RESEARCH

Using machine learning to predict habitat suitability of sloth bears at multiple spatial scales

Tahir Ali Rather[1,2]* , Sharad Kumar[1,2] and Jamal Ahmad Khan[1]

Abstract

Background: Habitat resources occur across the range of spatial scales in the environment. The environmental resources are characterized by upper and lower limits, which define organisms' distribution in their communities. Animals respond to these resources at the optimal spatial scale. Therefore, multi-scale assessments are critical to identifying the correct spatial scale at which habitat resources are most influential in determining the species-habitat relationships. This study used a machine learning algorithm random forest (RF), to evaluate the scale-dependent habitat selection of sloth bears (*Melursus ursinus*) in and around Bandhavgarh Tiger Reserve, Madhya Pradesh, India.

Results: We used 155 spatially rarified occurrences out of 248 occurrence records of sloth bears obtained from camera trap captures ($n = 36$) and scats located ($n = 212$) in the field. We calculated focal statistics for 13 habitat variables across ten spatial scales surrounding each presence-absence record of sloth bears. Large (> 5000 m) and small (1000–2000 m) spatial scales were the most dominant scales at which sloth bears perceived the habitat features. Among the habitat covariates, farmlands and degraded forests were the essential patches associated with sloth bear occurrences, followed by sal and dry deciduous forests. The final habitat suitability model was highly accurate and had a very low out-of-bag (OOB) error rate. The high accuracy rate was also obtained using alternate validation matrices.

Conclusions: Human-dominated landscapes are characterized by expanding human populations, changing land-use patterns, and increasing habitat fragmentation. Farmland and degraded habitats constitute ~ 40% of the landform in the buffer zone of the reserve. One of the management implications may be identifying the highly suitable bear habitats in human-modified landscapes and integrating them with the existing conservation landscapes.

Keywords: Bandhavgarh, *Melursus ursinus*, Multi-scale, Habitat selection, Random forest, Sloth bear, Species distribution models

* Correspondence: murtuzatahiri@gmail.com
[1]Department of Wildlife Sciences, Aligarh Muslim University, Aligarh, Uttar Pradesh 202002, India
[2]The Corbett Foundation, 81-88, Atlanta Building, Nariman Point, Mumbai, Maharashtra 400021, India

Introduction

Sloth bears are endemic to the Indian sub-continent. About 90% of their current range occurs in India (Dharaiya et al. 2016) from the Western Ghats to the forests of the Shivalik ranges along the foothills of the Himalayas (Yoganand et al. 2006). Despite being a widely distributed bear species, the sloth bear has a patchy distribution across 20 states in India. The reduction in their range is attributed to forest fragmentation, continuous habitat loss, and human-caused mortalities (Dharaiya et al. 2016). Though no reliable population estimates are available for sloth bears in India, the total occupied area was earlier estimated at 2,000,000 km^2 (Johnsingh 2003; Akhtar et al. 2004). More recently, Sathyakumar et al. (2012) and Puri et al. (2015) reported the occupied area for sloth bears in India might be higher than 400,000 km^2. Sloth bears are confined to five distinct bio-graphical regions in India, namely northern, north-eastern, central, south-eastern, and south-western (Garshelis et al. 1999; Johnsingh 2003; Yoganand et al. 2006; Sathyakumar et al. 2012; Dharaiya et al. 2016).

Animals are known to select habitat resources across a range of spatial scales. Multiple factors drive the species distribution, with each being most influential at a specific spatial scale; thus, the apparent habitat-species relationships may change across spatial scales (Wiens 1989). The inclusion of scales is vital for understanding the species-habitat relationships (Schaefer and Messier 1995; Shirk 2012; Wasserman et al. 2012; Sánchez et al. 2014). The concept of scale in ecology is believed to be much older (e.g., see Schneider 2001) and is now recognized as a central theme in spatial ecology (Schneider 1994; Schneider et al. 1997; Schneider 1998; Cushman and McGarigal 2004).

For sloth bears, the habitat selection varies with seasonal food availability at a small spatial scale (Joshi et al. 1995; Akhtar et al. 2004; Yoganand et al. 2006; Ratnayeke et al. 2007; Ramesh et al. 2012). In our study area, insects (ants and termites) form a substantial portion of the sloth bear diet (Rather et al. 2020a). The distribution of ants and termites that sloth bear feeds on is also likely to be determined by fine-scale variables. On a larger scale, the occurrence of the sloth bears will likely be determined by factors such as forest cover, habitat connectivity, proximity to the human habitation, and so on (Puri et al. 2015). Johnson (1980) pointed out that species depend for their essential life-history functions and decisions on habitat features across a range of spatial scales. Often, organisms interact with all structures in their environment. The environmental resources are characterized by their upper and lower limits, which define the distribution and fitness of the organism in their communities (Mayor et al. 2009). Fitness is greatly influenced by the scales at which organisms select

habitat resources (Mayor et al. 2009). The optimal scale for each habitat feature may occur anywhere across the structured environmental continuum on the landscape (Boyce et al. 2003; Mayor et al. 2007). For example, Schaefer and Messier (1995) found habitat selection by muskoxen (*Ovibos moschatus*) to be consistent across scales in a relatively homogenous habitat, and contrastingly habitat selection by elk was found to be scale-dependent in a more structured landscape of Rocky Mountains (Boyce et al. 2003). Likewise, predators and prey species select habitat variables at different spatial scales (Hostetler and Holling 2000). Some authors (Fisher et al. 2011) argue that body size alone best explains the dominant scale of habitat selection among terrestrial mammals with a direct relationship between the body size and extent of scale. Thus, habitat selection quantified at one scale is often insufficient to predict habitat selection at another scale (Mayor et al. 2009). Thus, single-scale habitat selection may fail to identify the factors determining the species-habitat relationships correctly and lead to biased inferences. Therefore, multi-scale assessments are critical to identifying the correct spatial scale at which habitat resources are most influential in determining the species-habitat relationships.

To date, no multi-scale habitat assessment of sloth bears has been attempted in India except a recent nationwide occupancy survey of sloth bears conducted at two spatial scales (Puri et al. 2015). Habitat features such as forest cover, terrain heterogeneity, and human population density were reported to be influential on a large scale (Puri et al. 2015). A similar multi-scale distribution assessment using the random forest algorithm was attempted for Himalayan brown bears (*Ursus arctos isabellinus*) across their range in Himalayas (Dar et al. 2021). The study showed that habitat selection in brown bears was scale-dependent and brown bears perceived the habitat features across multiple spatial scales. Likewise, habitat selection of brown bears in Northwest Spain was found to be sensitive to the scale at which habitat variables were evaluated (Sánchez et al. 2014). In another similar study using resource selection functions (RSFs), the habitat selection by grizzly bears was also found to be scale-dependent (Ciarniello et al. 2007). The importance of multi-scale assessment in determining the species-habitat relationships has been demonstrated in a wide range of species (e.g., Wan et al. 2017; Klaassen and Broekhuis 2018; Khosravi et al. 2019; Atzeni et al. 2020; Rather et al. 2020b, 2020c; Ash et al. 2021; Dar et al. 2021).

The habitat selection studies of sloth bears at fine-scale have been carried out across many regions of its range (e.g., Joshi et al. 1995; Akhtar et al. 2004; Yoganand et al. 2006; Ratnayeke et al. 2007; Ramesh et al. 2012). These studies indicate that moist and dry

deciduous forests, human presence, seasonal availability of food resources, and termites were critical factors determining the habitat associations of sloth bears. Likewise, Das et al. (2014) found that the mean number of termite mounds and trees positively influenced the sloth bear occurrence in the Western Ghats.
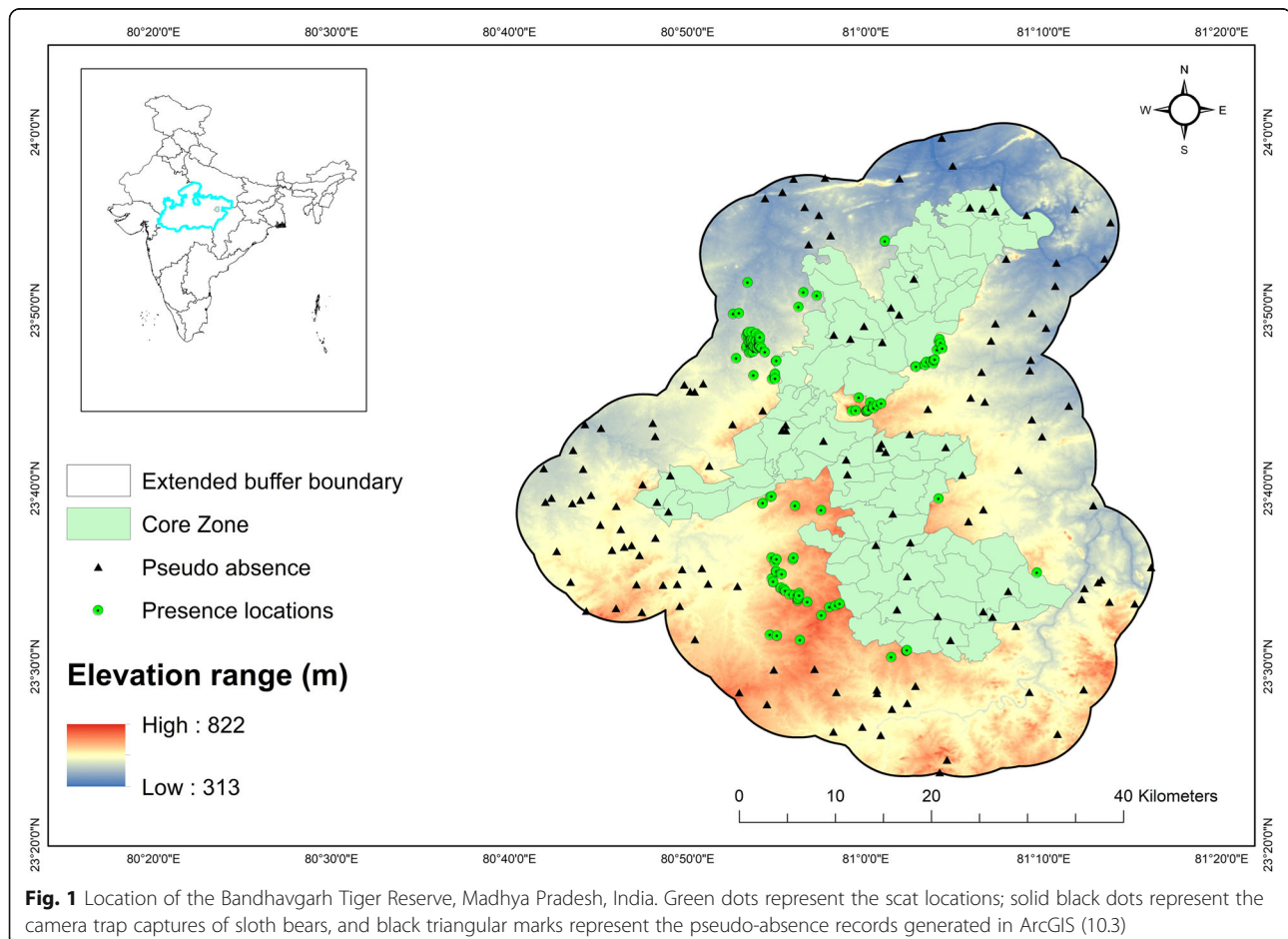
In this study, we used the random forest algorithm (Breiman 2001a, 2001b) to determine the habitat selection of sloth bears at multiple spatial scales in a largely anthropogenic region. Random forest is an ensemble of classification and regression trees (CART) based on bagging, which has generated considerable interest in the ecological community (Cutler et al. 2007). We aimed to evaluate the scale at which sloth bears respond to habitat variables. We hypothesized that sloth bears would respond to the habitat variables at various scales based on their ecological requirements. In achieving our objectives, we used random forest (RF), a highly accurate bagging classification algorithm with a suite of 13 habitat variables, to build a multi-scale suitability model for sloth bears. RF performs better when executed as classification rather than regression. Traditionally, logistic regression was the dominant statistical

approach in assessing multi-scale habitat associations (Hegel et al. 2010; McGarigal et al. 2016). RF is a non-parametric approach and does not assume independence. Thus, the inherent spatial bias associated with habitat selection data does not affect the model predictions significantly. RF produces accurate model predictions without overfitting (Breiman 2001a). RF is a bootstrap-based machine learning algorithm utilizing the decision tree-based bagging technique and has been reported to outperform traditional logistic regression approaches (Evans et al. 2011; Cushman et al. 2017; Cushman and Wassermann 2018) and resource selection function (Manly et al. 1993).

Materials and methods
Study area
The study was conducted in and around the Bandhavgarh Tiger Reserve (BTR), Madhya Pradesh, India (Fig. 1). The reserve's core zone includes the Panpatha Wildlife Sanctuary (PWS) in the North and Bandhavgarh National Park (BNP) in the South, with an area of 716 km^2. The surrounding buffer zone has an area of 820 km^2, adding the reserve's total size to

Fig. 1 Location of the Bandhavgarh Tiger Reserve, Madhya Pradesh, India. Green dots represent the scat locations; solid black dots represent the camera trap captures of sloth bears, and black triangular marks represent the pseudo-absence records generated in ArcGIS (10.3)

1536 km^2. The reserve is located between 23° 27′ 00″ and 23° 59′ 50″ north latitude and 80° 47′ 75″ to 81° 15′ 45″ east longitude in the Umaria, Shahdol, and Katni districts of Madhya Pradesh, in Central India. A detailed account of the study area is available in Rather et al. (2020b). The primary habitat types in the reserve are sal-dominated forests, sal-mixed forest, moist and dry deciduous forests, grasslands, riverine patches across the streams, and bamboo dominant forest patches across the slopes of the hillocks. The buffer zone is highly anthropogenic and consists of ∼ 160 villages. Approximately 40% of the land use category within the buffer zone is classified as agricultural fields interspersed with degraded forest patches (Supplemental Information 1). Fragmented and degraded territorial forest divisions further surround the buffer zone.

Sloth bear occurrence records and pseudo-absences

We used the scat locations of the sloth bears collected in the study area as species occurrence records. Scats were collected randomly as and when encountered within the study area between 2016 and 2018. Due care was observed to collect the scats in all significant habitats present within the study area. A detailed description of the sampling approach is available in Rather et al. (2020a). The additional species occurrence records were obtained from camera trap captures. The camera traps (Cuddeback™, Model C1) were deployed in 2 × 2 km grids overlaid the entire study area in ArcGIS (10.3). Camera trap sampling was carried out from 2016 to 2017. A total of 25 pair of camera traps was placed systematically within the buffer zone. Camera traps remained active 24 h a day, except for a few stations where the theft risk was high. Each camera trap session consisted of eight consecutive trap days/nights.

The main objective of the camera trap sampling was to estimate the density of the tiger and leopard within the study area (Rather et al. 2021). A total effort of 2211 trap nights resulted in 36 photo captures of the sloth bear. A total of 212 occurrences of the sloth bear were based on the scat locations, and 36 captures of the sloth bears were obtained during one year of camera trap sampling. We implemented spatial filtering using the SDM toolbox (v2.3) in ArcGIS (10.3) to remove the duplicated and aggregated occurrence records. Random forest is a highly accurate bagging algorithm and is not affected by model overfitting (Breiman 2001a). Out of 248 occurrence records, we retained a total of 155 spatially rarified occurrences of the sloth bear for further modeling. Out of 155 rarified occurrences, most of the records were retained from scats locations ($n = 130$),

and only 25 presence records were of the camera trap captures.

The actual species absence records of large animals are challenging to obtain. Thus, we created the pseudo-absence records for sloth bears in ArcGIS (10.3) using the following procedure. We created a circular buffer of a 500-m radius around each presence records (spatially rarified) and then generated 550 absence records in the first step. Any of the pseudo-absence points that occurred within these 500-m radius buffers around the presence locations were removed, and we considered only the pseudo-absences that occurred at least at the distance of 500 m from the presence locations to reduce spatial dependence. The imbalance between presence-absence classes has been proven to reduce the power of ensemble learners (Chawla 2005). Building on Chen et al. (2004) and Chawla (2005), we further removed the absence points to obtain an approximately balanced set of presence and absence records to avoid the problems arising due to imbalance data (Chawla et al. 2003). Finally, we retained a total of 155 spatially rarified presence records and an equal number of pseudo-absence points.

Predictors of sloth bear distribution

We considered the variables reported to be strong predictors of sloth bear distribution in the Indian subcontinent. The variables are based on the previous habitat selection studies of sloth bears (Joshi et al. 1995; Akhtar et al. 2004; Yoganand et al. 2006; Ratnayeke et al. 2007; Ramesh et al. 2012). Based on these studies, we limited the number of variables to 13 and did not consider the commonly used bio-climatic variables. We included topographic, vegetation (land cover classification), and anthropogenic variables in sloth bear habitat modeling (Table 1). We downloaded the digital elevation map (DEM) at 90-m resolution from Shuttle Radar Topography Mission (SRTM) elevation database (http://srtm.cs.cgiar.org). Other topographic features such as slope, aspect, and terrain ruggedness were derived from the elevation layer using surface analysis tools in the Spatial Analyst toolbox in ArcGIS (10.3). The land use land cover (vegetation layer) was obtained from the Indian Institute of Remote Sensing (IIRS, http://iirs.gov.in). We used the line density tool in ArcGIS (10.3) to calculate the road and river density within the study area's spatial extent at 1000 and 2000 m spatial scales. All the variables were resampled at the spatial resolution of 90 m using the SDM toolbox in ArcGIS (10.3). The choice of grain size or spatial resolution of variables is usually based on the data availability (Mayer and Cameron 2003) rather than species' ecology or the scale of the study. Bio-climatic variables were not included in the analysis due to their limited capability and relevance in

Table 1 Predictor variables included in the random forests modeling and the scales retained in the univariate scaling step of sloth bears in Bandhavgarh Tiger Reserve

Variable type	Variables	Optimal scale (m)	Abbreviations
Topographic	Elevation	2000	elevation2km
	Terrain ruggedness index	6000	rug6km
	River density	1000	river1km
Cover	NDVI in summer season	5000	ndvisum5km
	NDVI in winter season	6000	ndviwin6km
	NDVI in wet season	8000	ndviwet8km
Habitat composition	Dry deciduous forests	1000	drydec1km
	Moist deciduous forests	4000	moistdec4km
	Sal-dominated forests	5000	sal5km
Disturbance	Degraded habitat patches	8000	degraded8km
	Farmland	9000	farmland9km
	Road density	1000	road1km
	Human settlements	6000	settlement6km

determining the sloth bear distribution in a small study area.

Multi-scale data processing

We calculated the focal statistics for each variable across ten spatial scales surrounding each location (presence/pseudo-absence) using a moving window analysis with a focal statistic tool in ArcGIS (10.3). At each sloth bear presence-absence (PA) record, we calculated focal statistics for 13 variables (Table 1) using ten circular buffer radii. The radii of the circular buffers surrounding each PA record varied from 1000 m (smallest spatial scale) to 10,000 m (largest spatial scale). The focal statistics' output was the raster layers of each predictor variable at ten spatial scales and a .dbf file of extracted raster values around each PA location of sloth bear (Supplemental Information 2). In doing so, we extracted each of the 13 variables at ten spatial scales. In the next step, we ran a series of univariate RF models using the package randomForest in R (Liaw and Wiener 2002) for each variable across ten spatial scales (1000–10,000 m). The best scale was selected based on the lowest out-of-bag (OOB) error rate (McGarigal et al. 2016).

In univariate RF analysis, we used the PA record of sloth bear as a dependent variable. We executed the RF algorithm as classification while using each predictor variable separately at ten spatial scales calculated in the first step. This step was repeated 13 times for all variables to extract them at ten spatial scales. Thus, a total of 130 univariate RF models were constructed for 13 variables. In the final step, we selected the best scale having the lowest OOB error rate of each predictor variable among the ten spatial scales.

Since we were working with a relatively small data set, we used model improvement ratio (MIR) (Murphy et al. 2010) to measure each variable's relative predictive strength across ten scales. MIR is used to calculate the permuted variable importance represented by the mean decrease in OOB error rates, standardized from zero to one. The OOB error rates are often used to assess the predictive performance of RF models. A detailed discussion of OOB error rates can be found in Breiman (1996a, 1996b). In the next step, we built multivariate RF models using the sloth bear PA as a function of scale optimized predictor variables calculated during univariate RF analysis in R (R core team 2019).

We tested mutual correlation among all possible pairs of predictor variables using the R package rfUtilities (Evans and Murphy 2018). The highly correlated predictor variables ($r > 0.5$) were consequently removed from further analysis. To deal with the problems arising from model overfitting due to the small data set, we used the MIR technique as a model selection procedure. In the model selection process using MIR, the variables were subset using 0.10 increments of MIR values, and all variables above this threshold were retained for each model (Murphy et al. 2010). This subset was always performed on the original model's variable importance to avoid overfitting. Comparisons were made between each subset model, and the model with the lowest OOB error rate and lowest maximum within-class error was selected as the final model (Murphy et al. 2010). In the last step, the model predictions were created using the ratio of majority votes to create a probability distribution of sloth bear.

We also determined the minimum number of trees required by testing 10,000 bootstrap samples to examine

when OOB error rates ceased to improve. The OOB error rates stabilized between 1000–1500 trees (Supplemental Figure 1), and subsequently, in all our models, we used 2000 trees.

Model assessment

We assessed model fit by random permutations ($n = 99$) and cross-validation by adopting a resampling approach (Evans and Murphy 2018). For each validation, one tenth of the data was withheld as a validation set for every permutation. We obtained the following suite of performance matrices as model fit, specificity (proportion of observed negatives correctly predicted), sensitivity (proportion of observed positives correctly predicted), area under curve (AUC), the resource operating characteristic curve (ROC), Kappa statistics, and true skill statistic (TSS).

Results

A total of ten spatial scales (1000–10,000 m) for each predictor variable were chosen for the univariate analysis. For each predictor variable, the scale selection was based on the lowest OBB error rate except road and river density, where only two scales (1000, 2000 m) were retained for the multivariate model. In the final model, three scales at a small spatial extent (1000 m), one scale at intermediate spatial extent (4000 m), and three scales were selected at the broader spatial extent (> 5000 m) (Fig. 2).

Multivariate modeling and habitat suitability

We used MIR as an approach of variable selection in the multivariate RF model. Out of 13 original variables, only seven variables were retained in the final multivariate model (Fig. 3).

The RF model predicted 28% of the reserve's buffer area to be a suitable habitat for sloth bears, accounting for 43,669 ha. Suitable areas for sloth bears included sal-dominated, moist, and dry deciduous forests with water availability and moderate presence of roads. A substantial suitable area for sloth bears in the buffer zone also included degraded forest patches and farmlands (mosaic of natural vegetation and cropland). The highly suitable habitat for sloth bears was predicted in the Panpatha wildlife sanctuary in the north, which forms the reserve's core zone (Fig. 4). Suitable habitats were also located along the western part of the reserve in the buffer zone extending towards the reserve's southern boundary (Fig. 4).

Partial dependency plots

Farmlands (mosaic of natural vegetation and croplands) and degraded forest patches represent > 40% of the total buffer area and, expectedly, were predicted to be positively associated with sloth bear occurrence. Variables considered proxy of anthropogenic disturbances such as degraded habitats, farmlands, and road density were positively associated with sloth bear occurrences (Fig. 5). Variables such as sal forests and moist and dry deciduous forests had no apparent positive association with the sloth bear occurrences. The sloth bear occurrences were predicted at very low percentages of these available habitat types (Fig. 6). Moist deciduous forests, in particular, did not influence the predicted occurrences (Fig. 6).
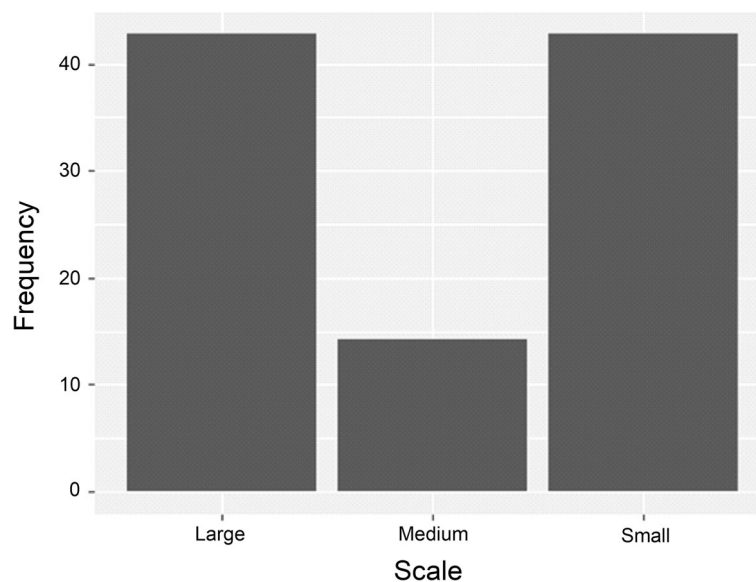

Fig. 2 Frequency of selected scales (in meters) across all variables for the random forest model
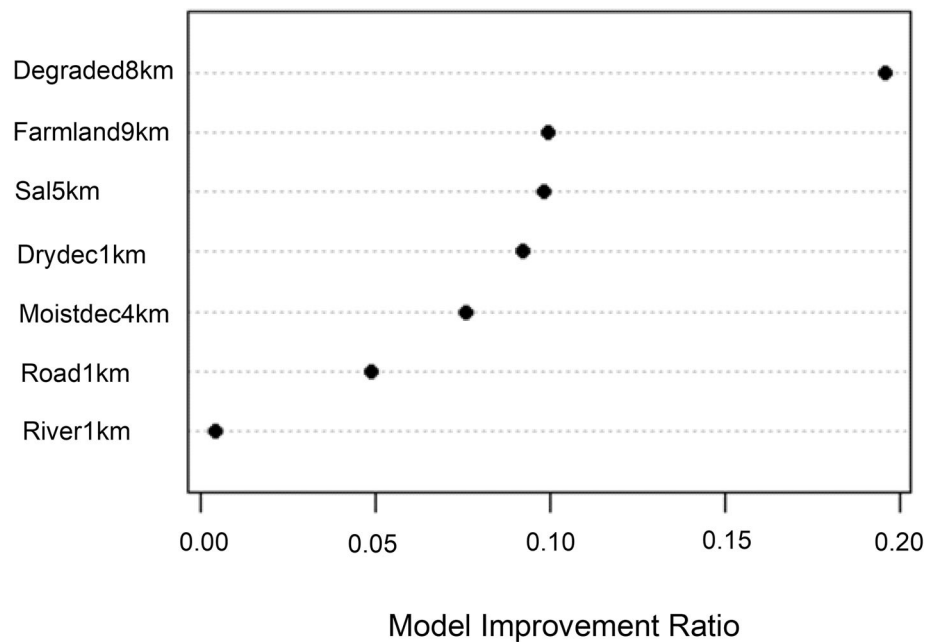
Fig. 3 Variable importance plot for scaled variables used in the multivariate random forest model of sloth bears based on model improvement ratio (MIR). The degraded forest was the most important variable, and the river density was the least important variable. Rest of the variables are listed in order of their relative importance to degraded forests. The *X*-axis represents the relative additional model improvement with the addition of each successive variable. Variables included are degraded8km, degraded forests; Farmland_9km, farmlands; sal5km, sal-dominated forests; drydec1km, dry deciduous forests; moistdec4km, moist deciduous forests; road1km, road density; and river1km, river density. The numerical value succeeding each variable represents the respective spatial scale

Model assessment

The model for predicting sloth bear occurrences was well supported and significant ($P < 0.001$). The model performed exceptionally well and had low model OOB error rates and high AUC, TSS, and Kappa statistic values (Table 2).

Discussion

Our results are consistent with similar studies arguing that habitat selection measured at one specific scale may be insufficient to predict that selection at another scale (Mayor et al. 2009). Similar studies for brown bears (Martin et al. 2012; Sánchez et al. 2014); Dar et al. 2021) and other species (Shirk 2012; Shirk et al. 2014; Wan et al. 2017; Klaassen and Broekhuis 2018) also support the scale-dependent habitat selection. Consistent with these studies, our results indicate that habitat selection occurs across the range of scales for sloth bears, thus supporting our hypothesis of scale-dependent habitat selection in sloth bears. In this study, habitat features such as access to water and travel routes used for daily ranging patterns were perceived at fine-scale corresponding to fourth-order selection of habitat variables (Johnson 1980). Likewise, the foraging patches such as sal forests and moist and deciduous forests may correspond to the third and second-order selection of habitat variables for sloth bears and so on.

The selection of habitat variables at different scales may also depend on the variation in the distribution of the habitat resources (Johnson 1980; Mayor et al. 2009). The spatial and seasonal variation in the availability of food resources may explain the high predicted occurrences of sloth bears in farmlands and degraded habitats. Farmlands and degraded habitats in our study area are characterized by large patches of invasive weed *Lantana camera*. Fruits of *Lantana camera* were consumed by sloth bears in the winter season, and the fruits of the most frequently occurring plant species were consumed in the summer season (Rather et al. 2020a). In winter, sloth bears primarily showed dependence on insects (ants and termites), *L. camera*, and *Ziziphus mauritiana*, all of which occurred at high abundance in the buffer zone. Thus, high predicted occurrences of sloth bears in disturbed habitats might have been due to the only food items available in such habitats during the winter season. Secondly, the farmlands and the degraded habitats represent ~ 40% of the reserve's buffer area, and thus a substantial portion of the sloth bear occurrences was recorded in such habitats. The *Lantana* patches are reportedly used as resting, denning, and foraging sites by sloth bears (Yoganand 2005; Akhtar et al. 2007; Ratnayeke et al. 2007). Lastly, under no circumstances does our study implicate increasing farmlands' area to conserve sloth bears in disturbed habitats.

Fig. 4 Predicted habitat suitability of sloth bears in and around Bandhavgarh Tiger Reserve. Red color indicates high suitability, and blue color indicates low suitability
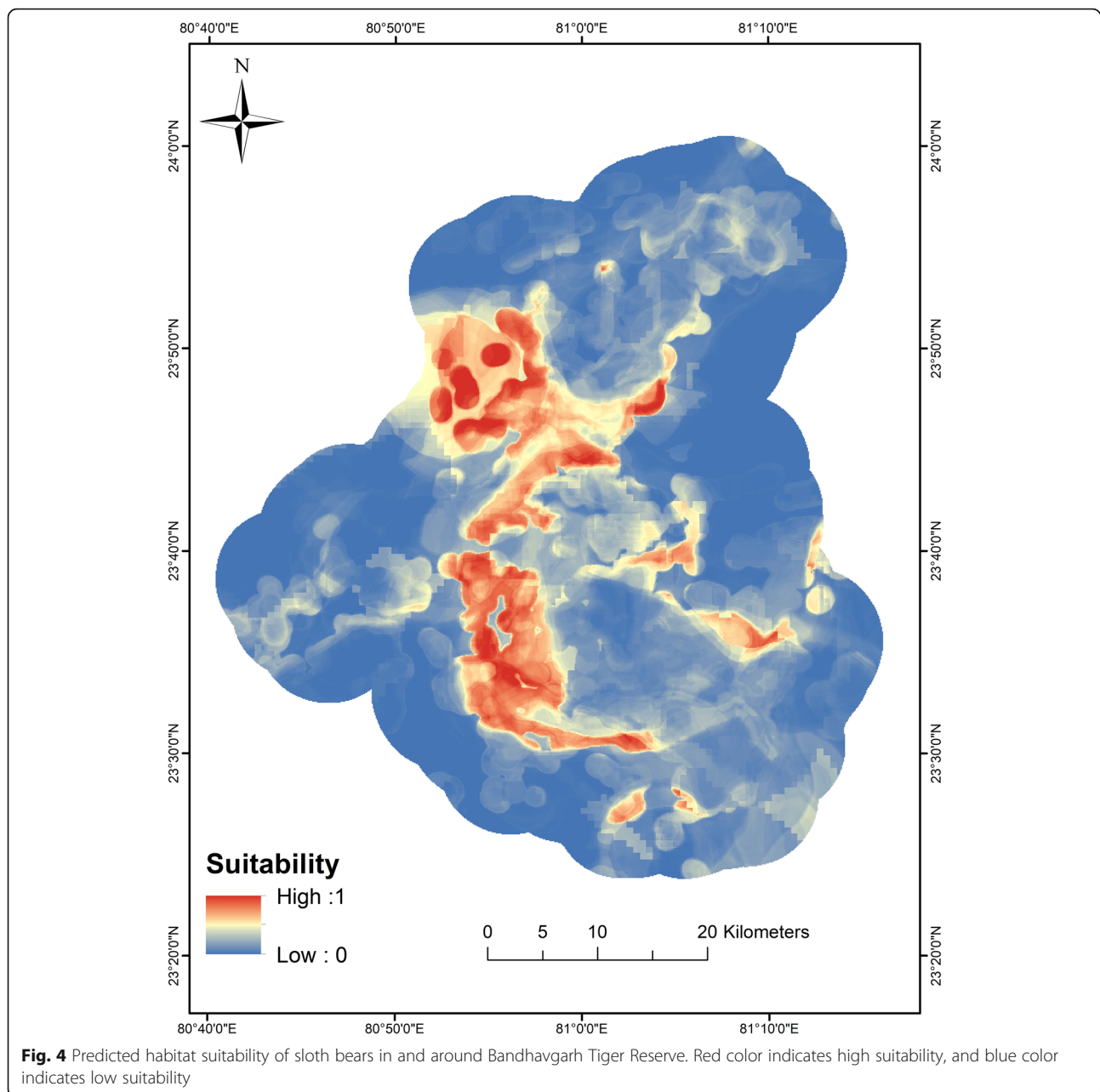
The habitat variables used in the multivariate model were based on the previous habitat selection studies of sloth bears (Joshi et al. 1995; Akhtar et al. 2004; Yoganand et al. 2006; Ratnayeke et al. 2007; Puri et al. 2015). Overall, sal, moist, and dry deciduous forests are positively associated with sloth bear occurrences across their range. However, in largely disturbed regions, these habitats represent only a small portion of the total area, thus making the species-habitat relationships complicate to predict or, in this case to conflict with previous studies. Therefore, our results are site-specific and make more sense when

applied to the disturbed regions. Puri et al. (2015) also point that sloth bears are not limited by protected areas and occur widely in unprotected, human-use habitats.

Only 28% of the total buffer area was predicted to be suitable for sloth bears. Like previous studies, suitable habitats were predicted to occur in sal, moist, and dry deciduous forests. However, these habitats were predicted to be weak determiners of sloth bear occurrence. We suspect this ambiguity to be related to the small percentage of these habitats in the buffer zone of the reserve. Positive association of sloth bears
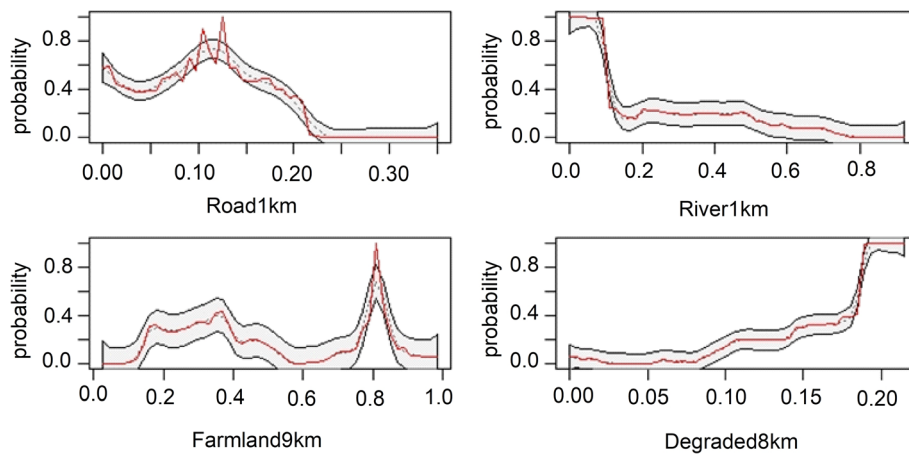
Fig. 5 Partial dependency plots for road density, river density, farmland, and degraded forest patches. The partial dependency plots represent the marginal effect of each habitat variable while keeping the effect of other variables at their average value. The shaded gray region represents 95% confidence intervals, and the red line indicates the average value. The *X*-axis represents the percentage of the variables ranging from 0 to 1%, and *Y*-axis represents the predicted probability of sloth bear occurrence

with farmlands and degraded habitats and thereof high suitability in such habitats may not be considered a general norm of sloth bear ecology. Sloth bears are reported to occur and use disturbed habitats across many areas of their range in India (Akhtar et al. 2004).

Species distribution models that relate species occurrence data to environmental variables are now essential tools in distributional and spatial ecology (Guisan and Zimmermann 2000; Elith et al. 2006; Drew et al. 2011). RF has been shown to perform better than other popular algorithms under the conditions of low occurrence data. The nationwide assessment of sloth bears using the traditional occupancy modeling approach conducted

by Puri et al. (2015) predicted sloth bear occurrences in Gir forests which are known not to harbor any sloth bear population. Likewise, Mi et al. (2017) implemented random forest for 33 records of Hooded Crane (*Grus monacha*), 40 records of white-naped crane (*Grus vipio*), and 75 records of black-necked crane (*Grus nigricollis*). They found that random forest performed exceptionally well than TreeNet, Maxent, and CART. Thus, comparatively low occurrence data used in this study would not have influenced model predictions largely. Our model assessment matrices also indicate better performance of the RF algorithm in producing accurate predictive maps under the conditions of low sampling intensity.
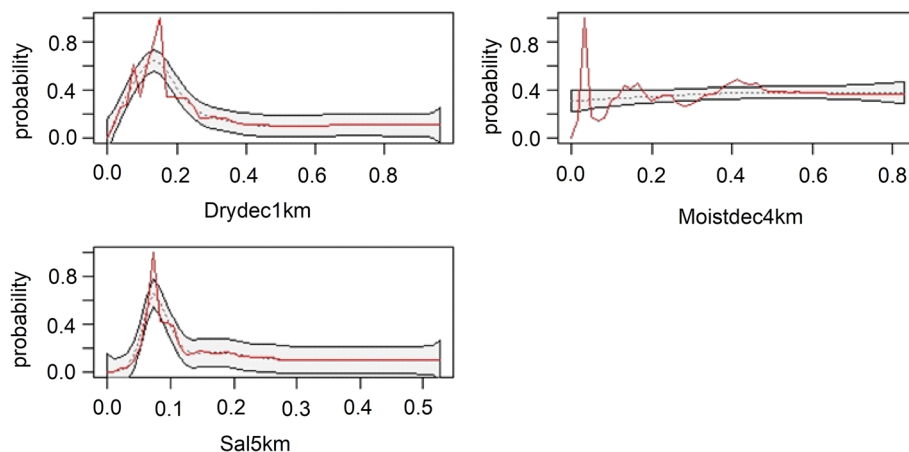

Fig. 6 Partial dependency plots for dry deciduous forests, moist deciduous forests, and sal-dominated forests. The shaded gray region represents 95% confidence intervals, and the red line indicates the average value. The *X*-axis represents the percentage of the variables ranging from 0 to 1%, and *Y*-axis represents the predicted probability of the sloth bear occurrence

Rather *et al. Ecological Processes*   (2021) 10:48

Page 10 of 12

Table 2 Model validation metrics including model OOB error, sensitivity, specificity, Kappa, TSS, AUC, and significance (*P*), for sloth bear model

Performance matrix	Value
Accuracy (PCC)	97.41%
Sensitivity	0.98[a]
Specificity	0.96[a]
Cohen's Kappa	0.94
True skill statistic (TSS)	0.94
Area under curve (AUC)	0.97[b]
Cross-validation Kappa	0.94
Cross-validation OBB error rate	0.03
Cross-validation error variance	2E.05
P value	0.001

[a]Supplemental Figure 2
[b]Supplemental Figure 3

Limitations, conclusion, and management implications

One of the significant limitations of our study is biased sampling in highly anthropogenic habitats, which may lead to conflicting results compared to other studies conducted in less disturbed areas. Thus, we recommend a caution when inferences are drawn from such studies. Nevertheless, this study still improves our understanding of the sloth bear habitat relationships on a multi-scale approach in a largely anthropogenic landscape. One of the management priorities should be identifying and protecting suitable habitats in disturbed regions and integrating the human-modified landscapes with the existing conservation landscape network as suggested by previous studies. Researchers may undertake the suitability assessments of sloth bears on a much larger scale in future.

Abbreviations
AUC: Area under curve; BTR: Bandhavgarh Tiger Reserve; OOB: Out-of-bag error; RF: Random forest; TSS: True skill statistic

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13717-021-00323-3.

Additional file 1: Supplemental Information 1. Area calculation – EP.

Additional file 2: Supplemental Information 2. PA data of Sloth bear.

Additional file 3: Supplemental Figure 1. Bootstrap Error convergence rate.

Additional file 4: Supplemental Figure 2. Area Under ROC Curve.

Additional file 5: Supplemental Figure 3. MDS Plot.

References
Akhtar N, Bargali HS, Chauhan NPS (2004) Sloth bear habitat use in disturbed and unprotected areas of Madhya Pradesh, India. Ursus 15(2):203–211. https://doi.org/10.2192/1537-6176(2004)015<0203:SBHUID>2.0.CO;2

Akhtar N, Bargali HS, Chauhan NPS (2007) Characteristics of sloth bear day dens and use in disturbed and unprotected habitat of North Bilaspur Forest Division, Chhattisgarh, Central India. Ursus 18(2):203–208. https://doi.org/10.2192/1537-6176(2007)18[203:COSBDD]2.0.CO;2

Ash E, Macdonald DW, Cushman SA, Noochdumrong A, Redford T, Kaszta Z (2021) Optimization of spatial scale, but not functional shape, affects the performance of habitat suitability models: a case study of tigers (*Panthera tigris*) in Thailand. Lands Ecol 36(2):455–474. https://doi.org/10.1007/s10980-020-01105-6

Atzeni L, Cushman SA, Bai D, Wang P, Chen KS, 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(14):7686–7712. https://doi.org/10.1002/ece3.6492

Boyce MS, Mao JS, Merrill EH, Fortin D, Turner MG, Fryxell JM, Turchin P (2003) Scale and heterogeneity in habitat selection by elk in Yellowstone National Park. Écoscience 10(4):421–431. https://doi.org/10.1080/11956860.2003.11682790

Breiman L (1996a) Out-of-bag estimation 1–13. https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf

Breiman L (1996b) Bagging predictors. Mach Learn 24(2):123–140. https://doi.org/10.1007/BF00058655

Breiman L (2001a) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Breiman L (2001b) Statistical modeling: the two cultures. Stat Sci 16:199–231

Chawla NV (2005) Data mining for imbalanced datasets: an overview. In: Maimon O, Rokach L (eds) Data mining and knowledge discovery handbook. Springer, Boston. https://doi.org/10.1007/0-387-25465-X_40

Chawla NV, Lazarevic A, Hall LO, Bowyer KW (2003) SMOTEboost: improving prediction of the minority class in boosting. In: Lavrac D, Gamberger L, Todorovski, H Blockeel (eds) PKDD 2003. 7th European conference on principles and practice of knowledge discovery in databases. Lecture notes in computer science Vol 2838. Springer, Berlin, pp 107–119.

Chen C, Liaw A, Breiman L (2004) Using random forest to learn imbalanced data. http://oz.berkeley.edu/users/chenchao/666.pdf

Ciarniello LM, Boyce MS, Seip DR, Heard DC (2007) Grizzly bear habitat selection is scale dependent. Ecol Appl 17(5):1424–1440. https://doi.org/10.1890/06-1100.1

Core Team R (2019) R: A language and environment for statistical computing. R Foundation for Statistical computing, Vienna. https://www.R-project.org/

Cushman SA, Macdonald EA, Landguth EL, Halhi Y, Macdonald DW (2017) Multiple-scale prediction of forest-loss risk across Borneo. Lands Ecol 32(8): 1581–1598. https://doi.org/10.1007/s10980-017-0520-0

Cushman SA, McGarigal K (2004) Patterns in the species–environment relationship depend on both scale and choice of response variables. Oikos 105(1):117–124. https://doi.org/10.1111/j.0030-1299.2004.12524.x

Cushman SA, Wasserman TN (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. In: Humphries GRW et al (eds) Machine learning for ecology and sustainable natural resource management. Springer, New York. https://doi.org/10.1007/978-3-319-96978-7_9

Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ (2007) Random forests for classification in ecology. Ecology 88(11):2783–2792. https://doi.org/10.1890/07-0539.1

Dar SA, Singh SK, Wan HY, Kumar V, Cushman SA, Sathyakumar S (2021) Projected climate change threatens Himalayan brown bear habitat more than human land use. Anim Conserv. https://doi.org/10.1111/acv.12671

Das S, Dutta S, Sen SJ, Babu H, Kumar A, Singh M (2014) Identifying regions for conservation of sloth bears through occupancy modelling in north-eastern Karnataka, India. Ursus 25(2):111–120. https://doi.org/10.2192/URSUS-D-14-00008.1

Dharaiya N, Bargali HS, Sharp T (2016) *Melursus ursinus*. The IUCN Red List of Threatened Species 2016:e.T13143A45033815. https://doi.org/10.2305/IUCN.UK.2016-3.RLTS.T13143A45033815.en

Drew CA, Wiersma Y, Huettmann F (2011) Predictive species and habitat modelling in landscape ecology: concepts and applications. Springer, London. https://doi.org/10.1007/978-1-4419-7390-0

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Mortiz C, Nakamura M, Nkazawa Y, Overton JM, Peterson AT, Philips SJ, Richardson K, Scachetti-pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29(2):129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Evans JS, Murphy MA (2018) rfUtilities. R package version. 2.1–3. https://cran.rproject.org/package=rfUtilities

Evans JS, Murphy MA, Holden ZA, Cushman SA (2011) Modeling species distribution and change using random forest. In: Drew CA (ed) Predictive species and habitat modeling in landscape ecology: concepts and applications. Springer, New York

Fisher JT, Anholt B, Volpe JP (2011) Body mass explains characteristic scales of habitat selection in terrestrial mammals. Ecol Evol 1(4):517–528. https://doi.org/10.1002/ece3.45

Garshelis DL, Joshi AR, Smith JLD, Rice CG (1999) Sloth bear conservation action plan. In: Servheen C, Herrero S, Peyton B (eds) Bears: Status survey and conservation action plan. International Union for the Conservation of Nature and Natural Resources, Gland, Switzerland. pp. 225–240

Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135(2):147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hegel TM, Cushman SA, Huettmann F (2010) Current state of the art for statistical modelling of species distributions. In: Cushman SA, Huettman F (eds) Spatial complexity, informatics and wildlife conservation. Springer, Tokyo, pp 273–312

Hostetler M, Holling CS (2000) Detecting the scales at which birds respond to structure in urban landscapes. Urban Ecosyst 4:25–54

Johnsingh AJT (2003) Bear conservation in India. J Bombay Nat Hist Soc 100:190–201

Johnson DH (1980) The comparison of usage and availability measurements for evaluating resource preference. Ecology 61(1):65–71. https://doi.org/10.2307/1937156

Joshi AR, Garsheils DL, Smith JLD (1995) Home ranges of sloth bears in Nepal: Implications for conservation. J Wildl Manage 59(2):204–214. https://doi.org/10.2307/3808932

Khosravi R, Hemani MR, Cushman SA (2019) Multi-scale niche modeling of three sympatric felids of conservation importance in central Iran. Lands Ecol 34: 2451–2467

Klaassen B, Broekhuis F (2018) Living on the edge: multiscale habitat selection by cheetahs in a human-wildlife landscape. Ecol Evol 8(15):7611–7623. https://doi.org/10.1002/ece3.4269

Liaw A, Wiener M (2002) Classification and regression by random forest. R News 2(3):18–22. https://cogns.northwestern.edu/cbmg/LiawAndWiener2002.pdf

Manly BFJ, McDonald LL, Thomas DL (1993) Resource selection by animals: statistical design and analysis for field studies. Chapman & Hall, London. https://doi.org/10.1007/978-94-011-1558-2

Martin J, Revilla E, Quenette PY, Naves J, Allaine D, Swenson JE (2012) Brown bear habitat suitability in the Pyrenees: transferability across sites and linking scales to make the most of scarce data. J Appl Ecol 49:621–631

Mayer AL, Cameron GN (2003) Consideration of grain and extent in landscape studies of terrestrial vertebrate ecology. Landsc Urban Plan 65(4):201–217. https://doi.org/10.1016/S0169-2046(03)00057-4

Mayor SJ, Schaefer JA, Schneider DC, Mahoney SP (2007) Spectrum of selection: new approaches to detecting the scale-dependent response to habitat. Ecology 88(7):1634–1640. https://doi.org/10.1890/06-1672.1

Mayor SJ, Schneider DC, Schaefer JA, Mahoney SP (2009) Habitat selection at multiple scales. Écoscience 16(2):238–247. https://doi.org/10.2980/16-2-3238

Mcgarigal K, Wan HY, Zeller KA, Timm BC, Cushman SA (2016) Multi-scale habitat modeling: a review and outlook. Lands Ecol 31:1161–1175

Mi C, Huettmann F, Guo Y, Han X, Wen L (2017) Why to choose random forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. PeerJ 5:e2849. https://doi.org/10.7717/peerj.2849

Murphy MA, Evans JS, Storfer A (2010) Quantifying *Bufo boreas* connectivity in Yellowstone National Park with landscape genetics. Ecology 91(1):252–261. https://doi.org/10.1890/08-0879.1

Puri M, Arivathsa A, Karanth KK, Kumar NS, Karanth KU (2015) Multiscale distribution models for conserving widespread species: the case of sloth bear *Melursus ursinus* in India. Divers Distrib 21(9):1087–1100. https://doi.org/10.1111/ddi.12390

Ramesh T, Kalle R, Sankar K, Qureshi Q (2012) Factors affecting habitat patch use by sloth bears in Mudumalai Tiger Reserve, Western Ghats, India. Ursus 23(1): 78–85. https://doi.org/10.2192/URSUS-D-11-00006.1

Rather TA, Kumar S, Khan JA (2020b) Multi-scale habitat modelling and predicting change in the distribution of tiger and leopard using random forest algorithm. Sci Rep 10(1):11473. https://doi.org/10.1038/s41598-020-68167z

Rather TA, Kumar S, Khan JA (2020c) Multi-scale habitat selection and impacts of climate change on the distribution of four sympatric meso-carnivores using random forest algorithm. Ecol Process 9:60. https://doi.org/10.1186/s13717-020-00265-2

Rather TA, Kumar S, Khan JA (2021) Density estimation of tiger and leopard using spatially explicit capture-recapture framework. PeerJ 9:e10634. https://doi.org/10.7717/peerj.10634

Rather TA, Tajdar S, Kumar S, Khan JA (2020a) Seasonal variation in the diet of sloth bears in Bandhavgarh Tiger Reserve, Madhya Pradesh, India. Ursus 31e12:1–8. https://doi.org/10.2192/URSUS-D-19-00013.2

Ratnayeke S, Van Manen FT, Padmalal UKGK (2007) Home ranges and habitat use of sloth bears *Melursus ursinus* in Wasgomuwa National Park, Sri Lanka. Wildlife Biol 13(3):272–284. https://doi.org/10.2981/0909-6396(2007)13[272: HRAHUO]2.0.CO;2

Sánchez MCM, Cushman SA, Saura S (2014) Scale dependence in habitat selection: the case of the endangered brown bear (*Ursus arctos*) in the Cantabrian Range (NW Spain). Int J Geogr Inf Sci 28(8):1531–1546. https://doi.org/10.1080/13658816.2013.776684

Sathyakumar S, Kaul R, Ashraf NVK, Mookherjee A, Menon V (2012) National Bear Conservation and Welfare Action Plan. Ministry of Environment and Forests, Wildlife Institute of India, and Wildlife Trust of India, India

Schaefer JA, Messier F (1995) Habitat selection as a hierarchy: the spatial scales of winter foraging by muskoxen. Ecography 18(4):333–344. https://doi.org/10.1111/j.1600-0587.1995.tb00136.x

Schneider DC (1994) Quantitative ecology: spatial and temporal scaling. Academic Press, San Diego

Schneider DC (1998) Applied scaling theory. In: Peterson DL, Parker VT (eds) Ecological scale. Columbia University Press, New York

Schneider DC (2001) The rise of the concept of scale in ecology: the concept of scale is evolving from verbal expression to quantitative expression.

BioScience 51(7):545–553. https://doi.org/10.1641/0006-3568(2001)051[0545:TROTCO]2.0.CO;2

Schneider DC, Walters R, Thrush S, Dayton PK (1997) Scale-up of ecological experiments: density variation in the mobile bivalve *Macomona liliana*. J Exp Mar Biol Ecol 216(1-2):129–152. https://doi.org/10.1016/S0022-0981(97)00093-2

Shirk AJ (2012) Scale dependency of American marten (*Martes americana*) habitat relationships. Biology and conservation of martens, sables, and fishers: a new synthesis. Cornell University Press, Ithaca

Shirk AJ, Raphael M, Cushman SA (2014) Spatio temporal variation in resource selection: insights from the American marten (*Martes americana*). Ecol Appl 24(6):1434–1444. https://doi.org/10.1890/13-1510.1

Wan HYI, Mcgarigal K, Ganey JL, Auret VL, Timm BC, Cushman SA (2017) Meta-replication reveals nonstationarity in multi-scale habitat selection of Mexican Spotted Owl. Condor 119(4):641–658

Wasserman TN, Cushman SA, Do W, Hayden J (2012) Multi scale habitat relationships of *Martes americana* in northern Idaho, USA. Research Paper RMRS-RP-94. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, p 21

Wiens JA (1989) Spatial scaling in ecology. Funct Ecol 3(4):385–397. https://doi.org/10.2307/2389612

Yoganand K (2005) Behavioural ecology of sloth bear (*Melursus ursinus*) in Panna National Park, Central India. PhD Thesis. Saurashtra University, India

Yoganand K, Rice CG, Johnsingh AJT, Seidensticker J (2006) Is the slothbear in India secure? A preliminary report on distribution, threats and conservation requirements. J Bombay Nat Hist Soc 103(2–3):172–181

Publisher's Note