



Not seeing the forest for the trees: Generalised linear model out-performs random forest in species distribution modelling for Southeast Asian felids

Luca Chiaverini^{a,*}, David W. Macdonald^a, Andrew J. Hearn^a, Żaneta Kaszta^a, Eric Ash^{a,b}, Helen M. Bothwell^{a,c}, Özgün Emre Can^d, Phan Channa^{a,e}, Gopalasamy Reuben Clements^{f,g,h}, Iding Achmad Haidir^{a,i}, Pyae Phyo Kyaw^{a,j}, Jonathan H. Moore^{k,l}, Akchousanh Rasphone^{a,m}, Cedric Kai Wei Tan^{a,n}, Samuel A. Cushman^{a,o}

^a Wildlife Conservation Research Unit, Department of Biology, University of Oxford, The Recanati-Kaplan Centre, Tubney House, Tubney, Oxon OX13 5QL, UK

^b Freeland Foundation, Bangkok, Thailand

^c Research School of Biology, Australian National University, Canberra, ACT, Australia

^d Department of Biology, Faculty of Science, Ankara University, Ankara, Turkey

^e General Directorate of Natural Protected Area, Ministry of Environment, Phnom Penh, Cambodia

^f Rimba, Kuala Lumpur, Malaysia

^g Department of Biological Sciences, Sunway University, Bandar Sunway, Malaysia

^h Jeffrey Sachs on Sustainable Development, Sunway University, Bandar Sunway, Malaysia

ⁱ Directorate of Conservation Area Planning, Directorate General of Natural Resources and Ecosystem Conservation, Ministry of Environment and Forestry, Jakarta, Indonesia

^j Wildlife Conservation Society – Myanmar Program, Yangon, Myanmar

^k School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China

^l School of Environmental Sciences, University of East Anglia, Norwich, UK

^m Wildlife Conservation Society – Lao PDR Program, Vientiane, Lao People's Democratic Republic

ⁿ School of Environmental and Geographical Sciences, University of Nottingham Malaysia, Semenyih, Malaysia

^o Rocky Mountain Research Station, United States Forest Service, Flagstaff, AZ, USA

ARTICLE INFO

Keywords:

Bootstrapping
Generalised linear model
Machine learning
Multi-scale
Random forest
Species distribution modelling

ABSTRACT

Species Distribution Models (SDMs) are a powerful tool to derive habitat suitability predictions relating species occurrence data with habitat features. Two of the most frequently applied algorithms to model species-habitat relationships are Generalised Linear Models (GLM) and Random Forest (RF). The former is a parametric regression model providing functional models with direct interpretability. The latter is a machine learning non-parametric algorithm, more tolerant than other approaches in its assumptions, which has often been shown to outperform parametric algorithms. Other approaches have been developed to produce robust SDMs, like training data bootstrapping and spatial scale optimisation. Using felid presence-absence data from three study regions in Southeast Asia (mainland, Borneo and Sumatra), we tested the performances of SDMs by implementing four modelling frameworks: GLM and RF with bootstrapped and non-bootstrapped training data. With Mantel and ANOVA tests we explored how the four combinations of algorithms and bootstrapping influenced SDMs and their predictive performances. Additionally, we tested how scale-optimisation responded to species' size, taxonomic associations (species and genus), study area and algorithm. We found that choice of algorithm had strong effect in determining the differences between SDMs' spatial predictions, while bootstrapping had no effect. Additionally, algorithm followed by study area and species, were the main factors driving differences in the spatial scales identified. SDMs trained with GLM showed higher predictive performance, however, ANOVA tests revealed that algorithm had significant effect only in explaining the variance observed in sensitivity and specificity and, when interacting with bootstrapping, in Percent Correctly Classified (PCC). Bootstrapping significantly explained the variance in specificity, PCC and True Skills Statistics (TSS). Our results suggest that there are systematic differences in the scales identified and in the predictions produced by GLM vs. RF, but that neither approach was consistently better than the other. The divergent predictions and inconsistent predictive abilities suggest that analysts should not assume machine learning is inherently superior and should test multiple methods. Our results

* Corresponding author.

E-mail address: luca.chiaverini@biology.ox.ac.uk (L. Chiaverini).

<https://doi.org/10.1016/j.ecoinf.2023.102026>

Received 13 October 2022; Received in revised form 9 February 2023; Accepted 10 February 2023

Available online 18 February 2023

1574-9541/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

have strong implications for SDM development, revealing the inconsistencies introduced by the choice of algorithm on scale optimisation, with GLM selecting broader scales than RF.

1. Introduction

Species distribution models (SDMs) are a powerful tool to derive spatially-explicit predictions of habitat suitability, by statistically relating species occurrence data at known locations with environmental features of those locations (Elith and Leathwick, 2009; Guisan and Thuiller, 2005). Several statistical approaches are available to produce SDMs (Guisan et al., 2017; Valavi et al., 2022), with regression methods (Guisan et al., 2002) in particular applied in the earlier stages of the SDMs' development, and machine learning methods receiving more attention in recent years (e.g., Evans and Cushman (2009); Cushman et al. (2017); Stupariu et al. (2021)).

Among regression models, generalised linear models (GLM) are extensions of linear model allowing non-linearity and non-constant variance of the data (Guisan et al., 2002; Hastie and Tibshirani, 1990). Consequently, GLM can deal with different families of data, making them an appropriate tool for analysing species-habitat relationships, which rarely are represented by normal distributions (Austin, 1987). GLM have long been the predominant algorithm used in SDMs (Elith and Leathwick, 2009; Guisan and Thuiller, 2005; Guisan and Zimmermann, 2000; McGarigal et al., 2016).

Among machine learning techniques, random forest (RF; Breiman (2001)) is the most commonly used algorithm (Stupariu et al., 2021). RF are Classification and Regression Tree (CART) based ensemble approaches, which overcome limitations of CARTs such as model overfitting. Additionally, RF can handle large numbers of predictor variables without being affected by multicollinearity, and can readily accommodate nonlinear relationships between predictor and response variables (Breiman, 2001; Evans and Cushman, 2009; Kumar et al., 2021). Furthermore, by permutation of variables, RF provides accessible measures of variable importance (Cutler et al., 2007; Evans and Cushman, 2009).

The applications of GLM and other parametric models are constrained by often neglected or violated assumptions, such as data independence, linearity and variable interaction. Therefore, non-parametric algorithms, like RF, in many cases show better performance in comparison to parametric methods (Cushman et al., 2017; Cushman and Wasserman, 2018; Kumar et al., 2021; Valavi et al., 2022). However, there are applications where RF perform poorly, often because non-parametric algorithms have little control over the shape of the fitted function (Roberts et al., 2017). Additionally, RF's performances are sometimes weak when using presence-absence data with an imbalance in favour of absences (e.g., Evans and Cushman (2009); Freeman et al. (2012)). Therefore, comparing the performances of GLM and RF is an important and contemporary issue to assess for developing reliable SDMs.

To produce accurate SDMs, not just the effects of different algorithms should be investigated, but also the effects of resampling techniques on model's training data (Efron, 1982; Freedman, 1981). Specifically, bootstrapping (i.e., random subsampling with replacement) the training data has been shown to increase models' precision by providing a combination of models, which reduces stochastic errors in estimation (e.g., Vaughan and Ormerod (2005); Hefley et al. (2014); Xu and Goodacre (2018)).

Additionally, species-habitat relationships occur at multiple spatial scales (Levin, 1992; Wiens, 1989). Therefore, scale optimisation has recently been an area of rapid development (Chiaverini et al., 2021; Vergara et al., 2016; Wan et al., 2019). Given that scale optimization is a critical step in the framework to produce reliable SDMs (McGarigal et al., 2016), it is important to understand how modelling algorithm and bootstrapping affect multiscale model optimisation.

Here, using presence-absence data of felid species sampled in three study regions in Southeast Asia (mainland, Borneo and Sumatra), we investigated nine hypotheses related to the performances and predictions of SDMs, and how these are affected by scale optimisation, choice of algorithm and application of data resampling. Specifically, we hypothesised that: (1) RF would outperform GLM in terms of their predictive performances; (2) predictive performances of the models trained with bootstrapped data would outperform non-bootstrapped models; (3) models trained with the same algorithm would be more correlated; (4) bootstrapping would have relatively little effect on correlation of the SDMs; (5) species that are more similar in size would have more similarities in the representative spatial scales at which predictor variables were selected; (6) models for the same species would have more similar scales than models for different species; (7) models for the same genus would have more similar scales than models for different genera; (8) models for species from the same study area would have more similar scales than species from different study areas; (9) models trained with the same algorithm would have more similar scales than models trained with a different algorithm.

2. Materials and methods

2.1. Data collection

Camera trap surveys were carried out in Southeast Asia between 2007 and 2016, covering the range of the two clouded leopard species (i.e., mainland clouded leopard (*Neofelis nebulosa*) in mainland Southeast Asia and Sunda clouded leopard (*Neofelis diardi*) in Borneo and Sumatra), which represented the original focus of our surveys. Sampling grids occurred mainly in protected areas, covering a wide altitudinal gradient and different environments. Grids consisted of approximately 80 camera stations placed 1.0–2.0 km from each other, with two cameras per station at ~40 cm above the ground, and deployed along forest trails and disused logging roads to maximise felids' detection success (Macdonald et al., 2018; Macdonald et al., 2019). Additionally, we incorporated data from camera trap studies conducted by Ash et al. (2021).

Even though the surveys were designed for sampling clouded leopards, we also collected data for Asiatic golden cat (*Catopuma temminckii*), marbled cat (*Pardofelis marmorata*) and leopard cat (*Prionailurus bengalensis*) in mainland, Borneo bay cat (*Catopuma badia*), marbled cat and Sunda leopard cat (*Prionailurus javanensis*) in Borneo, and Asiatic golden cat, marbled cat and Sunda leopard cat in Sumatra. To minimise overestimation bias, we excluded records of the same species at the same camera trap stations within 1 h, except when animals were individually recognisable and when sex and/or age classes were unambiguous.

2.2. Habitat covariates

We chose two preliminary sets of covariates to model habitat suitability for felids in the mainland and in the offshore extents of the study area, reflecting different habitat composition of the regions. Covariates were selected based on previous knowledge of felid ecological requirements in Southeast Asia (Hearn et al., 2018; Macdonald et al., 2018; Macdonald et al., 2019), and included environmental, anthropogenic, topographic and climatic covariates (Tables S1–S3).

To investigate the scales at which sampled felids select habitat features, we assessed each covariate at eight spatial scales, by using circular buffers of varying radii, centred on each camera trap location. For mainland, we investigated representative spatial scales by using circular buffers of 250 m, 500 m, 1000 m, 2000 m, 4000 m, 8000 m, 16,000 m and 32,000 m. For Borneo and Sumatra we used buffers of 250 m, 500 m,

1000 m, 2000 m, 4000 m, 6000 m, 8000 m and 10,000 m (Chiaverini et al., 2022; Macdonald et al., 2020).

At each of the aforementioned scales, we calculated biologically informative metrics by using FRAGSTATS (McGarigal et al., 2012) for discrete raster layers and ArcMap v10.6.1 (ESRI, 2018) for continuous raster layers: in FRAGSTATS, we implemented percentage of landscape (PLAND) metric, while in ArcMap we calculated focal mean (FM). Both the functions average the pixel values of the covariates' raw raster layers to obtain a smoothed version of the original layer. Additionally, PLAND function calculates the percentage of the focal covariate within the circular buffer, and this is why it has been preferred for discrete layers. Finally, we removed poorly sampled covariates occurring at <10% of camera trap stations.

2.3. Modelling framework

2.3.1. Generalised linear model – scale optimisation and covariates selection

To investigate the most representative spatial scales, we performed univariate generalised linear model (GLM), independently at each scale for each covariate, using the presence-absence data independently for each felid as the response variable. We subset the dataset by randomly selecting 80% of the camera traps to train the univariate GLM, leaving the remaining 20% for model validation. For each covariate, we selected the scale whose univariate GLM showed the lowest Akaike Information Criterion (AIC). We then checked the best-scaled covariates for multicollinearity by calculating Pearson's correlation index between all covariates pairs and, when two covariates were highly correlated (i.e., $|r| \geq 0.7$), by removing the one whose univariate GLM showed the lower adjusted-R² (Guisan and Zimmermann, 2000).

2.3.2. Generalised linear model – model training

We trained GLM independently for each felid species, by applying two different approaches. For the first approach (hereafter *non-bootstrapped GLM*), we trained the models by using the same dataset used to evaluate the best scales, composed of the 80% of the camera traps, and we produced species distribution probability surfaces for each sampled felid.

For the second approach (hereafter *bootstrapped GLM*) we subsampled the entire camera trap dataset by randomly selecting 20% of the presence and of the absence locations, trained a GLM and produced a species distribution probability surface. We reiterated this process 100 times with replacement, and produced the final species distribution probability surfaces by averaging the 100 probability surfaces, independently for each felid.

2.3.3. Generalised linear model – model validation

We evaluated the performances of the non-bootstrapped GLM by calculating the Area Under the ROC Curve (AUC), Percent Correctly Classified (PCC), sensitivity, specificity and True Skill Statistic (TSS) using the 20% camera traps retained from the training dataset.

To evaluate the bootstrapped GLM, for each of the 100 iterations, we calculated AUC, PCC, sensitivity, specificity and TSS using the data retained from the model training, composed of 80% of the presence and of the absence locations. We obtained the final metrics by averaging the results of 100 validation iterations. We implemented this framework independently for each felid. Validation metrics were calculated in R v3.5.1 (R Core Team, 2018) using the package *PresenceAbsence* (Freeman and Moisen, 2007).

2.4. Random forest – scale optimisation and covariates selection

Independently for each felid, we investigated the most representative scale for each covariate by performing univariate Random Forest (RF), using the full camera traps dataset and the presence-absence data as response variable. For each covariate, we selected the scale whose

univariate RF showed the lowest Out-Of-Bag (OOB) error.

2.4.1. Random forest – model training

Similar to the GLM approaches, for RF we used two different training data selection frameworks. For the first one (hereafter *non-bootstrapped RF*), we trained the RF by using the same training dataset used for the non-bootstrapped GLM, composed of randomly selected 80% of camera trap locations, and produced species distribution probability surfaces for each felid.

For the second approach (hereafter *bootstrapped RF*), we subsampled the entire camera trap dataset by randomly selecting 20% of the presence and of the absence locations. We then trained RF with the subsampled dataset and produced the species distribution probability surface. We reiterated this framework 100 times, replacing each time the data used to train the models, and produced 100 different probability surfaces. We then obtained the final species distribution probability surfaces by averaging the 100 probability surfaces, independently for each felid. Each individual RF was produced by bagging the training data, a procedure automatically implemented in the algorithm to reduce the variance that might otherwise characterise decision trees. Therefore, by bootstrapping also the training data of each RF, we implemented a two-stage bagging procedure.

2.4.2. Random forest – model validation

We evaluated the non-bootstrapped RF by calculating AUC, PCC, sensitivity, specificity and TSS using the 20% camera traps retained from the RF training.

To evaluate the bootstrapped RF, we used the retained dataset from the RF training, composed of the 80% of presence and absence locations, to calculate AUC, PCC, sensitivity, specificity and TSS for each iteration, independently for each species, using the package *PresenceAbsence* in R v3.5.1. To obtain the final metrics, we averaged the metric results of the 100 iterations.

2.5. Effects of algorithm and bootstrapping on predictive performances

To evaluate whether the algorithm and the bootstrapping affected the predictive performances of the models, we performed ANOVA tests. We performed the tests independently for each performance metric we calculated, and tested the effect of algorithm, bootstrapping and combination of algorithm and bootstrapping.

2.6. Effects of algorithm and bootstrapping on models similarity

To investigate the effects of algorithm and bootstrapping on the similarities between species distribution surfaces, we performed Mantel tests in R v3.5.1 using the package *ecodist* (Goslee and Urban, 2007). We first calculated Pearson's correlation indexes between the four models, for each species, for each study area, resulting in a total of twelve independent matrices (Tables S4-S15). Then, we calculated Mantel tests independently between these matrices and two binary matrices ("model matrices" sensu Legendre and Legendre (1998)) representing whether the models have been trained with GLM or RF (Table S16), and whether they have been bootstrapped or non-bootstrapped (Table S17). Finally, we averaged the results of the Mantel tests for each species, independently for algorithm and bootstrapping, to investigate their overarching effect on the species distribution probability surfaces.

2.7. Factors determining spatial scales selection

To evaluate the effects of different factors on the selection of the representative spatial scales, we performed Mantel tests in R v3.5.1 using the package *ecodist*. Specifically, we ran five simple Mantel tests between the difference in the ordinal scales among pairs of covariates across models and the difference in each of the factors we hypothesised to affect scale selection: body size (indexed to reflect the relative

differences between species), species, genus, study area and algorithm.

Additionally, we ran eight partial Mantel tests separating the effects of the methodological factor (i.e., algorithm) and of the four ecological factors (i.e., body size, species, genus and study area). We ran four partial Mantel tests between the difference in ordinal scales and difference in algorithm, partialling out the difference in each of the four ecological factors, sequentially. Then, we also ran four partial Mantel tests between the difference in ordinal scales and difference in each of the four ecological factors, partialling out the difference in algorithm.

3. Results

3.1. Data collection

Camera trap grids were deployed in 15 sampling locations in seven countries in the mainland study area (Table S18) and in 22 sampling locations in the Sunda Islands (15 in Borneo and 7 in Sumatra; Table S19). This sampling effort yielded a total of 1384 camera trap stations in mainland and 1544 in the Sunda Islands (801 in Borneo and 743 in Sumatra). We achieved a combined sampling effort of 115,389 trap nights in the mainland (mean = $83.4 \pm 2.6SE$) and of 138,515 trap nights in the Sunda Islands (77,687 in Borneo (mean = $97.0 \pm 2.1SE$) and 60,828 in Sumatra (mean = $81.9 \pm 0.7SE$)).

In the mainland we sampled clouded leopard at 235 camera trap stations, Asiatic golden cat at 148 stations, marbled cat at 116 stations and leopard cat at 345 stations (Table S20). In Borneo, we sampled Sunda clouded leopard at 154 camera trap stations, Borneo bay cat at 25

stations, marbled cat at 85 stations and Sunda leopard cat at 202 stations (Table S21). In Sumatra, we sampled Sunda clouded leopard at 101 camera trap stations, Asiatic golden cat at 102 stations, marbled cat at 39 stations and Sunda leopard cat at 23 stations (Table S22).

3.2. Effects of algorithm and bootstrapping on predictive performances

Predictive performances were substantially different comparing the species distribution probability surfaces produced with the four modelling frameworks (see *Model validation* in Supplementary Information; Tables S23-S27).

The ANOVA tests performed to assess whether the algorithm, the bootstrapping or their combination had significant effects on the models' predictive performances, revealed that there was a statistically significant difference in the PCC scores based on the bootstrapping ($F_{(1, 44)} = 9.72, p < 0.01$) and a less strong difference based on the interaction of algorithm and bootstrapping ($F_{(1, 44)} = 3.56, p < 0.1$; Fig. S1). We also found sensitivity to be significantly related to algorithm ($F_{(1, 44)} = 3.72, p < 0.1$; Fig. S2), and specificity to be related to bootstrapping ($F_{(1, 44)} = 8.53, p < 0.01$) and, more weakly, to algorithm ($F_{(1, 44)} = 3.30, p < 0.1$; Fig. S3). We also found TSS to be related to bootstrapping ($F_{(1, 44)} = 3.30, p < 0.1$; Fig. S4). We did not find any statistically significant difference for AUC (Fig. S5).

3.3. Effects of algorithm and bootstrapping on models similarity

Using the final sets of scale-optimised, uncorrelated covariates, we

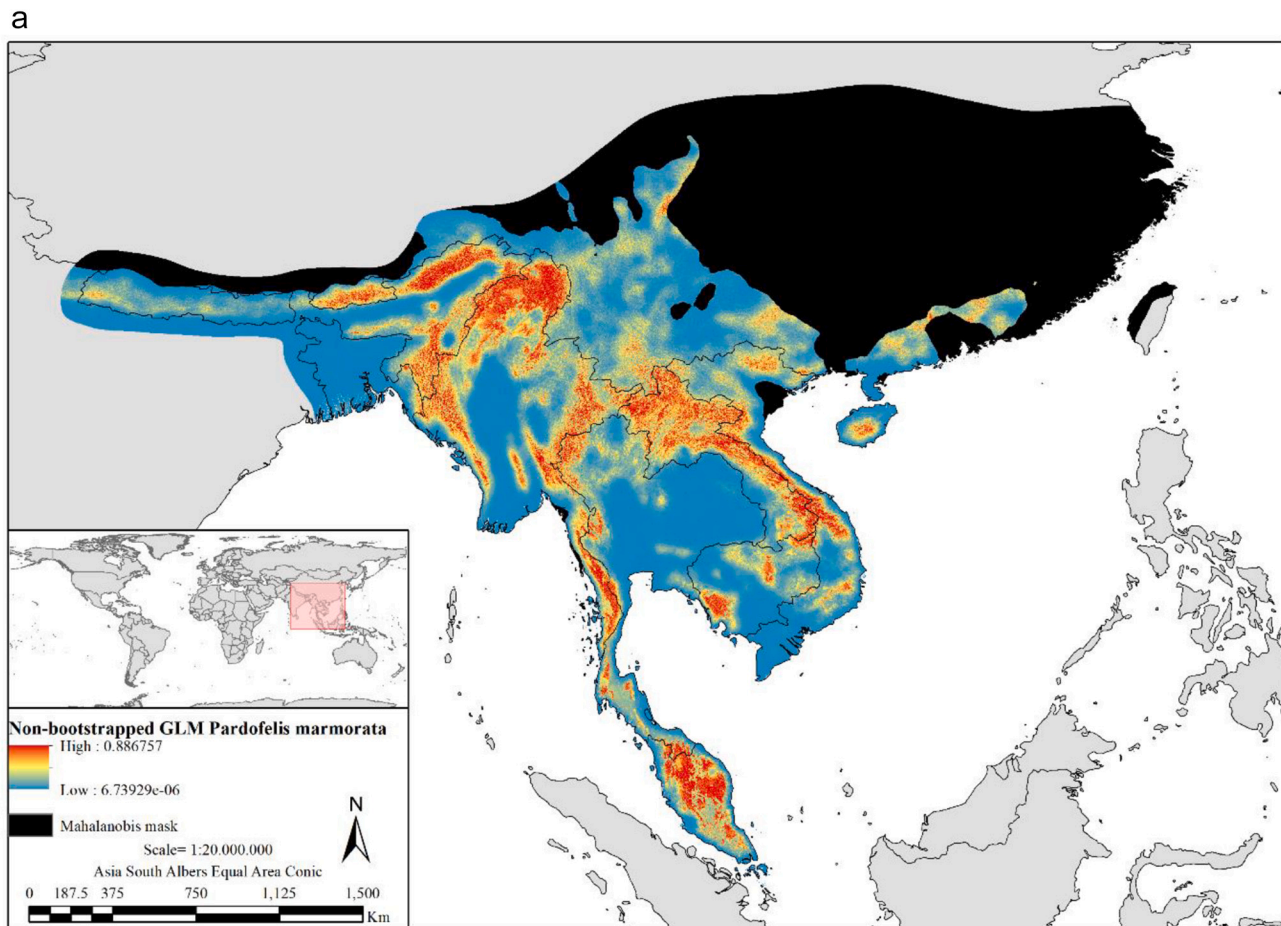


Fig. 1. Examples of different probability surfaces for marbled cat (*Pardofelis marmorata*) in mainland Southeast Asia, produced with (a) non-bootstrapped GLM, (b) bootstrapped GLM, (c) non-bootstrapped RF and (d) bootstrapped RF. We masked out the regions that were ecologically highly divergent from the sampled ones by producing a Mahalanobis mask (see *Mahalanobis mask* in Supplementary Information).

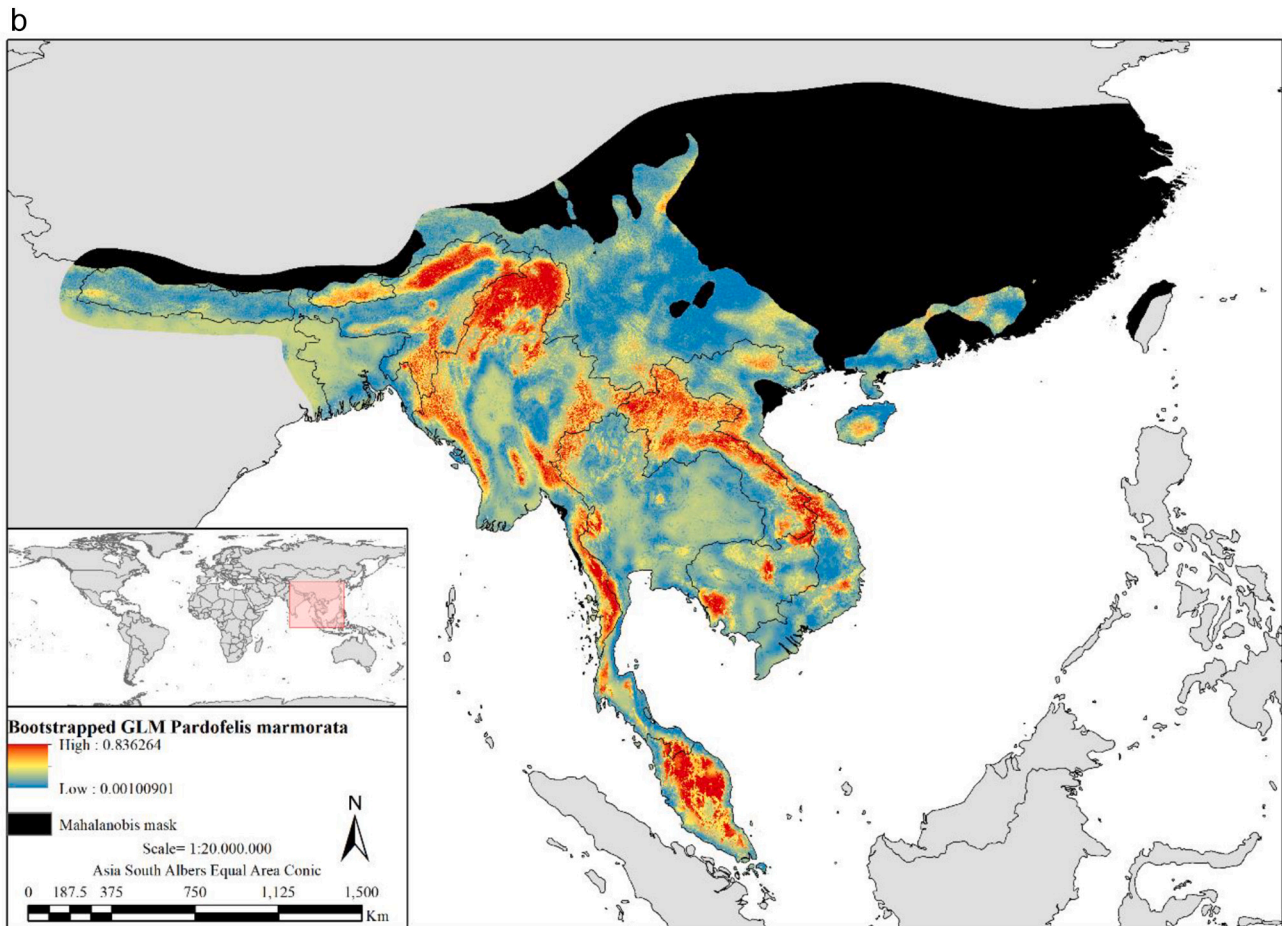


Fig. 1. (continued).

produced species distribution probability surfaces for the sampled felids by using the four modelling approaches (Fig. 1 and Figs. S6-S53).

The four models produced substantially different probability surfaces, showing sometimes very low correlation between models for the same species (see *Model predictions* in Supplementary Information; Tables S4-S15).

The Mantel tests, performed to evaluate whether there was a significant effect of algorithm and bootstrapping on the similarities observed between species distribution probability surfaces, showed a clear pattern in which there is strong effect of algorithm on the correlations in the output (mean Mantel = 0.84; Table 1). However, the Mantel tests showed also that there was no effect of bootstrapping on the correlations between probability surfaces (mean Mantel = -0.40; Table 1).

3.4. Factors determining spatial scale selection

The scale optimisation revealed that the most representative scales for each covariate, for each felid, varied substantially based on the algorithm applied, with GLM selecting overall consistently and substantially broader scales than RF (see *Scale optimisation and covariates selection* in Supplementary Information; Tables S1-S3). Mantel r value is a good indicator of relative effect size (Cushman et al., 2013b) and, based on this criterion, difference in algorithm had the strongest relationship with difference in ordinal scales among covariates across species models. The second highest relationship was for study area, which had approximately 70% of the effect size of difference in algorithm, based on magnitude of Mantel r . The third strongest relationship between ordinal scales and predictive factors was with species, which

showed approximately 20% of the effect size of difference in algorithm (Table 2).

The partial Mantel tests were carried out to determine whether the methodological differences (i.e., algorithm) were independent of the ecological differences (i.e., species' size, species, genus and study area) in relation to correlation between difference in ordinal scales and difference in predictive variables. The partial Mantel tests showed that all relationships identified in the single Mantel tests were independent of relationships with other factors. Indeed, in all cases the partial Mantel tests slightly increased the Mantel r values, showing that removing the co-varying effect of other factors strengthened the correlation between difference in ordinal scales and difference in predictive variables. The correlation between ordinal scales difference and algorithm difference was strongest, followed also in this case by study area, species, genus and species' size (Table 3).

4. Discussion

4.1. Similarity of prediction and model performance hypotheses

Of the two hypotheses related to model similarities and correlations, our results supported one of them. Specifically, the Mantel tests supported the hypothesis that there are strong differences between spatial predictions produced with GLM and RF, across levels of bootstrapping. Species distribution surfaces resulting from the application of these algorithms are used to prioritise areas for protection or to identify the areas richest in biodiversity (e.g., Kaszta et al. (2020); Penjor et al. (2021)), and differences in the model predictions based on algorithm applied, and not on the underlying ecological relationships, would add a

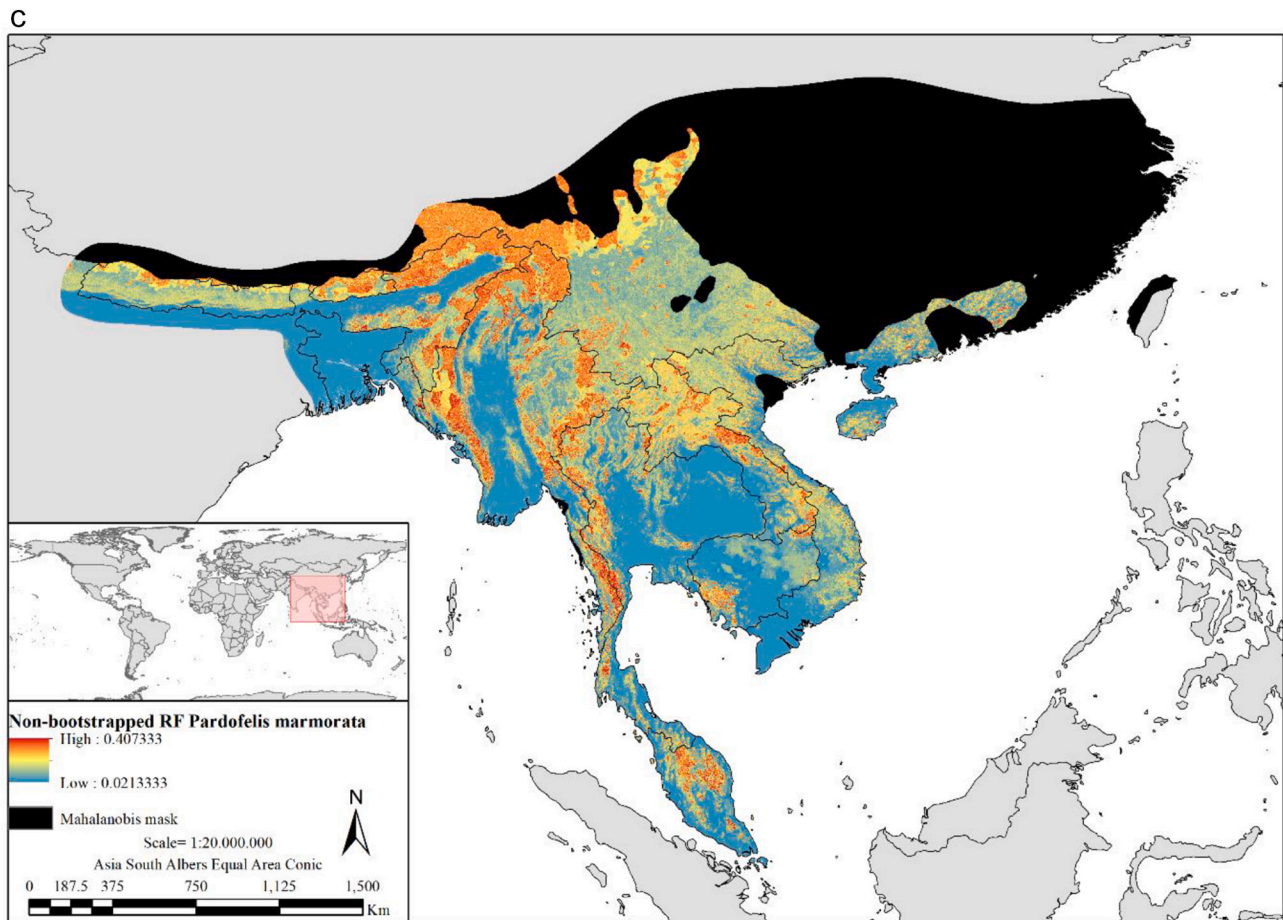


Fig. 1. (continued).

high level of uncertainty to interpreting the results of species distribution models.

Given this large systematic difference between model prediction by GLM and RF, it is important to identify a priori which method was superior to avoid the uncertainty and ambiguity of conflicting predictions. A number of past studies have suggested that RF is generally more robust, accurate and reliable than GLM for species distribution modelling (e.g., Cushman et al. (2017); Cushman and Wasserman (2018); Kumar et al. (2021)). Therefore, we hypothesised that RF would consistently out-perform GLM in terms of model performance and accuracy of predictions. However, our hypothesis was not supported. Nominally, GLM outperformed RF based on AUC in all three study areas for the majority of species. The only marginally significant differences due to the algorithm were observed in the scores of the sensitivity and specificity. However, the combined effect of algorithm and bootstrapping was also statistically significant in exemplifying variance in PCC scores.

We found our results surprising given the rich literature supporting RF as a much stronger modelling algorithm than GLM. Cushman and Wasserman (2018) compared the species distribution models for American marten (*Martes americana*) in northern USA, by using GLM and RF, concluding that RF produced the most supported model. Similarly, Kumar et al. (2021), in their effort to develop of a new technique to model species-habitat relationships (i.e., the *smoothing function*), also compared GLM and RF, demonstrating that the latter outperformed the former. However, there are cases in which RF has been reported to perform poorly. Specifically, poor predictive performances for RF have often been reported when models are trained with unbalanced presence-absence data, characterised by several absences that outnumber

presences (Evans and Cushman, 2009; Freeman et al., 2012). This is indeed the case of our data characterised, especially for some species, by high proportions of absences and, even though we corrected for imbalance data by implementing the RF class-balance modelling approach (Evans and Cushman, 2009), we believe that the effect of zero-inflation on RF should be further evaluated. Additionally, Valavi et al. (2021) reported that a critical factor that negatively affects the predictive performances of RF trained with presence-absence data is what they defined as “class overlap”, represented by presence and absence locations occurring in close geographic association in the same habitats. This circumstance is particularly frequent in our dataset, which was composed of camera traps that, when occurring within the same sampling regions, were deployed at 1.0–2.0 km from each other. Therefore, the unbalance between presences and absences, as well as the occurrence of presence and absence locations within the same geographic contexts, were both factors that likely influenced negatively the predictive performances of the RF, and future research is strongly recommended to better characterise the effects that the structure of datasets has on RF's predictive performances. Recently, Kumar et al. (2021) used a simulation approach to evaluate the effects of spatial sampling (random vs actual clustered camera traps), as well as the shape of the environmental manifold, nonlinearity, and interaction on GLM and RF modelling. That study found that RF consistently was accurate in describing the underlying species-environment relationship, even when data were clustered. Given that that study used the actual spatial distribution of samples used in a subset of the data used in this study, we infer that clustered sampling is not the major driver of the poorer than expected performance of RF.

Finally, even though we demonstrated that differences in the

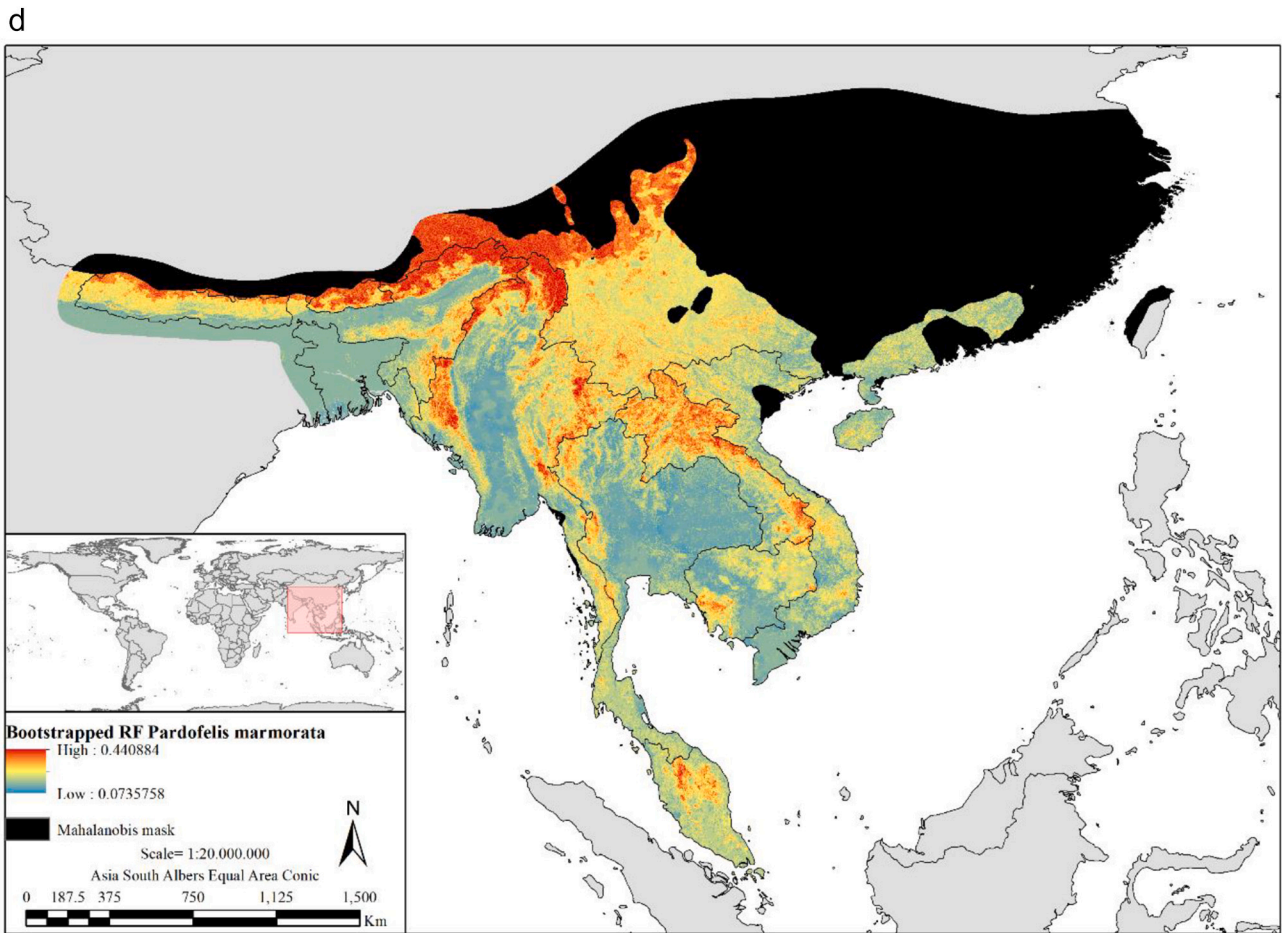


Fig. 1. (continued).

Table 1
Results of the Mantel tests to evaluate the effects of algorithm and bootstrapping on the correlations between species distribution probability surfaces.

Study areas	Species	Algorithm Mantel tests	Bootstrapping Mantel tests
Mainland	<i>Neofelis nebulosa</i>	0.98	-0.49
	<i>Catopuma temminckii</i>	0.24	-0.16
	<i>Pardofelis marmorata</i>	0.93	-0.47
	<i>Prionailurus bengalensis</i>	0.94	-0.46
	<i>Neofelis diardi</i>	0.99	-0.46
Borneo	<i>Catopuma badia</i>	0.93	-0.42
	<i>Pardofelis marmorata</i>	0.71	-0.23
	<i>Prionailurus javanensis</i>	0.96	-0.48
Sumatra	<i>Neofelis diardi</i>	0.64	-0.21
	<i>Catopuma temminckii</i>	0.90	-0.57
	<i>Pardofelis marmorata</i>	0.95	-0.34
	<i>Prionailurus javanensis</i>	0.96	-0.48
	Mean	0.84	-0.40

algorithms led to large and systematic differences in SDM predictions, the Mantel tests did not support our hypothesis that models trained with bootstrapped data were more similar to other bootstrapped models, across algorithms. Therefore, bootstrapping the training data of the

Table 2
Simple Mantel tests results for scales relationships.

Covariate	Mantel r	p
Algorithm	0.14	0.001
Study area	0.10	0.001
Species	0.03	0.001
Genus	0.003	0.15
Size	-0.0003	0.49

Table 3
Partial Mantel tests results for scales relationships.

Covariate	Partial	Mantel r	p
Algorithm	Study area	0.13590	0.001
Algorithm	Species	0.13521	0.001
Algorithm	Genus	0.13503	0.001
Algorithm	Size	0.13501	0.001
Study area	Algorithm	0.09804	0.001
Species	Algorithm	0.02928	0.001
Genus	Algorithm	0.00329	0.11
Size	Algorithm	0.00001	0.49

models, regardless of the algorithm, did not lead to significant differences from models trained with non-bootstrapped data. These results are corroborated by the weak support for our hypothesis that bootstrapping would produce SDMs with higher predicting power. Bootstrapping had only a slightly stronger effect on explaining the predictive performances, given that it had strong statistical power in explaining PCC and specificity scores and, to a lesser extent, TSS. This conflicts with our

expectation that bootstrapping would improve the model's ability to correctly classify presence-absence data by reducing single models' stochastic errors. However, an interesting behaviour of bootstrapping that we partially observed here, and which has already been demonstrated by, for example, Osawa et al. (2011), is that it improved SDMs' ability to discern true positive rates (i.e., sensitivity) and true negative rates (i.e., specificity). Consequently, TSS, calculated as "sensitivity + specificity - 1", benefitted from bootstrapping, showing significant difference from non-bootstrapped SDMs. Our results generally found that non-bootstrapped models had higher performance based on the majority of our performance measures than bootstrapped models for most species in all three study areas. This suggests that the large effort to conduct bootstrapped sampling to train SDMs may be misplaced and that generally bootstrapping does not improve model predictions.

4.2. Scaling hypotheses

The Mantel test analyses supported three of the five hypotheses related to spatial scales. First, contrary to predictions, our results suggested that there is no relationship between similarities in scales of relationship and body size based on species' weight. One would expect for a coherent guild, such as felids, that there would be strong and monotonic association between scales of habitat relationships and body size, with larger-bodied species generally related to habitat factors at broader scales. Large felids are expected to select habitat factors at broad scales, reflecting their mobility (Elliot et al., 2014) and home-range selection (Khosravi et al., 2019). It is not uncommon, however, to observe species selecting habitat factors at scales that are uncorrelated with their body sizes. Hearn et al. (2018), for example, highlighted that in Borneo, Sunda leopard cat, marbled cat and Sunda clouded leopard selected tree cover at increasing spatial scales, as one would expect based on their sizes, but that they selected elevation at spatial scales that would be unlikely based on their sizes, with Sunda clouded leopard and marbled cat selecting elevation at slightly >100 m, and the small Sunda leopard cat at almost 2 km. These examples suggest that simple ideas about the association between body sizes and scales of species-habitat relationships might be misplaced, and species-specific formal optimal scales optimisation has to be carried out to assess the representative spatial scales (McGarigal et al., 2016).

Second, we found significant relationships between species and scales, such that covariates for the same species across the three study areas had more similarities in scales than covariates for different species. This is expected if there are species-level associations with scales of relationships with habitat factors which are general and consistent (e.g., Cushman and McGarigal (2002)). However, our results suggested that there is no significant association between the scales of covariates selected and taxonomic genera, rejecting our third scaling hypothesis. This suggests that the species-level scales associations with habitat factors do not extend to the next higher level of taxonomic association. However, differences in the spatial scales between members of the same genus have been highlighted, for example, between mainland clouded leopard and Sunda clouded leopard (Macdonald et al., 2018; Macdonald et al., 2019). Of the covariates used to model suitable habitat for the two species, compound topographic index was selected at 500 m by mainland clouded leopard and 1 km by Sunda clouded leopard, and the difference became stronger looking at other shared covariates, with focal mean of slope position selected at 8 km by mainland clouded leopard and at 250 m by Sunda clouded leopard. Similarly, Vergara et al. (2016) found strong differences in the selection of spatial scales between pine marten (*Martes martes*) and stone marten (*Martes foina*) in the northern Iberian Peninsula, with some of the covariates assessed to model habitat suitability showing strong displacement between the two species (e.g., % crops = 16 km for pine marten and 4 km for beech marten, and % forests = 2 km for pine marten and 32 km for beech marten).

Our results indicated that there is strong relationship between scales of covariates and study area. This suggests that there are regional-scale

ecological or geographical factors that affect scales of habitat relationships across multiple species. This may reflect differences in regional limiting factors (e.g., Cushman et al. (2013a); Vergara et al. (2016)) that result in different factors at different spatial scales being the best predictors in different study areas. Multivariate assessment of the suitable habitat for terrestrial vertebrate fauna in Southeast Asia (Chiaverini et al., 2022; Macdonald et al., 2020) revealed, for example, that altitudinal gradients affected mainland biodiversity at much broader scale (32 km) than insular species, which are in turn affected by elevation only at a very fine scale (250 m). Differences between study areas have been highlighted also for felids: Atzeni et al. (2020) found that snow leopard (*Panthera uncia*) differently selected the response scale for limiting habitat factors across two landscapes in China.

Finally, we found an extremely strong correlation between difference in scales and difference in algorithm, supporting our last hypothesis. This relationship was almost twice as strong as between any other factor and difference in scales, suggesting that selection of spatial scales is differently affected by GLM and RF. GLM generally selected broader scales than RF, and the strength of the Mantel correlations suggests that these methods strongly differ in the scales they selected across models for the 12 species (4 species for 3 study areas) in Southeast Asia. Our result that GLM systematically predicts broader spatial scales than RF is consistent with several other studies that have compared scale selection in the two studies (e.g., Cushman et al. (2017); Cushman and Wasserman (2018)).

This finding has major implications for species distribution modelling using scale optimisation. Specifically, if distinct statistical methods consistently select different spatial scales, then the results of the scale optimisation process are unreliable and not consistently related to underlying ecological relationships. However, given this is an empirical analysis in which the true relationships between species and habitat factors across scales are not known, these results strongly suggest the need to use a controlled simulation study (e.g., Atzeni et al. (2020); Chiaverini et al. (2021); Kumar et al. (2021)) to systematically evaluate the performance of the GLM and RF in terms of correctly identifying the underlying scales of relationships. Chiaverini et al. (2021), for example, demonstrated that inconsistencies in the scale optimisation process were due to sampling bias of the training data. Therefore, since also in our study areas camera trap samplings were biased towards forested areas, protected areas and easily accessible areas, the discrepancies observed between different algorithms might be partly due to sampling bias, and further simulations are recommended.

A key point in machine learning is represented by the data, not only by the algorithms. The structure of the training data can have profound and critical effects on the outcomes of models trained with the RF algorithm. While this is an important point which is worthy of evaluation, our dataset was limited to a collection of camera trap data on several species which makes it difficult to comprehensively compare the effects of the data and its structure. We note that this is better explored with simulation, as in a recent paper from our group (Kumar et al., 2021) which indeed showed that the spatial pattern and shape of the environmental manifold itself affects performance of GLM and RF.

5. Conclusions

Our study used a vast empirical camera-trap data set for multiple felid species across three large geographical areas of Southeast Asia to test nine hypotheses about the effects of species distribution modelling algorithm and bootstrapping on model prediction and performance. The most important results of our study were that (1) there was not a clear pattern of RF being superior to GLM in model performance, (2) their predictions were spatially highly divergent and (3) there were large and systematic differences between GLM and RF in the scales of environmental variables they identified. These results are a cause of concern and caution in species distribution modelling. One of the premises of SDMs is that algorithms used provide objective and informative predictions of

the factors driving species-habitat relationships, and ability to accurately predict species distribution. If there are large and systematic differences in the scales of factors identified and the predicted probability surfaces produced between algorithms, irrespective of the underlying ecological relationships, then there is likely a much higher degree of uncertainty in SDM predictions than was previously appreciated. This suggests that much more care should be taken by analysts conducting SDM analyses to compare multiple methods to predict their particular data sets and select those that are shown to be most effective in that system and for those species. Altogether, these results are worrying since SDMs are frequently used to design conservation actions. We recommend further investigations of the effects of different factors on the outcomes of SDMs, preferably by applying simulation experiments in which the scales of the species-habitat relationships and the suitable habitats distribution are known.

Authors' contributions

LC and SAC conceived the ideas and designed the methodology. DWM conceived the larger framework on which the study is based and secured the funding. Individual field teams led by AJH, EA, ÖEC, PC, GRC, IAH, PPK, JHM, AR and CKWT collected the data. HMB helped collating the final dataset. LC analysed the data with substantial contribution of SAC. LC, with substantial contribution of SAC, DWM, AJH and ZK interpreted the results. LC led the writing, with substantial contribution of SAC, DWM, AJH and ZK. All authors contributed critically to the drafts and gave final approval for publication.

Funding

The majority of the team, as well as the data, were part of the core WildCRU effort supported principally by Robertson Foundation and Panthera grants to DWM.

Declaration of Competing Interest

The authors declare no conflicts of interest or competing interests.

Data availability

Given the extremely sensitive nature of species occurrence data with respect to illegal wildlife trade, locations of camera traps will not be made public to avoid further endangering the already threatened species. However, we welcome correspondence with scholars and conservations regarding collaborative use of the data to advance science and conservation of Southeast Asian felid species.

Acknowledgments

The data used in this analysis come from more than a decade-long camera-trapping study of clouded leopards and associated biodiversity funded principally by grants to DWM from the Robertson Foundation and Panthera. Primary findings are reported by the co-authors of Macdonald et al. (2018); (2019; 2020) and Chiaverini et al. (2022), all of whom, and the associated field teams, we gratefully acknowledge here. In particular, we thank the past and present leaders of WildCRU field teams, and our collaborators in each country, from west to east: Dr. Ewan Macdonald (UK, University of Oxford), Bhupendra Yadav (Nepal, WAP, DNPWC), Priya Singh (India, Ecosystems India), Saw Htun, Hla Naing, Kyaw Moe, Su Pan (Myanmar, WCS), Dr. Jan Kamler (Cambodia, FFI, WWF), Dr. Ahimsa Campos-Arceiz, Dr. Jamie Wadey, Gilmore Bolongon (Malaysia, DWNP, University of Nottingham (and colleagues co-authoring Tan et al. (2017))), Dr. D. Mark Rayan (Malaysia, WCS), Laurie Hedges (Malaysia, Rimba), Dr. Susan Cheyne (Kalimantan, Universitas Palangka Raya, Universitas Muhammadiyah Palangka Raya and Universitas Mularmawan Samarinda), Dr. Manabu Onuma (Japan,

NIES). We also wish to thank Freeland Foundation and Thailand's Department of National Parks, Wildlife and Plant Conservation. Without the field effort, and analyses that were undertaken by this large team, and its collaborators, to the original published papers, the current analysis would not have been possible.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecoinf.2023.102026>.

References

- Ash, E., Kaszta, Z., Noochdumrong, A., Redford, T., Chanteap, P., Hallam, C., Jaroensuk, B., Raksat, S., Srinoppawan, K., Macdonald, D.W., 2021. Opportunity for Thailand's forgotten tigers: assessment of the Indochinese tiger *Panthera tigris corbetti* and its prey with camera-trap surveys. *Oryx* 55, 204–211.
- Atzeni, L., Cushman, S.A., Bai, D.F., Wang, J., Chen, P.J., Shi, K., Riordan, P., 2020. Meta-replication, sampling bias, and multi-scale model selection: a case study on snow leopard (*Panthera uncia*) in western China. *Ecol. Evol.* 10, 7686–7712.
- Austin, M.P., 1987. Models for the analysis of species response to environmental gradients. *Vegetatio* 69, 35–45.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- 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. Model.* 450.
- Chiaverini, L., Macdonald, D.W., Bothwell, H.M., Hearn, A.J., Cheyne, S.M., Haidir, I., Hunter, L.T.B., Kaszta, Z., Macdonald, E.A., Ross, J., Cushman, S.A., 2022. Multi-scale, multivariate community models improve designation of biodiversity hotspots in the Sunda Islands. *Anim. Conserv.* 25, 660–679.
- Cushman, S.A., McGarigal, K., 2002. Hierarchical, multi-scale decomposition of species-environment relationships. *Landscape Ecol.* 17, 637–646.
- Cushman, S.A., Wasserman, T.N., 2018. Landscape applications of machine learning: comparing random forests and logistic regression in multi-scale optimized predictive modeling of American Marten Occurrence in Northern Idaho, USA. *Machine Learning for Ecology and Sustainable Natural Resource Management*. Springer, pp. 185–203.
- Cushman, S.A., Shirk, A.J., Landguth, E.L., 2013a. Landscape genetics and limiting factors. *Conserv. Genet.* 14, 263–274.
- Cushman, S.A., Wasserman, T.N., Landguth, E.L., Shirk, A.J., 2013b. Re-evaluating causal modeling with mantel tests in landscape genetics. *Diversity* 5, 51–72.
- Cushman, S.A., Macdonald, E., Landguth, E., Malhi, Y., Macdonald, D., 2017. Multi-scale prediction of forest loss risk across Borneo. *Landscape Ecol.* 32, 1581–1598.
- Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., 2007. Random forests for classification in ecology. *Ecology* 88, 2783–2792.
- Efron, B., 1982. The Jackknife, the Bootstrap, and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Elith, J., Leathwick, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Ann. Rev. Ecol. Evol.* 40, 677–697.
- Elliot, N.B., Cushman, S.A., Macdonald, D.W., Loveridge, A.J., 2014. The devil is in the dispersers: predictions of landscape connectivity change with demography. *J. Appl. Ecol.* 51, 1169–1178.
- ESRI, 2018. ArcGIS Desktop: Release 10.6.1. Environmental System Research Institute, Redlands, CA, USA.
- Evans, J.S., Cushman, S.A., 2009. Gradient modeling of conifer species using random forests. *Landscape Ecol.* 24, 673–683.
- Freedman, D.A., 1981. Bootstrapping regression models. *Ann. Stat.* 9, 1218–1228.
- Freeman, E., Moisen, G., 2007. PresenceAbsence: an R package for presence absence analysis. *J. Stat. Softw.* 23.
- Freeman, E.A., Moisen, G.G., Frescino, T.S., 2012. Evaluating effectiveness of down-sampling for stratified designs and unbalanced prevalence in random Forest models of tree species distributions in Nevada. *Ecol. Model.* 233, 1–10.
- Goslee, S.C., Urban, D.L., 2007. The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* 22, 1–19.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.
- Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135, 147–186.
- Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecol. Model.* 157, 89–100.
- Guisan, A., Thuiller, W., Zimmermann, N.E., 2017. *Habitat Suitability and Distribution Models: With Applications in R*. Cambridge University Press, Cambridge.
- Hastie, T., Tibshirani, R., 1990. *Generalized Additive Models*. Chapman and Hall, London.
- Hearn, A.J., Cushman, S.A., Ross, J., Goossens, B., Hunter, L.T.B., Macdonald, D.W., 2018. Spatio-temporal ecology of sympatric felids on Borneo. *PLoS One* 13.
- Hefley, T.J., Baasch, D.M., Tyre, A.J., Blankenship, E.E., 2014. Correction of location errors for presence-only species distribution models. *Methods Ecol. Evol.* 5, 207–214.
- Kaszta, Z., Cushman, S.A., Macdonald, D.W., 2020. Prioritizing habitat core areas and corridors for a large carnivore across its range. *Anim. Conserv.* 23, 607–616.

- Khosravi, R., Hemami, M.R., Cushman, S.A., 2019. Multi-scale niche modeling of three sympatric felids of conservation importance in central Iran. *Landsc. Ecol.* 34, 2451–2467.
- Kumar, S.U., Maini, P.K., Chiaverini, L., Hearn, A.J., Macdonald, D.W., Kaszta, Z., Cushman, S.A., 2021. Smoothing and the environmental manifold. *Ecol. Inform.* 66, Legendre, P., Legendre, L., 1998. *Numerical Ecology, Developments in Environmental Modelling 20*, 2nd English edition. ed. Elsevier, Amsterdam, p. 1 (online resource).
- Levin, S.A., 1992. The problem of pattern and scale in ecology. *Ecology* 73, 1943–1967.
- Macdonald, D.W., Bothwell, H.M., Hearn, A.J., Cheyne, S.M., Haidir, I., Hunter, L.T.B., Kaszta, A., Linkie, M., Macdonald, E.A., Ross, J., 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.
- Macdonald, D.W., Bothwell, H.M., Kaszta, Z., Ash, E., Bolongon, G., Burnham, D., Can, Ö. E., Campos-Arceiz, A., Channa, P., Clements, G.R., Hearn, A.J., Hedges, L., Htun, S., Kamler, J.F., Kawanishi, K., Macdonald, E.A., Mohamad, S.W., Moore, J., Naing, H., Onuma, M., Penjor, U., Rasphone, A., Mark Rayan, D., Ross, J., Singh, P., Tan, C.K.W., Wadey, J., Yadav, B.P., Cushman, S.A., 2019. Multi-scale habitat modelling identifies spatial conservation priorities for mainland clouded leopards (*Neofelis nebulosa*). *Divers. Distrib.* 0, 1–16.
- Macdonald, D.W., Chiaverini, L., Bothwell, H.M., Kaszta, Z., Ash, E., Bolongon, G., Can, O.E., Campos-Arceiz, A., Channa, P., Clements, G.R., Hearn, A.J., Hedges, L., Htun, S., Kamler, J.F., Macdonald, E.A., Moore, J., Naing, H., Onuma, M., Rasphone, A., Rayan, D.M., Ross, J., Singh, P., Tan, C.K.W., Wadey, J., Yadav, B.P., Cushman, S.A., 2020. Predicting biodiversity richness in rapidly changing landscapes: climate, low human pressure or protection as salvation? *Biodivers. Conserv.* 29, 4035–4057.
- McGarigal, K., Cushman, S.A., Ene, E., 2012. FRAGSTATS v4: spatial pattern analysis program for categorical and continuous maps. Computer software program produced by the authors at the University of Massachusetts. Amherst, MA. Available: <http://www.umass.edu/landeco/research/fragstats/fragstats.html>.
- 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.
- Osawa, T., Mitsuhashi, H., Uematsu, Y., Ushimaru, A., 2011. Bagging GLM: improved generalized linear model for the analysis of zero-inflated data. *Ecol. Inform.* 6, 270–275.
- Penjor, U., Kaszta, Z., 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.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guisera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schroder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F., Dormann, C.F., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40, 913–929.
- Stupariu, M.S., Cushman, S.A., Plesoianu, A.I., Patru-Stupariu, I., Furst, C., 2021. Machine learning in landscape ecological analysis: a review of recent approaches. *Landsc. Ecol.* 37, 1227–1250.
- Tan, C.K.W., Rocha, D.G., Clements, G.R., Brenes-Mora, E., Hedges, L., Kawanishi, K., Mohamad, S.W., Rayan, D.M., Bolongon, G., Moore, J., Wadey, J., Campos-Arceiz, A., Macdonald, D.W., 2017. Habitat use and predicted range for the mainland clouded leopard *Neofelis nebulosa* in Peninsular Malaysia. *Biol. Conserv.* 206, 65–74.
- Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guisera-Arroita, G., 2021. Modelling species presence-only data with random forests. *Ecography* 44, 1731–1742.
- Valavi, R., Guisera-Arroita, G., Lahoz-Monfort, J.J., Elith, J., 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecol. Monogr.* 92.
- Vaughan, I.P., Ormerod, S.J., 2005. The continuing challenges of testing species distribution models. *J. Appl. Ecol.* 42, 720–730.
- Vergara, M., Cushman, S.A., Urra, F., Ruiz-Gonzalez, 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.
- Wan, H.Y., Cushman, S.A., Ganey, J.L., 2019. Improving habitat and connectivity model predictions with multi-scale resource selection functions from two geographic areas. *Landsc. Ecol.* 34, 503–519.
- Wiens, J.A., 1989. Spatial scaling in ecology. *Funct. Ecol.* 3, 385–397.
- Xu, Y., Goodacre, R., 2018. On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *J. Anal. Test* 2, 249–262.