method (Phillips et al., 2006). MAXENT was used as a benchmark because it is the state-of-the-art and most widely used presence-only method and it generally achieves good performance (Elith et al., 2006; Elith and Graham, 2009). Also, default parameter values of MAXENT were fine-tuned using a large data set and thus users are relieved from the difficulty of tweaking model parameters (Phillips and Dudík, 2008). The MAXENT software (version 3.3.3k) with its default parameter settings was used in this study.

The habitat suitability map predicted using MAXENT with the 123 deer presence data is shown in Fig. 5. Noticeably, this map shows a spatial distribution pattern of winter deer habitat suitability that is very similar to the suitability map predicted using our approach (Fig. 4): patches of relatively high suitability areas were in the central, northwestern and southeastern parts of the park.

Running MAXENT on the 100 random splits, the average training AUC was 0.784 (SD = 0.021) and the average test AUC was 0.673 (SD = 0.027). The average training AUC of our approach was significantly higher than MAXENT (*paired t-test*,  $t = 3.632$ ,  $p < 0.001$ ). The average test AUC of our approach was significantly lower than MAXENT (*paired t-test*,  $t = -3.829$ ,  $p < 0.001$ ), though only by 0.009 units. Overall, our approach achieved performance that is comparable to MAXENT. But compared to the prediction-oriented MAXENT, one strength of our approach is that the modelled suitability-environment responses are more amenable to ecological interpretation (Section 3.4.1).

## 4. Discussion

### 4.1. Effects of accounting for resource availability

We anticipated that the proposed approach, which models habitat suitability using the presence PDF adjusted by background PDF to account for resource availability, should perform better than modelling habitat suitability purely based on the presence PDF. To investigate the

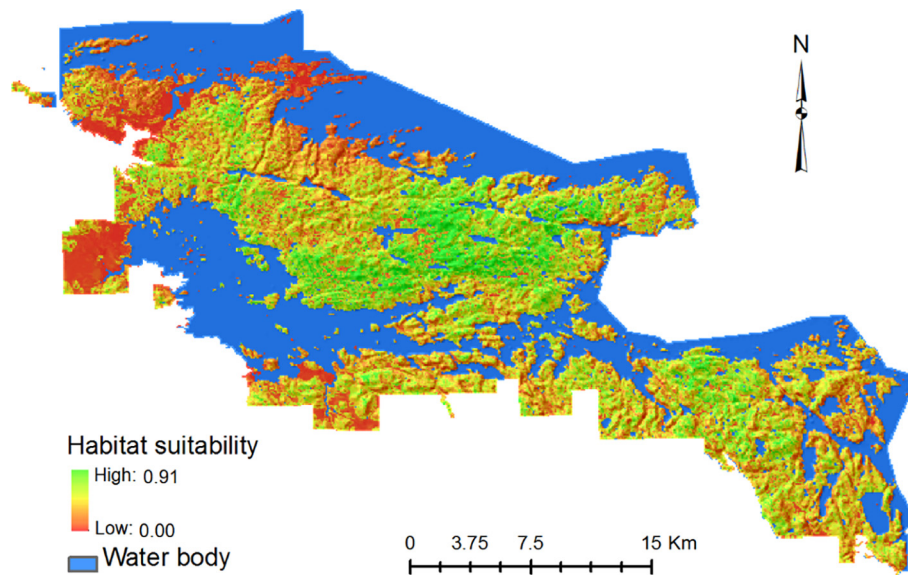


Fig. 5. Habitat suitability map of white-tailed deer in winter in Voyageurs National Park, MN, USA predicted using MAXENT.

effectiveness of accounting for resource availability, the proposed approach was compared against a variant that models habitat suitability by simply normalizing the presence PDF:

$$S'(x) = \frac{f_{\text{presence}}(x)}{\max(f_{\text{presence}}(x))} \quad (7)$$

where  $S'(x)$  is species habitat suitability with respect to environmental factor  $x$ ,  $f_{\text{presence}}(x)$  is the presence PDF, and  $\max(f_{\text{presence}}(x))$  is the maximum density value of the presence PDF.  $S'(x)$  is bounded within  $[0, 1]$ . In the variant approach, Eq. (7) was used to compute habitat suitability with respect to individual environmental factors (replacing Eq. (5)). The overall habitat suitability was computed using Eq. (6).

Running this variant approach on the 100 random splits, the average training AUC was 0.743 (SD = 0.030) and the average test AUC was 0.645 (SD = 0.023). The average training AUC was significantly lower than the original approach (*paired t-test*,  $t = -36.875$ ,  $p < 0.001$ ). The average test AUC was also significantly lower than the original approach (*paired t-test*,  $t = -7.501$ ,  $p < 0.001$ ). The average training AUC and test AUC of the original approach were 0.049 and 0.020 units higher than the variant approach, which suggests that accounting for resource availability did effectively improve model performance.

#### 4.2. Impact of the aggregation strategy

The “weighted average” strategy was adopted to aggregate suitability values with respect to individual environmental factors to determine the overall habitat suitability (Eq. (6)). Many habitat suitability modelling methods adopt this strategy to aggregate the effects of environmental factors on species habitat suitability, for example GLM (Guisan et al., 2002), ENFA (Hirzel et al., 2002) and MAXENT (Phillips et al., 2006). The “weighted average” strategy implicitly assumes that the ecological effect of one environmental factor on species can be compensated by other factors. An alternative aggregation strategy is the “limiting factor” strategy adopted by some other habitat suitability modelling methods (Li and Hilbert, 2008; Zhang et al., 2017a; Zhu et al., 2015). The “limiting factor” strategy assumes that the ecological effect of one environmental factor on species is indispensable and thus cannot be compensated by other factors.

To investigate the impact of the aggregation strategy, the proposed approach was compared against a variant that computes the overall

habitat suitability following the “limiting factor” aggregation strategy. The variant approach takes the minimum amongst the suitability values to individual environmental factors as the overall habitat suitability:

$$S'(x^1, x^2, x^j, \dots, x^m) = \min_{j=1,2,\dots,m} w_j S(x^j) \quad (8)$$

In the variant approach, Eq. (8) replaces Eq. (6) to compute the overall habitat suitability (environmental factors were equally weighted by 1). Habitat suitability with respect to individual environmental factors was computed using Eq. (5).

Running this variant approach on the 100 random splits, the average training AUC was 0.775 (SD = 0.034) and the average test AUC was 0.615 (SD = 0.038). The average training AUC was significantly lower than that of the original approach (*paired t-test*,  $t = -8.484$ ,  $p < 0.001$ ). The average test AUC was also significantly lower than the original approach (*paired t-test*,  $t = -14.757$ ,  $p < 0.001$ ). The average training AUC and test AUC of the original approach were 0.017 and 0.049 higher than the variant approach, which indicates that adopting the “weighted average” aggregation strategy achieved better model performance.

#### 4.3. Impact of environmental factors

Environmental factors used for modelling deer habitat suitability were primarily determined based on domain knowledge of the ecology of the species in the study area (Section 3.3). The selection of environmental factors may affect model performance. Experiments were conducted on subsets of the full set of six environmental factors to investigate such effects. One subset of environmental factors was obtained by excluding *elevation*, *tangent of slope*, *sine and cosine of aspect*, *snow depth*, or *land cover type* from the six environmental factors (five subsets in total; each consists of five environmental factors). Results of running the proposed KDE approach and MAXENT on the 100 random splits using each of the above subsets of environmental factors were shown in Table 2.

Compared to performance of the two approaches on the full set of environmental factors (KDE: average training AUC = 0.792, test AUC = 0.664; MAXENT: average training AUC = 0.784, test AUC = 0.673), the average training AUCs achieved on the subsets of environmental factors are consistently lower (*paired t-test*,  $p < 0.001$ ). The average test AUCs achieved on most subsets are also lower (*paired t-test*,  $p < 0.001$ ) except that the average test AUC is higher on the subset where *sine* and *cosine of aspect* were excluded (*paired t-test*,



**Table 2**  
Performance of the proposed KDE approach and MAXENT on the 100 random splits using subsets of environmental factors.

	Factor(s) excluded	Elevation		Tangent of slope		Sine and cosine of aspect		Snow depth		Land cover type	
		AUC	Average	SD	Average	SD	Average	SD	Average	SD	Average
KDE	Training	0.777	0.030	0.779	0.028	0.750	0.031	0.759	0.030	0.779	0.031
	Test	0.646	0.025	0.660	0.026	<b>0.670</b>	0.025	0.626	0.026	0.652	0.025
MAXENT	Training	0.774	0.022	0.764	0.024	0.749	0.022	0.769	0.021	0.769	0.021
	Test	0.668	0.026	0.661	0.027	<b>0.685</b>	0.024	0.659	0.026	0.666	0.025

$p < 0.005$ ). We also experimented with subsets of environmental factors excluding *sine* and *cosine of aspect* and other factors but found that model performance consistently degrades (compared to model performance on the subset of environmental factors excluding *sine* and *cosine of aspect*). In general, using subsets of the environmental factors degrades model performance. It is thus recommended using environmental factors identified based on domain knowledge as a starting point for habitat suitability modelling.

#### 4.4. Advantages of the approach

The proposed kernel density estimation approach offers a very flexible framework for modelling and mapping species habitat suitability from species presence-only data. *To begin with*, both quantitative (continuous) and qualitative (categorical) environmental factors can be used naturally in this approach. In many other habitat suitability modelling methods, a categorical factor is either converted to binary factors (MAXENT, Phillips et al., 2006) or replaced by quantitative factors derived from the original categorical factor (ENFA, Hirzel et al., 2002).

*Second*, our approach accounts for resource availability in modelling habitat suitability. Resource availability is an important consideration when modelling species habitat suitability from presence-only data (Section 2.1) but is in lack of consideration in many other methods (e.g., ENFA, Hirzel et al., 2002). The case study shows that accounting for resource availability did effectively improve model performance.

*Third*, our approach is a nonparametric approach that does not impose any assumptions on the shape of species suitability-environment responses. Compared to other methods (e.g., ENFA), our approach provides flexibility for habitat suitability modelling on species whose suitability-environment responses do not follow a presumed unimodal and symmetrical shape. In addition, the modelled suitability-environment responses approximate species environmental niches or species resource selection functions (Boyce et al., 2002; Johnson et al., 2006) and thus are amenable to ecological interpretation. This is a strength of our approach compared to other prediction-oriented modelling methods (e.g., MAXENT).

*Fourth*, our approach is flexible to incorporate environmental factor importance when modeling habitat suitability. By default, environmental factors are assumed equally important to species habitat suitability and are equally weighted (Eq. (6)). But if knowledge about the relative importance of different factors is available, varying weights can be assigned on environmental factors in our approach to model species habitat suitability.

*Finally*, sample selection bias is a common issue for species presence-only data (Phillips, 2008; Phillips et al., 2009). Within the framework of our approach, mechanisms of correcting for sample selection bias can be easily implemented. For example, in estimating species presence PDF over environmental factors, presence locations can be weighted by sampling effort to compensate for bias in species presence data (Zhu et al., 2015).

#### 4.5. Applicability of the approach for ecological monitoring

Understanding species distribution and environmental niches is crucial to many ecosystem and species management efforts, and ecological monitoring programs have been widely established to support such missions (Lindenmayer and Likens, 2010). Species presence-only data are collected and accumulated in many monitoring programs. Our approach is well suited for analyzing this kind of species data. The resultant species-environment responses and species habitat suitability maps in turn can be useful for ecosystem and species monitoring and for supporting management decision making.

## 5. Conclusions

This article presents a novel approach for modelling and mapping habitat suitability from species presence-only data. The approach models the relationship between species habitat suitability and environmental conditions using probability distributions of species presence over environmental factors. It accounts for resource availability by adjusting species presence probability distribution with background probability distribution. The probability distributions were estimated using nonparametric kernel density estimation and thus the approach imposes no assumptions on the shape of species suitability-environment relationships (e.g., linear, gaussian, unimodality, symmetry). Besides, the suitability-environment relationships be readily interpreted as species realized environmental niches or resource selection functions.

The case study of modelling and mapping winter habitat suitability for white-tailed deer at Voyageurs National Park demonstrated that the approach achieved good model performance. It fit training data fairly well and achieved better-than-random prediction that is comparable to the state-of-the-art MAXENT method. In addition, the suitability-environment responses modelled using our approach are more amenable to ecological interpretation compared to MAXENT. Experimental results also suggest that accounting for resource availability effectively improved performance of the approach. The approach offers a flexible framework for modelling and mapping species habitat suitability from species presence-only data. The resultant species-environment responses and species habitat suitability maps can be very useful for ecosystem and species monitoring and for supporting management decision making.

## Acknowledgements

This study was supported by grants from the National Natural Science Foundation of China (NSFC, No 41431177), National Basic Research Program of China (No 2015CB954102), Natural Science Research Program of Jiangsu (No 14KJA170001), PAPD, National Key Technology Innovation Project for Water Pollution Control and Remediation (No 2013ZX07103006), and Outstanding Innovation Team in Colleges and Universities in Jiangsu Province. Support to Guiming Zhang through the Whitbeck Graduate Dissertator Award from the Department of Geography, University of Wisconsin-Madison is appreciated. Supports to A-Xing Zhu through the Vilas Associate Award, the Hammel Faculty Fellow Award, the Manasse Chair Professorship



from the University of Wisconsin-Madison, and the “One-Thousand Talents” Program of China are greatly appreciated. Cheng-Zhi Qin is grateful for support through the “Excellent Young Scholars” project from National Natural Science Foundation of China. We thank pilot W. Erickson and observers B. Olson, R. Moen, and F. Swendsen for assistance with aerial surveys in 2009.

## References

- Bolvar-Cirín, B., Gallina, S., 2012. An optimal habitat model for the white-tailed deer (*Odocoileus virginianus*) in central Veracruz, Mexico. *Anim. Prod. Sci.* 52, 707–713. <http://dx.doi.org/10.1071/AN12013>.
- Boyce, M.S., Vernier, P.R., Nielsen, S.E., Schmiegelow, F.K.A., 2002. Evaluating resource selection functions. *Ecol. Modell.* 157, 281–300. [http://dx.doi.org/10.1016/s0304-3800\(02\)00200-4](http://dx.doi.org/10.1016/s0304-3800(02)00200-4).
- Brooks, T., 2004. Protected areas and species. *Conserv. Biol.* 18, 616–618.
- Brunsdon, C., 1995. Estimating probability surfaces for geographical point data: an adaptive kernel algorithm. *Comput. Geosci.* 21, 877–894. [http://dx.doi.org/10.1016/0098-3004\(95\)00020-9](http://dx.doi.org/10.1016/0098-3004(95)00020-9).
- Busby, J.R., 1991. BIOCLIM: a bioclimate analysis and prediction system. In: Margules, C.R., Austin, M.P. (Eds.), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Melbourne, Australia, pp. 64–68.
- Carpenter, G., Gislison, A.N., Winter, J., 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* 2, 667–680. <http://dx.doi.org/10.1007/bf00051966>.
- Chefaoui, R.M., Lobo, J.M., 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecol. Modell.* 210, 478–486. <http://dx.doi.org/10.1016/j.ecolmodel.2007.08.010>.
- Cobb, M.A., Gogan, P.J.P., Kozie, K.D., Olexa, E.M., Lawrence, R.L., Route, W.T., 2004. Relative spatial distributions and habitat use patterns of sympatric moose and white-tailed deer in Voyageurs National Park, Minnesota. *Alces* 40, 169–191.
- De’Ath, G., 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. *Ecology* 83, 1105–1117.
- Elith, J., Graham, C.H., 2009. Do they? how do they? why do they differ? on finding reasons for differing performances of species distribution models. *Ecography (Cop.)* 32, 66–77. <http://dx.doi.org/10.1111/j.1600-0587.2008.05505.x>.
- Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species’ distributions from occurrence data. *Ecography (Cop.)* 29, 129–151.
- Elith, J., Leathwick, J., 2007. Predicting species distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Divers. Distrib.* 13, 265–275. <http://dx.doi.org/10.1111/j.1472-4642.2007.00340.x>.
- Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Syst.* 40, 677–697. <http://dx.doi.org/10.1146/annurev.ecolsys.110308.120159>.
- Engler, R., Guisan, A., Rechsteiner, L., 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *J. Appl. Ecol.* 41, 263–274. <http://dx.doi.org/10.1111/j.0021-8901.2004.00881.x>.
- ESRI (Environmental Systems Research Institute), 2013. ArcGIS Desktop 10.1.
- Farber, O., Kadmon, R., 2003. Assessment of alternative approaches for bioclimatic modeling with special emphasis on the Mahalanobis distance. *Ecol. Modell.* 160, 115–130. [http://dx.doi.org/10.1016/s0304-3800\(02\)00327-7](http://dx.doi.org/10.1016/s0304-3800(02)00327-7).
- Ferrier, S., Watson, G., Pearce, J., Drielsma, M., 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodivers. Conserv.* 11, 2275–2307.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24, 38–49.
- Franklin, J., Miller, J.A., 2009. *Mapping Species Distributions: Spatial Inference and Prediction*. Cambridge University Press, Cambridge.
- Fry, J.A., Xian, G., Jin, S., Dewitz, J.A., Homer, C.G., Limin, Y., Barnes, C.A., Herold, N.D., Wickham, J.D., 2011. Completion of the 2006 national land cover database for the conterminous United States. *Photogramm. Eng. Remote Sensing* 77, 858–864.
- Gable, T.D., Windels, S.K., Olson, B.T., 2017. Estimates of white-tailed deer density in Voyageurs National Park: 1989–2016. *Natural Resource Report NPS/VOYA/NRR—2017/1427*. Fort Collins, Colorado.
- Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., Tyler, D., 2002. The national elevation dataset. *Photogramm. Eng. Remote Sensing* 68, 5–11.
- Gillespie, T.W., Foody, G.M., Rocchini, D., Giorgi, A.P., Saatchi, S., 2008. Measuring and modelling biodiversity from space. *Prog. Phys. Geogr.* 32, 203–221. <http://dx.doi.org/10.1177/0309133308093606>.
- Gogan, P.J.P., Kozie, K.D., Olexa, E.M., Duncan, N.S., 1997. Ecological status of moose and white-tailed deer at Voyageurs National Park, Minnesota. *Alces* 33, 187–201.
- Gogan, P.J.P., Route, W.T., Olexa, E.M., Thomas, N., Kuehn, D., Podruzny, K.M., 2004. *Gray Wolves in and Adjacent to Voyageurs National Park, Minnesota: Research and Synthesis 1987–1991*. National Park Service.
- Graham, C.H., Ferrier, S., Huettmann, F., Moritz, C., Peterson, A.T., 2004a. New developments in museum-based informatics and applications in biodiversity analysis. *Trends Ecol. Evol.* 19, 497–503. <http://dx.doi.org/10.1016/j.tree.2004.07.006>.
- Graham, C.H., Ron, S.R., Santos, J.C., Schneider, C.J., Moritz, C., 2004b. Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution (N. Y.)* 58, 1781–1793. <http://dx.doi.org/10.1111/j.0014-3820.2004.tb00461.x>.
- Gu, W.D., Swihart, R.K., 2004. Absent or undetected? effects of non-detection of species occurrence on wildlife-habitat models. *Biol. Conserv.* 116, 195–203. [http://dx.doi.org/10.1016/s0006-3207\(03\)00190-3](http://dx.doi.org/10.1016/s0006-3207(03)00190-3).
- Guisan, A., Edwards Jr, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Modell.* 157, 89–100. [http://dx.doi.org/10.1016/s0304-3800\(02\)00204-1](http://dx.doi.org/10.1016/s0304-3800(02)00204-1).
- Guisan, A., Zimmerman, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Modell.* 135, 147–186. [http://dx.doi.org/10.1016/s0304-3800\(00\)00354-9](http://dx.doi.org/10.1016/s0304-3800(00)00354-9).
- Guo, Q., Kelly, M., Graham, C.H., 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecol. Modell.* 182, 75–90. <http://dx.doi.org/10.1016/j.ecolmodel.2004.07.012>.
- Hanberry, B.B., He, H.S., Palik, B.J., 2012. Pseudoabsence generation strategies for species distribution models. *PLoS One* 7, e44486. <http://dx.doi.org/10.1371/journal.pone.0044486>.
- Hirzel, A.H., Arlettaz, R., 2003. Modeling habitat suitability for complex species distribution by environmental-distance geometric mean. *Environ. Manage.* 32, 614–623.
- Hirzel, A.H., Haussler, J., Chessel, D., Perrin, N., 2002. Ecological-niche factor analysis: how to compute habitat-suitability maps without absence data? *Ecology* 83, 2027–2036.
- Hirzel, A.H., Lay, G. Le, 2008. Habitat suitability modelling and niche theory. *J. Appl. Ecol.* 45, 1372–1381. <http://dx.doi.org/10.1111/j.1365-2664.2008.01524.x>.
- Johnson, C.J., Nielsen, S.E., Merrill, E.H., McDonald, T.L., Boyce, M.S., 2006. Resource selection functions based on use-availability data: theoretical motivation and evaluation methods. *J. Wildl. Manage.* 70, 347–357.
- Jones, M.C., Marron, J.S., Sheather, S.J., 1996. Progress in data-based bandwidth selection for kernel density estimation. *Comput. Stat.* 11, 337–381.
- Kerr, J.T., Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. *Trends Ecol. Evol.* 18, 299–305. [http://dx.doi.org/10.1016/S0169-5347\(03\)00071-5](http://dx.doi.org/10.1016/S0169-5347(03)00071-5).
- Leathwick, J.R., Austin, M.P., 2001. Competitive interactions between tree species in New Zealand’s old-growth indigenous forests. *Ecology* 82, 2560–2573. [http://dx.doi.org/10.1890/0012-9658\(2001\)082\[2560:CIBTSI\]2.0.CO;2](http://dx.doi.org/10.1890/0012-9658(2001)082[2560:CIBTSI]2.0.CO;2).
- Leibold, M.A., 1995. The niche concept revisited: mechanistic models and community context. *Ecology* 76, 1371–1382.
- Li, J., Hilbert, D.W., 2008. LIVES: a new habitat modelling technique for predicting the distribution of species’ occurrences using presence-only data based on limiting factor theory. *Biodivers. Conserv.* 17, 3079–3095. <http://dx.doi.org/10.1007/s10531-007-9270-7>.
- Lindenmayer, D.B., Likens, G.E., 2010. The science and application of ecological monitoring. *Biol. Conserv.* 143, 1317–1328. <http://dx.doi.org/10.1016/j.biocon.2010.02.013>.
- Mac Nally, R., 2000. Regression and model-building in conservation biology, biogeography and ecology: the distinction between—and reconciliation of—“predictive” and “explanatory” models. *Biodivers. Conserv.* 9, 655–671.
- Moen, A.N., 1976. Energy conservation by white-tailed deer in the winter. *Ecology* 57, 192–198.
- NOHRSC (National Operational Hydrologic Remote Sensing Center), 2004. Snow Data Assimilation System (SNODAS) data products at National Snow and Ice Data Center (NSIDC), Version 1 [WWW Document]. doi: 10.7265/N5TB14TC.
- Ozoga, J.J., Gysel, L.W., 1972. Response of white-tailed deer to winter weather. *J. Wildl. Manage.* 36, 892–896.
- Ozoga, J.J., Verme, L.J., 1970. Winter feeding pattern of penned white-tailed deer. *J. Wildl. Manage.* 34, 431–439.
- Phillips, S.J., 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography (Cop.)* 31, 272–278. <http://dx.doi.org/10.1111/j.0906-7590.2008.5378.x>.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190, 231–259. <http://dx.doi.org/10.1016/j.ecolmodel.2005.03.026>.
- Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography (Cop.)* 31, 161–175.
- Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecol. Appl.* 19, 181–197. <http://dx.doi.org/10.1890/07-2153.1>.
- Pressey, R.L., 2004. Conservation planning and biodiversity: assembling the best data for the job. *Conserv. Biol.* 18, 1677–1681. <http://dx.doi.org/10.1111/j.1523-1739.2004.00434.x>.
- Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London, UK.
- Stockwell, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. *Int. J. Geogr. Inf. Sci.* 13, 143–158.
- Stokland, J.N., Halvorsen, R., Støa, B., 2011. Species distribution modelling—effect of design and sample size of pseudo-absence observations. *Ecol. Modell.* 222, 1800–1809. <http://dx.doi.org/10.1016/j.ecolmodel.2011.02.025>.
- Telesco, R., Van Manen, F.T., Clark, J.D., Cartwright, M.E., 2007. Identifying Sites for Elk Restoration in Arkansas. *J. Wildl. Manage.* 71, 1393–1403. <http://dx.doi.org/10.2193/2005-673>.
- Thorn, J.S., Nijman, V., Smith, D., Nekaris, K.A.I., 2009. Ecological niche modelling as a technique for assessing threats and setting conservation priorities for Asian slow lorises (*Primates: Nycticebus*). *Divers. Distrib.* 15, 289–298. <http://dx.doi.org/10.1111/j.1472-4642.2008.00535.x>.

- Thuiller, W., Richardson, D.M., Pyšek, P., Midgley, G.U.Y.F., Hughes, G.O., Rouget, M., 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Glob. Chang. Biol.* 11, 2234–2250.
- Vanderwaal, K.L., Windels, S.K., Olson, B.T., Vannatta, J.T., Moen, R., 2015. Landscape influence on spatial patterns of meningeal worm and liver fluke infection in white-tailed deer. *Parasitology* 142, 706–718. <http://dx.doi.org/10.1017/S0031182014001802>.
- Viña, A., Bearer, S., Zhang, H., Ouyang, Z., Liu, J., 2008. Evaluating MODIS data for mapping wildlife habitat distribution. *Remote Sens. Environ.* 112, 2160–2169. <http://dx.doi.org/10.1016/j.rse.2007.09.012>.
- Walker, P.A., Cocks, K.D., 1991. HABITAT: a procedure for modelling a disjoint environmental envelope for a plant or animal species. *Glob. Ecol. Biogeogr. Lett.* 1, 108–118.
- Warren, D.L., 2012. In defense of “niche modeling”. *Trends Ecol. Evol.* 27, 497–500. <http://dx.doi.org/10.1016/j.tree.2012.03.010>.
- Wiley, E.O., McNyset, K., Peterson, T., Robins, R., Stewart, A., 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. *Oceanography* 16, 120–127. <http://dx.doi.org/10.5670/oceanog.2003.42>.
- Windels, S.K., 2014. 2013 Voyageurs National Park moose population survey report, Natural Resource Data Series NPS/VOYA/NRDS—2014/644. Fort Collins, Colorado.
- Windels, S.K., Olson, B.T., 2017. Voyageurs National Park moose population survey report: 2017. Nat. Resour. Rep. NPS/VOYA/NRR—2017/1455. Natl. Park Serv. Fort Collins, Color. USA.
- Wisn, M.S., Guisan, A., 2009. Do pseudo-absence selection strategies influence species distribution models and their predictions? an information-theoretic approach based on simulated data. *BMC Ecol.* 9, 8. <http://dx.doi.org/10.1186/1472-6785-9-8>.
- Zhang, G., Zhu, A.-X., Huang, Z.-P., Ren, G., Qin, C.-Z., Xiao, W., 2017a. Validity of historical volunteered geographic information: Evaluating citizen data for mapping historical geographic phenomena. *Trans. GIS* 1–16. <http://dx.doi.org/10.1111/tgis.12300>.
- Zhang, G., Zhu, A.-X., Huang, Z.-P., Xiao, W., 2017b. A heuristic-based approach to mitigating positional errors in patrol data for species distribution modeling. *Trans. GIS* 1–15. <http://dx.doi.org/10.1111/tgis.12303>.
- Zhu, A.-X., Zhang, G., Wang, W., Xiao, W., Huang, Z.-P., Dunzhu, G.-S., Ren, G., Qin, C.-Z., Yang, L., Pei, T., Yang, S., 2015. A citizen data-based approach to predictive mapping of spatial variation of natural phenomena. *Int. J. Geogr. Inf. Sci.* 29, 1864–1886. <http://dx.doi.org/10.1080/13658816.2015.1058387>.