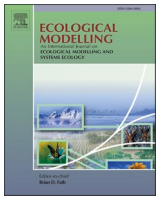
Ecological Modelling 493 (2024) 110749

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel



Differentially biased sampling strategies reveal the non-stationarity of species distribution models for Indian small felids

Divyashree Rana [a,*], Caroline Charão Sartor [b], Luca Chiaverini [b,c], Samuel Alan Cushman [b], Żaneta Kaszta [b,d], Uma Ramakrishnan [a], David W. Macdonald [b]

[a] National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560065, India
[b] Wildlife Conservation Research Unit, Department of Biology, University of Oxford, The Recanati-Kaplan Centre, Tubney House, Tubney, Oxon OX13 5QL, UK
[c] Rewilding Apennines, Via San Giorgio 5, 67055, Gioia dei Marsi (AQ), Italy
[d] Department of Biological Sciences, Northern Arizona University (NAU), Flagstaff, AZ 86011, United States of America

ARTICLE INFO

Keywords:
Species distribution modelling
Datatype
Sampling extent
Environmental space
Small cats

ABSTRACT

Species Distribution Models (SDMs) have been used extensively to understand species-habitat relationships and design conservation strategies. The ability to train these models using a wide variety of datasets and modelling algorithms has led to their wide applicability across systems. However, the ease of modelling also leads to their use as off-the-shelf models without a detailed investigation of the data and their suitable end-use application. The effect of various modelling parameters on inferences has been explored, however, their interaction with training data type is limited. We used country-wide data for four sympatric Indian small cat species to understand the sensitivity of SDMs to data types, sampling extents and their interaction. Our results reveal the non-stationarity of models with varying modelling parameters. The extent of the training dataset had major implications on the inferences and interacted strongly with the type of dataset used. The divergent distribution of the target species revealed that the effect of sampling extent was more pronounced for species that have limited distribution within the predictive extent. Lastly, our results highlight the significance of sampled environmental space in explaining the non-stationarity of the model outputs.

1. Introduction

Species Distribution Models (SDMs) are widely used powerful tools that find applications in understanding fundamental ecological relationships and predicting species occurrence, which can then help develop conservation strategies (Bellamy et al., 2013; Rana et al., 2022). By identifying the factors that drive organism occurrence, SDMs provide insights into the habitat niche and ecological factors limiting species distribution (Hegel et al. 2010), while contributing with means to identify important areas for conservation prioritisation (Penjor et al. 2021; Macdonald et al. 2018). These applications make them an important tool for modelling the distribution and habitat niche of rare species (Razgour et al., 2011), predicting the possible expanse of invasive species or the distribution of species under future climate and landscape change scenarios (Couce et al., 2013; Saranya et al., 2021). Such applications have resulted in extensive use of SDMs, with over 6000 studies using this tool over the past two decades (Araújo et al., 2019).

SDMs are models relating species' empirical observation with variables best predicting species presence, based either on statistically or theoretically derived response surfaces (Guisan and Zimmermann, 2000). Several ways exist to obtain empirical observation of species. The flexibility and wide applicability of SDMs can be attributed to their utilization of data from designed in-situ surveys as well as opportunistic data to train these models. Specifically, the ability to use opportunistic species occurrence data from the literature and public platforms like GBIF, eBird, and iNat makes these models relevant for multiple species and spaces. Alternatively, a widely used method to gain empirical data on species occurrence is camera trapping (Meek et al., 2014; Rovero et al., 2013). More recently, empirical data from diverse sources have also found application in Species Distribution Modelling. This includes species occurrence data from eDNA surveys (Carraro et al., 2018; Neto et al., 2020) and acoustic monitoring (Desjonquères et al., 2022; Frasier et al., 2021). The flexibility of input data and modelling techniques make SDMs an appealing tool for conservation. However, a strong association between data type, model output, and end-use applicability

* Corresponding author.
E-mail address: divyashreer@ncbs.res.in (D. Rana).

https://doi.org/10.1016/j.ecolmodel.2024.110749
Received 19 January 2024; Received in revised form 16 April 2024; Accepted 2 May 2024
Available online 11 May 2024

has been demonstrated, which is often overlooked (Guillera-Arroita et al., 2015). Additionally, the ease of modelling complex relationships poses the risk of poor or inappropriate outcomes when wrong or inadequate training data are used, or when the modelling technique fails to accurately depict the species-habitat relationships, leading to flawed patterns and understanding (Chiaverini et al., 2021, 2023).

Hence, despite the wide utility in conservation biology, like most mathematical models, SDMs can lead to interpretation errors, resulting in erroneous generalisations about the species biology (Collart et al., 2023; Lee-Yaw et al., 2022). Off-the-shelf use of SDMs without understanding the assumptions of the model and limitations of the input data runs a risk of inferring patterns describing the data, instead of the species' biology. There exist multiple guidelines and cautionary notes for using SDMs (Hijmans, 2012; Tessarolo et al., 2014; Walting et al., 2015; Sillero et al., 2021), however, their application in real ecological datasets is often poorly explored. Multiple studies have looked at the sensitivity of model output on modelling frameworks like the choice of pseudo-absence distribution (Barbet-Massin et al., 2012; Hazen et al., 2021; Senay et al. 2013; Hysen et al., 2022), spatial autocorrelation in occurence data (Crase et al., 2012; Miller, 2012; Václavík et al., 2012), biased sampling of environmental space (Phillips et al., 2009), and choice of validation metric (Konowalik and Nosol, 2021; Valavi et al., 2018). Studies have also found models highlighting broad-scale and fine-scale factors affecting species distribution based on sampling strategy and extent (Razgour et al., 2011). Despite multiple studies exploring different modelling frameworks, research aimed at understanding the influence of datatype (Guillera-Arroita et al., 2015) and congruence in the model outputs using independent datasets for multiple species is limited. Mostly, model performance is based on evaluating a random subset of the data. This internal cross-validation is often termed as independent validation, however, studies have emphasized the critical value of truly independent data for accurate validation (Lee-Yaw et al., 2022). Additionally, the model's sensitivity to the type of data and its interaction with the sampling extent of the data is unknown.

Here, we aim to understand the sensitivity of model performance and output to different data types and sampling extents using scale-optimised SDMs for four ecologically different felids found in India. Our target species are widespread in India but with differences in their ecological niche and geographic range within the country. Weighing less than 15 kg, these small cat species from the genus *Felis* (*F. chaus*) and *Prionailurus* (*P. viverrinus, P. rubiginosus, P. bengalenses*) are important mesopredators in different ecosystems in the country. Sympatric with each other, they often share space with larger, more charismatic species like tigers and elephants. Owing to their elusive nature, there have been limited studies on these species, with our understanding being mostly informed by bycatch data from camera traps targetting larger sympatric species (Srivatsha et al., 2015; Chatterjee et al. 2020). In such cases, SDMs driven by species records varied in space and time for predictive spatial modelling act as robust methodological tools to provide insights into the ecology of species and design dedicated conservation strategies. Hence, these species represent cases that generally require and utilize SDMs to assist in the conservation of rare and elusive species. This provides us with a unique opportunity to understand and comment on the sensitivity of these frequently used models to the training data and model parameters, prompting these questions: 1. How does sampling extent influence the model performance and outputs?; 2. Is congruence in model outputs from different datasets governed by overlap in their sampled environmental space?; 3. Are presence-background as well as presence-absence data equally sensitive to change in sampling extents?; 4. How do species distribution patterns within the predictive domain interact with sampling extents to affect model outputs? We predict sampling extent would significantly influence model performance and output. The congruence in model outputs was predicted to be governed by an overlap in their sampled environmental space, with a larger overlap in datasets leading to more congruent model outputs. However, models built using datasets consisting of randomly generated pseudo-absences were predicted to showcase higher sensitivity to sampling extent changes, more so for range-restricted species. Hence, utilizing the natural differences in species' distribution within the sampled and predicted geographical space, the study would allow an understanding of the differential effect of sampling extent and data type on these sympatric small cat species.

2. Materials and methods

2.1. Species and occurrence datasets

We focused on four species of sympatric small cats found in India - jungle cat (JC, *Felis chaus*), rusty-spotted cat (RSC, *Prionailurus rubiginosus*), fishing cat (FC, *Prionailurus viverrinus*), and leopard cat (LC, *Prionailurus bengalensis*). With similar ecological roles, these sympatric mesopredators share similar dispersal distances and home range sizes. JC and RSC have wide distribution, covering most of the geographical area of the country, as they are associated with dry and open habitats (Sunquist and Sunquist, 2002). Given their association with open habitats, they are also found in modified landscapes including agricultural fields. LC and FC, on the other hand, show a strong association with closed-canopy forests and wetlands, respectively, leading to a smaller range across India (Sunquist and Sunquist, 2002). As the majority of the Indian peninsula is dry with natural and modified open landscapes, the distribution of LC and FC is patchy and towards the extremities of the country compared to the widely distributed JC and RSC. Despite differences in their distribution, the sampling extent of the study covers the entire or a significant portion of the global range of each species.

We collected two independent datasets of species records from published resources. The first dataset consisted of detection/non-detection (bycatch) data from the All India Tiger Estimation monitoring survey carried out in 2018–19 across the tiger reserves in India (Jhala et al., 2020). Details about the camera trap survey design including grid size, distance between trap stations, and number of trap nights can be found in Jhala et al. (2020). As the focal species of this study are sympatric with tigers across many parts of the country, the by-catch data from these surveys provided extensive records for small cats. These data were gained from a total of 14,147 trap stations with varying species detection at each station ($n = 3732$ JC, $n = 655$ RSC, $n = 623$ FC, $n = 1003$ LC). This data with detection and non-detections across the country is henceforth referred to as **Presense-Absence (PA)** data.

The second dataset comprises species-specific presence records from published literature and open repositories. A systematic search of published articles and reports was conducted on Google Scholar with combinations of keywords "Species name" (both common and scientific name for each species) and "India", and the first 100 results were extracted for further processing. Exact coordinates of species detection were extracted from studies provided either in-text or in geo-referenced figures, including tables and maps. Additionally, species record data from the citizen science platforms "Global Biodiversity Information Facility" and the "India Biodiversity Portal" were included. All these presence records ($n = 351$ JC, $n = 254$ RSC, $n = 659$ FC, $n = 96$ LC) were limited to the political extent of India. Compared to PA data which was collected from systematic surveys, this data is opportunistic and could entail inherent sampling bias. Spatial sampling bias correction approaches are numerous, but context-specific (Baker et al., 2024). Hence, these corrections were not incorporated into the processing of these datasets to keep them comparable for understanding the effect of sampling extent. Additionally, for training SDMs, pseudo-absences of ten times the occurrence records were randomly generated around these secondary presence records for each species with a minimum distance of 4000 m between points using the 'Create Random Points' tool in ArcMap10.8 (ESRI, 2018). The proportion of presence/pseudo-absences points was decided based on previous studies that demonstrated that

large samples of spatially random pseudo-absences are known to yield the most reliable model predictions, especially in the case of logistic regression models (Barbet-Massin et al., 2012; Iturbide et al. 2015). The minimum distance between pseudo-absence points corresponds to the largest scale for covariate calculation (explained below) and the median home range size for targeted small cats. This data, consisting of presence records and background points, is hereafter referred to as **Presence-Background (PB)** data.

For each species, datasets were compiled at two geographic extents: (i.) Within the political boundary of the country, hereafter referred to as "**India**" extent; and (ii.) Focusing on the specific range of each species within the country, hereafter referred to as "**Range**" extent. As presence points were invariably inside the species range, only absence and pseudo-absence data varied with changes in the geographical extent.

2.2. Landscape variables

To understand the habitat suitability of different target species, landscape variables were collated from multiple openly available datasets. Guided by previous studies on felids in Southeast Asia, the selected variables ($n = 23$, Table 1) included indicators of geomorphological, anthropogenic, biological, and climatic features within the study area (Chiaverini et al., 2022; Chiaverini et al., 2023; Macdonald et al., 2018). The variables with no values for over 90 % of the species occurrence sites were dropped. The retained list of variables used in the study is given in Table 1. All layers were reprojected to 500 m resolution for downstream processing. The species-environment interactions are known to be scale-dependent (Levin, 1992). Hence, the effect of each covariate was assessed at eight spatial scales: 500 m, 1000 m, 1500 m, 2000 m, 2500 m, 3000 m, 3500 m and 4000 m. For this multiscale assessment, each layer was recalculated for the focal mean value in a circular moving window for each scale using the 'Focal statistics' tool in ArcMap10.8, resulting in eight rescaled rasters for each predictor variable.

2.3. Modelling framework

We built independent SDMs for each species and dataset as described here briefly. Each of the four target species had a combination of two independent datasets (PA and PB) clipped to two geographical extents (India and Range). Hence, each species had four datasets which were used to build four species distribution models. For each dataset, the optimal set of predictors was selected based on univariate scale selection followed by the removal of correlated variables. With an optimal set of scale-selected uncorrelated variables, species distribution was modelled using Generalized Linear Models (GLMs) with binomial error structures. The details for each step of the modelling framework have been described below.

2.3.1. Scale optimization and covariate selection

Species select habitat features at different spatial scales and therefore we implemented a multiscale model optimisation (Macdonald et al., 2018; Sarkar et al. 2018; Vergara et al., 2016; Whitenack et al., 2023). The covariate data for each occurrence dataset were generated by extracting raster values at each species occurrence site. We performed univariate binomial regression for every dataset to identify the scale at which each variable showed the strongest effect on each species. For each predictor variable, the scale with the lowest Akaike Information Criterion (AIC) value was retained for the multicollinearity screening step. The multicollinearity between scale-selected covariates was checked using Pearson's correlation coefficient. If two variables were found to be correlated ($r > |0.7|$), the variable with the lowest AIC for univariate GLMs was retained. The above-mentioned scale optimization and selection were performed based on the methods described in Chiaverini et al. (2023).

Table 1
List of environmental variables used to model species distributions.

S. No.	Category	Layer name	Resolution (m)	Source
1	Geomorphological	Elevation	90	https://srtm.csi.cgiar.org/
2		Topographic position index	90	https://doi.org/10.1038/s41597-020-0479-6
3		Compound topographic index		
4		Roughness		
5		Soil organic carbon	250	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169748
6		Bulk density		
7		Cation exchange capacity		
8		Clay		
9		Sand		
10		Nitrogen		
11		Average annual soil water content	1000	https://cgiarcsi.community/data/global-high-resolution-soil-water-balance/
12		Surface water percent	30	https://doi.org/10.1016/j.rse.2020.111792
13	Anthropogenic	Human population density	100	www.worldpop.org
14		Global modification index	1000	doi:10.1111/gcb.14549
15	Biological	Normalized difference vegetation index	250	MOD13Q1.061 Terra Vegetation Indices 16-Day Global 250m
16		Spectral variability vegetation index svvi	500	MOD09A1.061 Terra Surface Reflectance 8-Day Global 500m
17		Gross primary production	500	MOD17A2H.061: Terra Gross Primary Productivity 8-Day Global 500m
18		Percentage tree cover	30	https://glad.earthengine.app/view/global-forest-change
19		Tree landcover class	100	doi:10.5281/zenodo.5571936
20	Climatic	Mean annual temperature	1000	https://doi.org/10.1038/sdata.2017.122
21		Temperature seasonality		
22		Annual precipitation		
23		Precipitation seasonality		

2.3.2. Species distribution models

The distribution for each species was modelled using GLMs with a binomial error structure (logistic regression). Multiple algorithms have been used in SDMs, but strong evidence indicating the strength of GLMs in similar species and ecosystems exists (Chiaverini et al. 2023; Cushman et al. in press). The four datasets (combination of data type - PA and PB, and geographic extent - India and Range) per species, as described above, were used to independently model the species distribution for each of the target small cat species. Each dataset was subsetted randomly, of which 80 % of the presences and absences/pseudo-absences were used to train the models and 20 % was

retained for subsequent model validation. An appropriate scale for each variable was selected, followed by removing correlated variables. Using the training data and scale-selected uncorrelated predictor variables, multivariate GLMs were generated to build species-specific SDMs. The approach generated multiple models, and the best model, with the best set of variables, was selected based on the lowest AIC value. Altogether, for each species, four best models were selected, i.e., one for each dataset.

2.4. Model prediction and validation

Each model was validated twice - (i.) Self-validation (validation using the 20 % test data from the same dataset; for example, a model built with PA training data validated using PA test data), and (ii.) Cross-validation (validation using the 20 % test data from the independent dataset; for example, the model built with PA training data validated using PB test data). Validation was based on the Area Under the ROC Curve (AUC), True Skill Statistic (TSS), and Cohen's Kappa statistics for each model.

Each model yielded a subsequent species distribution prediction. These predictions, independent of the geographic extent of the training dataset, were produced for the entire country. An ensemble prediction was also generated for each species using an AUC-weighted average of predictions from the best models selected with PA and PB datasets, at each geographic extent. Hence, in total, six species distribution predictions were generated for each species, using the PA training dataset (PA model), the PB training dataset (PB model), and their AUC-weighted average (average) for two geographic extents (India and Range).

2.5. Diagnostic analysis

As each species had multiple models predicting habitat suitability within the country, diagnostic analysis was conducted to explore differences among environmental variation across the datasets to potentially explain variation among the model predictions. According to our second prediction, differences in the sampled environmental space by different datasets would lead to contrasting model predictions for the same species. For exploring these differences we used Principal Component Analysis (PCA) to describe differences in the sampled environmental space of each dataset and explore its impact on the AUC validation metrics. We calculated Minimum Convex Polygons (MCP) to estimate the spread of the data on the first two PC axes. The area within the MCP was compared across different datasets for each species. Additionally, we used niche overlap metrics - Schoener's D and Hellinger's I which allow the quantification of environmental spread incorporating both occurrence and background data (Warren et al., 2008). Lastly, to compare model predictions, Pearson's correlation coefficients were calculated across the pixels of the prediction maps for the different models for each species. Model comparisons using MCP and Pearson's correlation were conducted to compare (i.) Dataset types - PA vs. PB, and (ii.) Sampling extents - India vs. Range. In the case of the comparison between sampling extents, the correlation was calculated for the entire predicted rasters (within the extent of India) and also restricted to each species range within India.

3. Results

3.1. Species distribution models, validation, and predictions

The selected models were self and cross-validated and their validation metrics (AUC, TSS and Kappa) are presented in Table 2. The metrics were found to be correlated ($r > 0.7$), indicative of a better fit of the models based on AUC metrics compared to TSS and Kappa. Given their correlation, results and inferences are based on a single widely used metric, the AUC. Self-validation of the models showed, on average, 1.3 times higher performance than cross-validation for all species at both geographic extents. For all models, irrespective of species and sampling extent, the cross-validation values were very low for both types of datasets, with average scores of AUC = 0.6. In comparison, model performance was higher using self-validation both for PA and PB, with an average of 0.81 and 0.76, respectively. The same patterns were found for all species, with self-validation performing better than cross-validation with higher AUC values across datasets.

Table 2
Validation metrics for best models for each species with different sampling extent. Each model was validated using 20 % test data from the same dataset (self validation) and independent dataset (cross validation).

Species	Extent	Model	Validation	AUC	TSS	Kappa
JC	India	PA	Self	0.81	0.47	0.42
		PA	Cross	0.53	0.09	0.05
		PB	Self	0.71	0.28	0.23
		PB	Cross	0.65	0.25	0.2
	Range	PA	Self	0.81	0.47	0.42
		PA	Cross	0.52	0.08	0.03
		PB	Self	0.71	0.36	0.23
		PB	Cross	0.65	0.2	0.21
RSC	India	PA	Self	0.75	0.38	0.14
		PA	Cross	0.62	0	0
		PB	Self	0.9	0.54	0.51
		PB	Cross	0.59	0.14	0.03
	Range	PA	Self	0.73	0.33	0.19
		PA	Cross	0.6	0	0
		PB	Self	0.89	0.55	0.59
		PB	Cross	0.56	0.12	0.03
FC	India	PA	Self	0.9	0.03	0.01
		PA	Cross	0.71	0	0
		PB	Self	0.6	0.45	0.15
		PB	Cross	0.66	0.47	0.08
	Range	PA	Self	0.83	0.39	0.4
		PA	Cross	0.65	0.17	0.16
		PB	Self	0.89	0.6	0.5
		PB	Cross	0.54	0.21	0.1
LC	India	PA	Self	0.8	0.5	0.23
		PA	Cross	0.72	0.43	0.21
		PB	Self	0.68	0.21	0.05
		PB	Cross	0.66	0.14	0.1
	Range	PA	Self	0.88	0.52	0.44
		PA	Cross	0.5	0.15	0.05
		PB	Self	0.72	0.55	0.28
		PB	Cross	0.56	0.08	0.09

The model performance also varied across sampling extent for different dataset types. However, the effect of the sampling extent was not consistent across the species. Comparing between sampling extents, India and Range, AUC values were fairly similar for the widely distributed species, JC and RSC, but presented notable differences for FC and LC (Table 2). More specifically, for FC and LC, the AUC values were usually higher when species were modelled within their range, except in the case of FC, which presented a higher AUC when modelled within India using the PA dataset. The differences in AUC self-validation values of models for JC and RSC (mean = 0.009, sd = 0.0075) were lower than range-restricted species FC and LC (mean = 0.12, sd = 0.11) when compared across the two sampling extents.

Irrespective of the sampling extent of the training data, self-validation of models built using PA data showed higher AUC values in most of the cases. However, models built using PB data had higher AUC values for FC at the range extent and for RSC at both geographic extents. Similar to the effect of sampling extent, the difference between performance using different data types - PA vs. PB, was more pronounced for range-restricted species - FC and LC (mean = 0.16, sd = 0.1), when compared to widely distributed species JC and RSC (mean = 0.12, sd = 0.03). In general, the India-extent models had better model performance for JC and RSC while the Range-extent models performed better for LC and FC for either of the datasets. Differences were more pronounced when comparing model performance between data types than between extents.

The best models for each dataset were used to predict suitable habitats for the species at the extent of the country. The predictions for each species varied across sampling extents and datasets. These differences have been showcased with the example of the species with the highest incongruency, the FC (Fig. 1).

3.2. Diagnostic analysis

3.2.1. Comparing niche overlap

The proportion of overlapping areas of MCPs and niche overlap metrics - Schoener's D and Hellinger's I were used as a metric to compare shared covariate space between different datasets. On average, a high overlap (75 %) was observed in the sampled environmental space of the same dataset clipped to different sampling extents. However, the proportion of overlap was found to be different across species, with JC presenting the highest overlap (0.93) and FC the lowest (0.46), with RSC (0.84) and LC (0.77) in between the extremes (Figure 2A; Supplementary Material, Figure C1-C4). The overlap was lower for PB datasets (Fig. 2A - blue bars, Figure S1-S4 Supplementary Material) across all species when compared to the PA datasets (Fig. 2A - red bars). This difference in extent overlaps between data types was higher for FC (0.61 for PA, 0.32 for PB) than for any other species. Similar trends were observed using the niche overlap metrics, with high overlap in sampled environmental space across different sampling extents (Schoener's D (average = 0.81), Hellinger's I (average = 0.93)). FC showed the least overlap amongst all the species, similar to MCP proportion overlap. However, the significant difference between PA and PB datasets was not observed.

Additionally, when comparing across independent datasets, on average, there was less than 50 % overlap in the sampled covariate space used to train respective models (Fig. 2B, Figure S1-S4 Supplementary Material). The mean overlap between datasets across species was 0.44. However, the overlap varied between presence and absence/pseudo-absence points. Overlap in the environmental space sampled by the presence points was lower than absence/pseudo-absence, with only one species (RSC = 0.65) showcasing overlap higher than the mean. Contrastingly, the overlap between presence points was extremely low for FC (0.21) and LC (0.16) (Fig. 2B). Niche overlap metrics also revealed low overlap between PA and PB datasets (Schoener's D (average = 0.4), Hellinger's I (average = 0.61)), similar to the MCP proportion overlap.

3.2.2. Comparing SDM predictions

Similar to the AUC model performance values, models built with the varying geographical extent of training data yielded starkly different predictions. These differences were most pronounced for the range-restricted species, FC and LC (Table 3). Comparing datasets with different sampling extents, JC and RSC showed a high correlation ($r > 0.76$) between predictions, whereas FC and LC showed a low correlation ($r < 0.63$) (Table 3). However, this was not the case for LC when using the PA dataset, in which the correlation values were considerably high ($r \geq 0.8$). Additionally, the correlation between rasters from models trained with different sampling extents was similar when considering the predicted values across the whole country or clipping to the species extent.

Prediction rasters obtained from independent datasets of the same species showed a very weak correlation ($-0.17 < r < 0.42$) with each other (Table 4). These weakly correlated predictions were shared across species, irrespective of their distribution patterns. Finally, when comparing the prediction of the different datasets with the averaged models, FC and LC presented high correlation values for both datasets regardless of the extent. However, this was not seen for the other two species. For both extents, JC presented a high correlation between the model produced with the PA dataset and the average ($r > 0.92$), but a low correlation between the PB model and the average ($r < 0.51$), while for RSC the opposite was found ($r < 0.45$ for PA and $r > 0.86$ for PB).

4. Discussion

The divergent distributions of the different study species within the country provided us with a natural experiment to explore the effects of

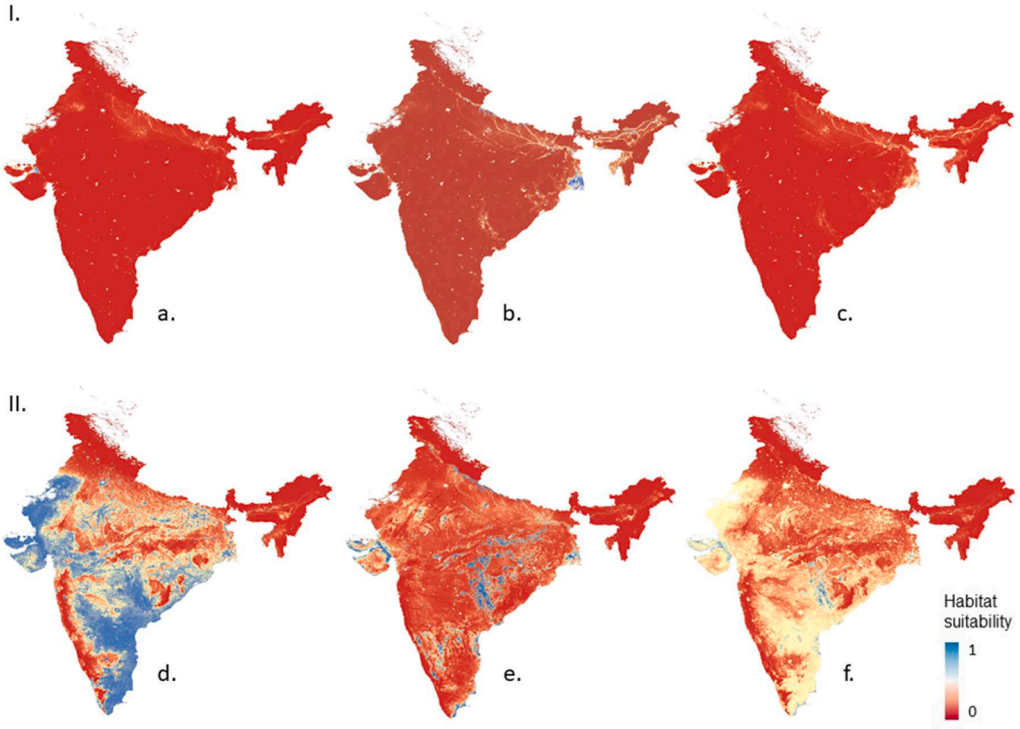


Fig. 1. Species distribution models of the fishing cat in India using models with different sampling extent (I. India extent, II. Range extent) and datasets (a and d - PA, b and e - PB, c and f - average).

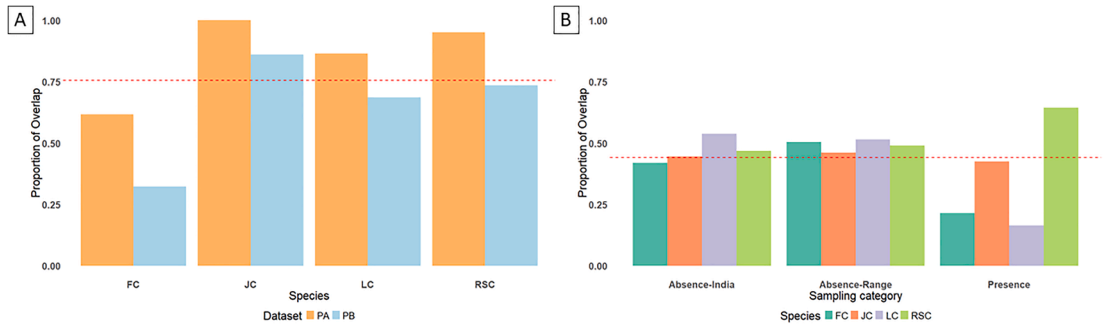


Fig. 2. The proportion of overlap between sampled environmental space used to train models with distinct A. Sampling extents - India and Range, B. Datasets - PA and PB data. The blue dotted line represents the mean value across datasets.

Table 3
Correlation between predicted rasters comparing sampling extents - India and Range. Correlations were done for the raster pixels within India or clipped to respective species extent.

Between different extents	PA				PB			
	JC	RSC	FC	LC	JC	RSC	FC	LC
Within India	1	0.8	0.04	0.9	0.91	0.89	0.32	0.56
Within Range	1	0.76	0.13	0.8	0.93	0.95	0.63	0.39

Table 4
Correlation comparing predictions from independent datasets - PA, PO, and average.

Between different datatypes	India				Range			
	JC	RSC	FC	LC	JC	RSC	FC	LC
PA - PB	0.19	-0.06	0.39	0.37	0.03	-0.17	0.42	0.33
PA - average	0.94	0.45	0.72	0.83	0.92	0.32	0.78	0.76
PB - average	0.51	0.86	0.92	0.82	0.41	0.87	0.89	0.86

range extent, type of sampling, and the prediction of species distribution models. Hence, with two independent datasets for multiple. species across a large geographical space, this study explores how training data type (PB vs PA) interacts with the study domain (i.e. India vs Range) and affects model performance. We found that the sampling extent of the dataset has critical implications for model performance and predictions and interacts strongly with the kind of training dataset used. This effect was more prominent in Presence-Background models, which are composed of occurrence records and pseudo-absences data. The study was unique in comparison across different small cat species with varying distribution patterns across the predictive space. The effect of sampling extent and datatype was more conspicuous for range-restricted species like FC and LC as compared to wide-ranging JC and RSC. Lastly, we found sampled environmental space to be non-overlapping across datasets, explaining the witnessed incongruence in predictions.

Owing to their power to discern species-environment associations, SDMs find wide applicability in ecological and conservation science. Despite the rapid proliferation of modelling approaches, SDMs are often used as off-the-shelf models without consideration of the modelling framework and training data characteristics. Considerable research has focused on the effects of multiple-scale optimization methods (McGarigal et al. 2016; Chiaverini et al. 2022), background data selection, and spatial autocorrelation on SDMs outputs and their inferences. However, limited studies compare model predictions and performance for the same species using independent datasets (Bradter et al., 2018; Henckel et al., 2020; Mair et al., 2017; Pruhsmeier et al., 2021). Some studies showed comparable results from independent data sources while others found high incongruency in the resulting occurrence patterns. Despite evidence of the sensitivity of species distribution models to training data characteristics (e.g. PB vs PA) and predictive domain, there have been very few analytical comparisons of the factors driving these outputs. Additionally, to our knowledge, the sensitivity of the model predictions to data type and its interaction with geographical training extent has not been explored.

4.1. Sampling extent

In line with our first prediction, sampling extent (i.e., in the entire country or only within the species range distribution within the country), had a notable impact on model performance. However, the impact of sampling extent varied across species. The effect of varying extents was more prominent on the range-restricted species (FC and LC) as inferred based on both low model validation and correlation between predictions. However, the AUC values show little change with distinct sampling extent for widely distributed species across the country (JC and RSC). Although AUC values are shown to be sensitive to sampling extents (Lobo et al., 2008), the lack of unidirectional trends of increasing AUC at larger extents across datasets, suggests underlying species-specific differences. As suggested by our fourth prediction, species distribution within the prediction extents seems to play a role in differing model outputs.

Since JC and RSC are adapted to the dry landscapes of the Indian subcontinent (Sunquist and Sunquist, 2002), they have a wide distribution across the country, inhabiting multiple habitat types. Thus, the absences and pseudo-absences sampled across India were environmentally similar to the ones sampled within the species range, resulting in more similar models. Conversely, LC and FC have restricted distributions in the country with disjunct populations which are found in narrow patches towards the extremities of the country with specific environmental conditions. In these cases, using the entire country to sample the absences and pseudo-absences resulted in sampling environmentally distinct areas, generating significantly different models, with the PB models that could have overfitted absences. The difference in model performance for range-restricted species was mirrored by a weak correlation from predictions from different models for the same species. The stark differences in model predictions were most prominent for FC (Fig. 1), which has the smallest and most fragmented distribution within the country.

Multiple studies have demonstrated similar effects of sampling extent affecting both model validation as well as model predictions (Anderson and Raza, 2010; El-Gabbas and Dormann, 2018; Vasquez et al., 2021). Despite the evidence, the choice of extent of the background data sampling as well as the predictive domain is often arbitrary. Although some studies use biological limits like known occurrences for a species or a group of species (Tôrres et al., 2012; Bellamy et al., 2013), most studies use geopolitical limits to define their sampling extents (Henckel et al., 2020; Collart et al., 2023; Da Re et al., 2023). Our study comparing biological and geopolitical extents for each species highlights the significance of a seldom overlooked characteristic of the training

dataset, the sampling extent. Our results suggest that the choice of sampling extent becomes more critical in cases where the niche of the species is much narrower than the environmental gradient present within the predictive geopolitical extent.

4.2. Data type differences

Validation of the SDM models is based on testing the model outputs using an independent partition of the available dataset. Here, along with using the usual approach of an independent partition of the training dataset (self-validation), we also used a partition of a truly independent dataset of the same species to test the model performance (cross-validation). Interestingly, despite showing high model predictive performance based on self-validation, cross-validation was found to be poor (Table 2). All species presented a similar amount of overlap (~50 %) of the environmental space among absence and pseudo-absence points indicating that, overall both PA and PB sampling strategies successfully sampled the background data. However, the environmental overlap of input presence points varied greatly across species, indicating different sampling strategies sampled unique and often non-overlapping niches of species biology. As the niche overlap metrics, Schoener's D and Hellinger's I, take into account both the presence and background data, differences across species weren't observed. However, FC and LC presented the lowest overlap in the environmental space of presence points when comparing the data types, with less than 25 % overlap. JC, however, had ~40 % overlap of presence points between data types and RSC ~60 %. Still, despite this striking difference, the cross-validation found between the models produced by distinct data types was similar to all species. This is contrary to our prediction, where we hypothesized that cross-validation would be higher when the datasets presented higher overlap in the sampled environmental space. Our results suggest, that data types comprising different environmental extents, shall produce distinct models, regardless of the magnitude of difference between the datasets. Specifically, the PA dataset is from camera traps placed systematically across the country providing better geographic coverage but limited to protected areas. The PB dataset, on the other hand, is a collation of published records on the species occurrence which could be from outside protected areas as well. Habitat type in terms of disturbance experienced has been shown to affect predicted distributions, with more stable habitats leading to higher SDM accuracy compared to regularly altered habitats (Collart et al., 2023; Marshall et al., 2015). This can be extended to partly explain poor cross-validation across our datasets, as one of them consists of species occurrence restricted to protected areas that offer more stable habitats contrary to the other dataset.

High self-validation and poor cross-validation indicate that models have a high predictive ability to explain the data they have been trained with, but less power to explain the ecological niche of the species. In other words, the inferences from these SDMs are limited in their scope to reveal general ecological patterns for the species. These results are further strengthened by the very low or even negative correlation values estimated between the predictions, demonstrating that the models generated from both data types are remarkably different. This highlights the importance of uniformly sampling species' habitats and considering this when predicting the models outside the sampled area. Underlining the significance of capturing environmental gradient, studies have recommended uniformly sampling environmental space instead of geographical space for generating pseudo-absence or background data (Varela et al., 2014; Hattab et al., 2017; Chiaverini et al., 2021; Perret and Sax, 2022; Da Re et al., 2023).

The limited studies comparing SDM predictions from multiple truly independent datasets have demonstrated mixed evidence for cross-validation across datasets (Gaulke et al., 2023; Whitenack et al., 2023). Summarising patterns, a recent comprehensive review of studies comparing predictions from independent occurrence datasets demonstrated that only 50 % of the studies show support for the accurate predictive performance of SDMs when validated with independent datasets (Lee-Yaw et al., 2022). Interestingly, most studies that found comparable results using independent datasets are focused on avian taxa, and aimed to understand the importance of opportunistically collected citizen science data when compared to systematic survey data. The growing number of bird enthusiasts and ease of bird sightings as compared to mammals or reptiles could be a contributing factor to the widespread coverage of opportunistic data yielding patterns similar to survey data. Despite limited comparative studies and contrasting patterns, even fewer studies explore the factors leading to matching or contrasting predictions from the SDMs.

4.3. Interaction of study extent and data type

Although sampling extent had an impact on model performance across species and datasets, this impact was more pronounced for presence-background datasets and range-restricted species (Table 2). This is in accordance with our third hypothesis, stating that the effect of the study extent would be stronger on the PB dataset. As algorithms used to produce SDMs require presence as well as absence data to train suitability models; random pseudo-absences or background data were generated for the PB dataset. As these absences were generated randomly within the species distributions or the boundaries of India, the sampling extent used in Presence-background modelling has a major impact on the ecological conditions sampled by the training data. This was observed in our analyses, where a higher difference in AUC values was observed in PB data when changing the sampling extent of the dataset as compared to PA data with fixed or pre-defined absences (Table 2). Similarly, when comparing across extents, model predictions obtained using PB data were on average less correlated than those obtained using PA datasets (Table 3). The sensitivity of the PB dataset and model to sampling characteristics has been documented in the literature (VanDerWal et al., 2009; El-Gabbas and Dormann, 2018). Additionally, highlighting the significance of sampling extent, studies have shown extrapolative prediction outside the training environmental niche space leading to erroneous predictions (e.g., Thuiller et al., 2004, Yates et al., 2018). Similarly, our results also highlight the sensitivity of the PB datasets to changes in modelling parameters.

Studies have shown that species traits like habitat specialization as well as habitat characteristics like habitat type and disturbance affect model performance (Collart et al., 2023; Marshall et al., 2015; Regos et al., 2019). Model predictions using either dataset for JC and RSC were highly correlated between different sampling extents, whereas the opposite trend was observed for FC and LC (Table 3). This pattern was mirrored by the sampled covariate environmental space where LC and FC had low overlap with changing sampling extents, specifically for PB datasets (Fig. 2A). Overall, a high average overlap of 70 % was observed when comparing sampled environmental space across different sampling extents. On the contrary, when comparing across data types, on average the overlap in sampled environmental space was less than 50 % (Fig. 2B). This low overlap correlates with poor model cross-validation as well as weak correlation between model predictions for the same species using independent datasets (Table 4). This indicates that different datasets are capturing unique variations regarding the biology of the species, hence leading to different predictions and low cross-validation scores. For wide-ranging species, with varying environmental conditions, biased sampling could result in non-stationarity in ecological models. Previous studies specifically in the field of landscape genetics have demonstrated that the strength of the species-habitat relationship could vary based on the sampled area within the species range (Short Bull et al., 2011; Vergara et al., 2017; Kaszta et al., 2021). This non-stationarity would result in variable signatures discerned by models built using different datasets, specifically in cases like ours with non-overlapping sampled environmental space. We expand these results by showing that not only the area sampled by differing datasets but also the study extent can affect our understanding

of the environmental limiting factors and ecological niche of a species. Differential species range within the country and the proportion of sampled environmental niches seem to play an important role in determining the predictive power of modelling their distributions.

5. Conclusions

With the growing applications of SDMs in understanding species biology and informing conservation actions, critical assessment of the predictive accuracy of the models is essential. To our knowledge, the exploration of the effect of datatypes on the modelling framework on model output is limited. By comparing independent datasets for sympatric small cat species, our study found evidence for high-performing models showing poor correlation across independent datasets for the same species. Our results highlight the role of sampled environmental space in explaining the incongruencies between model outputs. The low overlap between sampled covariate spaces across datasets explains poor cross-validation. This finding stresses that the one-fit-all modelling approach could lead to incorrect inferences. Adequate exploration of data type and species distribution within the study extent is critical. These considerations are especially important when dealing with presence-background data and species with restricted distributions. Using independent data on multiple species, our study underscores the importance of data exploration in light of sampling strategy while ensuring adequate sampling of the predictive environmental domain.

CRediT authorship contribution statement

Divyashree Rana: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Caroline Charão Sartor:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Conceptualization. **Luca Chiaverini:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Formal analysis, Conceptualization. **Samuel Alan Cushman:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Formal analysis, Conceptualization. **Żaneta Kaszta:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Uma Ramakrishnan:** Writing – review & editing, Supervision. **David W. Macdonald:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

DR's time at WildCRU and the involvement of WildCRU personnel in this research were funded by a grant to DWM from the Robertson Foundation. DR would specifically like to thank Imran for his critical comments on the draft.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2024.110749.

References

A Lee-Yaw, J., L McCune, J., Pironon, S., N Sheth, S., 2022. Species distribution models rarely predict the biology of real populations. Ecography 6, e05877.
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. J. Biogeogr. 37 (7), 1378–1393.
Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Rahbek, C, 2019. Standards for distribution models in biodiversity assessments. Sci. Adv. 5 (1), eaat4858.
Baker, D.J., Maclean, I.M., Gaston, K.J., 2024. Effective strategies for correcting spatial sampling bias in species distribution models without independent test data. Divers. Distrib. 30 (3), e13802.
Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W., 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods Ecol. Evol. 3, 327–338.
Bellamy, C., Scott, C., Altringham, J., 2013. Multiscale, presence-only habitat suitability models: fine-resolution maps for eight bat species. J. Appl. Ecol. 50 (4), 892–901.
Bradter, U., Mair, L., Jönsson, M., Knape, J., Singer, A., Snäll, T., 2018. Can opportunistically collected Citizen Science data fill a data gap for habitat suitability models of less common species? Methods Ecol. Evol. 9 (7), 1667–1678.
Carraro, L., Hartikainen, H., Jokela, J., Bertuzzo, E., Rinaldo, A., 2018. Estimating species distribution and abundance in river networks using environmental DNA. Proc. Natl Acad. Sci. 115 (46), 11724–11729.
Chatterjee, N., Nigam, P., Habib, B., 2020. Population density and habitat use of two sympatric small cats in a central Indian reserve. PLoS ONE 15 (6), e0233569.
Chiaverini, L., Wan, H.Y., Hahn, B., Cilimburg, A., Wasserman, T.N., Cushman, S.A., 2021. Effects of non-representative sampling design on multi-scale habitat models: flammulated owls in the Rocky Mountains. Ecol. Modell. 450, 109566.
Chiaverini, L., Macdonald, D.W., Bothwell, H.M., Hearn, A.J., Cheyne, S.M., Haidir, I., Cushman, S.A., 2022. Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands. Anim. Conserv. 25 (5), 660–679.
Chiaverini, L., Macdonald, D.W., Hearn, A.J., Kaszta, Ż., Ash, E., Bothwell, H.M., Cushman, S.A., 2023. Not seeing the forest for the trees: generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids. Ecol. Inform. 75, 102026.
Collart, F., Broennimann, O., Guisan, A., Vanderpoorten, A., 2023. Ecological and biological indicators of the accuracy of species distribution models: lessons from European bryophytes. Ecography e06721.
Couce, E., Ridgwell, A., Hendy, E.J., 2013. Future habitat suitability for coral reef ecosystems under global warming and ocean acidification. Glob. Chang. Biol. 19 (12), 3592–3606.
Crase, B., Liedloff, A.C., Wintle, B.A., 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. Ecography 35 (10), 879–888.
Da Re, D., Tordoni, E., Lenoir, J., Lembrechts, J.J., Vanwambeke, S.O., Rocchini, D., Bazzichetto, M., 2023. USE it: uniformly sampling pseudo-absences within the environmental space for applications in habitat suitability models. Methods Ecol. Evol. 14 (11), 2873–2887.
Desjonquères, C., Villén-Pérez, S., De Marco, P., Márquez, R., Beltrán, J.F., Llusia, D., 2022. Acoustic species distribution models (aSDMs): a framework to forecast shifts in calling behaviour under climate change. Methods Ecol. Evol. 13 (10), 2275–2288.
El-Gabbas, A., Dormann, C.F., 2018. Wrong, but useful: regional species distribution models may not be improved by range-wide data under biased sampling. Ecol. Evol. 8 (4), 2196–2206.
ESRI, 2018. ArcGIS Desktop: Release 10.6.1. Environmental Systems Research Institute, Redlands, CA, USA.
Frasier, K.E., Garrison, L.P., Soldevilla, M.S., Wiggins, S.M., Hildebrand, J.A., 2021. Cetacean distribution models based on visual and passive acoustic data. Sci. Rep. 11 (1), 8240.
Gaulke, S.M., Hohoff, T., Rogness, B.A., Davis, M.A., 2023. Sampling methodology influences habitat suitability modeling for chiropteran species. Ecol. Evol. 13 (6), e10161.
Guillera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., Gordon, A., Kujala, H., Lentini, P.E., Wintle, B.A., 2015. Is my species distribution model fit for purpose? Matching data and models to applications. Glob. Ecol. Biogeogr. 24 (3), 276–292.
Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Modell. 135 (2–3), 147–186.
Hattab, T., Garzón-López, C.X., Ewald, M., Skowronek, S., Aerts, R., Horen, H., Lenoir, J., 2017. A unified framework to model the potential and realized distributions of invasive species within the invaded range. Divers. Distrib. 23 (7), 806–819.
Hazen, E.L., Abrahms, B., Brodie, S., Carroll, G., Welch, H., Bograd, S.J., 2021. Where did they not go? Considerations for generating pseudo-absences for telemetry-based habitat models. Mov. Ecol. 9, 1–13.
Hegel, T.M., Cushman, S.A., Evans, J., Huettmann, F., 2010. Current State of the Art For Statistical Modelling of Species Distributions. Spatial complexity, informatics, and wildlife conservation, pp. 273–311.
Henckel, L., Bradter, U., Jönsson, M., Isaac, N.J., Snäll, T., 2020. Assessing the usefulness of citizen science data for habitat suitability modelling: opportunistic reporting versus sampling based on a systematic protocol. Divers. Distrib. 26 (10), 1276–1290.
Hijmans, R.J., 2012. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. Ecology 93 (3), 679–688.
Hysen, L., Nayeri, D., Cushman, S., Wan, H.Y., 2022. Background sampling for multi-scale ensemble habitat selection modeling: does the number of points matter? Ecol. Inform. 72, 101914.

Iturbide, M., Bedia, J., Herrera, S., del Hierro, O., Pinto, M., Gutiérrez, J.M., 2015. A framework for species distribution modelling with improved pseudo-absence generation. Ecol. Modell. 312, 166–174.
Jhala, Y.V., Qureshi, Q., Nayak, A.K.(eds), 2020. Status of tigers, copredators and prey in India, 2018. National Tiger Conservation Authority. Government of India, New Delhi, and Wildlife Institute of India, Dehradun.
Kaszta, Ż., Cushman, S.A., Slotow, R., 2021. Temporal non-stationarity of path-selection movement models and connectivity: an example of African Elephants in Kruger National Park. Front. Ecol. Evol. 9, 553263.
Konowalik, K., Nosol, A., 2021. Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. Sci. Rep. 11 (1), 1482.
Levin, S.A., 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. Ecology 73 (6), 1943–1967.
Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17 (2), 145–151.
Macdonald, D.W., Bothwell, H.M., Hearn, A.J., Cheyne, S.M., Haidir, I., Hunter, L.T., Cushman, S.A., 2018. Multi-scale habitat selection modeling identifies threats and conservation opportunities for the Sunda clouded leopard (Neofelis diardi). Biol. Conserv. 227, 92–103.
Mair, L., Harrison, P.J., Jönsson, M., Löbel, S., Nordén, J., Siitonen, J., Snäll, T., 2017. Evaluating citizen science data for forecasting species responses to national forest management. Ecol. Evol. 7 (1), 368–378.
Marshall, L., Carvalheiro, L.G., Aguirre-Gutiérrez, J., Bos, M., de Groot, G.A., Kleijn, D., Biesmeijer, J.C., 2015. Testing projected wild bee distributions in agricultural habitats: predictive power depends on species traits and habitat type. Ecol. Evol. 5 (19), 4426–4436.
McGarigal, K., Wan, H.Y., Zeller, K.A., Timm, B.C., Cushman, S.A., 2016. Multi-scale habitat selection modeling: a review and outlook. Landsc. Ecol. 31, 1161–1175.
Meek, P.D., Ballard, G., Claridge, A., Kays, R., Moseby, K., O'brien, T., Townsend, S., 2014. Recommended guiding principles for reporting on camera trapping research. Biodivers. Conserv. 23, 2321–2343.
Miller, J.A., 2012. Species distribution models: spatial autocorrelation and non-stationarity. Prog Phys Geogr 36 (5), 681–692.
Neto, J.G.D.S., Sutton, W.B., Spear, S.F., Freake, M.J., Kéry, M., Schmidt, B.R., 2020. Integrating species distribution and occupancy modeling to study hellbender (Cryptobranchus alleganiensis) occurrence based on eDNA surveys. Biol. Conserv. 251, 108787.
Penjor, U., Kaszta, Ż., Macdonald, D.W., Cushman, S.A., 2021. Prioritizing areas for conservation outside the existing protected area network in Bhutan: the use of multi-species, multi-scale habitat suitability models. Landsc. Ecol. 36, 1281–1309.
Perret, D.L., Sax, D.F., 2022. Evaluating alternative study designs for optimal sampling of species' climatic niches. Ecography 2022 (1).
Pruhsmeier, H.N., McGrann, M.C., Graham, J., 2021. Combined use of data from avian surveys along the pacific crest trail with biodiversity repositories to model habitat suitability throughout northern California. IdeaFest: Interdisc. J. Creative Works Res. Cal Poly Humboldt 5 (1), 3.
Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J., Ferrier, S., 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol. Appl. 19 (1), 181–197.
Rana, D., Samad, I., Rastogi, S., 2022. To a charismatic rescue: designing a blueprint to steer Fishing Cat conservation for safeguarding Indian wetlands. J. Nat. Conserv. 68, 126225.
Razgour, O., Hanmer, J., Jones, G., 2011. Using multi-scale modelling to predict habitat suitability for species of conservation concern: the grey long-eared bat as a case study. Biol. Conserv. 144 (12), 2922–2930.
Regos, A., Gagne, L., Alcaraz-Segura, D., Honrado, J.P., Domínguez, J., 2019. Effects of species traits and environmental predictors on performance and transferability of ecological niche models. Sci. Rep. 9 (1), 4221.
Rovero, F., Zimmermann, F., Berzi, D., Meek, P., 2013. Which camera trap type and how many do I need?" A review of camera features and study designs for a range of wildlife research applications. Hystrix Italian J. Mammal. 24 (2), 148–156.
Saranya, K.R.L., Lakshmi, T.V., Reddy, C.S., 2021. Predicting the potential sites of Chromolaena odorata and Lantana camara in forest landscape of Eastern Ghats using habitat suitability models. Ecol. Inform. 66, 101455.
Sarkar, M.S., Pandey, A., Singh, G., Lingwal, S., John, R., Hussain, A., Rawal, R.S., 2018. Multiscale statistical approach to assess habitat suitability and connectivity of common leopard (Panthera pardus) in Kailash Sacred Landscape, India. Spatial Stat. 28, 304–318.
Senay, S.D., Worner, S.P., Ikeda, T., 2013. Novel three-step pseudo-absence selection technique for improved species distribution modelling. PLoS ONE 8 (8), e71218.
Sillero, N., Barbosa, A.M., 2021. Common mistakes in ecological niche models. Int. J. Geogr. Inform. Sci. 35 (2), 213–226.
Short Bull, R.A., Cushman, S.A., Mace, R., Chilton, T., Kendall, K.C., Landguth, E.L., Luikart, G, 2011. Why replication is important in landscape genetics: american black bear in the Rocky Mountains. Mol. Ecol. 20 (6), 1092–1107.
Srivathsa, A., Parameshwaran, R., Sharma, S., Karanth, K.U., 2015. Estimating population sizes of leopard cats in the Western Ghats using camera surveys. J. Mammal. 96 (4), 742–750.
Sunquist, M., Sunquist, F., 2002. Wild Cats of the World. The University of Chicago Press.
Tessarolo, G., Rangel, T.F., Araújo, M.B., Hortal, J., 2014. Uncertainty associated with survey design in Species Distribution Models. Divers. Distrib. 20 (11), 1258–1269.
Thuiller, W., Brotons, L., Araújo, M.B., Lavorel, S., 2004. Effects of restricting environmental range of data to project current and future species distributions. Ecography 27 (2), 165–172.
Tôrres, N.M., De Marco, P., Santos, T., Silveira, L., de Almeida Jácomo, A.T., Diniz-Filho, J.A., 2012. Can species distribution modelling provide estimates of population densities? A case study with jaguars in the Neotropics. Divers. Distrib. 18 (6), 615–627.
Václavík, T., Kupfer, J.A., Meentemeyer, R.K., 2012. Accounting for multi-scale spatial autocorrelation improves performance of invasive species distribution modelling (iSDM). J. Biogeogr. 39 (1), 42–55.
Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G., 2018. blockCV: An r package For Generating Spatially Or Environmentally Separated Folds For K-Fold Cross-Validation of Species Distribution Models. Biorxiv, 357798.
VanDerWal, J., Shoo, L.P., Graham, C., Williams, S.E., 2009. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? Ecol. Modell. 220 (4), 589–594.
Varela, Sara, Anderson, Robert P., García-Valdés, Raúl, Fernández-González, Federico, 2014. Environmental filters reduce the effects of sampling bias and improve predictions of ecological niche models. Ecography 37 (11), 1084–1091.
Vasquez, V.L., de Lima, A.A., dos Santos, A.P., Pinto, M.P., 2021. Influence of spatial extent on habitat suitability models for primate species of Atlantic Forest. Ecol. Inform. 61, 101179.
Vergara, M., Cushman, S.A., Urra, F., Ruiz-González, A., 2016. Shaken but not stirred: multiscale habitat suitability modeling of sympatric marten species (Martes martes and Martes foina) in the northern Iberian Peninsula. Landsc. Ecol. 31, 1241–1260.
Vergara, M., Cushman, S.A., Ruiz-González, A., 2017. Ecological differences and limiting factors in different regional contexts: landscape genetics of the stone marten in the Iberian Peninsula. Landsc. Ecol. 32, 1269–1283.
Warren, D.L., Glor, R.E., Turelli, M., 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. Evolution (N Y) 62 (11), 2868–2883.
Watling, J.I., Brandt, L.A., Bucklin, D.N., Fujisaki, I., Mazzotti, F.J., Romanach, S.S., Speroterra, C., 2015. Performance metrics and variance partitioning reveal sources of uncertainty in species distribution models. Ecol. Modell. 309, 48–59.
Whitenack, L.E., Snell Taylor, S.J., Tomcho, A., Hurlbert, A.H., 2023. Comparing multiscale, presence-only habitat suitability models created with structured survey data and community science data for a rare warbler species at the southern range margin. PLoS ONE 18 (4), e0275556.
Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Sequeira, A.M., 2018. Outstanding challenges in the transferability of ecological models. Trends Ecol. Evol. (Amst.) 33 (10), 790–802.