Title: A test of species' mobility hypothesis in ecological niche modeling

Running title: Mobility hypothesis in ENM

Author: Xiao Feng[1*]

[1]Department of Geography, Florida State University, Tallahassee, Florida, USA

[*]Correspondence to fengxiao.sci@gmail.com

Abstract

Aim: Ecological niche model (ENM) or species distribution model is a modeling technique broadly used in ecology and biogeography and is increasingly used in decision-making regarding land use and biodiversity conservation. The methodology behind ENM applications is critical for model accuracy. One critical question that every ENM study faces is how to define a model training area. Theories have suggested designing a training domain based on species' dispersal ability for improved model performance (species' mobility hypothesis). While this idea has been well perceived, there still lacks direct quantitative evidence that whether this approach leads to optimal model performance. Here I conducted a modeling experiment to investigate species' mobility hypothesis.

Location: North and South America

Time period: 1950-present

Major taxa studied: hummingbird

Methods: The modeling experiment was based on a group of hummingbird species. A series of spatial buffers (from 5 to 5000km) were created around occurrences, where background data were sampled and used as input for model calibration. The models calibrated with spatial buffers were compared with models calibrated with training domains that considered species' dispersal abilities (bioM).

Results: The experiment showed that model performance increased when the size of the training domain was larger, though the model performance reached saturation when size of the training domain passed a certain threshold. The model performance based on bioM was comparable to the saturation performance of models when spatial buffers were used.

Main conclusions: This study provided positive evidence that supports the species' mobility hypothesis that designing a training domain based on species' dispersal ability could lead to optimal or near-optimal model performance. When no information of dispersal is available, modelers may use a tuning strategy to identify the size of the training domain for optimized model performance.

Keywords: BAM, GIS, SDM, species' mobility hypothesis, Maxent

Introduction

Ecological niche model or species distribution model is a modeling technique broadly used in the field of ecology and biogeography. Ecological niche model uses the associations between species' point observations and environmental conditions of those locations to estimate species' ecological niche and potential distributions. Typical applications of ecological niche model include: species' range shifts under climate change (Blowes *et al.*, 2019), range reductions due to habitat loss and anthropogenic disturbance (Doughty et al. 2016), biological invasions (Park and Potter 2015), and the conservation of rare or endangered species (Hannah *et al.*, 2020).

While the ecological niche model has had rapid developments over the past decade and is increasingly used in decision-making regarding land use and biodiversity conservation (Araújo *et al.*, 2019), the theory and critical assumptions behind this technique also warrant our attention. The ecological niche theory forms the basis of the ecological niche model, and the more commonly adopted concept of ecological niche is the one by Hutchinson -- a set of environmental conditions permitting the species to live or to exist indefinitely (Hutchinson, 1957). More recently, Soberon and Peterson (2005) conceptualized a BAM framework that greatly influenced the development and use of ecological niche models. The BAM framework classifies the various factors that determine species' distributions as biotic interactions (B), abiotic conditions (A), and dispersal ability (or mobility; M), and a species is expected to be present in accessible areas with suitable sets of abiotic conditions and biotic interactions.

It is critical to use theories to guide the practices of ecological niche modeling to achieve the fullest potential (Peterson & Soberón, 2012). One important assumption in ecological niche modeling is the equilibrium status between environmental conditions and species' occurrences (Araújo & Pearson, 2005), in other words, a species is present in all suitable locations and absent from unsuitable locations (Soberon & Peterson, 2005). Considering species' dispersal ability (or mobility/M in BAM) has been broadly acknowledged and used in designing the training domain (or modeling domain) to better fit the equilibrium assumption. When true absence data are unavailable (Mackenzie, 2005), selecting pseudo-absences or background data from the training domain has been broadly adopted in applications of ecological niche modeling using a variety of algorithms, such as Maxent (Phillips et al. 2009) and generalized linear model (Wintle et al. 2005). In particular, when species' dispersal ability is considered in designing a training domain, the selected background data are expected to represent the environmental conditions that a species has explored or been exposed to but did not preferably select compared to the environmental conditions of occurrences.

The idea of designing a training domain based on species' dispersal ability has been well perceived and used in the current literature, but there still lacks direct quantitative evidence that whether this approach leads to optimal model performance. For example, Anderson and Raza (2010) compared models using a large training domain and a smaller training domain directly surrounding the localities of the focal species. They concluded that the smaller training domain led to more realistic predictions of species' potential distributions, though this was mainly based on expert opinions of the species' natural history and biogeographic knowledge of the region.

Barve et al. (2011) conducted simulations of virtual species and training domains of varied sizes, and demonstrated the positive association between size of training domain and (the potentially inflated) the area under the receiver operating characteristic curve (ROC AUC). They provided insightful discussion of the role of considering species' dispersal ability in model training, validation, comparison, though there was no direct demonstration of those effects. Saupe et al. (2012) used virtual species simulations to compare models calibrated with training domains based on species' dispersal ability and models calibrated with much broader regions, but the conclusion that the former provided more reliable model performance than the latter was not reached. Their interpretation was that the information of species' dispersal ability did not directly inform the model fitting process, but only indirectly influenced the model fitting through the selection of pseudo-absences or background data. Owens et al. (2013) proposed a potential modification of how to design a training domain based on species' dispersal ability; that is to further refine the training domain based on sampling effort, and one simple example is to exclude areas from the training domain that have not been sampled thus can not produce occurrence records. Cooper and Soberon (2018) implemented a series of ecological niche models for hummingbirds based on training domains that reflect topography, ecoregions and known occurrences, refined by significant abiotic barriers of dispersal (e.g. rivers, deep valleys, crests of mountains). They concluded that restricting a training domain to a theoretically accessible area can improve model performance; however, the model assessment was conducted at a stacked level (i.e. community composition) instead of individual species level, whereas the latter was more commonly used in assessing the performance of ecological niche models. To summarize, there is no doubt of the importance of species' dispersal ability in designing a training domain and its implications in model training and validation, but, to my understanding, there is no direct quantitative evidence supporting the optimal model performance by considering species' dispersal ability.

The design of the training domain can also be thought from the perspective of modeling algorithms, which may lead to different expectations of the optimal training domain. As discussed in Saupe et al. (2012), the information of species' dispersal is used "indirectly" by modeling algorithms, and a training domain is used to generate pseudo-absences or background data, which are used differently as input by different algorithms. Take the popular algorithm of Maxent as an example, background data can be randomly sampled from the training domain. Maxent uses the environmental conditions of the background data to characterize the environmental profile of the training domain. The model prediction of the (relative) probability of presences depends on the contrast between the environmental profile of the background data and occurrences (Merow *et al.*, 2013). In this manner, the hypothesized optimal training domain would be the one that helps the modeling algorithm best distinguish occurrences vs. background data. If so, the question of whether the training domain based on species' dispersal ability is optimal becomes whether this training domain helps the algorithm distinguish occurrences vs. background data.

Following previous explorations, here I continue to investigate the role of species' dispersal ability in defining a training domain, size of a training domain, and model performance. I proposed three hypotheses regarding the relationship between training domain and model performance (Fig. 1). The unimodal hypothesis predicts that model performance shows a unimodal relationship with size of the training domain, whereas a medium sized training domain would lead to an optimal model performance. The linear hypothesis predicts that the model performance increases with increased size of training domain, whereas a larger domain provides more useful information thus better model performance. The saturation hypothesis predicts increased model performance with increased size of training domain, though there is a bottleneck for the amount of useful information obtained from the training domain, thus model performance will reach saturation when the training domain reaches a certain size. The species' mobility hypothesis predicts that the training domain designed from species' mobility would lead to optimal model performance. The species' mobility hypothesis could be compatible with the unimodal and saturation hypotheses whereas the model performance based on bioM could be either at the peak of the unimodal curve or the plateau of the saturation curve. Here I designed a modeling experiment to test these hypotheses. The experiment used a series of training domains with different size and training domains delineated based on species' mobility (Cooper & Soberón, 2018). Specifically, I aim to investigate two questions: 1) How would different sized training domains affect model performance? 2) How are the models calibrated with training domains that considered species' dispersal abilities (bioM) compared with models calibrated with different sized training domains?

Methods

Occurrence data

The occurrences of hummingbird species were obtained from Cooper & Soberón (2018). Briefly, this dataset included occurrences of all hummingbird species (Aves: Trochilidae) that were available from GBIF and eBird in December 2013. This dataset has been filtered and curated by (Cooper & Soberón, 2018) (2018) based on multiple criteria, including taxonomy, ebird observation effort, spatial clustering, and spatial outlier. Here I only kept species with adequate (≥100) spatially unique occurrences as a way to guarantee reliable model performance (Proosdij *et al.*, 2016). Eight species that are mainly distributed in islands were excluded from this study, as their restricted range conflict with the study design when large spatial buffers were used (see next section). The final dataset included 87 species.

Training domain

Two approaches were used to design the training domains, where background points were randomly sampled to reflect the environmental profile of the study area (Fig. 2). In the first approach, the training domain was designed to reflect each species' accessible aera (or mobility in BAM framework; here termed bioM). The bioM for each species was manually delineated by (Cooper & Soberón, 2018) (2018) to reflect topography, ecoregions and known occurrences, refined by significant abiotic barriers of dispersal (e.g. rivers, deep valleys, crests of mountains). The second approach created a series of 26 spatial buffers of each species' occurrences to approximate accessible areas with varied dispersal capacities (here termed spatial buffers). The sizes of the spatial buffers were 5km, 7km, 9km, from 10km to 100km with increment of 10km, from 100km to 1000km with increment of 100km, and from 1000 to 5000 with increment of 1000km. The bioM used here represents the hypothesized accessible area, and is expected to lead to optimal model performance based on the species' mobility hypothesis (Barve *et al.*, 2011; Saupe *et al.*, 2012; Owens *et al.*, 2013). In contrast, the spatial buffers cover the locations "around" known occurrences and assume uniform dispersal capacity along all directions.

Model training

Four bioclimatic variables that represent the climatic extremes that often constrain species distributions were used to train models; the variables were the same as those used in Cooper & Soberón (2018). The four variables were: bio10 (mean temperature of the warmest quarter), bio11 (mean temperature of the coldest quarter), bio16 (precipitation of the wettest quarter), and bio17 (precipitation of the driest quarter). The climatic data at the resolution of 2.5 arc-minutes were downloaded from Worldclim version 1.4 (Hijmans *et al.*, 2005). From each training domain, ten thousand pixels were randomly selected as background points used for model training. All pixels would be used if the training domain has less than ten thousand pixels.

The models were trained using Maxent, a broadly used modeling algorithm in the field of ENM/SDM, using maxnet package version 0.1.4. Linear, quadratic, and hinge features were selected to allow adequate model complexity. Default regularization parameters were used. Fine tuning of model performance was not implemented, as the training domain was the objective of investigation. Clamping was turned off, as its effect could depend on the point where clamping begins (Qiao *et al.*, 2018).

The occurrences of each species were separated into four sets to perform cross-validations with three sets for training and one set for testing. The separation was performed with random separation and environmental blocking, to mimic the scenario of with or without model extrapolation. The four sets of occurrences generated by random separation generally had similar spatial extent and environmental conditions, thus no or very little extrapolation would be involved in model prediction. The environmental blocking grouped occurrences into clusters based on their Euclidean distance in the environmental space (Valavi *et al.*, 2019), thus strict or combinational extrapolation could be involved during model prediction (Qiao *et al.*, 2018). Environmental blocking was performed using the blockCV package (version 2.1.4) (Valavi *et al.*, 2019). Two methods were used to generate background data for model evaluation: random background data from the largest buffer (5000km), reflecting a scenario when a calibrated model is projected to a large spatial extent (e.g. a continent), and random background data from bioM, reflecting a scenario when the model is projected to an area that species is known to occur. Model performance was measured by sensitivity, specificity, and true skill statistics (TSS) at the threshold of 5% training omission rate.

Regression analyses

Stepwise regression was performed to analyze the relationship between buffer size (independent variable) and model performance (dependent variable) using segmented package 1.6-0 (Muggeo, 2003). Preliminary analysis of the relationship showed an increase of model performance with buffer size, though the pattern was commonly accompanied with performance saturation when the buffer size passed a certain threshold. Logistic regression was not used here, because the logistic curve is monotonic (continuously increasing or decreasing) that would not allow the fitting of a unimodal trend. Instead, two segments stepwise regression was used, as it allows the fitted curve to have two segments with different slopes. The buffer size was log10 transformed so that the independent variable could be more evenly sampled. The model performance based on bioM was compared with model performance based on a spatial buffer that was similar to bioM in size (area) using t test; model performance based on bioM was also compared with model performance based on a spatial buffer that had the highest evaluation score among all buffers using t test. These analyses were implemented respectively for each species and the mean of all species, different data separation methods, and different model performance indices. The analyses were performed in R (version 4.1.2).

Results

The model evaluations were based on background data from the largest buffer or bioM, and the two methods showed slightly different but similar patterns. When background points generated from the largest buffer were used for model evaluation, the buffer size used to define the modeling domain was positively associated with model performance (sensitivity, specificity, and TSS). The model performance increased more rapidly during the initial increase of buffer size (e.g from 5km to 50km), after which the model performance gradually reached saturation (Fig. 3a). The same trend was found in random separation and environmental block separation, though environmental block separation generally showed lower sensitivity and TSS but higher specificity compared with random separation. The turning point of the stepwise regression was estimated to be 42-154 km based on mean model performance of all species, or between 42 and

130km based on the median of model estimations of individual species (Table 1). The model sensitivity based on bioM was comparable to that of the saturation stage when using spatial buffers (Fig. 3a). There was no significant difference in sensitivity between models based on bioM and models based on spatial buffers with similar size as bioM (Table 2), or between models based on bioM and models based on spatial buffers that had highest sensitivity (Table 3). For a small number of cases, there were significant difference in specificity and TSS between between models based on bioM and models based on spatial buffers with similar size as bioM (Table 2); the number of cases were doubled when comparing models based on bioM to models based on spatial buffers that had highest sensitivity (Table 3). Similar pattern was found for different occurrence separation methods (Fig. 3a).

The patterns were the same for sensitivity when background points generated from the bioM were used for model evaluation, but specificity showed slightly different patterns (Fig. 3b). Specificity showed a decreasing trend with buffer size when buffer size reached the turning point (Fig. 3b). The combined effect of sensitivity and specificity led to a more stable saturation status for TSS after buffer size reached the turning point (Fig. 3b). Also, when using background points generated from the bioM for model evaluation (instead of the largest buffer), most of the significant differences in specificity and TSS (between models based on bioM and spatial buffer) disappeared (Tables 2 & 3).

Discussion

Which hypothesis is supported?

This study conducted an experiment to investigate the relationship between size of training domain and model performance, as well as whether defining training domain based on species mobility leads to optimal model performance. Overall, I concluded that the saturation hypothesis was supported (Fig. 1); i.e. model performance (here, sensitivity and TSS) showed a rapid increase when buffer size, and stabilized at a certain level after buffer size reached the turning point (Fig. 3). The results of specificity showed a trend of increase and decrease when bioM was used in model evaluation, seemingly support the unimodal hypothesis, but together with the results of specificity when largest buffer was in model evaluation, it can be interpreted as that the background data used in model training would have the highest specificity in recognizing itself; in other words, highest specificity occurs when the same background data are used in training and testing. Interestingly, the model performance based on species' mobility (bioM) was almost always similar to the saturation performance obtained from spatial buffers, suggesting bioM provided reasonably good training domains, thus providing positive evidence to support the species' mobility hypothesis.

Is it reasonable to use spatial buffers as training domain?

In view of considering species' mobility in the application of ENM/SDM, multiple approaches of designing the training domain have been used in literature, such as using ecoregions, convex hull of species' occurrences, or expert range maps. Building spatial buffers around known occurrences is another approach that is commonly used in literature. It provides a relatively simple way to delineate an area around known occurrences as the training domain. The size of the spatial buffer (or radius) can be set to reflect the dispersal ability of a species, thus better meet the theoretical consideration of species' mobility. The spatial buffer could lead to relatively smoothed boundaries (e.g. a circle) that assume uniform dispersal ability along all directions and do not capture fine scale dispersal barriers, compared with fine tuned area based on dispersal simulations (Machado-Stredel *et al.*, 2021).

The spatial buffer appcoach is a relatively simple method, but is it a reasonable approach for defining the training domain? The experiment showed models built with small spatial buffers had lowest sensitivity, suggesting using background points immediately next to known occurrences does not provide enough useful information for the model to recognize the difference between the environmental conditions a species preferred vs. background conditions. On the other hand, models with intermediate and extremely large buffers had similar sensitivity, suggesting a larger buffer can provide more useful information for model calibration, but this improvement saturates at a certain point. Therefore, the buffer method can be considered as a reasonable way to construct a training domain when an appropriate size is used.

A caveat to note is that, when using a spatial buffer to generate background data, an extremely large buffer could potentially lead to inflated model specificity, and may mislead model evaluation or model comparison. There are several reasons for this caveat. First, when specificity is calculated from background (pseudo-absence) data, its interpretation becomes the proportion of areas predicted as absence, thus it should be given less weight compared with sensitivity, which is calculated from presence data (Peterson *et al.*, 2008). Second, the value of specificity can depend on the selection of background data used in evaluation. Background data from a larger area can yield higher specificity, and subsequently inflated TSS or AUC values, compared with that of a smaller area for the same model.

Does species' mobility provide the optimal training domain?

The experiment provided positive evidence to support the hypothesized role of species' mobility in defining the training domain of ENM. The model sensitivity based on species' mobility (bioM) generally fell in the saturation sensitivity obtained from spatial buffers. In most cases, the model based on bioM was not significantly different from that based on similar sized buffers or buffers with the highest evaluation index. This suggested that bioM can help define a training domain that leads to optimal or near-optimal performance (especially sensitivity), though not necessarily superior to spatial buffers.

Implications for future studies

Ecological niche modelers always face the question of how to define the training domain. The ecological theory suggests the training domain shall consider species' dispersal ability, though the information and knowledge of dispersal are largely inadequate (Driscoll *et al.*, 2014), thus limiting the implementation of this idea in ecological niche model applications. Two findings of this study (i.e. the stratution relationship of spatial buffers and model performance, and the similarity between bioM and the stratution performance of spatial buffers) suggests that, when there is no information of species' dispersal, modelers may explore model performance across a series of training domains based on different sized spatial buffers. This method may help identify the buffer size that leads to the stratution performance among many different sized buffers. This strategy is conceptually similar to other model tuning strategies, such as feature and regularization selection in Maxent (Muscarella *et al.*, 2014; Cobos *et al.*, 2019) and stepwise variable selection in multiple regression (Efroymson). If a rough estimate of dispersal distance across the time of interest is known, this information may also be used as the radius of the spatial buffer and to be further compared with buffers with different sizes. The conclusions here were based on cross-validations of the occurrence and dispersal data of a group of hummingbird species, and future studies can further investigate the role of dispersal ability in ecological niche modeling with other taxons that have adequate knowledge of dispersal as well as independent testing data.

References:

Anderson, R.P. & Raza, A. (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus *Nephelomys*) in Venezuela. *Journal of Biogeography*, **37**, 1378–1393.

Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E. & Rahbek, C. (2019) Standards for distribution models in biodiversity assessments. *Science Advances*, **5**.

Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S.P., Peterson, A.T., Soberón, J. & Villalobos, F. (2011) The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling*, **222**, 1810–1819.

Blowes, S.A., Supp, S.R., Antão, L.H., Bates, A., Bruelheide, H., Chase, J.M., Moyes, F., Magurran, A., McGill, B., Myers-Smith, I.H., Winter, M., Bjorkman, A.D., Bowler, D.E., Byrnes, J.E.K., Gonzalez, A., Hines, J., Isbell, F., Jones, H.P., Navarro, L.M., Thompson, P.L., Vellend, M., Waldock, C. & Dornelas, M. (2019) The geography of biodiversity change in marine and terrestrial assemblages. *Science*, **366**, 339–345.

Cobos, M.E., Peterson, A.T., Barve, N. & Osorio-Olvera, L. (2019) kuenm: an R package for detailed development of ecological niche models using Maxent. *PeerJ*, **7**, e6281.

Cooper, J.C. & Soberón, J. (2018) Creating individual accessible area hypotheses improves stacked species distribution model performance. *Global Ecology and Biogeography*, **27**, 156–165.

Doughty, C.E., Wolf, A., Morueta‐Holme, N., Jørgensen, P.M., Sandel, B., Violle, C., Boyle, B., Kraft, N.J.B., Peet, R.K., Enquist, B.J., Svenning, J., Blake, S. & Galetti, M. (2016) Megafauna extinction, tree species range reduction, and carbon storage in Amazonian forests. *Ecography*, **39**, 194–203.

Driscoll, D.A., Banks, S.C., Barton, P.S., Ikin, K., Lentini, P., Lindenmayer, D.B., Smith, A.L., Berry, L.E., Burns, E.L., Edworthy, A., Evans, M.J., Gibson, R., Heinsohn, R., Howland, B., Kay, G., Munro, N., Scheele, B.C., Stirnemann, I., Stojanovic, D., Sweaney, N., Villaseñor, N.R. & Westgate, M.J. (2014) The trajectory of dispersal research in conservation biology. Systematic review. *PloS ONE*, **9**, e95053.

Efroymson Multiple regression analysis. *Mathematical methods for digital computers*.

Feng, X., Merow, C., Liu, Z., Park, D.S., Roehrdanz, P.R., Maitner, B., Newman, E.A., Boyle, B.L., Lien, A., Burger, J.R., Pires, M.M., Brando, P.M., Bush, M.B., McMichael, C.N.H., Neves, D.M., Nikolopoulos, E.I., Saleska, S.R., Hannah, L., Breshears, D.D., Evans, T.P., Soto, J.R., Ernst, K.C. & Enquist, B.J. (2021) How deregulation, drought and increasing fire impact Amazonian biodiversity. *Nature*, **597**, 516–521.

Feng, X. & Papeş, M. (2015) Ecological niche modelling confirms potential north-east range expansion of the nine-banded armadillo (*Dasypus novemcinctus*) in the USA. *Journal of Biogeography*, **42**, 803–807.

Hannah, L., Roehrdanz, P.R., Marquet, P.A., Enquist, B.J., Midgley, G., Foden, W., Lovett, J.C., Corlett, R.T., Corcoran, D., Butchart, S.H.M., Boyle, B., Feng, X., Maitner, B., Fajardo, J., McGill, B.J., Merow, C., Morueta-Holme, N., Newman, E.A., Park, D.S., Raes, N. & Svenning, J.-C. (2020) 30% land conservation and climate action reduces tropical extinction risk by more than 50%. *Ecography*, **43**, 943–953.

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.

Hutchinson, G.E. (1957) Concluding Remarks. *Cold Spring Harbor symposia on quantitative biology*, **22**, 415–427.

Machado-Stredel, F., Cobos, M.E. & Peterson, A.T. (2021) A simulation-based method for selecting calibration areas for ecological niche models and species distribution models. *Frontiers of Biogeography*, **13**. DOI: https://doi.org/10.21425/F5FBG48814

Mackenzie, D.I. (2005) Was it there? Dealing with imperfect detection for species presence/absence data+. *Australian & New Zealand journal of statistics*, **47**, 65–74.

Merow, C., Smith, M.J. & Silander, J.A., Jr (2013) A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, **36**, 1058–1069.

Muggeo, V.M.R. (2003) Estimating regression models with unknown break-points. *Statistics in Medicine*, **22**, 3055–3071.

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.

Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C.M., Myers, C.E. & Peterson, A.T. (2013) Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling*, **263**, 10–18.

Park, D.S. & Potter, D. (2015a) A reciprocal test of Darwin's naturalization hypothesis in two mediterranean-climate regions. *Global Ecology and Biogeography*, **24**, 1049–1058.

Park, D.S. & Potter, D. (2015b) Why close relatives make bad neighbours: phylogenetic conservatism in niche preferences and dispersal disproves Darwin's naturalization hypothesis in the thistle tribe. *Molecular Ecology*, **24**, 3181–3193.

Peterson, A.T., Papeş, M. & Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213**, 63–72.

Peterson, A.T. & Soberón, J. (2012) Species distribution modeling and ecological niche modeling: getting the concepts right. *Natureza & Conservação*, **10**, 102–107.

Proosdij, A.S.J., Sosef, M.S.M., Wieringa, J.J. & Raes, N. (2016) Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, **39**, 542–552.

Qiao, H., Feng, X., Escobar, L.E., Peterson, A.T., Soberón, J., Zhu, G. & Papeş, M. (2018) An evaluation of transferability of ecological niche models. *Ecography,* **42**, 521-534.

Saupe, E.E., Barve, V., Myers, C.E., Soberón, J., Barve, N., Hensz, C.M., Peterson, A.T., Owens, H.L. & Lira-Noriega, A. (2012) Variation in niche and distribution model performance: The need for a priori assessment of key causal factors. *Ecological Modelling*, **237-238**, 11–22.

Soberon, J. & Townsend Peterson, A. (2005) Interpretation of Models of Fundamental Ecological Niches and Species' Distributional Areas. *Biodiversity Informatics*, **2**. DOI: https://doi.org/10.17161/bi.v2i0.4

Valavi, R., Elith, J., Lahoz-Monfort, J.J. & Guillera-Arroita, G. (2019) block CV : An r package for generating spatially or environmentally separated folds for k ‐fold cross‐validation of species distribution models. *Methods in Ecology and Evolution*, **10**, 225–232.

Data Accessibility Statement: The data supporting the results will be archived in a public repository and the data DOI will be included at the end of the article.

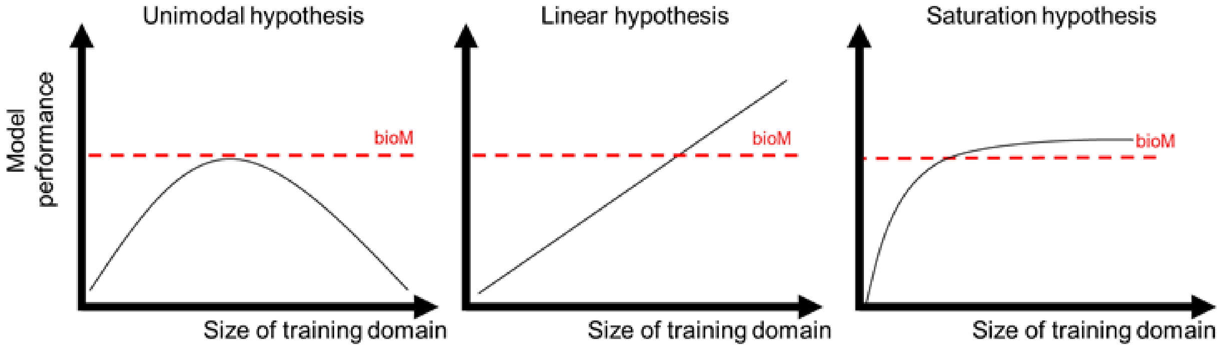

Figure 1. The hypothesized relationship between training domain and model performance. The black curve in each panel represents the unimodal, linear, or saturation relationship between increased size of training domain and model performance. The red dashed line represents the expected model performance based on a training domain that is delineated based on species' mobility (bioM).

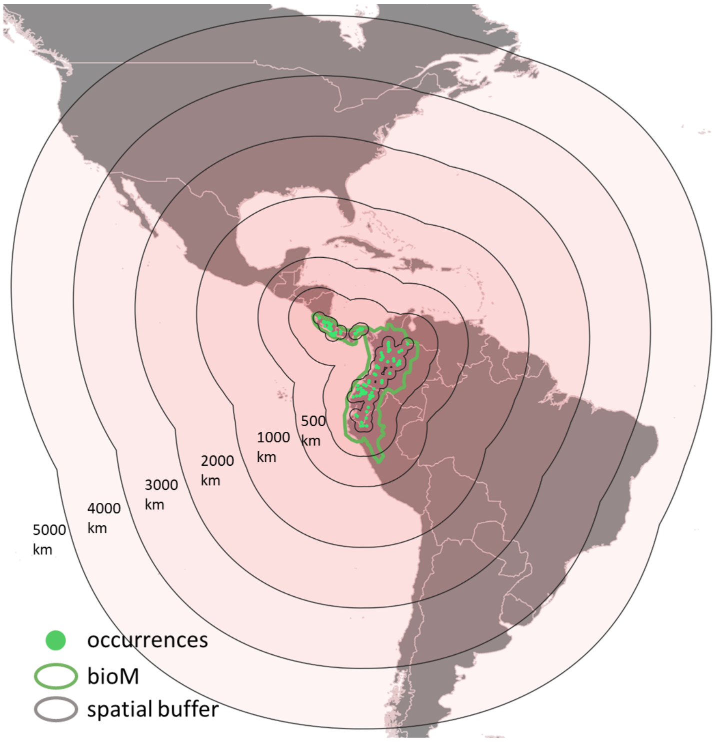

Figure 2. Comparison of training domains based on species' accessible aera (bioM; green polygon) and spatial buffers of occurrences (gray polygons). The green points are occurrences of *Eutoxeres aquila*. A total of 26 sized buffers are used in this study, and this figure only shows examples of 100, 500, 1000, 2000, 3000, 4000, 5000km buffers.

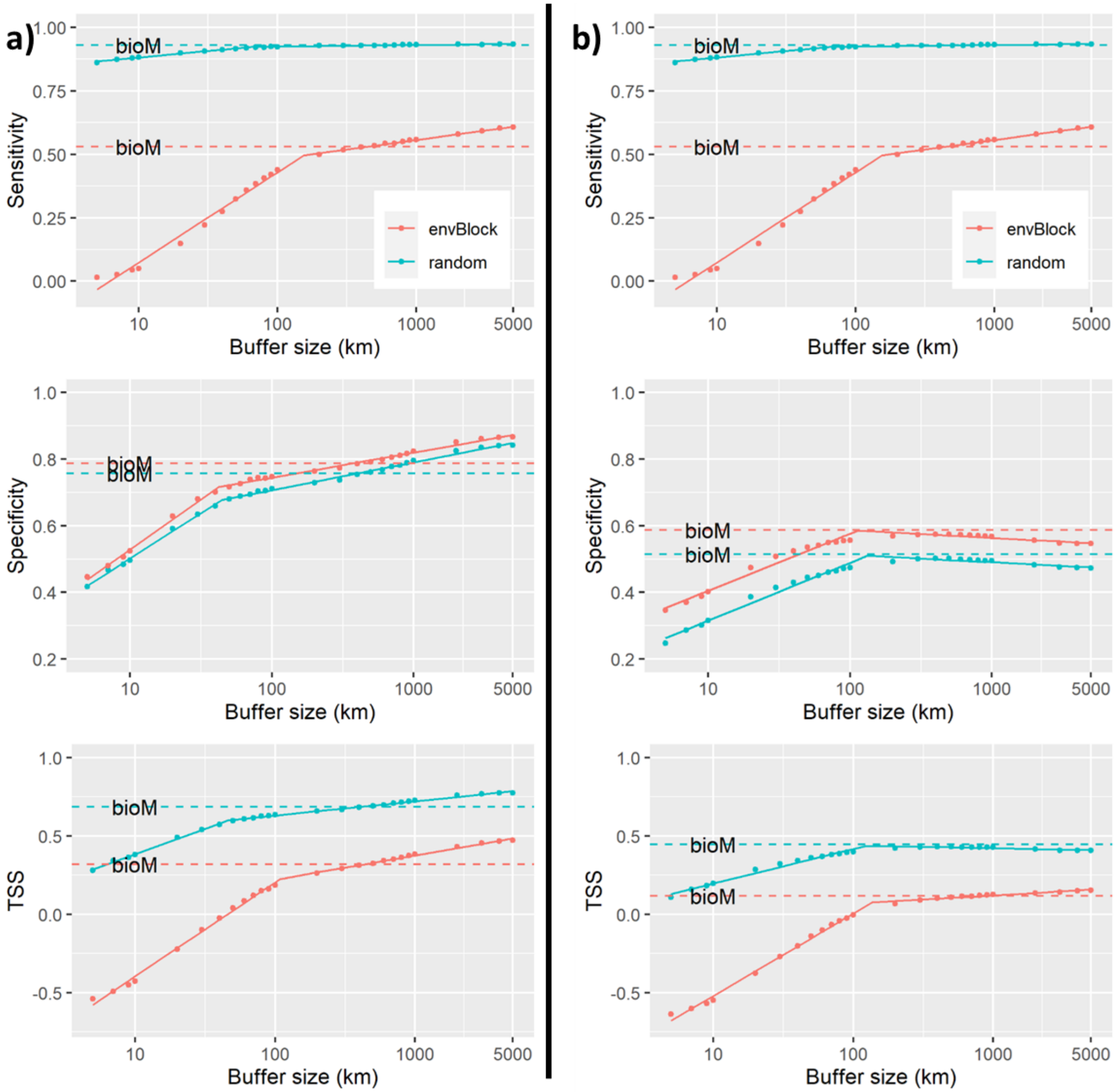

Figure 3. Model performance averaged across all species. Panel a) shows the sensitivity, specificity, and TSS based on background data from the largest spatial buffer, while those indices in panel b) are based on background data from bioM. Occurrence separation method is represented by different colors: red for environmental block and blue for random. The solid lines represent the fitted stepwise regressions of buffer size and model performance. The horizontal dashed lines represent model performance based on bioM.

Table 1. Turning point of model performance along buffer size estimated from the two-segment stepwise regression. The turning point is estimated from the stepwise regression of mean model performances of all species. It is also shown as the median of the turning points estimated from model performances of individual species.

	Estimated turning point, when background data from the largest buffer is used		Estimated turning point, when background data from bioM is used	
	random	environmental block	random	environmental block
Sensitivity	mean: 65km median: 58km	mean: 154km median: 124km	mean: 65km median: 59km	mean: 154km median: 130km
Specificity	mean: 44km median: 47km	mean: 42km median: 42km	mean: 45km median: 50km	mean: 42km median: 42km
TSS	mean: 46km median: 46km	mean: 108km median: 85km	mean: 46km median: 50km	mean: 108km median: 87km

Table 2. Number of cases (species) when a model calibrated with bioM is significantly different from the model based on a spatial buffer of similar size.

	Evaluation based on background data from the largest buffer		Evaluation based on background data from bioM	
	random	environmental block	random	environmental block
Sensitivity	0/87	0/87	0/87	0/87
Specificity	21/87	7/87	8/87	0/87
TSS	12/87	0/87	8/87	0/87

Table 3. Number of cases (species) when a model calibrated with bioM is significantly different from the model based on a spatial buffer that leads to the highest evaluation index.

	Evaluation based on background data from the largest buffer		Evaluation based on background data from bioM	
	random	environmental block	random	environmental block
Sensitivity	0/87	0/87	0/87	0/87
Specificity	61/87	16/87	10/87	0/87
TSS	40/87	1/87	10/87	12/87