**Supplementary Material**

**Methods**

We cleaned community science occurrences by removing duplicate records, those with no georeferenced location, and any locations over water, resulting in the 33 occurrences to use in the calibration models. Additionally, community science data were checked for geolocation accuracy and any locations that were deemed biologically implausible due to being outside the species’ known habitat requirements were removed from the analysis. Only locations recorded from 1990 onwards were used to match the timeframe of the environmental raster data, whilst retaining sufficient sample size for modelling (van Proosdij *et al.* 2016). We omitted occurrences originating from the eBird database (*n =* 5, Sullivan *et al.* 2009) from the GBIF download because of uncertainty in location accuracy from the eBird checklist system.

We tested spatial clustering in our point locations using Nearest Neighbour Index (NNI) to evaluate the potential level of sampling bias. NNI is the ratio of the observed distance divided by the expected distance between neighbours in a hypothetical random distribution. NNI < 1 indicates spatial clustering, with values of 1 indicating random dispersion, and those closer to 2 indicating regular dispersion. Despite the advantages gained from applying spatial filtering to occurrence data in SDMs (Aiello-Lammens *et al.* 2015), we opted to retain all 33 cleaned records without spatial filtering because the effects of biased sampling were minimal in our dataset (NNI = 0.879, *z* = -1.33, *p =* 0.182). By retaining all records, we aimed to capture the widest possible range of species-habitat associations because of the low sample size of our occurrence data and the high environmental heterogeneity of the study area.

Continuous Boyce Index (CBI) is consistent with a Spearman correlation (*rs*) with CBI values ranging from -1 to +1, with positive values indicating predictions consistent with observed presences, values close to zero no different than a random model, and negative values indicating areas with frequent presences having low environmental suitability. CBI was calculated using 20 % test data with a moving window for threshold-independence and 101 defined bins in the R package enmSdm (Smith 2019). We Partial ROC ratios range from 0 – 2 with 1 indicating a random model. Function parameters were set with a 10 % omission error rate, and 1000 bootstrap replicates on 50 % test data to determine significant () pROC values >1.0 in the R package ENMGadgets (Barve & Barve, 2013).

**References**

Aiello‐Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B. & Anderson, R.P. (2015) spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography.* 38: 541-545.

Barve, N. & Barve, V. (2013). ENMGadgets: tools for pre and post processing in ENM workflows. https://github.com/narayanibarve/ENMGadgets.

Smith, A.B. (2019). enmSdm: Tools for modeling niches and distributions of species. R package v0.3.4.6. https://github.com/adamlilith/enmSdm/

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*. 142: 2282-2292.

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

**Supplementary Tables**

**Table S1.** Model parameters from the best supported candidate logistic regression habitat covariate model (#6) for the Madagascar Serpent-eagle using Akaike’s Information Criterion corrected for small sample sizes. Bold indicates significant coefficient estimates *p* < 0.01.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Parameter | β | SE | *z* | Lower CI | Upper CI |
| Intercept | **-4.054** | 1.738 | -2.332 | -8.325 | -1.592 |
| NDVI | 2.128 | 1.988 | 1.070 | -1.617 | 6.324 |
| Evergreen Forest | **1.802** | 0.654 | 2.757 | 0.666 | 3.289 |
| Heterogeneity | 1.004 | 0.584 | 1.722 | -0.046 | 2.304 |
| Leaf Area Index | -0.213 | 1.159 | -0.184 | -2.503 | 2.173 |

**Table S2.** Multi-collinearity test using stepwise elimination Variance Inflation Factor (VIF) analysis. Covariates with VIF < 2 have low correlation with other covariates, and thus are suitable for inclusion in calibration models when further evaluated for ecological relevance.

|  |  |  |  |
| --- | --- | --- | --- |
| Covariate | *R*2 | Tolerance | VIF |
| Heterogeneity | 0.476 | 0.524 | 1.909 |
| NDVI | 0.399 | 0.601 | 1.664 |
| Evergreen Forest | 0.274 | 0.726 | 1.377 |
| Leaf Area Index | 0.257 | 0.743 | 1.347 |

**Table S3.** Model selection metrics for all candidate models using all three feature classes from Akaike’s Information Criterion corrected for small sample sizes. RM = regularization multiplier (β), FC = feature classes, LQH = Linear, Quadratic, Hinge.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | RM | FC | AICc | ΔAICc |
| 15 | 3.0 | LQH | 755.971 | 0.335 |
| 18 | 3.5 | LQH | 756.899 | 1.263 |
| 27 | 5.0 | LQH | 757.504 | 1.869 |
| 12 | 2.5 | LQH | 757.813 | 2.177 |
| 21 | 4.0 | LQH | 757.936 | 2.300 |
| 24 | 4.5 | LQH | 759.075 | 3.439 |
| 6 | 1.5 | LQH | 761.538 | 5.902 |
| 9 | 2.0 | LQH | 764.654 | 9.018 |
| 3 | 1.0 | LQH | 776.307 | 20.671 |

**Supplementary Figures**

****

**Figure S1.** Habitat covariates used in the penalized Species Distribution Models for the Madagascar Serpent-eagle.

****

**Figure S2.** Beta coefficient paths for the optimal penalized logistic regression model where each curve corresponds to a covariate linear term. The paths of each coefficient term are plotted against the L1-norm (lasso or elastic net) of the whole coefficient vector as lambda (the amount defining the level of coefficient shrinkage)varies. The upper axis indicates the number of non-zero coefficients at the current lambda which is the effective degrees of freedom for the lasso or elastic net.

****

**Figure S3.** Beta coefficient paths for the optimal penalized logistic regression model where each curve corresponds to a covariate linear term. The paths of each coefficient term are plotted against the log-lambda of the whole coefficient vector as lambda (the amount defining the level of coefficient shrinkage)varies. Log-lambda on the y-axis indicates the log of the optimal value of lambda which minimizes the prediction error. This lambda value will give the most accurate model. The upper axis indicates the number of decreasing non-zero coefficients at the current lambda which is the effective degrees of freedom for the elastic net.

****

**Figure S4.** Beta coefficient paths for the optimal penalized logistic regression model where each curve corresponds to a covariate linear term The paths of each coefficient term are plotted against the fraction deviance explained on the training data. The upper axis indicates the number of non-zero coefficients at the current lambda which is the effective degrees of freedom for the elastic net.