Biodiversity Informatics, 20, 2026, pp. 74-80

THE ENVIRONMENTAL REPRESENTATIVENESS EFFECT IN SPECIES DISTRIBUTION MODEL EVALUATION WITHOUT ABSENCE DATA

Alberto Jiménez-Valverde

Department of Biogeography and Global Change, Museo Nacional de Ciencias Naturales (MNCN), CSIC, Madrid, Spain

Abstract. In species distribution modeling (SDM), the environmental representativeness effect hinders the comparison and generalization of discrimination statistics, as their values are context-dependent and may vary without reflecting actual differences in model accuracy. To address this issue, a harmonization approach based on the uniform distribution of suitability values has been proposed, demonstrating effectiveness when true absence data are available. However, in most cases, models solely rely on presence records and use background points instead of true absences, posing additional validation challenges. This study simulates habitat suitability and presence-absence data, and evaluates the robustness of several validation indices: background-based AUC (AUC_b), its harmonized version ($uAUC_b$), and two variations of the Boyce index. The results show that the Boyce index remains unaffected by the representativeness effect, whereas AUC_b varies across scenarios, confirming the influence of the representativeness effect on SDM results with background data. Harmonization through $uAUC_b$ successfully makes values comparable, but its reliability depends on sample size, requiring at least 100 presences and 10,000 background points.

Key words: background data, Boyce index, evaluation, ecological niche models, uniform *AUC*

Introduction

In species distribution modelling (SDM; Franklin 2009; Peterson et al. 2011), a critical step is model validation, typically assessing how accurately species occurrence data are predicted using discrimination statistics (Fielding and Bell 1997). Jiménez-Valverde et al. (2013) cautioned that discrimination measures estimated on different evaluation datasets may not be directly comparable, because their values depend on the statistical distribution of suitability scores, which in turn is shaped by the distribution of predictor variables. They termed this phenomenon the environmental representativeness effect, which implies that discrimination statistic values are inherently context dependent. Consequently, a model may exhibit high or low performance only due to the statistical distribution of the suitability scores in the evaluation dataset rather than on its intrinsic accuracy. This means that a given species may respond to the environment in the same way as modelled across different territories, yet the evaluation statistic may yield different values in each case. While this behavior is expected from those statistics that measure how well the instances of presence are discriminated against those of absence (i.e., they function as intended), it complicates comparisons and significantly influences the conclusions that can be drawn (Jiménez-Valverde 2022).

Harmonizing discrimination statistics, defined as recalculating these metrics after weighing the cases in the evaluation dataset to impose a uniform distribution of suitability values, has been shown to appropriately account for the representativeness effect (Jiménez-Valverde 2022, 2025a). This method has been demonstrated to perform well specifically when true absences are available for evaluation, but in SDM this is rarely the case. Modelers typically have only presence records for the species but lack information on where the species is absent. Even when absence data are available, their reliability is often questionable (Lobo et al. 2010). As a result, researchers often have to select a random set of locations within the study area with no information on the presence or absence of the target species and use these as absences, the so-called background points (Stockwell and Peters 1999; Phillips et al. 2006). How harmonization handles the representativeness effect in scenarios without a gold standard for validation (i.e., presences and true absences) remains unclear (Jiménez-Valverde 2022).

Using discrimination statistics in presence-background data scenarios has been questioned on numerous occa-

---

* Correspondence: email: alberto.jimenez@mncn.csic.es

```
$n$ = 100000                      # total dataset size
scenarios = {a, b, c}             # HS distributions (see Fig. 2)
$n_1$ = {15, 100, 1000}           # number of presences
$n_b$ = {100, 1000, 10000}        # number of background points

FOR each scenario in scenarios
        FOR each combination ($n_1$, $n_b$)
                FOR iterations = 1 to 10000
                        1. Generate $n$ habitat suitability values ($hs_i$)
                        2. Generate species presence-absence data ($sp_i$)
                        3. Build the validation dataset (sample $n_1$
                        presences + $n_b$ background points)
                        4. Calculate $AUC_b$, $uAUC_b$, BW and BS
                END FOR
        END FOR
END FOR
```

Figure 1. Pseudocode summarizing the simulation procedure used in this study. Habitat suitability values (*HS*) were simulated under three alternative distributions (scenarios a, b, c) representing different environmental representativeness scenarios. Species occurrences were then generated using a Bernoulli process with probability equal to *HS*, resulting in perfectly calibrated predictions. Validation datasets were constructed by sampling presences (n_1) and background points (n_b) with different sample sizes. For each scenario and validation dataset, the procedure was repeated 10000 times and four validation statistics were calculated: AUC_b, uniform AUC_b ($uAUC_b$), and two versions of the Boyce index (*BW* and *BS*).

sions (Boyce et al. 2002; Lobo et al. 2008; Peterson et al. 2008; Jiménez-Valverde 2012; Smith 2013; Leroy et al. 2018). Essentially, since the background data is a mixture of true absence and presence data, the true negative rate (proportion of correctly predicted absences) and the commission error rate (proportion of absences incorrectly predicted as presences) are unreliable. Thus, for instance, the maximum possible value of the very well-known Area Under the ROC Curve (*AUC*) statistic (Krzanowski and Hand 2009) is no longer 1 and depends on the actual but unknown distribution of the species (Wiley et al. 2003). Yet, one could argue that, despite these limitations and for practical purposes, the *AU*C could still be used to rank models, with a higher *AUC* indicating a better model than a lower one (Phillips et al. 2006), at least when comparing models for the same species to ensure that the range area remains constant. Despite its challenging and nuanced interpretation, the *AUC* remains a common choice among modelers in presence-background scenarios, which are the most frequent in species distribution modeling. It is thus important to understand how its harmonization to account for the representativeness effect -the so-called uniform *AUC* or *uAUC*- works in this type of presence-background frameworks (Jiménez-Valverde 2022). However, it is important to note that the *uAUC* does not address the absence of a true gold standard when reliable absence data are unavailable.

Methods

Since the purpose of this study is to analyze the behavior of various validation indices and how they respond to the environmental representativeness effect, I directly simulated habitat suitability (*HS*) values and species occurrences with a known relationship between them, rather than following the traditional approach of first generating virtual species, then sampling them, followed by parameterizing distribution models, and finally validating them. The procedure followed in this study is simpler and eliminates many sources of uncertainty and variation that may arise with the traditional method. Idealized simulations are necessary to understand the theoretical behavior of the method. The overall simulation workflow is summarized in Fig. 1. Following the general procedure of Jiménez-Valverde (2022), I simulated *HS* values bounded between 0 and 1 using the truncnorm package (Mersmann et al. 2018) for R (R Development Core Team 2024), and I considered three different scenarios, setting the sample size (n) to 100,000: ***Scenario a***: cases with low and high *HS* values were more frequent than cases with medium values (Fig. 2A). To generate hs_i, $n/2$ random numbers were generated from two truncated normal distributions bounded between 0 and 1 and with $\mu_1 = 0$, $\mu_2 = 1$ and $\sigma_1 = \sigma_2 = 0.15$, and concatenated. ***Scenario b***: *HS* follows a hump-shaped distribution (Fig. 2B). To generate hs_i, n random numbers were generated from a truncated normal distribution bounded between 0 and 1 and with $\mu = 0.5$ and $\sigma = 0.2$. ***Scenario c***: cases with high and medium *HS* values were less frequent than cases with low *HS* values (Fig. 2C). To generate hs_i, $n/2$ random numbers were generated from a truncated normal (with $\mu = 0$ and $\sigma = 0.05$) and from a uniform distribution, both bounded between 0 and 1, and concatenated.

I also considered three sample size levels for the validation presence set ($n_1 = \{15, 100, 1000\}$) and the validation background dataset ($n_b = \{100, 1000, 10{,}000\}$), resulting in nine different validation datasets.

For each scenario and validation dataset, I ran 10,000 iterations. In each iteration, first, a vector hs_i was simulated. Then, a vector sp_i with the presence-absence data was generated by drawing n values from a Bernoulli distribution, where the probability of presence was determined by the corresponding hs_i value. In this way, the simulated predictions were perfectly calibrated (i.e., the species respond to the environment in the same way). The *AUC* values calculated from the presence-absence data were approximately 0.930, 0.716, and 0.909 for scenarios a, b, and c, respectively, evidencing the environmental representativeness effect, that is, the response of the species is the same but the *AUC* values differ simply because the distributions of *HS* are dissimilar. Next, out of the n cases (each case has two associated values, hs_i *and* sp_i), n_1 were randomly extracted under the condition that they corresponded to a presence; from the remaining $n\text{-}n_1$ cases, n_b were randomly sampled. Both subsets were combined to form the validation dataset with which four statistics were finally calculated: the AUC_b (I use the subscript to differentiate this *AUC* value calculated with background data from the reference *AUC* value calculated with presence-absence data of each scenario, see above), the uniform AUC_b ($uAUC_b$) as proposed in Jiménez-Valverde (2025a), the Boyce index estimated with the moving window approach (*BW*) as proposed in Hirzel et al. (2006), and the Boyce index estimated with the smoothing approach (*BS*) as proposed in Liu et al. (2024). I estimated the Boyce index since it is the recommended measure for background data scenarios (Boyce et al. 2002; Hirzel et al 2006). The AUC_b was estimated non-parametrically by the trapezoidal rule

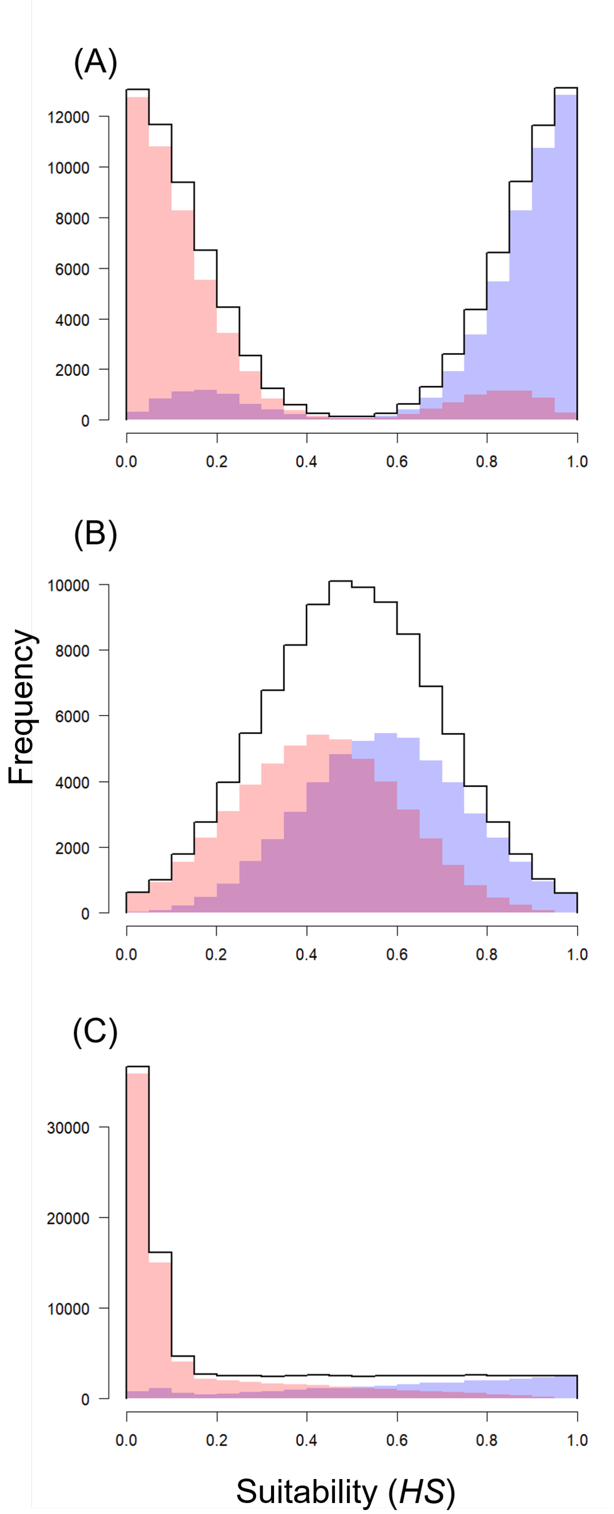

Figure 2. Examples of distributions of suitability values (*hs*, black line; *hs* | *sp* = 1, in blue; *hs* | *sp* = 0, in red). Scenario A represents situations where low and high suitability values are more frequent than cases with medium values. Scenario B represents situations where medium suitability values are more frequent than cases with low and high suitability values. Scenario C represents situations where high and medium suitability values are less frequent than cases with low values.

(Hanley and McNeil 1982). The $uAUC_b$ was estimated by partitioning the validation dataset into 10 bins according to the values of hs_i. Each case was assigned a weight equal to one divided by the number of observations in the corresponding bin, and the direct weighted trapezoidal estimation procedure was then applied (see Jiménez-Valverde 2025a for details). Both AUC_b and $uAUC_b$ were calculated using the function *AUCuniform.2()* of the vandalico package (Jiménez-Valverde 2025b). *BW* was estimated by using the ecospat (Broennimann et al. 2024) package with default values, which in this study implies a width of the moving window equal to 0.1 and a resolution of the moving window equal to 100. *BS* was estimated with the thin plate regression splines method using the mgcv package (Wood 2025); since all the smoothing methods evaluated by Liu et al. (2024) yielded comparable results, I selected one for the analysis and used a simplified version of the R code provided by the authors. Finally, boxplots were generated to interpret the differences between scenarios and validation datasets.

For assessing the results, it is necessary to know the expected value of the AUC_b under a uniformly distributed and perfectly calibrated *HS*. So, if the *HS* values in the full dataset are uniformly distributed between 0 and 1, and because the model is perfectly calibrated, the probability that a randomly selected presence case has an *HS* value higher than a randomly selected background case (i.e., the definition of the *AUC*) is exactly equal to its *HS* value. The *AUC* is therefore the mean *HS* value among presence cases. When plotting the density of *HS* values for presence cases against *HS*, this density has a triangular, increasing shape; the mean *HS* value among presences corresponds to the horizontal position of the barycenter of this triangle, which is 2/3 (~0.667) (Diccionarios Temáticos Vox 2019). Thus, this is the value we would expect to obtain for the $uAUC_b$ in the three scenarios (in the same way that it is 5/6 or ~0.833 for the *uAUC*; see Diamond 1992 and Jiménez-Valverde et al. 2013).

The R script used to perform the simulations of this study is available in Jiménez-Valverde (2026) for reproducibility.

Results and Discussion

The environmental representativeness effect is also evident with background data, as reflected in the varying AUC_b values across scenarios (Fig. 3). Interestingly, the effect may differ from that observed with absence data. For instance, scenario c exhibits the highest average AUC_b value, whereas its *AUC* falls between those of the other two scenarios (see Methods section). This highlights that the AUC_b does not reflect ranking capacity *sensu stricto* (i.e., the probability of assigning a higher *HS* value to a presence than to an absence), but rather a ranking of presences *versus* data that may include suitable locations in an unknown proportion. Scenario c represents a case where unsuitable areas are disproportionately represented in the validation data (Fig. 2C), suggesting that the effect of the spatial extent on the validation of the models (the "there are no elephants in Antarctica" effect, Lobo et al. 2010) might be exacerbated when using background data.

The harmonization procedure successfully makes the values of the discrimination statistics comparable, as evidenced by the $uAUC_b$ values being more similar compared to AUC_b across scenarios (Fig. 3). The precision of the estimated $uAUC_b$ is affected by sample size, especially by n_1; a value of 15 yields highly imprecise estimations in all scenarios and n_b levels. The lowest level on n_b produces imprecise estimations in all scenarios and n_1 levels. Bias, i.e., the departure of the averaged $uAUC_b$ values from 2/3, shows less clear patterns than precision. For instance, increasing n_1 tends to increase bias, which is especially noticeable for $n_b = 100$ and 1000, although not all scenarios are affected in the same way (compare scenarios a and c). Both presences and background points should be sufficiently numerous so that their distribution across hs_i bins approximates the underlying distributions of occurrences and available conditions; otherwise, the resulting $uAUC_b$ may be biased. This requirement is analogous to the presence-absence situation, in which observations within each hs_i bin must reflect the true prevalence of the species. In both situations, departures from these conditions distort the empirical frequencies used in the bin-based estimation procedure, which in turn biases the estimated *uAUC*. The most accurate estimations, in terms of both precision and bias, occur with $n_b = 10000$ and $n_1 = 100$ or 1000. The fact that the average $uAUC_b$ approaches 2/3 supports the validity of the harmonization method in the presence-background data scenarios.

The Boyce index quantifies the monotonic increment with *HS* of the ratio between the frequency of species observations in predicted *HS* classes and the expected frequency under a random distribution (P/E; Boyce et al. 2002), thus assessing the model's ability to discriminate between areas of higher and lower suitability. This index is unaffected by the environmental representativeness effect, as evidenced by its constant value of 1 (Fig. 3), which is the expected value given that all the scenarios correspond to well-calibrated *HS* values. For the same reason, Liu et al. (2024) explained that the Boyce index is insensitive to the extent of the unsuitable area, a key issue in SDM (Lobo et al. 2008; Anderson and Raza 2010; Barve et al. 2011). Note that the extent problem is a special case of the environmental representativeness effect (Jiménez-Valverde et al. 2013;

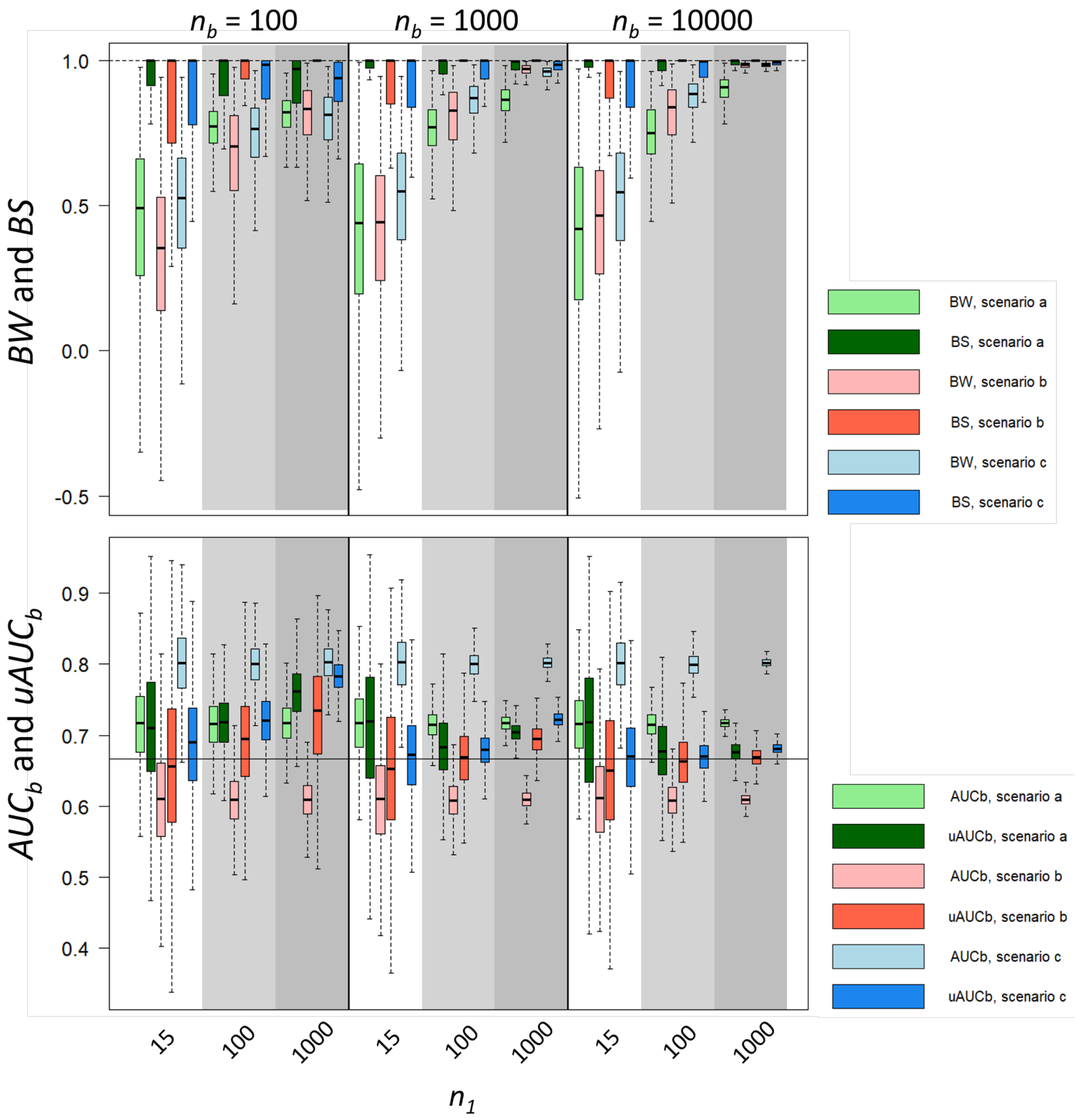

Figure 3. Top: median (horizontal marks), interquartile range (boxes), and 1.5 × spread (whiskers) of the Boyce index calculated with a moving window (*BW*, light colors) and the smoothing method (*BS*, dark colors) for the three different scenarios (scenario a in green, scenario b in red, scenario c in blue; see Fig. 2) and different sample sizes (n_1, number of presence data; n_b, number of background data). Bottom: same for the background-based *AUC* (AUC_b; light colors) and the uniform (harmonized) AUC_b ($uAUC_b$). Horizontal line marks the expected value of 2/3 (see text for details).

Jiménez-Valverde 2022).

Two additional results regarding the Boyce index are worth noting, both aligning with previous findings (Liu et al. 2024). First, *BS* is noticeably more accurate than *BW* (Fig. 3). Second, while both indices are influenced by sample size, *BS* is affected to a much lesser extent and primarily by n_b, whereas *BW* is more strongly impacted and mainly depends on n_1 (Fig. 3). This is because *BW* relies on data within each moving window to estimate P/E, whereas *BS* is calculated using the entire dataset (Liu et al. 2024).

Conclusions

The environmental representativeness effect hampers the comparison and generalization of discrimination statistic values. It affects presence-absence as well as presence-background statistics. Harmonizing the calculation of the $uAUC_b$ successfully improves the comparability of results, although sample sizes of $n_1 \geq 100$ and $n_b = 10{,}000$ are needed for reliability. While background data can be easily generated, obtaining sufficient presence data is challenging for rare species, making their model validation inherently difficult. Note, however, that while the $uAUC_b$ statistic accounts for the environmental representativeness effect, other issues associated with using the *AUC* without a gold standard (Jiménez-Valverde 2012) may still be relevant.

The Boyce index might be the preferred statistic when absence data are unavailable, as it avoids the theoretical issues associated with the *AUC* (Lobo et al. 2008; Peterson et al. 2008; Jiménez-Valverde 2012) and is insensitive to the representativeness effect. Its implementation through statistical smoothing, recently proposed by Liu et al. (2024), appears to be the recommended approach, as it exhibits less bias and greater precision than the traditional Boyce index, especially at low sample sizes. Under these circumstances, the Boyce index may be preferable to AUC_b (see also Liu et al. 2024), especially for rare species, for which the harmonization of AUC_b may yield unreliable results. However, contradictory results reported by Jiménez and Soberón (2020) suggest that the performance of the Boyce index may depend on specific data characteristics, and therefore a definitive recommendation cannot yet be made.

The simulations represent an idealized scenario intended to isolate the environmental representativeness effect and evaluate the theoretical behavior of the statistics. Additional complexities present in empirical datasets (e.g., detection error, spatial autocorrelation, or sampling bias) were not included, as they would introduce additional sources of variation unrelated to the mechanism examined here. Overall, the results highlight the importance of accounting for the environmental representativeness effect when interpreting discrimination statistics, either through harmonization or by using an index intrinsically insensitive to this effect, and provide practical guidance on the appropriate use of $uAUC_b$ and the Boyce index in species distribution model evaluation. Finally, It should be noted that harmonization facilitates the comparison of discrimination metrics across different contexts, but it does not by itself ensure greater ecological validity of the models.

Acknowledgments

This study was funded by the CliMod project (PID2023-147672NB-I00) financed by MCIU/AEI/10.13039/501100011033/FEDER, UE, and by the grant 2024ICT253 supported by CSIC. Lucia Maltez reviewed the English.

Competing Interests

The author has declared no competing interests.

Data availability

No new data were used or generated. The code used for the simulations is available at Figshare (Jiménez-Valverde 2026).

References

Anderson, R. P., and A. Raza. 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. https://doi.org/10.1111/j.1365-2699.2010.02290.x

Barve, N., V. Barve, A. Jiménez-Valverde, A. Lira-Noriega, S. P. Maher, A. T. Peterson, J. Soberón, and F. Villalobos. 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

Boyce, M. S., P. R. Vernier, S. E. Nielsen, and F. K. A. Schmiegelow. 2002. Evaluating resource selection functions. Ecological Modelling 157: 281-300. https://doi.org/10.1016/S0304-3800(02)00200-4

Broennimann, O., Di Cola, V., and Guisan, A. 2024. ecospat: Spatial Ecology Miscellaneous Methods. R package version 4.1.1. Available at https://CRAN.R-project.org/package=ecospat.

Diamond, G.A. 1992. What price perfection? Calibration and discrimination of clinical prediction models. Journal of Clinical Epidemiology 45: 85-89. https://doi.org/10.1016/0895-4356(92)90192-P

Diccionarios Temáticos Vox. 2019. Matemáticas. Colección Vox 10. Larousse Editorial, Barcelona.

Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation 24: 38-49. https://doi.org/10.1017/S0376892997000088

Franklin, J. 2009. Mapping species distributions. Cambridge University Press.

Hanley, J. A., and B. J. McNeil. 1982. The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. Radiology 143: 29-36. https://doi.org/10.1148/radiology.143.1.7063747.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., and Guisan, A. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecological Modelling 199: 142-152. https://doi.org/10.1016/j.ecolmodel.2006.05.017.

Jiménez, L., and J. Soberón. 2020. Leaving the area under the receiving operating characteristic curve behind: An evaluation method for species distribution modelling applications based on presence-only data. Methods in Ecology and Evolution 11: 1571-1586. https://doi.org/10.1111/2041-210X.13479.

Jiménez-Valverde, A. 2012. Insights into the area under the receiver operating characteristic curve as a discrimination measure in species distribution modelling. Global Ecology and Biogeography 21: 498-507. https://doi.org/10.1111/j.1466-8238.2011.00683.x

Jiménez-Valverde, A. 2022 The uniform AUC: Dealing with the representativeness effect in presence–absence models. Methods in Ecology and Evolution 13: 1224-1236. https://doi.org/10.1111/2041-210X.13826.

Jiménez-Valverde, A. 2025a. Refining uniform discrimination metrics: towards a case-by-case weighting evaluation in species distribution models with presence-absence data. Ecology and Evolution 15: e72573. https://doi.org/10.1002/ece3.72573

Jiménez-Valverde, A. 2025b. vandalico: Evaluation of presence-absence models. R package version 0.2.0. Available at https://cran.r-project.org/web/packages/vandalico/index.html.

Jiménez-Valverde, A. 2026. R script for reproducing SDM validation analyses without absence data under the environmental representativeness effect. figshare. Software. https://doi.org/10.6084/m9.figshare.31942269.

Jiménez-Valverde, A., Acevedo, P., Barbosa, A.M., Lobo, J.M., and Real, R. 2013. Discrimination capacity in species distribution modelling depends on the representativeness of the environmental domain. Global Ecology and Biogeography 22: 508-516. https://doi.org/10.1111/geb.12007.

Krzanowski, W.J., and D.J. Hand. 2009. ROC curves for continuous data. Chapman and Hall, Boca Raton, FL.

Leroy, B., Delsol, R., Hugueny, B., Meynard, C. N., Barhoumi, C., Barbet-Massin, M., and Bellard, C. 2018. Without quality presence–absence data, discrimination metrics such as TSS can be misleading measures of model performance. Journal of Biogeography, 45: 1994-2002. https://doi.org/10.1111/jbi.13402

Liu, C., Newell, G., White, M., Machunter, J. 2024. Improving the estimation of the Boyce index using statistical smoothing methods for evaluating species distribution models with presence-only data. Ecography 2025: e07218. https://doi.org/10.1111/ecog.07218.

Lobo, J.M., Jiménez-Valverde, A., and 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

Lobo, J. M., Jiménez-Valverde, A., and Hortal, J. 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography 33: 103-114. https://doi.org/10.1111/j.1600-0587.2009.06039.x

Mersmann, O., Trautmann, H., Steuer, D., Bornkamp, B. 2018. truncnorm: Truncated normal distribution. R package version 1.0-9. Available at https://CRAN.R-project.org/package=truncnorm.

Peterson, A.T., Papeş, M., and Soberón, J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modelling. 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. L., and Araújo, M. B. 2011. Ecological niches and geographic distributions. Princeton University Press.

Phillips, S. J., Anderson, R. P., and 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

R Development Core Team. 2024. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available at https://CRAN.R-project.org.

Smith, A. B. 2013. On evaluating species distribution models with random background sites in place of absences when test presences disproportionately sample suitable habitat. Diversity and Distributions19: 867-872. https://doi.org/10.1111/ddi.12031

Stockwell, D. R. B., and D. P. Peters. 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

Wiley, E.O., McNyset, K.M., Peterson, A.T., Robins, C.R., and Stewart, A.M. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography 16: 120-127. https://doi.org/10.5670/oceanog.2003.42

Wood, S. 2025. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.9-3. Available at https://cran.r-project.org/web/packages/mgcv.