

ORIGINAL RESEARCH

Ecological niche modeling re-examined: A case study with the Darwin's fox

Luis E. Escobar¹  | Huijie Qiao² | Javier Cabello^{3,4} | A. Townsend Peterson⁵

¹Department of Fish and Wildlife Conservation, Virginia Tech, Blacksburg, VA, USA

²Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China

³Centro de Conservación de la Biodiversidad Chiloé-Silvestre, Los Lagos, Chile

⁴Facultad de Medicina Veterinaria, Universidad San Sebastián, Puerto Montt, Chile

⁵Biodiversity Institute, University of Kansas, Lawrence, KS, USA

Correspondence

Huijie Qiao, Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China.
Email: qiaohj@ioz.ac.cn

Funding information

National Key R&D Program of China, Grant/Award Number: 2017YFC1200603; National Natural Science Foundation of China, Grant/Award Number: 31772432

Abstract

Many previous studies have attempted to assess ecological niche modeling performance using receiver operating characteristic (ROC) approaches, even though diverse problems with this metric have been pointed out in the literature. We explored different evaluation metrics based on independent testing data using the Darwin's Fox (*Lycalopex fulvipes*) as a detailed case in point. Six ecological niche models (ENMs; generalized linear models, boosted regression trees, Maxent, GARP, multivariable kernel density estimation, and NicheA) were explored and tested using six evaluation metrics (partial ROC, Akaike information criterion, omission rate, cumulative binomial probability), including two novel metrics to quantify model extrapolation versus interpolation (E-space index I) and extent of extrapolation versus Jaccard similarity (E-space index II). Different ENMs showed diverse and mixed performance, depending on the evaluation metric used. Because ENMs performed differently according to the evaluation metric employed, model selection should be based on the data available, assumptions necessary, and the particular research question. The typical ROC AUC evaluation approach should be discontinued when only presence data are available, and evaluations in environmental dimensions should be adopted as part of the toolkit of ENM researchers. Our results suggest that selecting Maxent ENM based solely on previous reports of its performance is a questionable practice. Instead, model comparisons, including diverse algorithms and parameterizations, should be the *sine qua non* for every study using ecological niche modeling. ENM evaluations should be developed using metrics that assess desired model characteristics instead of single measurement of fit between model and data. The metrics proposed herein that assess model performance in environmental space (i.e., E-space indices I and II) may complement current methods for ENM evaluation.

KEYWORDS

ecological niche modeling, evaluation, *Lycalopex fulvipes*, Maxent, ROC AUC

1 | INTRODUCTION

Ecological niche models (ENMs) have been employed as a predictive tool in diverse research applications, including studies of distributional ecology, biological conservation, climate change effects,

evolution, and spatial epidemiology (Peterson et al., 2011). ENM approaches link occurrence data with environmental variables based on a correlative approach to build a representation of a species' ecological requirements. Numerous algorithms have been used to create ENMs, which provide geographic outputs that approximate

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2018 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

distributional areas of species (Franklin & Miller, 2009). Ecological niches are manifested in environmental spaces that comprise sets of abiotic variables that shape the species' potential occurrence; niches translate into geographic distributions according to the combined effects of the distribution of the abiotic conditions, biotic interactions, and accessibility by dispersal (Soberón & Peterson, 2005). Even though records of distribution of a species may be abundant, they may be biased, characterizing just a portion of the species' niche, limited by biotic factors (e.g., interspecific competition), dispersal constraints, biased in sampling effort, or simply the existence of sets of conditions on relevant landscapes (Soberón & Peterson, 2005).

Such disjunctions between the fundamental niche and what is observable of it occur when distributional limits are set chiefly or entirely by dispersal considerations, termed the "Wallace's Dream" scenario (Qiao, Soberón, & Peterson, 2015; Saupe et al., 2012). This idea refers to Alfred Russel Wallace, who realized that geographic barriers often limit species to circumscribed regions. The "Wallace's Dream" scenario describes situations in which a species' distributional potential is circumscribed by barriers to dispersal rather than by unsuitable conditions. Take, for instance, the case of the shrub, *Acacia mearnsii*, which was originally restricted by geographic barriers to southeastern Australia and Tasmania. However, this species has a broad potential distribution, and, once introduced to regions beyond the barriers that originally confined it, spread across the Americas, Europe, Asia, Africa, New Zealand, and Pacific and Indian Ocean islands, making it one of the most successful invaders, in light of its capacity to establish populations in new regions worldwide. In Wallace's Dream situations (i.e., species' distributions constrained by dispersal rather than by environmental conditions), ENMs aiming to estimate a species' fundamental niche and in turn its potential distribution lack necessary contrasts for adequate model calibration and, as a consequence, make erroneous conclusions of the species true potential and generally lack predictive ability (Owens et al., 2013; Saupe et al., 2012).

Appropriate evaluation of ENM predictions requires considerable preparation and care: that evaluation samples be independent and that each be representative of the population under study (Hurlbert, 1984). These assumptions are violated when ENMs are evaluated without accounting for spatial autocorrelation and sampling bias implicit in data from real species or when random points are used to replace absence records, which is frequent (Guillera-Arroita et al., 2015), although often not appreciated. Models evaluated using incorrect metrics and nonindependent data generate incorrect or incomplete results (Lobo, Jiménez-Valverde, & Real, 2008). Strikingly, robust model assessments are quite rare in ENM, and users too often trust software without understanding or assessing predictive performance (Joppa et al., 2013). Even the important and highly cited work of Elith et al. (2006), which assessed different ENMs based on a large-scale suite of species and regions, and found that some methods outperformed others, was based on data susceptible to bias (Kadmon, Farber, & Danin, 2004), used biologically unrealistic and mathematically weak evaluation metrics (Lobo et al., 2008; Peterson, Papeş, & Soberón, 2008), and explored only a single feature of model performance (i.e., omission error; see below).

Even today, ENM evaluation methods are limited and restricted to geographic dimensions (Muscarella et al., 2014), even in spite of the fact that ecological niches are manifested in environmental space. The limited availability of metrics for robust model evaluation is alarming given how often ENMs are used to map organisms of high public interest such as agents of infectious diseases, agricultural pests, and endangered species (Peterson et al., 2011). Recent studies suggest that different ENM methods can differ in their performance under diverse circumstances (Qiao et al., 2015), such that no single "best" ENM likely exists, signaling the need for a critical examination of the unquestioning use of particular methods by modelers (e.g., Bhatt et al., 2013). Hence, it is critical to develop new evaluation metrics that can assess diverse characteristics of ENMs, including the amount of interpolation and extrapolation, that is, prediction inside or outside the range of environmental values observed, respectively. In this study, we explored different evaluation metrics using a particular Wallace's Dream case study, the Darwin's Fox (*Lycalopex fulvipes*, Martin 1837), and its likely geographic distribution. We assessed diverse ENM (generalized linear models, boosted regression trees, Maxent, GARP, multivariable kernel density estimation, and NicheA) and evaluation metrics (partial ROC, Akaike information criterion, omission rate, cumulative binomial probability, and E-space indices I and II) to test how predictive performance may vary or differ in behavior based on the evaluation metric employed.

2 | METHODS

2.1 | Case study

We used data from the Darwin's Fox, the only endemic canid of Chile, known from the southern temperate forests along the Pacific coast (Yahnke, 1995). The species was reported by Charles Darwin in 1834 on Chiloé Island and was long considered as an island endemic. However, in 1990, a mainland population was reported at Nahuelbuta National Park to the north, 550 km from the island population (Medel, Jiménez, Jaksić, Yáñez, & Armesto, 1990).

The Darwin's Fox faces important conservation challenges in terms of conflicts with human settlements and dramatic habitat loss (Jiménez, Lucherini, & Novaro, 2008; Sillero-Zubiri, 2004). A recent survey on Chiloé Island, home to the largest population of the fox (~250 individuals), revealed that >85% of local citizens had negative attitudes toward the fox (Molina-Espinosa, 2011). The Darwin's Fox is threatened by illegal hunting, apparent competition with other fox species and domestic dogs, lack of ex situ management, and limited knowledge of the basic biology of the species (Baillie, Hilton-Taylor, & Stuart, 2004; Jaksić, Jiménez, Medel, & Marquet, 1990; Jiménez & McMahon, 2004; Sillero-Zubiri, 2004). With an estimated global population of ~375 individuals and limited knowledge of areas for potential reintroduction, Darwin's Fox is considered to be the canid species at highest risk of extinction globally (Escobar, 2013).

Hence, biological and ecological processes affecting the endangered Darwin's Fox remain poorly understood, complicating efforts to guide field research and policies for its conservation

(Sillero-Zubiri, 2004). The geographic separation of the Chiloé Island and Nahuelbuta populations, due to past (e.g., Chacao Channel dividing Chiloé Island and the mainland) or recent (e.g., extirpation) circumstances, provides an ideal situation in which to test the performance of ecological niche modeling approaches under a situation approximating a “Wallace’s Dream” configuration. Thus, ENMs based on this configuration make necessary assessment of diverse evaluation metrics, which can help to identify good candidate models for a Wallace’s Dream scenario: In our particular case study, we prioritize ENMs providing high interpolation and low extrapolation to avoid an exaggerated potential range and focus conservation efforts while accounting for the species’ observed environmental tolerances.

2.2 | Occurrence data

Considering the limited knowledge on this species, we collected data on Darwin’s Fox occurrences from diverse sources, including the available literature in English, German, and Spanish; natural history museum collections data online; and camera-trap observations and field observations from our long-term studies (see summary in Appendix S1). We separated occurrences into three population groups: Nahuelbuta (“northern”), Chiloé Island (“southern”), and new records of the species’ occurrence (“central”; Figure 1a). The Nahuelbuta population (47 occurrence sites) was termed D_n , the Chiloé Island population (108 occurrences) D_s , and the new records

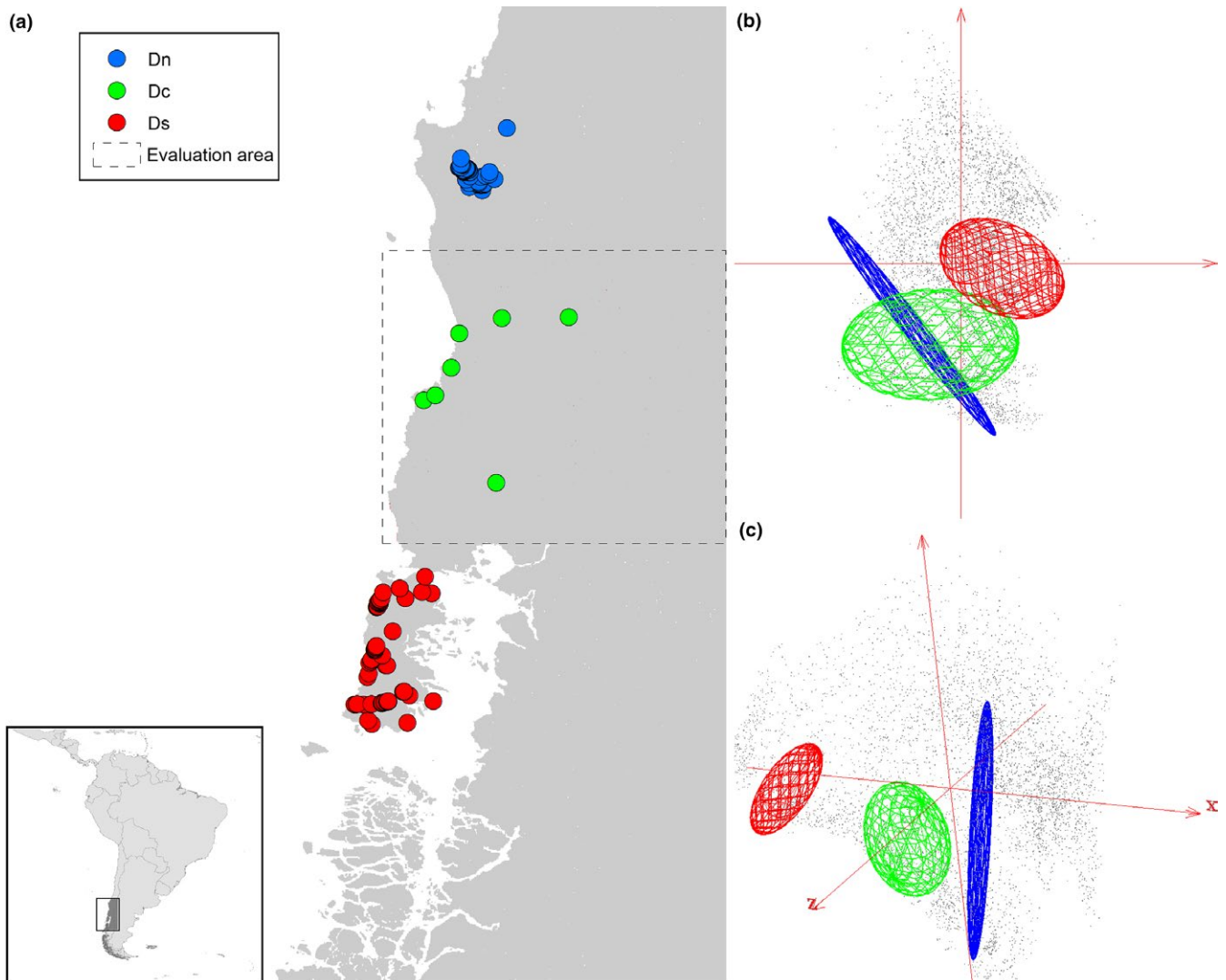


FIGURE 1 Darwin’s Fox (*Lycalopex fulvipes*) occurrences in G and E spaces. (a) Occurrences in geographic space (G) in southern Chile (see inset), including a population in Nahuelbuta National Park (D_n) in the northern part of the known species’ range (blue points), new records in the central part of the species’ distribution (D_c ; green points), and the southern populations on Chiloé Island (D_s ; red points). The evaluation area, not used in model calibration, is denoted by the dashed line. (b) Occurrences in a two-dimensional environmental space (E), with the same color scheme as in panel A; axes are the first two principal components of the 19 bioclimatic variables, and gray points are the environmental background of the study area. Note the environmental overlap between the blue and green ellipsoids in this bidimensional environmental space (84.4% of the variability in the climatic data). (c) Occurrences in a three-dimensional environmental space (E) with symbolism as in geography. Note that in this higher dimensionality space, no environmental overlap exists between blue and green ellipsoids (94.2% of the variability in the climate data)

(7 occurrences) D_c . To reduce pseudoreplication, occurrences were resampled across a $2.5 \times 2.5'$ grid, which reduced numbers of records to 42 for D_n and 61 for D_s . Figure 1 shows known occurrences of the species in both geographic (**G**) and environmental (**E**) spaces.

To assess model predictions, we merged D_n and D_s as a data set for model calibration, as these populations were recognized much earlier than the intervening populations D_c . In fact, the D_c occurrences included in this study are here published for the first time in the scientific literature (Appendix S1). D_c was used for model validation because of its geographic and environmental independence from the D_s and D_n populations (i.e., D_c populations may occur in different areas and under different climatic conditions; Figure 1) and in light of the interest inherent in the question of the relative continuity of populations in between the northern and southern populations.

2.3 | Study area and environmental variables

The extent of the model calibration area has key impacts on ENM results (Barve et al., 2011). We designed our study area based on the prior knowledge of the species' current distribution across the region 70.5 to 74°W and 38.5 to 41°S, resulting in 18,526 cells (Figure 1a). Specifically, we delimited the study area in the south to include Chiloé Island, in the north to include the Nahuelbuta reserve, to the west by the Pacific Ocean, and to the east by the crest of the Andes Mountains.

Environmental dimensions used to model the species' ecological niche included 19 "bioclimatic" variables with a grid resolution of $2.5 \times 2.5'$ (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005). Because many of the environmental variables were highly correlated, we developed a principal component analysis (PCA) and retained components sufficient to explain $\geq 90\%$ of the total variation in the climatic data: The first three principal components accumulated 94.2% of information and were used in analyses.

2.4 | Ecological niche models

We explored predictive power of six ENMs that can be classified into two functional groups. The first group employs correlational approaches that use occurrences and background data to characterize environmental landscapes. The second group of ENMs uses presence data only in generating model outputs.

Models based on occurrences and environmental background data (background **E**) included the maximum entropy method in Maxent v3.3.3k (Phillips, Anderson, & Schapire, 2006), generalized linear models (GLM; MacCullagh & Nelder, 1989) in BIOMOD2 (Thuiller, Lafourcade, Engler, & Araújo, 2009), boosted regression trees (BRT; Elith, Leathwick, & Hastie, 2008) in dismo (Hijmans, Phillips, Leathwick, & Elith, 2012), and the genetic algorithm for rule-set production (GARP; Stockwell, 1999) with the best subsets procedure in openModeller 1.5 (de Souza Muñoz et al., 2011). Models in this group generate outputs with continuous values. Models using only presence data included niche-centroid distance estimation via minimum-volume ellipsoid approaches in NicheA v.3.0 (Qiao,

Peterson, Soberón, Campbell, Ji, & Escobar, 2016) with continuous outputs and hypervolume multivariable kernel density estimation (KDE; Blonder, Lamanna, Violle, & Enquist, 2014) in R (R Core Team 2016) using the package hypervolume (Blonder, 2015) with a binary output. In all cases (i.e., for each of the modeling algorithms), all models were calibrated using default parameters to allow easy replication, fair comparisons with other studies (Elith et al., 2006), and to reproduce customary applications by the broader community of users—however, for illustrative purposes only, a more detailed calibration of Maxent models is described briefly in the Section 4. To facilitate some of the evaluations, all models were set to generate outputs in binary format (i.e., suitable/unsuitable) based on two common threshold values (see below).

Models were calibrated using D_n and D_s and then evaluated using D_c . Evaluations were conducted in a two-dimensional geographic space (i.e., latitude and longitude; **G**) and also in a three-dimensional environmental space (**E**) defined by the first three components derived from the climate data. During evaluation, we assessed three features of the models: abilities of models to predict correctly the independent D_c occurrences, the fit of the model with the calibration data, and levels of interpolation and extrapolation.

2.5 | Model evaluation in geography

2.5.1 | Evaluation of continuous models

Models were evaluated based on their spatial fit with the calibration data and on correct prediction of independent evaluation occurrence data across an evaluation area (D_c ; Figure 1a, green points and dashed line, respectively). To assess continuous-output models, we used two metrics: the Akaike information criterion (AIC; Johnson & Omland, 2004; Burnham, Anderson, & Huyvaert, 2011; Warren & Seifert, 2011) and partial area under the receiver operating characteristic curve (pROC; Peterson et al., 2008). Details of our implementation of the two evaluation approaches are in the paragraphs that follow.

AIC is a penalized likelihood criterion expressed as $AIC = 2 \cdot K - 2 \cdot \ln(L)$, where L is the maximized value of the likelihood function for a model and K is the number of parameters employed in the model (Burnham et al., 2011). For Maxent models, K is extracted from the "lambdas" file (Warren & Seifert, 2011). GLM were quadratic and without interaction terms and, based on the first three principal components, we set $K = 6$.

Computing K for BRT and GARP is more complex. Generally, the K of a classification model (regression tree model or genetic algorithm) is calculated as $K = N \times (p + (1/2p(p+1)) + c)$, where N is the number of nodes on the regression tree (Sain & Carmack, 2002) or the number of individuals per population for a genetic algorithm (Kosakovsky, Mannino, Gravenor, Muse, & Frost, 2007; Yoshimoto, Moriyama, & Harada, 1999), p is the dimensionality of the variables, and c is the penalty for each data-based split. Here, we generated 400 populations in GARP model and used 1,000 regression trees in the BRT models; we set N to the average number of individuals

among the 400 populations and nodes among the 1,000 trees. We used $p = 3$ because the first three principal components were used as the **E** background. Because every individual in a population or every node on a tree splits the records into three groups (i.e., true, false, and unknown), we set c to 3.

Traditional model evaluation approaches (e.g., Elith et al., 2006) involved receiver operating characteristic (ROC) analyses, which have been criticized based on equal weighting of omission and commission errors, consideration of irrelevant predictions, and other issues (Lobo et al., 2008; Peterson et al., 2008); true absences with which to estimate commission error are generally lacking at coarse geographic scales (see Peterson et al., 2011). Hence, we used the alternative pROC metric developed for ENM evaluations (Peterson et al., 2008). This metric assesses the relationship between omission error for independent points and the proportion of area predicted as suitable for the species, but only under conditions of low omission error. AUC ratios (the partial AUC divided by random expectations) range from 0 to 2, with values of 1 representing random performance (Peterson, 2012; Peterson et al., 2008). This evaluation was carried out in pROC software (Barve, 2008) using the continuous output in the evaluation area and evaluation occurrences D_c , with 100 replicate analyses and $\alpha = 0.05$. A detailed explanation of pROC procedures can be found in Appendix S2.A.

2.5.2 | Evaluation of binary models

To compare all six ENMs, including those with binary results only (i.e., KDE), we used omission rate (OR; proportion of evaluation occurrences D_c predicted incorrectly by binary models), proportion of area predicted suitable in the evaluation area, and cumulative binomial probability (CBP; test based on the omission rate and the proportion of area predicted as suitable in the evaluation area; Escobar et al., 2015; Peterson, 2012). Continuous outputs from Maxent, NicheA, GARP, NicheA, GLM, and BRT were converted to binary based on two thresholding values based on omission error tolerances from the calibration occurrences (Peterson et al., 2011). First, we used a threshold based on the minimum predicted value of all calibration occurrences (D_s and D_n)—aka minimum training presence, which represents 0% omission error in the calibration occurrences. Additionally, we evaluated models based on a threshold of 5% omission error; that is, we removed the 5% of calibration occurrences, in D_s and D_n , with the lowest predicted values. OR and CBP allowed us to measure Type I error for every model prediction. To reduce potential Type II error, results of low OR associated with large areas identified as suitable were identified.

2.6 | Model evaluation in environmental space

We designed two new evaluation metrics that are applied in environmental dimensions: E-space indices I and II. E-space index I assesses the amount of environmental interpolation and extrapolation in predictions: Environmental interpolation is prediction of environmental values that are within the range of environmental values of the

occurrences used for model calibration (blue points in Figure 2). In contrast, extrapolation refers to prediction of environmental values beyond the range of values represented among the occurrence data (red points in Figure 2). We used all available occurrences (D_s , D_n , D_c ; black points in Figure 2) to estimate a best-fit (minimal volume) ellipsoid (MVE; black ellipsoid in Figure 2) in a three-dimensional environmental space (Van Aelst & Rousseeuw, 2009; details in Appendix S2.B) composed of the first three principal components. This MVE was then used as the observed niche with which interpolation and extrapolation were evaluated. Analyses in E-space were carried out based on unique environmental combinations; thus, we ignored duplicate points with identical values in environmental space. We estimated frequency of extrapolation as the number of unique environmental values predicted outside of the ellipsoid. Interpolation was the number of unique environmental combinations predicted

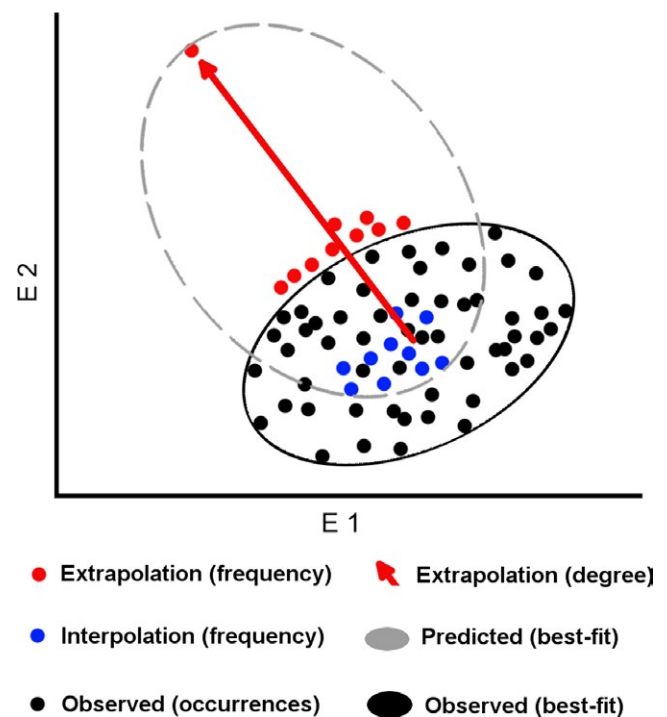


FIGURE 2 New performance metrics for environmental space. Available occurrences were displayed in a two-dimensional environmental space (black points) to estimate a minimum-volume ellipsoid resembling the observed niche (black ellipsoid). *E-space index I*: Model prediction values were categorized as frequency of interpolation (i.e., number of points predicted inside the observed range; blue points) and frequency of extrapolation (i.e., number of points predicted outside the observed range; red points). *E-space index II*: This metric compares model's fit with degree of extrapolation. Model fit was measured as the Jaccard index of similarity between the volume of the modeled niche (gray ellipsoid) versus the volume of the observed niche (gray ellipsoid). This metric also included the degree of extrapolation as the niche distance between the centroid of the observed niche and the most distance prediction of the modeled niche: Note the arrow indicating maximum extrapolation distance between central values of observed data (black points) and most extrapolative values predicted by the model (farthest red point)

inside the ellipsoid (Appendix S2.B). This approach allowed us to discriminate between failed models (i.e., high extrapolation, low interpolation) and successful models (i.e., low extrapolation, high interpolation), according to our criteria of biological realism (Owens et al., 2013; Peterson et al., 2011).

E-space index II is a more complex metric related to the level or intensity of extrapolation and the fit of the model with the data available. For example, two ENMs might have similar frequencies or amounts of extrapolation (e.g., both models with 10 values predicted inside of the observed range of 10–20°C, and 10 predicted outside; see above E-space index I), but the degree of extrapolation between these models could be dramatically different, making models biologically incompatible (e.g., extrapolation to values of 21°C vs. 100°C; Figure 2). Thus, because a metric measuring this *degree* of extrapolation in terms of the observed values (i.e., distance of points from the MVE) would be useful, we measured the degree of extrapolation as the Euclidian distance between the centroid and the most distant prediction away from the MVE (Figure 2; Appendix S2.B). To assess overall fit of the model to the data in environmental space, we also measured similarity between the observed niche (i.e., MVE) and niches predicted by the different models using the Jaccard similarity index (Jaccard, 1912; Figure 2; Appendix S2.B). Finally, once models were evaluated with independent data, final predictions were developed using all the occurrences available for more informed models of the species truly potential. Binary models of suitability were generated using the threshold based on 0% omission error described above, and binary maps were summed to generate an ensemble summarizing areas with agreement among models.

3 | RESULTS

The environmental space defined by the first three principal components explained 94.2% of overall variance; the first two dimensions explained 84.4%. Interestingly, the D_c occurrences overlapped environmentally with northern populations D_n from a 2D environmental perspective (Figure 1b), but environmental overlap between D_c and D_n was nil in the 3D space (Figure 1c). Darwin's Fox populations in the northern and southern areas of the range provided distinct environmental information to the models (i.e., no environmental overlap manifested between D_s and D_n). D_c occurrences filled a portion of the environmental gap between $D_s + D_n$, and occupied the broadest environmental space (green ellipsoid in Figure 1b,c). D_c niche volume was 5.56, compared to D_n and D_s volumes of 2.87 and 0.54, respectively. We found heterogeneous frequencies of occurrences across the species' temperature range: When mean temperature was extracted using all available occurrences, Darwin's Foxes were not found across the entire temperature range available in the study area, but rather clustered in specific temperature intervals (Appendix S3.A).

Our analyses focused on the predictive capabilities of models calibrated based on incomplete data (i.e., using only D_n and D_s ; Figures 3–7). According to AIC values from models with incomplete

data (Table 1), the default parameter Maxent model had the best fit, followed by GLM. However, when all occurrences were employed (D_s , D_n , D_c) to calibrate final models, GLM provided the best fit to the data based on AIC.

Using independent evaluation data (D_c), the models' pROC ratios provided detailed quantitative information on model outputs, allowing us to detect inconsistent models (i.e., high variation; e.g., GLM, BRT; Figure 3). However, pROC showed low discrimination ranking high models with very broad (e.g., GLM) and very narrow predictions (e.g., BRT; Figures 3 and 4). On average, based on pROC, GLM, BRT, and Maxent had good predictive capabilities compared to random models, whereas NicheA and GARP had low performance in predicting independent evaluation data (Figures 3 and 4).

Threshold values based on 0% and 5% omission error changed considerably for GLM, BRT, and GARP (Appendix S3.B). OR evaluations with and omission error threshold of 0% revealed that GLM and Maxent predictions anticipated all independent evaluation points (i.e., zero omission error; Table 2), but by predicting extensive areas as suitable (Figure 4). On the other hand, BRT identified only 3% and KDE 21% of the evaluation area as suitable, but were unable to anticipate any of the evaluation occurrences D_c (Figure 4; Table 2). NicheA and GARP identified narrower areas as suitable (41.2% and 28.3%, respectively), but failed to anticipate four of seven occurrences, predictions that were not better than chance according to the CBP test. Hence, based on the CBP metric with a threshold of 0% omission error, only GLM and Maxent provided better-than-random predictions (Table 2). When models were thresholded based on 5% omission error, GLM and Maxent dramatically reduced the area predicted suitable (Table 2; Appendix S3.C), which could be associated with the positive skew distribution of their predicted values (Appendix S3.B). For example, most Maxent predictions had low values, so that small

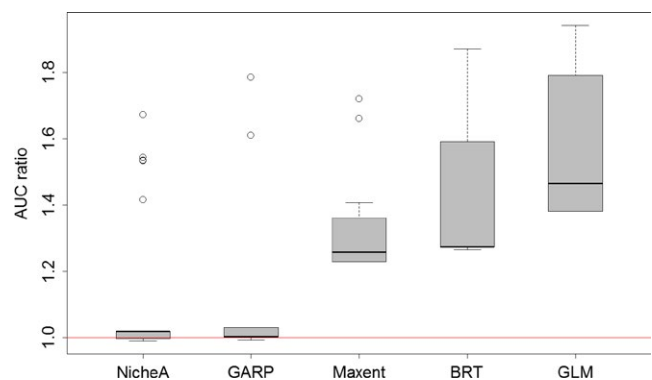


FIGURE 3 Ecological niche model evaluation in central populations (D_c) using partial ROC ratios based on northern and southern ($D_s + D_n$) populations. Boxplots denote AUC ratios in 100 replicates using 50% of evaluation occurrences in each replicate and 5% of omission error. The red line denotes a null distribution of AUC ratios under which predictions are not better than by random expectations. GLM, generalized linear model; BRT, boosted regression trees; Maxent, maximum entropy; GARP, genetic algorithm for rule-set production; KDE, hypervolume multivariable kernel density estimation; NicheA, minimum-volume ellipsoid

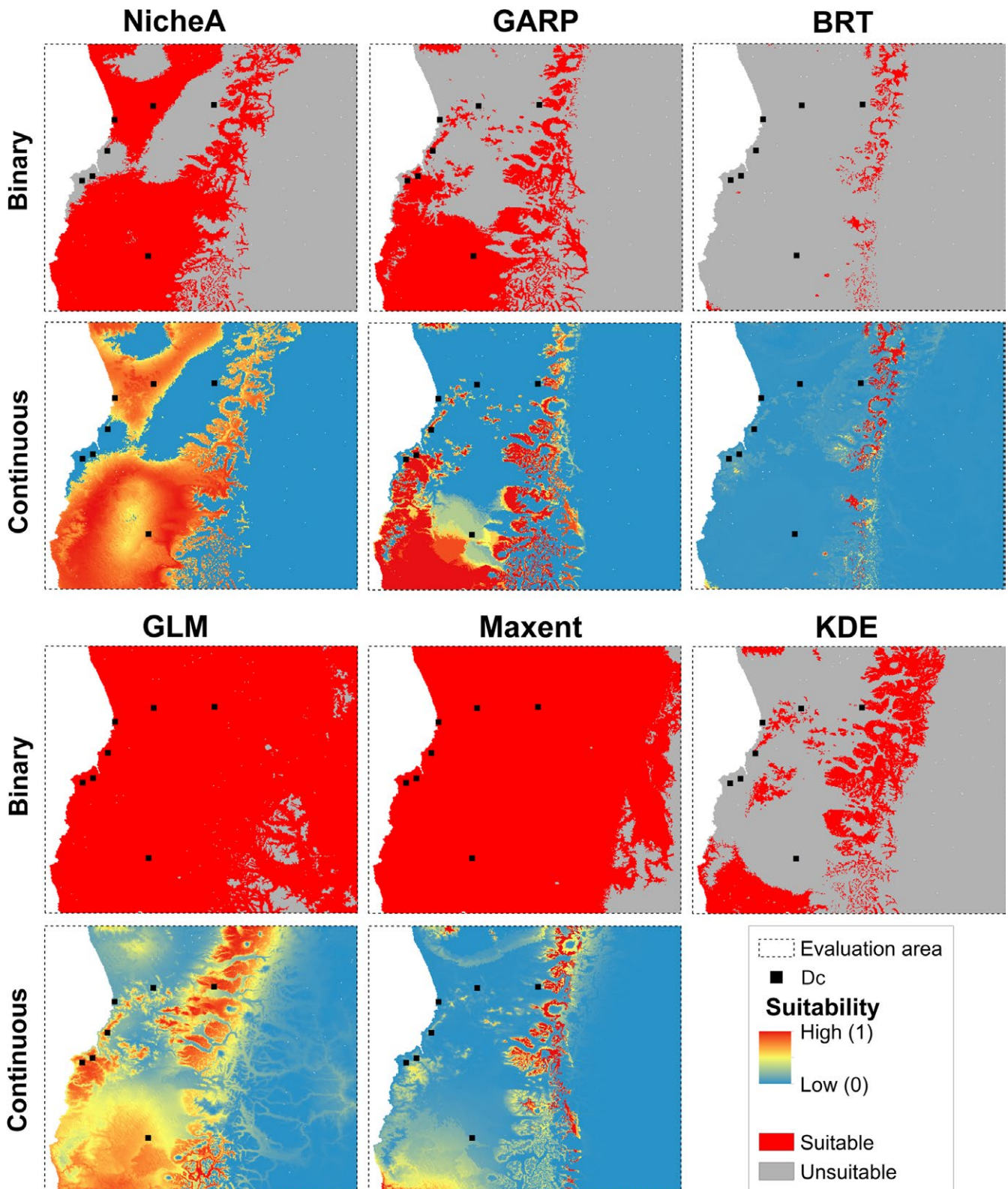


FIGURE 4 Continuous and binary models of Darwin's Fox in the evaluation area (threshold = 0%). Models calibrated using northern and southern ($D_s + D_n$) Darwin's Fox occurrences, projected in the evaluation area (dashed line). Independent occurrences from the central population (D_c ; black squares) are used to evaluate the model in terms of predictions in continuous (i.e., range of colors; highly suitable = red, unsuitable = blue) and binary outputs (i.e., suitable = red, unsuitable = gray). Binary models were generated based on 0% omission error from calibration occurrences. NicheA, minimum-volume ellipsoid; GARP, genetic algorithm for rule-set production; BRT, boosted regression trees; GLM, generalized linear model; Maxent, maximum entropy; KDE, hypervolume multivariable kernel density estimation

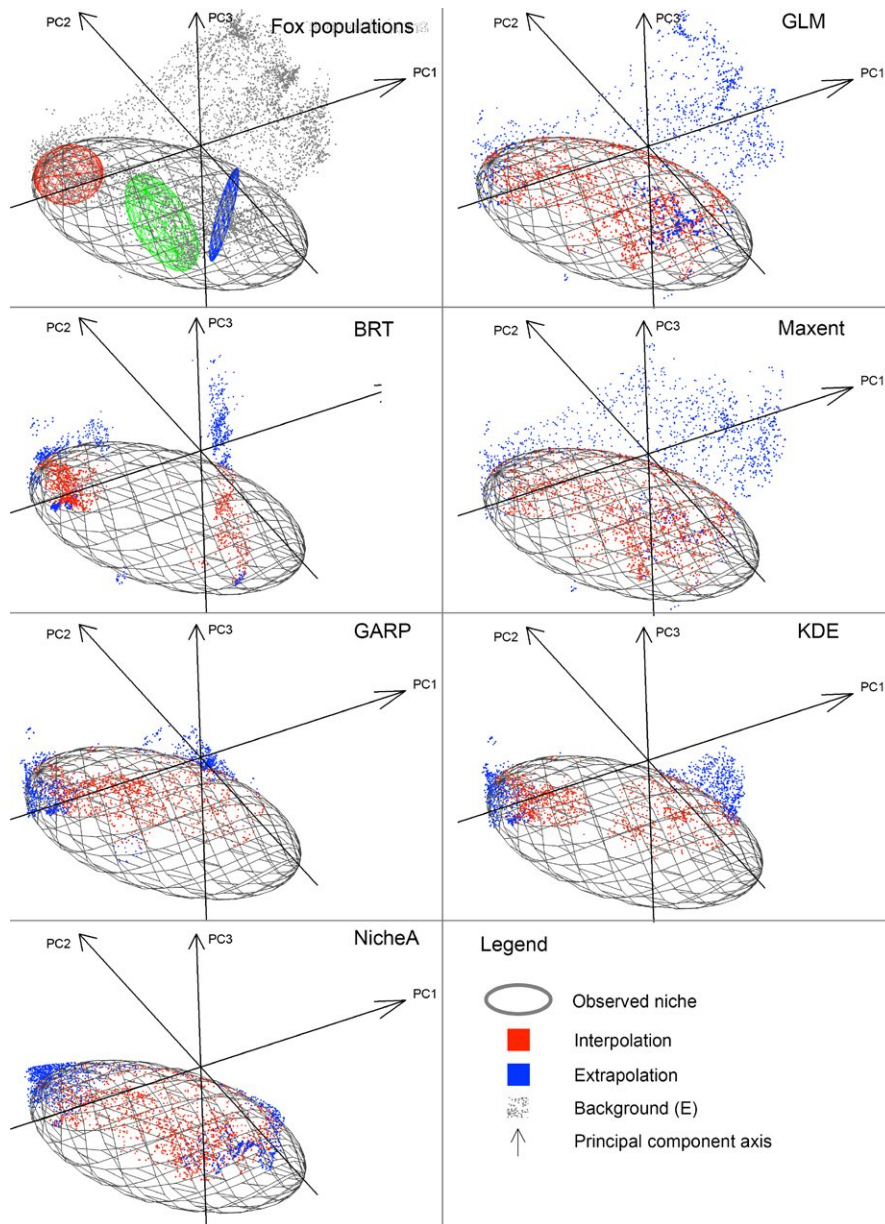


FIGURE 5 Model evaluations based on interpolation and extrapolation in environmental space (threshold = 0%). Top left: Darwin's Fox populations, from the northern (blue ellipsoid), central (green ellipsoid), and southern (red ellipsoid); populations were enclosed to generate observed ecological niche hypotheses; the environmental background is shown in this panel as gray points. Subsequent panels: Predictions were categorized according to environmental interpolation (red points) as predictions inside the ellipsoid and environmental extrapolation (blue points; see Section 2) as predictions outside the ellipsoid; the environmental background is not shown in these panels for better visualization of models output. GLM, generalized linear model; BRT, boosted regression trees; Maxent, maximum entropy; GARP, genetic algorithm for rule-set production; KDE, hypervolume multivariable kernel density estimation; NicheA, minimum-volume ellipsoid. Note that predictions by some models resemble the background cloud (e.g., GLM and Maxent), suggesting that all the conditions available in the model calibration area were predicted suitable by the model via model interpolation (red points) or extrapolation (blue points)

increments in the threshold will assign broad areas as unsuitable. Under this threshold, only GLM provided predicted better than by chance based on CBP metric (Table 2).

Metrics of ENM performance in environmental space showed that different models predicted different levels of interpolation and extrapolation, information not captured by traditional metrics (e.g., AIC; Figure 5 and Appendix S3.D). Based on E-space index I, ENMs showed different frequency of interpolation and extrapolation (Figure 6). Based on a threshold of 0% omission error, models with highest frequency of environmental extrapolation were GLM and Maxent (Figure 6a); BRT had the lowest amount of extrapolation. NicheA showed high amounts of environmental interpolation, similar to Maxent, but much lower environmental extrapolation, similar to BRT. GARP, followed by KDE, had intermediate amounts of interpolation and low extrapolation in predictions, suggesting that these models balanced interpolation and extrapolation better

(Figure 6a). However, even though GARP and KDE were similar in interpolation and extrapolation frequencies, they predicted suitability under different environmental values and extents, as visualized in environmental space (Figure 6a). E-space index I also showed that some ENMs are more affected than others by the threshold value selected. For example, while BRT and GLM remained consistent, Maxent was the most affected by variations in threshold values, with a radical change from very high interpolation and extrapolation to very low interpolation and extrapolation (Figure 6a vs. b).

When E-space index II was considered, we found that under a 0% omission threshold, GLM and Maxent predicted suitability in environmental conditions considerably beyond observed values (Figure 7a). GLM and Maxent models also showed the lowest environmental overlap with observed occurrences, revealing high extrapolation and low fit to observations when environmental dimensions are considered. NicheA provided the best fit between

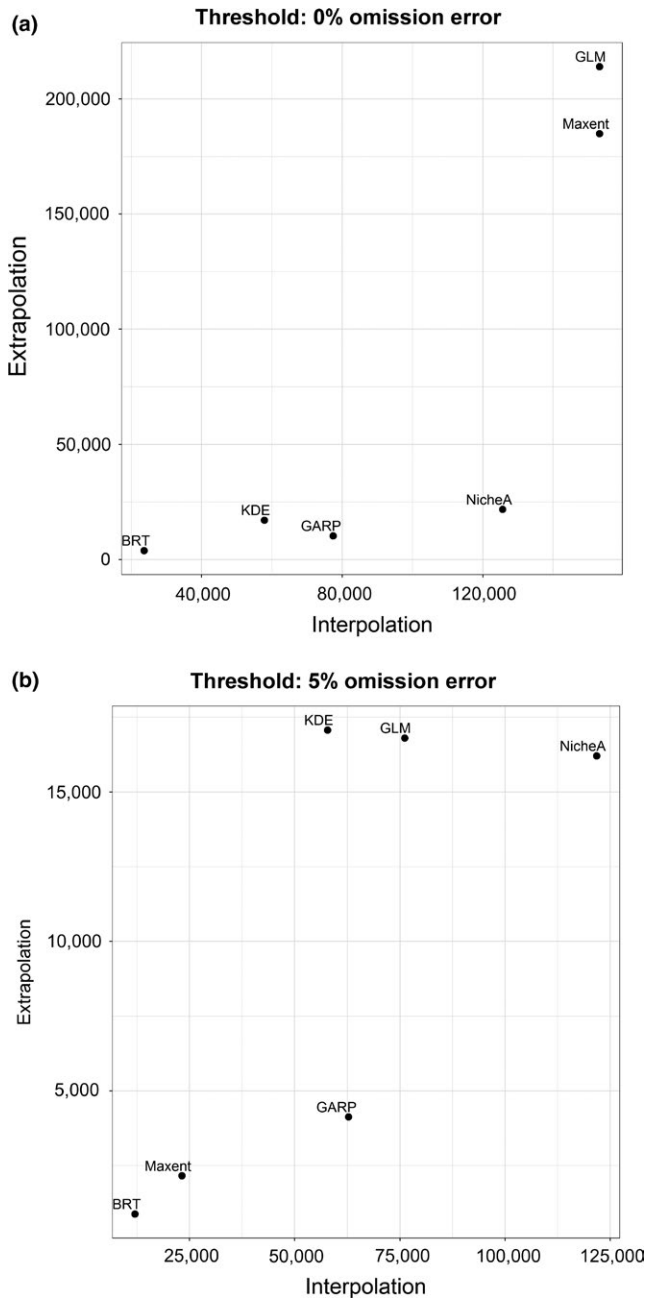


FIGURE 6 Model performance evaluation in environmental dimensions based on the *E-space index I*. Frequency of environmental interpolation and extrapolation. Models were plotted in terms of environmental interpolation (x-axis) and extrapolation (y-axis) to compare their performance under both circumstances. Units are number of environmental points from principal component variables with predicted values inside (i.e., interpolation) or outside the ellipsoid of the observed niche (i.e., model extrapolation). (a) Binary models based on a threshold of 0% omission error in the calibration occurrences. (b) Binary models based on a threshold of 5% omission error in the calibration occurrences. GLM, generalized linear model; BRT, boosted regression trees; Maxent, maximum entropy; GARP, genetic algorithm for rule-set production; KDE, hypervolume multivariable kernel density estimation; NicheA, minimum-volume ellipsoid

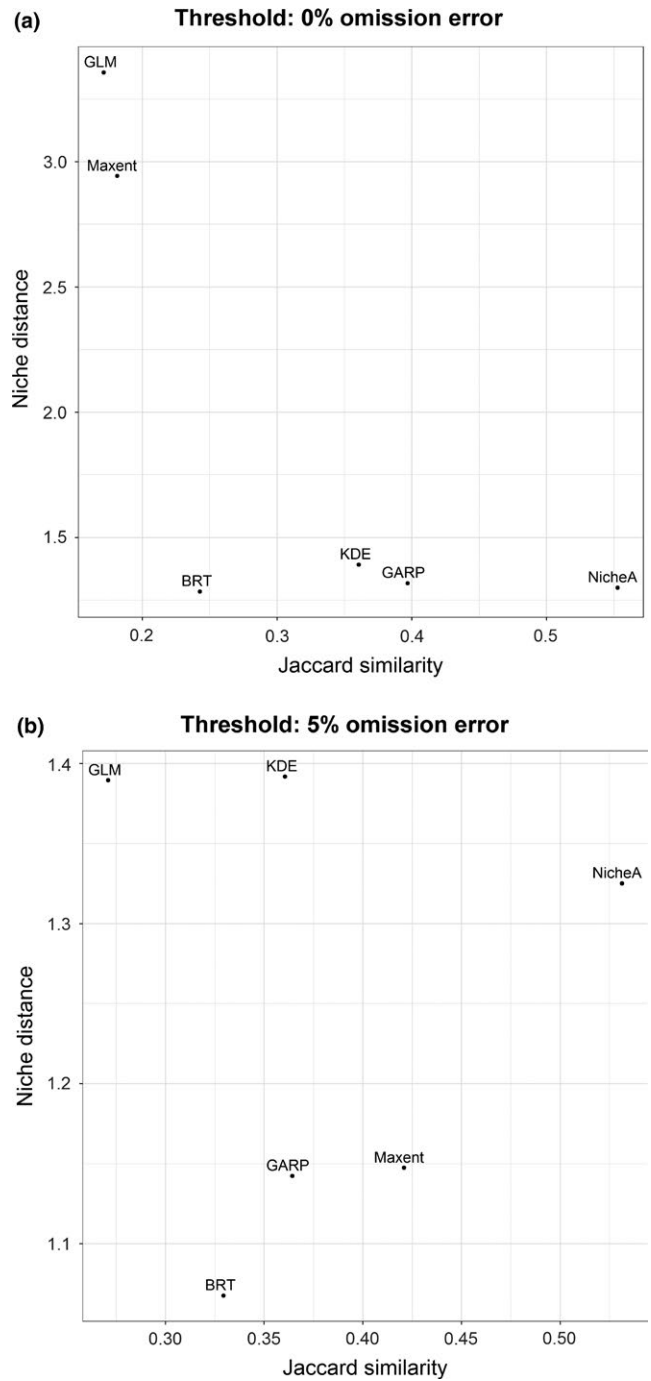


FIGURE 7 Model performance evaluation in environmental dimensions based on the *E-space index II*. Model fit in terms of niche overlap measured by Jaccard similarity and the degree of extrapolation as the niche distance between the occurrences and niche center. Models were plotted in terms of Jaccard similarity (x-axis) and degree of extrapolation in terms niche distance (y-axis; see Figure 2). (a) Binary models based on a threshold of 0% omission error in the calibration occurrences. (b) Binary models based on a threshold of 5% omission error in the calibration occurrences. GLM, generalized linear model; BRT, boosted regression trees; Maxent, maximum entropy; GARP, genetic algorithm for rule-set production; KDE, hypervolume multivariable kernel density estimation; NicheA, minimum-volume ellipsoid

TABLE 1 Akaike information criterion (AIC) values for models calibrated using northern and southern populations (i.e., $D_n + D_s$) and all occurrences available ($D_n + D_c + D_s$)

	K	Ln(likelihood)	AIC
Calibration models $D_n + D_s$			
Maxent	57	-1061.09	2236.17
GLM	6	-1206.42	2424.85
BRT	192	-1031.91	2447.82
GARP	96	-1132.65	2457.30
Final models $D_n + D_c + D_s$			
GLM	6	-1187.17	2386.34
Maxent	55	-1173.17	2456.33
GARP	96	-1155.53	2503.06
BRT	192	-1133.39	2650.78

predictions and observations as measured with Jaccard similarity; NicheA and BRT showed the lowest degree of extrapolation. KDE and GARP showed moderate similarity between predictions and observations and low degree of extrapolation as expressed as niche distance (Figure 7b). Using a 5% omission error threshold reduced considerably the exaggerated degree of extrapolation for GLM and Maxent as expressed by distance of outlier predictions in environmental space (Figure 7b; Appendix S3.D).

In geographic space and using 0% omission error threshold, models with high extrapolation (e.g., GLM, Maxent) were visualized as broad areas of potential distribution for Darwin's Fox. GLM and Maxent were considerably impacted geographically by increments in the omission error threshold to 5%, shrinking dramatically the area predicted suitable (Appendix S3.C). ENMs with limited extrapolation and interpolation (e.g., BRT, KDE, GARP) identified more specific areas as suitable even when calibrated with all available occurrences (Figure 4 and Appendix S3.E). Summing all of the binary models calibrated using all available occurrences for the final ENM ensemble approach showed areas of high and low agreement of models, mostly clustered in areas of known occurrence, and extending across the Andean mountain valleys and onto the plateau (Appendix S3.E).

Model calibrated using	Omission rate		Area predicted suitable		CBP (<i>p</i> value)	
	0%	5%	0%	5%	0%	5%
$D_n + D_s$						
GLM	0.00	0.30	0.96	0.30	<.001	<.001
Maxent	0.00	1.00	0.92	0.03	<.001	>.05
BRT	1.00	1.00	0.03	0.01	>.05	>.05
KDE	1.00	NA	0.21	NA	>.05	NA
NicheA	0.57	0.57	0.41	0.40	>.05	>.05
GARP	0.57	0.71	0.28	0.20	>.05	>.05

CBP, Cumulative binomial probability; GLM, generalized linear model; BRT, boosted regression trees; Maxent, maximum entropy; GARP, genetic algorithm for rule-set production; KDE, hypervolume multivariable kernel density estimation; NicheA, minimum-volume ellipsoid.

4 | DISCUSSION

Many previous studies have attempted to assess predictive performance of ecological niche modeling methods regarding their predictive ability, overfitting, and accuracy (Rangel & Loyola, 2012). Others have argued that a model ensemble approach is a parsimonious way to deal with ENM-based variation, avoiding the decision of choosing one ENM over another (Araújo & New, 2007); however, using a single model is the common practice (Qiao et al., 2015). For instance, the work of Elith et al. (2006) has seen over 5,200 citations, but its conclusions are rarely questioned, even given known weaknesses in the evaluation metrics used (Golicher, Ford, Cayuela, & Newton, 2012; Lobo et al., 2008; Peterson, Papeş, & Eaton, 2007; Peterson et al., 2008). In this study, we have begun to address key issues that have been generally neglected in selecting ENMs for studies (but see Diniz-Filho et al., 2009; Buisson, Thuiller, Casajus, Lek, & Grenouillet, 2010; Terrible et al., 2012; de Oliveira, Araújo, Rangel, Alagador, & Diniz-Filho, 2012; de Oliveira, Rangel, Lima-Ribeiro, Terribile, & Diniz-Filho, 2014; de Oliveira et al., 2015; Collevatti et al., 2012, 2013). Model performance varied dramatically among ENMs and depending on the evaluation metric employed, making multimetric comparisons and careful consideration of the needs of each particular study a critical element in the analytical process and final model selection.

Under ideal conditions, species will occupy a continuous portion of environmental space that reflects their fundamental ecological niches (Soberón & Nakamura, 2009). For most species, however, such conditions rarely exist, and Wallace's Dream scenarios may appear, in which the true dimensions of the niche that are observable are limited due to other factors. Here, the historical distribution of the Darwin's Fox is an ideal example of a Wallace's Dream scenario, with isolated populations occupying different environmental spaces. Novel environmental values representing previously unknown sectors of the species' fundamental ecological niche were illuminated using novel reports of Darwin's Fox populations in the central parts of the species' range (D_c). This new information effectively filled both environmental and geographic gaps between northern and southern populations (Figure 1). We also found that some ENMs may

TABLE 2 True omission error evaluations based on validation data from the novel population D_c ($n = 7$). Binary maps based on an a priori percentage of omission error tolerance of 0% and 5% in the calibration data D_n and D_s

fail to reconstruct species' niches under Wallace's Dream scenarios, which was manifested by models' inability to predict D_c populations.

GLM filled consistently and substantially the environmental gap in the observable parts of the species' ecological niche (Figures 4 and 5), but at the cost of dramatic extrapolation beyond the environmental conditions occupied by the species (Figures 5 and 6). High model extrapolation may be undesirable, as it assumes that the species can survive under conditions outside of the range of conditions under which it has been observed to maintain populations, sometimes far outside known conditions. Such extrapolation can be biologically unrealistic; for example, in some cases, models anticipate suitability at 100°C, which is implausible for most species (Owens et al., 2013).

On the other hand, model overfitting, expressed in our environmental space metrics as low interpolation and extrapolation (e.g., BRT), is also likely to be biologically unrealistic. For example, why would a species not be able to survive under intermediate conditions otherwise contained inside its environmental range (e.g., see Appendix S3.A)? Basic physiology suggests that species will be able to survive under intermediate conditions, as physiological responses tend to be bell-shaped in terms of response of suitability to environmental conditions, rather than bimodal, intolerance to intermediate environmental conditions (Austin, Cunningham, & Fleming, 1984; Birch, 1953; Maguire, 1973; Qiao, Escobar, et al., 2016).

Under this thinking framework, we would seek an ENM method with high interpolation but low extrapolation, at least for the needs of this study. Some methods performed poorly under some evaluation metrics (e.g., AIC), but may fulfill the requirements of this study, such as NicheA, GARP, and KDE. We suggest that the evaluation metric should be selected based on the model feature desired and the use intended (Soberón & Peterson, 2005). For example, some research questions may require prioritizing model overfitting expressed as low interpolation and low extrapolation, whereas others may require predictions that are not overfitted and that are inclusive of broad suitable conditions. For example, an overfit model would be desirable in cases attempting to identify suitable areas for reintroductions of rare species, while broad models may be desirable for searches for last populations of possibly extinct species. Hence, assessing modeling methods in terms of diverse metrics should be a common practice in view of the abilities of different metrics to assess different model features. In this vein, AIC corrected by sample sizes (AICc) could be included in the set of evaluation metrics for ENMs developed from small number of occurrences (Warren & Seifert, 2011). We also noted that our small number of occurrences affected dramatically the thresholding. For example, removing 5% of calibration occurrences with the lowest predicted values, instead of 0%, resulted in Maxent models that were markedly different.

Detailed parameterizations instead of default configurations may impact fit of models to the data in interesting ways (de Oliveira et al., 2017). In particular, selection of Maxent models based on information criteria such as AIC is emerging as a popular new paradigm in ecological niche modeling (Warren & Seifert, 2011). This practice, however, is still under intense exploration

and experimentation (Muscarella et al., 2014), and the biological significance of such "best" models remains understudied. We recall the words of Samuel Karlin, an American mathematician who stated, "The purpose of models is not to fit the data but to sharpen the questions." As regards the present study, our focus was on developing useful evaluation metrics, rather than on detailed parameterization of models, which have been treated elsewhere (Muscarella et al., 2014; Peterson et al., 2011).

Still, as many readers will be curious about the effects of detailed parameterization on the sort of results that we have presented in this contribution, we explored a more detailed parameterization of Maxent (Appendix S4). We assessed 1,220 candidate models and found that the optimal AICc metrics do not coincide with default parameters. That is, detailed model parameterization helped to generate models with better fit with the data and less complexity (57 parameters for the default model vs. 23 for the optimized model). In terms of other metrics, however, default and optimized models did not differ markedly (e.g., omission rate, area predicted suitable, cumulative binomial probability), such that the resulting distribution maps did not differ noticeably (Appendix S4).

Based on diverse model evaluation criteria (amount of extrapolation, amount of interpolation, degree of extrapolation, model fit with the data, pROC, cumulative binomial test, and omission rate), we found that no single ENM achieved the highest scores in all metrics. For the particular application examined in this study, we preferred models presenting low extrapolation and high interpolation, criteria that were chosen a priori. Under this condition, NicheA was a good candidate model in terms of low extrapolation, high interpolation, moderate omission rate, high similarity to the observed niche, and low degree of extrapolation, but at the cost of a non-significant p -value as estimated based on a one-tailed binomial test. For an expanded discussion of the results on the distributional ecology of Darwin's Fox, see Appendix S3.E.

Ecological niche models are usually designed and assessed from a geographic perspective (e.g., Radosavljevic & Anderson, 2014). However, our results suggest that such interpretations hold relatively limited information and should be taken with caution; models should rather be analyzed in both environmental and geographic spaces. What is more, complications arising from spatial autocorrelation and nonindependence of points in geographic spaces further complicate geographic only evaluations. More highly dimensional environmental spaces may be still more informative in such explorations (Figure 1).

We encourage a future reanalysis of the original work of Elith et al. (2006), based on the same data sets and model outputs, but under different and diverse evaluation metrics. Such a reanalysis would determine whether that study's conclusions are consistent under the same assumptions and parameters, but in the context of different evaluation metrics. Such a re-evaluation exercise would increase the transparency and good practices behind one of the foundational studies in ecological niche modeling. We note that ENMs in this study used default parameters to allow replicability and fair

comparisons, but more detailed parameterizations may generate different outputs. To facilitate the replication of our study, we have incorporated the scripts of the metrics as Appendix S2.B. However, development of a formal software package including ENM evaluations in environmental space is warranted.

5 | CONCLUSIONS

AUC and AIC have dominated protocols for ENM model evaluation for at least a decade; however, such metrics are limited in information and may fail to evaluate some properties of the desired model (Qiao et al., 2015). Researchers should establish clear and delimited a priori assumptions and desired model characteristics (Peterson, 2006); based on these decisions, researchers can select ENM methods and evaluation metrics that address their requirements. We found that model evaluations in environmental dimensions were highly informative to guide model selection and interpretation. Our proposed E-space metrics of extrapolation and interpolation in the environmental space offer a useful enrichment to more customary characterization of model predictions. Future research on these metrics should include development of standardized indices to make studies comparable.

ACKNOWLEDGMENTS

Huijie Qiao was supported by the National Key R&D Program of China (No. 2017YFC1200603) and National Natural Science Foundation of China (No. 31772432).

CONFLICT OF INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

L.E.E. conceived the study. H.Q. and L.E.E. performed the analyses and wrote the manuscript. H.Q. developed the evaluation metrics. J.C. collected data and wrote the manuscript. A.T.P. co-wrote the manuscript. All authors reviewed the manuscript.

ORCID

Luis E. Escobar  <http://orcid.org/0000-0001-5735-2750>

REFERENCES

- Araújo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, 22, 42–47. <https://doi.org/10.1016/j.tree.2006.09.010>
- Austin, M. P., Cunningham, R. B., & Fleming, P. M. (1984). New approaches to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Plant Ecology*, 55, 11–27. <https://doi.org/10.1007/BF00039976>
- Baillie, J., Hilton-Taylor, C., & Stuart, S. N. (2004). *2004 IUCN Red List of Threatened Species: A global species assessment*. Cambridge, UK: World Conservation Union.
- Barve, N. (2008). *Tool for partial-ROC*. Version 1.0. Biodiversity Institute, Lawrence. Retrieved from <http://kuscholarworks.ku.edu/dspace/handle/1808/10059> [accessed 9 September 2016]
- Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., ... Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, 222, 1810–1819. <https://doi.org/10.1016/j.ecolmodel.2011.02.011>
- Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., ... Hay, S. I. (2013). The global distribution and burden of dengue. *Nature*, 496, 504–507. <https://doi.org/10.1038/nature12060>
- Birch, L. C. (1953). Experimental background to the study of the distribution and abundance of insects: III. The relation between innate capacity for increase and survival of different species of beetles living together on the same food. *Evolution*, 7, 136–144. <https://doi.org/10.1111/j.1558-5646.1953.tb00072.x>
- Blonder, B. (2015). *Hypervolume: High-dimensional Kernel density estimation and geometry operations*. R package version 1.2.2. Retrieved from <http://CRAN.R-project.org/package=hypervolume>.
- Blonder, B., Lamanna, C., Violle, C., & Enquist, B. J. (2014). The n-dimensional hypervolume. *Global Ecology and Biogeography*, 23, 595–609. <https://doi.org/10.1111/geb.12146>
- Buisson, L., Thuiller, W., Casajus, N., Lek, S., & Grenouillet, G. (2010). Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16, 1145–1157. <https://doi.org/10.1111/j.1365-2486.2009.02000.x>
- Burnham, K. P., Anderson, D. R., & Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65, 23–35. <https://doi.org/10.1007/s00265-010-1029-6>
- Collevatti, R. G., Terribile, L. C., de Oliveira, G., Lima-Ribeiro, M. S., Nabout, J. C., Rangel, T. F., & Diniz-Filho, J. A. F. (2013). Drawbacks in palaeodistribution modelling: The case of South American seasonally dry forests. *Journal of Biogeography*, 40, 345–358. <https://doi.org/10.1111/jbi.12005>
- Collevatti, R. G., Terribile, L. C., Lima-Ribeiro, M. S., Nabout, J. C., de Oliveira, G., Rangel, T. F., ... Diniz-Filho, J. A. F. (2012). A coupled phylogeographical and species distribution modelling approach recovers the demographical history of a Neotropical seasonally dry forest tree species. *Molecular Ecology*, 21, 5843–5863.
- de Oliveira, G., Araújo, M. B., Rangel, T. F., Alagador, D., & Diniz-Filho, J. A. F. (2012). Conserving the Brazilian semi-arid (Caatinga) biome under climate change. *Biodiversity Conservation*, 21, 2913–2916. <https://doi.org/10.1007/s10531-012-0346-7>
- de Oliveira, G., Lima-Ribeiro, M. S., Terribile, L. C., Dobrovolski, R., Telles, M. P. C., & Diniz-Filho, J. A. F. (2015). Conservation biogeography of the Cerrado's wild edible plants under climate change: Linking biotic stability with agricultural expansion. *The American Journal of Botany*, 102, 870–877. <https://doi.org/10.3732/ajb.1400352>
- de Oliveira, G., Rangel, T. F., Lima-Ribeiro, M. S., Terribile, L. C., & Diniz-Filho, J. A. F. (2014). Evaluating, partitioning, and mapping the spatial autocorrelation component in ecological niche modeling: A new approach based on environmentally equidistant records. *Ecography*, 37, 637–647. <https://doi.org/10.1111/j.1600-0587.2013.00564.x>
- de Oliveira, S. V., Romero-Alvarez, D., Martins, T. F., dos Santos, J. P., Labruna, M. B., Gazeta, G. S., ... Gurgel-Gonçalves, R. (2017). Amblyomma ticks and future climate: Range contraction due to climate warming. *Acta Tropica*, 176, 340–348. <https://doi.org/10.1016/j.actatropica.2017.07.033>
- de Souza Muñoz, M. E., de Giovanni, R., de Siqueira, M. F., Sutton, T., Brewer, P., Scachetti-Pereira, R., ... Perez Canhos, V. (2011).

- OpenModeller: A generic approach to species' potential distribution modelling. *Geoinformatica*, 15, 111–135. <https://doi.org/10.1007/s10707-009-0090-7>
- Diniz-Filho, J. A. F., Bini, L. M., Rangel, T. F., Loyola, R. D., Hof, C., Nogueira-Bravo, D., & Araújo, M. B. (2009). Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate change. *Ecography*, 32, 897–906. <https://doi.org/10.1111/j.1600-0587.2009.06196.x>
- Elith, J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., ... Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129–151. <https://doi.org/10.1111/j.2006.0906-7590.04596.x>
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Escobar, L. E. (2013). Conservation from heaven: Remote sensing and open access tools to guide biodiversity conservation. In L. E. Escobar (Ed.), *New hope for conservation* (pp. 13–27). Beijing, China: Peking University.
- Escobar, L. E., Ryan, S. J., Stewart-Ibarra, A. M., Finkelstein, J. L., King, C. A., Qiao, H., & Polhemus, M. E. (2015). A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Tropica*, 149, 202–211. <https://doi.org/10.1016/j.actatropica.2015.05.028>
- Franklin, J., & Miller, J. A. (2009). *Mapping species distributions: Spatial inference and prediction*. Cambridge, UK: Cambridge University Press.
- Golicher, D., Ford, A., Cayuela, L., & Newton, A. (2012). Pseudo-absences, pseudo-models and pseudo-niches: Pitfalls of model selection based on the area under the curve. *International Journal of Geographical Information Science*, 8816, 2049–2063. <https://doi.org/10.1080/13658816.2012.719626>
- Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., ... Wintle, B. A. (2015). Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography*, 24, 276–292. <https://doi.org/10.1111/geb.12268>
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965–1978. [https://doi.org/10.1002/\(ISSN\)1097-0088](https://doi.org/10.1002/(ISSN)1097-0088)
- Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2012). *dismo: Species distribution modeling*. R package version 0.7-17. Retrieved from <https://cran.r-project.org/web/packages/dismo/index.html>
- Hurlbert, S. H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, 54, 187–211. <https://doi.org/10.2307/1942661>
- Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytologist*, 11, 37–50. <https://doi.org/10.1111/j.1469-8137.1912.tb05611.x>
- Jaksic, F. M., Jiménez, J. E., Medel, R. G., & Marquet, P. A. (1990). Habitat and diet of Darwin's fox (*Pseudalopex fulvipes*) on the Chilean mainland. *Journal of Mammalogy*, 71, 246–248. <https://doi.org/10.2307/1382176>
- Jiménez, J. E., Lucherini, M., & Novaro, A. J. (2008). *Pseudalopex fulvipes* (Darwin's Fox). IUCN 2013 IUCN Red List of Threatened Species Version. IUCN.
- Jiménez, J. E., & McMahon, E. (2004). *Darwin's fox Pseudalopex fulvipes* (Martin, 1837) critically endangered – CR: C2a(ii). Cambridge, UK: IUCN/SSC Canid Specialist Group.
- Johnson, J. B., & Omland, K. S. (2004). Model selection in ecology and evolution. *Trends in Ecology & Evolution*, 19, 101–108. <https://doi.org/10.1016/j.tree.2003.10.013>
- Joppa, L. N., McInerney, G., Harper, R., Salido, L., Takeda, K., O'Hara, K., ... Emmott, S. (2013). Troubling trends in scientific software use. *Science*, 340, 814–815. <https://doi.org/10.1126/science.1231535>
- Kadmon, R., Farber, O., & Danin, A. (2004). Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. *Ecological Applications*, 14, 401–413. <https://doi.org/10.1890/02-5364>
- Kosakovsky, P. S. L., Mannino, F. V., Gravenor, M. B., Muse, S. V., & Frost, S. D. (2007). Evolutionary model selection with a genetic algorithm: A case study using stem RNA. *Molecular Biology and Evolution*, 24, 159–170.
- Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17, 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. Boca Raton, FL: Chapman & Hall/CRC. <https://doi.org/10.1007/978-1-4899-3242-6>
- Maguire, B. Jr (1973). Niche response structure and the analytical potentials of its relationship to the habitat. *The American Naturalist*, 107, 213–246. <https://doi.org/10.1086/282827>
- Medel, R. G., Jiménez, J. E., Jaksic, F. M., Yáñez, J., & Armesto, J. J. (1990). Discovery of a continental population of the rare Darwin's fox, *Dusicyon fulvipes* (Martin, 1837) in Chile. *Biological Conservation*, 51, 71–77. [https://doi.org/10.1016/0006-3207\(90\)90033-L](https://doi.org/10.1016/0006-3207(90)90033-L)
- Molina-Espinosa, M. (2011). *Dieta y Uso de Hábitat del Huillín (Lontra provocax) en Ambientes de Agua Dulce y su Relación con Comunidades Locales en el Bosque Templado Lluvioso, Isla Grande de Chiloé, Chile*. DVM thesis, Universidad Mayor.
- Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5, 1198–1205. <https://doi.org/10.1111/2041-210X.12261>
- Owens, H. L., Campbell, L. P., Dornak, L. L., Saupe, E. E., Barve, N., Soberón, J., ... Peterson, A. T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, 263, 10–18. <https://doi.org/10.1016/j.ecolmodel.2013.04.011>
- Peterson, A. T. (2006). Uses and requirements of ecological niche models and related distributional models. *Biodiversity Informatics*, 3, 59–72.
- Peterson, A. T. (2012). Niche modeling—Model evaluation. *Biodiversity Informatics*, 8, 41.
- Peterson, A. T., Papeş, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: A comparison of GARP and Maxent. *Ecography*, 30, 550–560. <https://doi.org/10.1111/j.0906-7590.2007.05102.x>
- Peterson, A. T., Papeş, M., & Soberón, J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, 213, 63–72. <https://doi.org/10.1016/j.ecolmodel.2007.11.008>
- Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martínez-Meyer, E., Nakamura, M., & Araújo, M. B. (2011). *Ecological niches and geographic distributions*. Princeton, NJ: Princeton University.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Qiao, H., Escobar, L. E., Saupe, E. E., Ji, L., & Soberón, J. (2016). A cautionary note on the use of hypervolume kernel density estimators in ecological niche modelling. *Global Ecology and Biogeography*, 26, 1066–1070.
- Qiao, H., Peterson, A. T., Soberón, J., Campbell, L. P., Ji, L., & Escobar, L. E. (2016). NicheA: Creating virtual species and ecological niches in multivariate environmental scenarios. *Ecography*, 39, 805–813. <https://doi.org/10.1111/ecog.01961>

- Qiao, H., Soberón, J., & Peterson, T. A. (2015). No silver bullets in correlative ecological niche modeling: Insights from testing among many potential algorithms for niche estimation. *Methods in Ecology and Evolution*, 6, 1126–1136. <https://doi.org/10.1111/2041-210X.12397>
- R Development Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org> [accessed 13 September 2016]
- Radosavljevic, A., & Anderson, R. P. (2014). Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography*, 41, 629–643. <https://doi.org/10.1111/jbi.12227>
- Rangel, T. F., & Loyola, R. D. (2012). Labeling ecological niche models. *Natureza & Conservação*, 10, 119–126. <https://doi.org/10.4322/natcon.2012.030>
- Sain, R., & Carmack, P. S. (2002). Boosting multi-objective regression trees. *Computing Science and Statistics*, 34, 232–241.
- Saupe, E., Barve, V., Myers, C., Soberón, J., Barve, N., Hensz, C., ... Lira-Noriega, A. (2012). Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling*, 237, 11–22. <https://doi.org/10.1016/j.ecolmodel.2012.04.001>
- Sillero-Zubiri, C. (2004). *Canids: Foxes, wolves, jackals, and dogs: Status survey and conservation action plan*. Cambridge, UK: Osprey Publishing.
- Soberón, J., & Nakamura, M. (2009). Niches and distributional areas: Concepts, methods, and assumptions. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 19644–19650. <https://doi.org/10.1073/pnas.0901637106>
- Soberón, J., & Peterson, A. T. (2005). Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics*, 2, 1–10.
- Stockwell, D. (1999). The GARP modelling system: Problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, 13, 143–158. <https://doi.org/10.1080/136588199241391>
- Terrible, L. C., Lima-Ribeiro, M. S., Araújo, M. B., Bizao, N., Collevatti, R. G., Dobrovolski, R., ... Diniz-Filho, J. A. F. (2012). Areas of climate stability of species ranges in the Brazilian Cerrado: Disentangling uncertainties through time. *Natureza & Conservação*, 10, 152–159. <https://doi.org/10.4322/natcon.2012.025>
- Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD – A platform for ensemble forecasting of species distributions. *Ecography*, 32, 369–373. <https://doi.org/10.1111/j.1600-0587.2008.05742.x>
- Van Aelst, S., & Rousseeuw, P. (2009). Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1, 71–82. <https://doi.org/10.1002/wics.19>
- Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: The importance of model complexity and the performance of model selection criteria. *Ecological Applications*, 21, 335–342. <https://doi.org/10.1890/10-1171.1>
- Yahnke, C. J. (1995). Metachromism and the insight of Wilfred Osgood: Evidence of common ancestry for Darwin's fox and the Sechura fox. *Revista Chilena de Historia Natural*, 68, 459–467.
- Yoshimoto, F., Moriyama, M., & Harada, T. (1999). Automatic knot placement by a genetic algorithm for data fitting with a spline. *Shape Modeling and Applications, 1999. Proceedings of the International Conference on Shape Modeling and Applications*. IEEE Computer Society. Washington, DC.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Escobar LE, Qiao H, Cabello J, Peterson AT. Ecological niche modeling re-examined: A case study with the Darwin's fox. *Ecol Evol*. 2018;8:4757–4770. <https://doi.org/10.1002/ece3.4014>