Evaluation of species distribution models for estimating animal dark diversity

Camilo Matus-Olivares  
Universidad de La Frontera  https://orcid.org/0000-0002-2436-8374

Jaime Carrasco  
Universidad de Chile  https://orcid.org/0000-0003-4123-4228

José Luis Yela  
Universidad de Castilla-La Mancha  https://orcid.org/0000-0003-1371-8495

Paula Meli  
Universidad de Concepción  https://orcid.org/0000-0001-5390-7552

Andres Weintraub  
Universidad de Chile  https://orcid.org/0000-0003-0440-0311

Fulgencio Lisón ( flison@udec.cl )  
Universidad de Concepción  https://orcid.org/0000-0003-1481-3750

Research Article

Keywords: Chiroptera, ensemble models, Lepidoptera, overfitting, SDM, threshold models

Posted Date: November 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1100842/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Aim

Applying wide and effective sampling of animal communities is rarely possible due to the associated costs and the use of techniques that are not always efficient. Thus, many areas have a faunistic hidden diversity we denote Animal Dark Diversity (ADD), defined as the diversity that is present but not yet detected plus the diversity defined by Pärtel et al. (2011) that is not (yet) present despite the area's favourable habitat conditions. We evaluated different species distribution model types (SDM techniques) on the basis of three requirements for ADD estimate reliability: 1) estimated spatial patterns of ADD do not differ significantly from other SDM techniques; 2) good predictive performances; and 3) low overfitting.

Location

Iberian Peninsula.

Taxon

Chiroptera and Noctuoidea (Lepidoptera)

Methods

We used distribution data for 25 species of bats and 352 species of moths. We evaluated eleven SDM techniques using biomod2 package implemented in the R software environment. We fitted the various SDM techniques to the data for each species and compared the resulting ADD estimates for the two animal groups under three threshold types.

Results

The results demonstrated that estimated ADD spatial patterns vary significantly between SDM techniques and depend on the threshold type. They also showed that SDM techniques with overfitting tend to generate smaller ADD sizes, thus reducing the possible species presence estimates. Among the SDMs studied, the ensemble models delivered ADD geographic patterns more like the other techniques while also presenting a high predictive performance for both faunal groups. However, the Ensemble Model Committee Average (ECA) performed much better on the sensitivity metric than all other techniques under any of the thresholds tested. In addition, ECA stood out clearly from the other ensemble model techniques in displaying low-medium overfitting.

Main conclusions

SDM techniques should no differ among each other in their ADD estimations, have good predictive performances and exhibit low overfitting. Furthermore, to reduce estimate uncertainty it is suggested that the threshold type be one that transforms high values of presences probabilities into binary information and furthermore that the SDM technique have a sensitivity bias, as otherwise the estimates will perform better for species absence in cases where it is not in fact known whether a species is truly absent.

Introduction

Understanding the geographical distribution patterns of animal species communities is a key objective in ecology and for conservation policies implementation. However, in many cases, determining such patterns is rendered impossible by the difficulties involved in monitoring the different taxa at large scales (Margules et al., 2002). The consequent undersampling of some areas has meant that there are regions for which animal diversity is still unknown. Also, in some cases species diversity is hidden owing to the application of no-suitable sampling techniques that are unable to detect certain species' presence (Bailey et al., 2014; MacKenzie et al., 2018; Pärtel, 2014). Nevertheless, there is another type of hidden diversity called "dark diversity," and defined by Pärtel et al. (2011) as the biodiversity that is absent from a site (true absence) yet could potentially inhabit it given the site's favourable environmental conditions. This concept does not apply to fauna, however, since true absence in the case of animals is never 100% certain due to detection problems (MacKenzie et al., 2018). In some regions there are thus three specific hidden diversity types, each one forming a part of the area's overall hidden diversity. Since the first two cited types are typical of animal inventories, we can group them under the heading of Animal Dark Diversity (ADD), defined as the diversity present in a given area but has not yet detected plus the dark diversity as defined by Pärtel and col.

The distribution patterns of these two ADD types could be determined through extensive monitoring and the implementation of better sampling techniques, but this would involve a considerable cost in resources, both human and material. An alternative approach would be to estimate the ADD distribution patterns using mathematical tools as proxies to determine a site's potential biodiversity. Good estimates of ADD would enable potential and as yet unknown high biodiversity areas for which conservation and management plans are needed to be identified and prioritized. Such estimates could also be used to pursue other conservation goals (e.g., to identify relatively high ADD areas regarding conservation problems, rare species, invasive species, etc.).

Studies have shown that the methods proposed for estimating dark diversity have performed well and could be successfully applied to the ADD concept. However, the main focus of these analyses has been to evaluate estimation techniques based on species co-occurrence information (e.g., Brown et al., 2019;
Carmona & Pärtel, 2021; Lewis et al., 2016; Ronk et al., 2016). Other methods, grounded upon species’ environmental affinities, have received little attention
(Carmona & Pärtel, 2021). Furthermore, approaches relying on species co-occurrence have generally been applied to taxonomic plant groups where good knowledge of species distribution usually exists. However, when complete information on each species’ distribution is not available, environmental affinities methods such as species distribution models (SDMs) appear to be more effective (Beissinger et al., 2016; Wilkinson et al., 2019).

Indeed, SDMs are an important tool in ecology, biogeography and conservation (Peterson et al., 2011), allowing researchers to explore the relationships between species’ geographical occurrence probabilities given certain environmental variables (Thuiller et al., 2009). There are various modelling techniques for SDM development, ranging from traditional statistics models such as generalized linear models (McCullagh, 1984) and modern methods based on machine learning such as random forest (Breiman, 2001) to formulations based on a combination of different individual modelling techniques known as ensemble models (Araujo & New, 2007).

One way to estimate ADD using SDM techniques is by transforming the estimated presence probability of each species derived from its individual SDM into binary data (presence/absence) and then stacking the predictions. The ADD at each site is then given by the number of absent species according to the data inputted to calibrate and test the SDMs, whose results indicate presences. This procedure is analogous to the dark diversity estimation method using SDMs applied by Ronk et al. (2016).

The ADD size so estimated will depend on the threshold type used to transform the occurrence probabilities given that the false positive count of each SDM constructed for each species varies with the probabilities transformation method employed (Fielding & Bell, 1997; Jiménez-Valverde & Lobo, 2007). ADD sizes will also depend on the degree of overfitting of the SDM technique. An overfitted model is one that has incorporated much residual variation as part of its structure (Bumham & Anderson, 2002) and will therefore predict very well with training data but not with testing data (Mutasa et al., 2020). Thus, SDM techniques with a high degree of overfitting will lead to very low ADD sizes given that the number of false positives predicted from training data (the larger part of the data from which ADD predictions will be made) will be low compared to those SDM techniques with less overfitting.

Although there are several SDM techniques that can be used to estimate ADD, there are no previous studies in which ADD estimates—or dark density estimates generally—based on different techniques have been analysed and compared. Consequently, more work is needed on how SDM technique choice can produce different spatial distributions of ADD in geographic space and whether these differences vary across several threshold types. In order to be optimal, an SDM technique must not only demonstrate good predictive performance but also produce ADD spatial pattern estimates that do not significantly depart from those obtained with other SDM techniques, since otherwise its results would simply be the particular set of estimates that particular technique happens to generate. Furthermore, the optimal technique should not produce models with significant overfitting as this would exclude many false positives due to its inclusion of a high degree of residual variation. Therefore, we should suggest that it is desirable for an SDM technique to deliver the highest number of false positives, as long as it accurately reflects the species’ environmental suitability (i.e., the technique has good predictive performance).

Bats and moths are two faunal groups with great biodiversity but due to their characteristics (nocturnal habits, elusive, cryptic, and rare species, different methodologies of study, etc) usually are studied using SDMs to understand their distribution patterns (e.g. Lisón & Calvo, 2013) and resolve many ecological issues. Therefore, they are a good model to explore and evaluate the implementation of ADD in these groups. In light of the foregoing, our aim in this study is to compare and assess ADD estimates derived from a range of SDM techniques using different threshold types while taking account of the degree of overfitting and its effects. Specifically, we propose to (1) compare ADD geographic distribution pattern estimates obtained from different SDM techniques and threshold types; (2) evaluate the techniques’ predictive capacities and degrees of overfitting and show how the latter is related to ADD size estimates; and (3) discuss the ability of these techniques to predict ADD in biogeography and conservation biology contexts.

### Material And Methods

#### Bat and moth distribution data and environmental variables

Our study was conducted using presence/absence distribution data on bat and moth (quadrifine Noctuoidea) species in the Iberian Peninsula nested in 10x10 km² UTM cells (6169 total cells). The bat distribution dataset was constructed from the distribution maps in “Atlas y Libro Rojo de los Mamíferos Terrestres de España” (Gisbert & Palomo, 2007) and “Atlas de dos morcegos de Portugal” (Rainho et al., 2013). These data were updated with new records at the same resolution from personal surveys and references (see Lisón et al., 2015; Lisón & Sánchez-Fernández, 2017). As regards the *Eptesicus serotinus* and *E. isabellinus* species of bats, we considered the distribution of the latter to be limited to the south and south-eastern regions of the Iberian peninsula (Andalusia and Murcia) (Gisbert & Palomo, 2007; Lisón, 2015), with the former species deemed to be found in the remaining areas. In fact, however, there is some overlap between the two areas (Santos et al., 2014).

Faunistic data on quadrifine Noctuoidea, as defined by (Yela et al., 2011; Zahiri et al., 2011), were derived from the FAUNOCIB database, which is part of the GeoBrink platform (Yela et al., 2011). The database contains some 170,000 published and validated faunistic records of the 850 species of Euteliidae, Erebidae, Nolidae and Noctuidae known to be present in the Iberian Peninsula and the Balearic Islands. Data on rare species (less than 40 records) were not included given that for estimating their geographical distribution, models types other than those used here are recommended (Breiner et al., 2015).

Data on rare species (less than 40 records) were not included except for one bat species with 33 records, which we decided to include due to the small number of bat species when compared to the number of moth species. However, these presence records still allow us to an appropriate use of SDM, since when the presence records are higher than 30, models accuracy is often good (Hernandez et al., 2006; Stockwell & Peterson, 2002; Wisz et al., 2008).

Land-use variables were obtained following the procedure due to Lisón et al. (2015) and Lisón & Sánchez-Fernández (2017). We employed the land-use map provide by the CORINE Land Cover Project (see www.eea.europa.eu and https://land.copernicus.eu/pan-european/corine-land-cover/clc-2006). Land-use cover
was reclassified from the 44 initial categories to 16 definitive ones (see Lisón & Sánchez-Fernández, 2017). For each such category, we calculated the percentage of surface area included in each cell. In addition, we extracted 44 environmental and geographical variables from BIOMOD for use with each cell following (Barbosa et al., 2012). The variables used to model are available in Supplementary Material Table S1.

**SDM modelling specifications**

We evaluated eleven different SDM techniques, of which eight were individual ones and the remaining three were ensemble model types that included the other eight (see Table S2). The individual techniques are varied and include both classical and modern models (Table S2). The ensemble models were the ones that are most popular among SDM users (Hao et al., 2019, 2020).

All of the SDM techniques were studied using the biomod2 package implemented in the R software environment (Thuiller et al., 2009). The package was chosen due to the variety of SDM techniques it supports and because it seems to be the most popular among users who prefer ensemble models for estimating species presence (Hao et al., 2019).

With each SDM technique we used biomod2's predetermined hyperparameters. Although their configuration could affect predictive performance (Hao et al., 2020), indications are that biomod2 users tend to employ the default hyperparameter settings (Hao et al., 2019) so to maintain consistency with existing studies we adopted the same practice. In any case, the default values have proven to give good results (e.g., Hao et al., 2020).

We fit the SDMs to each species through cross-validation using a randomly selected 80% of the data for calibration (i.e., fitting or training) and the remaining 20% for validation (i.e., testing). We repeated this step ten times and then averaged the estimated probabilities for each UTM cell to obtain a single probability and thus reduce the variance in the final results.

**Animal dark diversity (ADD) estimation and spatial patterns comparison**

We estimated the ADD for each UTM cell under each SDM technique and threshold type to be the number of species estimated as present by the models when the input dataset indicated absences for them. To binarize the SDM presences/absences, three threshold types were tested: MaxKappa (Monserud & Leemans, 1992), which maximizes Cohen's kappa coefficient; MaxTSS (Allouche et al., 2006), which maximizes the True Skills Statistic; and MeanProb (Cramer, 2003), which is the average of the presence probability estimates. Following Scherrer et al. (2018), who in turn based themselves on Liu et al. (2005) and Nenzén & Araújo (2011), the three may be respectively classified by technique as single-index-based, sensitivity and specificity combined, and predicted-probabilities-based. The threshold values were determined using the optimal.thresholds function of the PresenceAbsence package in R (Freeman & Moisen, 2008). We tested five thousand cut-off thresholds with MaxKappa and MaxTSS to find the optimal value.

The ADD spatial patterns generated by all combinations of SDM technique and threshold type for each animal group were checked for correlation using Spearman's coefficient ($\rho$). The results were considered to indicate high correlation when $\rho$ was $\geq 0.7$.

**SDM predictive performance and overfitting evaluation**

The SDM techniques' predictive performance was evaluated using three metrics: area under the receiver operating characteristic curves (AUC), sensitivity (i.e., proportion of correctly predicted presences or true positive rate) and specificity (i.e., proportion of correctly predicted absences or true negative rate). AUC shows how well a model discriminates species presences from absences. It measures the quality of the SDM’s predictions irrespective of a single cut-off threshold. Sensitivity and specificity depend on a single cut-off threshold, so they were calculated using the cut-off threshold values used to make the ADD estimates.

SDM technique overfitting was evaluated by the faunal group as the difference, under each metric, between the predictive performance calculated on the calibration data and that calculated on the testing data. Since the predictive performance of an overfitted SDM technique is expected to be better with the calibration data, the most overfitted techniques were considered those for which the predictive performance differences were the greatest. The relationship between overfitting and ADD size for each faunal group was assessed using scatter plots. For each SDM technique, the expected value of ADD in a UTM cell was plotted by threshold type against the mean overfitting value calculated by the AUC. A trend line obtained using simple linear regression was added to the graph of each plot.

**Results**

Data were analysed for 25 species of bats and 352 species of moths. With ten calibrations per species and technique, this meant that the number of SDMs fitted was 41470, of which 2750 were for bats and 38720 for moths. For each SDM technique, 3770 models were calibrated. Presence estimates stacked by SDM technique and threshold type resulted in 66 maps of ADD geographical distributions, 33 for each faunal group (see Figure 1, Supplementary Material Figure S1, S2, S3, S4 and S5).

In general terms, the ADD size estimates varied from one UTM cell to the next depending on the SDM technique and threshold type (Figures 2, 3 and 4). The MaxKappa threshold yielded smaller sizes, particularly in the case of moths (Figure 2 and 4). For both animal groups, when MaxKappa and MaxTSS thresholds were used, the SDM technique that generated the highest ADD size estimates was GLM whereas with the MeanProb threshold, the highest estimates were produced by the ECA ensemble method (Figure 2 and 4). At the other extreme, when using MaxKappa and MaxTSS thresholds the lowest estimates were obtained by the RF technique, while with the MeanProb threshold, this was the case for both the RF and FDA techniques on bats and FDA only on moths (Figure 2 and 4).
The correlation analyses of the different SDM techniques showed that for both animal groups, the geographic distribution patterns of the ADD sizes estimated by the ensemble models were more closely correlated with those of the other techniques than the latter were with each other (Figure 3). The estimated patterns of several of the techniques were highly correlated when MaxTSS and MeanProb thresholds were used (p > 0.7; Fig. 3), but with the MaxKappa threshold, only the ensemble models were highly correlated in both animal groups. RF was the technique that estimated the most particular ADD patterns when MaxKappa and MaxTSS thresholds were used, but with MeanProb this changed drastically, its estimates in this case being highly correlated with those of the other techniques, particularly for moths.

Most of the SDM techniques managed a good general predictive performance with the testing data (AUC > 0.7) for both animal groups (Table 1). RF achieved the best AUC values for bats, followed in declining order by EMP EWMP and ECA. However, RF also exhibited the highest overfitting for bats on this metric (Table 1). ECA scored the highest AUC value for moths, with EMP and EWMP in second and third spots, whereas for this faunal group RF again displayed the most overfitting together with GBM (Table 1).

GLM was the individual SDM technique with the least overfitting for both animal groups while among ensemble models, ECA's overfitting was the lowest by a clear margin. Also, ECA was respectively 0.05 points above and 0.08 points below the values for the techniques with the lowest and highest overfitting in bats, and 0.01 points below and 0.08 points above the lowest and highest overfitting techniques in the case of moths. Thus, on this criterion ECA is medium-low for bats and low for moths.

ECM performed very well on the sensitivity metric under any threshold type. It was the third best on overfitting with all threshold types except MeanProb, where it was the best of all, but on their performance the three ensemble models varied greatly. GLM and FDA presented the best overfitting on sensitivity under MaxKappa, and GLM and CTA did the same under MaxTSS for both faunal groups, while with the MeanProb threshold, GLM and CTA were the best only for bats (Table 2).

Regarding specificity, under MaxKappa, RF obtained the best values for both animal groups while under MaxTSS, RF performed the best for bats, and CTA did so for moths. However, RF was also the technique with the highest overfitting on this metric for bats under any threshold type (Table 2). When MeanProb was used, the best specificity was achieved by FDA technique for both bats and moths. GLM obtained the best overfitting under any threshold type for either animal group (Table 2).

Finally, the results showed a clearly negative relationship between the mean overfitting of SDM techniques and their expected ADD values in a UTM cell under the MaxKappa and MaxTSS thresholds (Figure 4). However, under the same two types, there were some differences regarding which techniques were associated with a larger expected ADD size concerning its overfitting (Figure 4). Under MeanProb, by contrast, no relationship was observed between mean overfitting and expected ADD values (Figure 4).

**Discussion**

In this study, we identified an optimal SDM technique for making ADD estimates as one that meets three key requirements: 1) its geographic distribution patterns do not significantly differ from those of other techniques; 2) it has good predictive performance; and 3) it displays a low degree of overfitting.

As regards the first requirement, the ensemble models tested in this study were the best option. They proved to significantly reduce the differences among individual SDM techniques in the ADD estimates under all of the threshold types tested. Those differences may be explained by the fact that each one of them produces a “noise” and a “signal” between presence probability estimates and environmental variables relationships as a result of its optimization methods (Araújo & New, 2007; Pearson et al., 2006). However, ensemble models can disjoint such noise to some degree, which improves their prediction quality (Araújo & New, 2007; Dormann et al., 2018) and consequently also the reliability of their estimated ADD patterns.

Considering now the SDM techniques with better general predictive performances (on AUC values), our results showed that ensemble models were better in both faunal groups, a finding that is consistent with other studies (e.g., Bouska et al., 2015; Farhadinia et al., 2015; Fletcher et al., 2016; Folmer et al., 2016; Hao et al., 2020; Marmion et al., 2009). ECA displayed a higher level of sensitivity than the other SDM techniques under any threshold type tested here but did not do so on specificity. This means that ECA did not explain species absences any better than the other techniques. Nevertheless, since we do not know if the species are truly absent in those areas, a technique that explains animal species absences better is not, in fact, the most appropriate. A more judicious choice would be a technique that is biased towards indicating presences, even if it means some cost in specificity. For our study this meant that the most suitable technique in terms of predictive performance on ADD estimates was the ECA ensemble model.

On overfitting as measured by AUC and sensitivity, ECA's results were noticeably lower than the other ensemble models. This result, plus the technique's impressive predictive performance (especially on sensitivity) and the fact that its ADD spatial pattern estimates did not differ significantly from those of the other SDM techniques all point to ECA as the best option for making ADD estimates under any of the threshold types we tested.

Our results further showed that SDM techniques with low overfitting values tended to estimate higher ADD sizes. Therefore, overfitting should be given due consideration in judging a technique's ADD estimates, for if it is not, much information on possible species presences in certain geographic regions could be lost. In other words, predictive performance should not be the only factor in the evaluation of SDM techniques for estimating ADD. To illustrate this point, recall that for bats, our results showed RF as having the best AUC on the testing data, which might at first suggest it was the best ADD estimation technique. However, it was also found that RF had a perfect classification on training data, with large differences between predictions on testing data (i.e., high overfitting). Other studies have also reported this problem with RF, and have advised that SDM technique overfitting should be investigated before deciding which model is best (Carlson et al., 2016). Yet others have declared RF to be better than techniques such as ensemble models, but precisely in those cases overfitting was not taken into account (e.g., Balestrieri et al., 2016; Boulanger et al., 2016; Koo et al., 2017; Murphy et al., 2015).
In most instances where an SDM is used, binarization of the presences probabilities is implemented with the MaxTSS threshold (Scherrer et al., 2018), possibly due to its balancing of sensitivity and specificity in the estimates regardless of species prevalence (Jiménez-Valverde & Lobo, 2007). However, when estimating ADD, thresholds such as MaxKappa that are biased towards better estimates for specificity could be a good alternative since they tend to transform the probabilities with higher values into presences, particularly when species prevalence is low (Jiménez-Valverde & Lobo, 2007). As a result, MaxKappa tends to reveal the most probable areas for species presence, thus reducing ADD estimate uncertainty. On the other hand, it can be extremely biased towards improving specificity, as was the case in our study for moths, leading to very low ADD size estimates. By contrast, the MeanProb threshold tends to obtain better estimates on sensitivity (Cramer, 2003), which is consistent with our results. It therefore has a tendency to transform probabilities with low values into presences, which is not ideal for ADD estimation given that the final estimates will then be characterized by considerable uncertainty.

In our view, thresholds should be used that transform high probabilities into presences, as this will lead to good ADD estimates. However, they must still be controlled using a tool that determines the threshold cut-offs so that only, say, those probabilities above the 75th percentile of their probability distribution are transformed into presences.

The Animal Dark Diversity (ADD) could be a tool with numerous applications in biogeography and conservation biology. Our results showed that it is possible to get good ADD estimates when ensemble models are implemented, taking into account the threshold used. The ADD areas could be used as a proxy of interesting areas where is possible to find species never recorded before. The ADD maps could be implemented into planning monitoring campaigns of species which are determinant, especially those with conservation problems, and which need intensive surveys for characteristic habits (nocturnal, cryptic, elusive species, etc.). The possibility to apply this novel concept through SDMs will help the ecologists and decision-makers to decide where they should focus their efforts, reducing the conservation costs.

Final Remarks
Understanding biodiversity patterns across different geographical areas is necessary in order to implement effective conservation plans, especially in the context of rapid evolution due to climate change. Our results suggest that the ECA ensemble model is the best option for animal dark diversity estimates, but the key conclusion is that whichever SDM technique is chosen, it must possess good predictive performance (especially in sensitivity) and low overfitting while generating animal dark diversity geographic pattern estimates that do not differ significantly from those generated by SDM techniques. Furthermore, to reduce uncertainty in the estimates, the threshold type should be one that transforms high presence probabilities into binary information. To conclude, the findings presented here should help improve animal dark diversity estimation for applications in conservation biology.

Declarations

Acknowledgements
We thank Ángeles Haz for her help with the English translation and Kenneth Rivkin for correcting and improving the text to achieve a native English level.

Author Contribution
Conceptualization, F.L. and C.M-O.; methodology, C.M-O., J.C., and F.L.; validation, C.M-O., J.C., and F.L.; formal analysis, C.M-O., J.C., and F.L.; investigation, C.M-O., J.C., and F.L.; resources, funding and project administration, F.L., J.C. and A.W.; writing—original draft preparation, C.M-O., J.C., and F.L.; writing—review and editing, C.M-O., J.C., J.L.Y., PM., A.W. and F.L. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest
The authors declare no conflict of interest.

Funding
F.L. and C.M-O were supported by ANID-FONDECYT Iniciación Nº 11180514 (Ministerio de Ciencia de Chile). C.M-O. was supported partially by Beca de Estudio de Doctorado (Universidad de La Frontera). J.C. was supported by ANID-FONDECYT Postdoctoral Nº 3210311 (Ministerio de Ciencia de Chile). P.M. was supported by ANID-FONDECYT Iniciación Nº 11191021 (Ministerio de Ciencia de Chile). J.L.Y was supported by the Junta de Comunidades de Castilla-La Mancha research scheme – European Regional Development Fund (ERDF), project SBPLY/17/180501/000492. A. W. was supported by ANID-FONDECYT Regular Nº 1170381.

References
Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. Journal of Applied Ecology, 43(6), 1223–1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x
Araujo, M., & New, M. (2007). Ensemble forecasting of species distributions. Trends in Ecology & Evolution, 22(1), 42–47. https://doi.org/10.1016/j.tree.2006.09.010
Bailey, L. L., MacKenzie, D. I., & Nichols, J. D. (2014). Advances and applications of occupancy models. Methods in Ecology and Evolution, 5(12), 1269–1279. https://doi.org/10.1111/2041-210X.12100
Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximising predictive performance of species distribution models. *Ecography, 43*(4), 549–558. https://doi.org/10.1111/ecog.04890

Hastie, T. J., & Tibshirani, R. J. (2017). *Generalized additive models*. Routledge.

Hastie, T., Tibshirani, R., & Buja, A. (1994). Flexible Discriminant Analysis by Optimal Scoring. *Journal of the American Statistical Association, 89*(428), 1255–1270. https://doi.org/10.1080/01621459.1994.10476866

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography, 29*(5), 773–785.

Jiménez-Valverde, A., & Lobo, J. M. (2007). Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica, 31*(3), 361–369. https://doi.org/10.1016/j.actao.2007.02.001

Koo, K. A., Park, S. U., Kong, W.-S., Hong, S., Jang, I., & Seo, C. (2017). Potential climate change effects on tree distributions in the Korean Peninsula: Understanding model & climate uncertainties. *Ecological Modelling, 353*, 17–27. https://doi.org/10.1016/j.ecolmodel.2016.10.007

Lewis, R. J., Szava-Kovats, R., & Pärtel, M. (2016). Estimating dark diversity and species pools: An empirical assessment of two methods. *Methods in Ecology and Evolution, 7*(1), 104–113. https://doi.org/10.1111/2041-210X.12443

Lisón, F. (2015). Murciélago hortelano meridional—*Eptesicus isabellinus*. In Salvador, A. & Barja, I. (Eds.), *Enciclopedia Virtual de los Vertebrados Españoles*. Museo Nacional de Ciencias Naturales.

Lisón, F., & Calvo, J. F. (2013). Ecological niche modelling of three pipistrellate bat species in semiarid Mediterranean landscapes. *Acta Oecologica, 47*, 68–73. https://doi.org/10.1016/j.actao.2013.01.002

Lisón, F., & Sánchez-Fernández, D. (2017). Low effectiveness of the Natura 2000 network in preventing land-use change in bat hotspots. *Biodiversity and Conservation, 26*(8), 1989–2006. https://doi.org/10.1007/s10531-017-1342-8

Lisón, F., Sánchez-Fernández, D., & Calvo, J. F. (2015). Are species listed in the Annex II of the Habitats Directive better represented in Natura 2000 network than the remaining species? A test using Spanish bats. *Biodiversity and Conservation, 24*(10), 2459–2473. https://doi.org/10.1007/s10531-015-0937-1

Liu, C., Berry, P. M., Dawson, T. P., & Pearson, R. G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography, 28*(3), 385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x

MacKenzie, D. I., Nichols, James D., Royle, J. Andrew, Pollock, Kenneth H., & Hines, James E. (2018). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence* (Second edition). Academic Press, an imprint of Elsevier.

Margules, C. R., Pressey, R. L., & Williams, P. H. (2002). Representing biodiversity: Data and procedures for identifying priority areas for conservation. *Journal of Biosciences, 27*(4), 309–326. https://doi.org/10.1007/BF02704962

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., & Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Biodiversity and Distributions, 15*(1), 59–69. https://doi.org/10.1111/j.1472-4642.2008.00491.x

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research, 16*(3), 285–292. https://doi.org/10.1016/0377-2217(84)90282-0

Monserud, R. A., & Leemans, R. (1992). Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling, 62*(4), 275–293. https://doi.org/10.1016/0304-3800(92)90003-W

Murphy, C. A., Grenouillet, G., & García-Berthou, E. (2015). Natural abiotic factors more than anthropogenic perturbation shape the invasion of Eastern Mosquitofish (*Gambusia holbrooki*). *Freshwater Science, 34*(3), 965–974. https://doi.org/10.1086/681948

Mutasa, S., Sun, S., & Ha, R. (2020). Understanding artificial intelligence based radiology studies: What is overfitting? *Clinical Imaging, 65*, 96–99. https://doi.org/10.1016/j.clinimag.2020.04.025

Nenzén, H. K., & Araújo, M. B. (2011). Choice of threshold alters projections of species range shifts under climate change. *Ecological Modelling, 222*(18), 3346–3354. https://doi.org/10.1016/j.ecolmodel.2011.07.011

Pärtel, M. (2014). Community ecology of absent species: Hidden and dark diversity. *Journal of Vegetation Science, 25*(5), 1154–1159. https://doi.org/10.1111/jvs.12169

Pärtel, M., Szava-Kovats, R., & Zobel, M. (2011). Dark diversity: Shedding light on absent species. *Trends in Ecology & Evolution, 26*(3), 124–128. https://doi.org/10.1016/j.tree.2010.12.004

Pearson, R. G., Thuiller, W., Araújo, M. B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T. P., & Lees, D. C. (2006). Model-based uncertainty in species range prediction. *Journal of Biogeography, 33*(10), 1704–1711. https://doi.org/10.1111/j.1365-2699.2006.01460.x
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 University Press.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling, 190*(3–4), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026

Rainho, A., Alves, P., Amorim, F., & Marques, J. T. (2013). *Atlas dos morcegos: De Portugal continental* (Instituto da Conservação da Natureza e das Florestas, Ed.).

Ripley, B. D. (2007). *Pattern recognition and neural networks*. Cambridge university press.

Ronk, A., de Bello, F., Fibich, P., & Pärtel, M. (2016). Large-scale dark diversity estimates: New perspectives with combined methods. *Ecology and Evolution, 6*(17), 6266–6281. https://doi.org/10.1002/ece3.2371

Santos, H., Juste, J., Ibáñez, C., Palmeirim, J. M., Godinho, R., Amorim, F., Alves, P., Costa, H., de Paz, O., Pérez-Suarez, G., Martínez-Alos, S., Jones, G., & Rebelo, H. (2014). Influences of ecology and biogeography on shaping the distributions of cryptic species: Three bat tales in Iberia: Shaping of Bat Cryptic Distribution in Iberia. *Biological Journal of the Linnean Society, 112*(1), 150–162. https://doi.org/10.1111/bij.12247

Stockwell, D. R., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling, 148*(1), 1–13.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography, 32*(3), 369–373. https://doi.org/10.1111/j.1600-0587.2008.05742.x

Wilkinson, D. P., Golding, N., Guillera-Arroita, G., Tingley, R., & McCarthy, M. A. (2019). A comparison of joint species distribution models for presence–absence data. *Methods in Ecology and Evolution, 10*(2), 198–211. https://doi.org/10.1111/2041-210X.13106

Yela, J. L., Zahiri, R., Wahlberg, N., Ronkay, L., & Zilli, A. (2011). Phylogenetic overview of Noctuidae sensu lato. *Noctuidae Europaeae, 13*, 17–22.

Zahiri, R., Kitching, I. J., Lafontaine, J. D., Mutanen, M., Kaila, L., Holloway, J. D., & Wahlberg, N. (2011). A new molecular phylogeny offers hope for a stable family level classification of the Noctuoidea (Lepidoptera). *Zoologica Scripta, 40*(2), 158–173.

### Tables

**Table 1. Summary of predictive performance by SDM technique and animal group.** AUC.TrD and AUC.TeD are the average AUC values for the model training and testing data, respectively. AUC.Ov is the techniques’ mean overfitting, calculated as the difference between AUC.TrD and AUC.TeD. Values in parentheses are the standard errors.

| SDM  | Bats       | Moths          |
|------|------------|----------------|
|      | AUC.TrD    | AUC.TeD        | AUC.Ov | AUC.TrD | AUC.TeD | AUC.Ov |
| ANN  | 0.77 (0.07) | 0.71 (0.07)    | 0.06 (0.05) | 0.8 (0.07) | 0.69 (0.09) | 0.11 (0.08) |
| CTA  | 0.73 (0.08) | 0.68 (0.08)    | 0.04 (0.04) | 0.67 (0.08) | 0.62 (0.09) | 0.06 (0.07) |
| FDA  | 0.84 (0.04) | 0.77 (0.05)    | 0.06 (0.05) | 0.81 (0.05) | 0.71 (0.08) | 0.1 (0.07) |
| GAM  | 0.85 (0.07) | 0.8 (0.07)     | 0.05 (0.07) | 0.94 (0.04) | 0.82 (0.07) | 0.11 (0.08) |
| GBM  | 0.91 (0.06) | 0.82 (0.06)    | 0.09 (0.05) | 0.97 (0.02) | 0.82 (0.08) | 0.15 (0.08) |
| GLM  | 0.82 (0.08) | 0.79 (0.08)    | 0.02 (0.03) | 0.89 (0.04) | 0.82 (0.07) | 0.06 (0.06) |
| M.Ph | 0.9 (0.07)  | 0.8 (0.07)     | 0.11 (0.05) | 0.94 (0.08) | 0.81 (0.09) | 0.14 (0.07) |
| RF   | 1 (0)      | 0.85 (0.06)    | 0.15 (0.06) | 1 (0)    | 0.85 (0.07) | 0.15 (0.07) |
| ECA  | 0.91 (0.04) | 0.84 (0.06)    | 0.07 (0.03) | 0.96 (0.02) | 0.89 (0.05) | 0.07 (0.04) |
| EMP  | 0.95 (0.03) | 0.84 (0.06)    | 0.12 (0.05) | 0.99 (0.01) | 0.87 (0.06) | 0.13 (0.05) |
| EWMP | 0.96 (0.03) | 0.84 (0.06)    | 0.12 (0.05) | 0.99 (0.01) | 0.87 (0.06) | 0.13 (0.05) |

**Table 2. Summary of SDM technique predictive performance by threshold type and animal group for cases that depend on a single threshold cutoff value.** Sen.TrD and Spe.TrD are the average sensitivity and specificity for the model training data, respectively, while Sen.TeD and Spe.TeD are the same for the
models’ testing data. Sen.O and Spe.O are the sensitivity and specificity overfitting, calculated as the difference between Sen.TrD and Sen.TeD, and Spe.TrD and Spe.TeD, respectively. Values in parentheses are the standard errors.
Table 1

|     | M.Ph  | RF    | ECA   | EMP   | EWMP  |
|-----|-------|-------|-------|-------|-------|
|     | (12)  | (0)   | (6.7) | (5.8) | (5.6) |
|     | 88.4  | 100   | 92.2  | 95.5  | 95.9  |
|     | (14.1)| (8.2) | (6.7) | (8.4) | (8.5) |
|     | 72.4  | 80.8  | 85.8  | 78.1  | 78.2  |
|     | (9.3) | (3.2) | (8.1) | (5.7) | (5.6) |
|     | 74.9  | 90.3  | 69.7  | 77.8  | 77.8  |
|     | (9.8) | (3.2) | (6.6) | (5.7) | (5.6) |
|     | 73.2  | 73.9  | 67.3  | 74.4  | 74.4  |
|     | (11.6)| (5.5) | (7)   | (6.8) | (6.7) |
|     | 16.1  | 19.2  | 6.4   | 17.4  | 17.6  |
|     | (1.9) | (8.2) | (4.9) | (8.4) | (8.5) |
|     | 1.7   | 16.4  | 2.4   | 3.3   | 3.5   |
|     | (1.8) | (6.3) | (2.1) | (2.2) | (2.2) |
|     | 911   | 100   | 99.4  | 99.8  | 99.8  |
|     | (16.8)| (0.5) | (1.5) | (1.3) | (1.8) |
|     | 65.1  | 76.7  | 92.3  | 76.1  | 76.3  |
|     | (19.5)| (12.5)| (7.4) | (11.7)| (11.6)|
|     | 82.7  | 89.3  | 68.6  | 82.7  | 82.7  |
|     | (7.2) | (1.6) | (7.9) | (4.9) | (4.9) |
|     | 82.3  | 79.7  | 66.2  | 81 (5.1)| 80.9 (5) |
|     | (7.2) | (3.7) | (8.1) | (5.1) | (5) |
|     | 25.9  | 23.2  | 7.2   | 23.7  | 23.5  |
|     | (14.6)| (12.5)| (7.2)| (11.6) | (11.6) |
|     | 0.4   | 9.6   | 2.5   | 1.8   | 1.8   |
|     | (1.2) | (2.6) | (2.4) | (1.7) | (1.7) |

Figures

Figure 1

Geographic distribution of estimated Animal Dark Diversity by each SDM technique for bats in the Iberian Peninsula using the MaxTSS threshold type (see Figs. S1, S2, S3, S4 and S5 for ADD estimates using the MaxKappa and MeanProb thresholds for both bats and moths).
Figure 2

Boxplot of estimated Animal Diversity sizes in each UTM cell of the Iberian Peninsula by SDM technique, threshold type and animal group.
Figure 3

Boxplots of the degree of correlation (p) between SDM technique estimates of Animal Dark Diversity in the Iberian Peninsula by threshold type and animal group. The dashed lines mark the boundary between low and high correlation (p = 0.7). The ensemble models are shown in purple.
Figure 4

Scatterplots of expected Animal Dark Diversity in a UTM cell and mean overfitting by SDM technique, threshold type and animal group. The dashed lines are the expected values of the relationship between the variables using simple linear regression.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SUPPLEMENTARYMATERIAL.docx