Supplementary Information S1: Habitat suitability modeling method reported as per the ODMAP protocol

Table 1. Habitat suitability modeling method reported as per the ODMAP protocol.

|  |  |
| --- | --- |
| ***ODMAP element*** | **Contents** |
| **OVERVIEW** |  |
| *Authorship* | **Authors:** Rajashekhar Niyogi, Mriganka Shekhar Sarkar, Poushali Hazra, Masidur Rahman, Subham Banerjee, and Robert John |
| **Contact email:** robert.john@iiserkol.ac.in |
| **Title:** Connectivity conservation of small ungulates in a human dominated landscape |
| **DOI:** |
| *Model Objective* | **Objective:** Mapping/Interpolation |
| **Target output:** Continuous habitat suitability index |
| *Taxon* | The focal species include 4 antelopes of the family Bovidae. They are Blackbuck (*Antilope cervicapra*), Chinkara (*Gazella bennettii*), Four-horned antelope (*Tetracerus quadricornis*), and Nilgai (*Boselaphus tragocamelus*) |
| *Location* | Central India (including some parts of three Indian states: Madhya Pradesh, Uttar Pradesh, and Chhattisgarh) |
| *Scale of analysis* | **Spatial extent (Lon/Lat):** Longitude: 78.06° E - 82.85° E, Latitude: 22.64° N - 25.40° N  Covering 89,398 km2. Study area map included in the main text of the article (Fig 1). |
| **Spatial resolution:** 2×2 km grids |
| **Temporal resolution:** N/A, **Temporal extent:** 1 year |
| **Type of extent boundary:** Administrative boundary |
| *Biodiversity data overview* | **Observation type:** Wildlife monitoring data and mortality records |
| **Response/Data type:** Presence data |
| *Type of predictors* | Vegetation, topographic, disturbances |
| *Conceptual model / Hypotheses* | We used vegetation, terrain, and human disturbances as predictors for antelope habitat distribution. |
| *Assumptions* | Species–environment equilibrium. Availability of all important predictors. |
| *SDM algorithms* | **Algorithms:** SDMs were fitted using Generalized Linear Model (GLM), Generalized Additive Model (GAM), Random Forest (RF), Gradient Boosting Machine (GBM), Multivariate Adaptive Regression Splines (MARS), Support Vector Machine (SVM), and Flexible Discriminant Analysis (FDA) |
| **Model complexity:** Not controlled quantitatively |
| **Model averaging:** We used weighted averages of the model outputs, with an AUC of above 0.7 |
| *Model workflow* | **Software:** This analysis was conducted using R (version 3.6; R Core Team, 2010) with the ‘sdm’ package |
| **Code availability:** On supplementary information S5 |
| **Data availability:** On the data repository named 'BioStudies'. Link given on main text. |
| **DATA** |  |
| *Biodiversity data* | **Taxon names:** *Antilope cervicapra, Gazella bennettii, Tetracerus quadricornis, Boselaphus tragocamelus* |
| **Taxonomic reference system:** not specified |
| **Ecological level:** Population level |
| **Data source:** Madhya Pradesh State Forest Department |
| **Sampling Design:** The species presence dataset was created by compiling data from three different sources, which include wildlife monitoring data from forest department, wildlife mortality records, and direct visual observations.  *a.* *Wildlife monitoring data:* We obtained written information from forest rangers about the location of the target species, in their respective ranges. A forest range is an administrative unit with a mean size of about 180 km2. Most forest ranges collect some form of wildlife monitoring data, using methods that widely differs from one range to another. These include data from standardized line transects and sign surveys, camera-trap detections, and opportunistic observation records. Most of these are designed and implemented at a local level. We obtained this monitoring data from 65% of the forest areas in the landscape. (See fig: S1). This data was compiled at the compartment level, which is the smallest forest management unit with a mean area of about 2.5 km2.  *b.* *Wildlife mortality records:* We obtained the locations of antelopes from the mortality records compiled by the state forest department. We obtained it from Madhya Pradesh state, which makes up 91.78% of the study landscape. These records are available online on the MP forest department website [1]. We only used the records that had location data appended to it. This data was compiled at the compartment level for cases reported inside the forest areas and at the village level for cases reported outside forest areas (see Fig: 4). Village boundary maps were obtained online from NASA- CEDAC [2]. Mean size of a village in the landscape is slightly less than 4 km2 .  *c. Direct visual observation in the field*: We verified the above-mentioned species location data using direct field observations, at the range level. Nothing contrary to the compiled records was found. Field verification of the obtained data at a lower spatial resolution was not possible due to resource constraints.  The final species presence dataset was compiled by extracting the data from the forest compartment and the village shapefiles into 2×2 km (4 km2) grids. The data was extracted from the compartment and village shapefiles to the grids that intersected with their respective centroids. As this is only the creation of a ‘presence only’ dataset we assume that there would be no significant inconsistencies due to this simplistic extraction method.  The final species presence dataset was re-verified with the already published large scale surveys [3] and other studies in the landscape [4]. Nothing contrary to the existing literature was found. |
| **Sample size:** The study area extent contains a total of 23,051 2X2 km cells. Geospatial data recorded for all the cells. Number of presence cells for blackbuck, chinkara, four-horned antelope and nilgai are 83, 762, 209 and 1638, respectively. |
| **Absence data:** Equal number of background points were randomly generated in ArcMap |
| *Data partitioning* | Subsampling- using 85% of the data as training and 15% as test for chinkara, four-horned antelope and nilgai. Subsampling using 95% of data as training and 5% as test for blackbuck, due to a low number of data points. |
| *Predictor variables* | **Predictor variables:** Described the main text of the research article (section 4.1) |
| **MODEL** |  |
| *Variable pre-selection* | Compromise between covariate availability and ecological relevance as indirect proxy of species distributions. |
| *Multicollinearity* | Estimated the variance inflation factor for the predictors. Collinearity was assessed using the vifcor function. The default threshold value of 0.9 was used. |
| *Model settings* | Default settings in the sdm package. |
| *Model estimates* | Variable importance and significance were estimated through a separate analysis. It was not included in this habitat suitability modelling exercise. |
| *Model averaging/ensembles* | We used weighted averages of the model outputs, with an AUC of above 0.7. TSS values wereused for model averaging. |
| *Non-independence analyses* | None |
| **ASSESSMENT** |  |
| *Performance statistics* | Performance statistics estimated on validation data (from data subsampling) |
| *Plausibility check* | No plausibility checks conducted. |
| **PREDICTION** |  |
| *Prediction output* | It depicts a habitat suitability index for each grid of the landscape. It is used to identify the areas containing the most suitable habitats for the focal species. Maps were binarized, followed by MSPA and used as the basis for graph-theory-based habitat connectivity analysis |
| *Uncertainty quantification* | N/A |

**References**

1. Madhya Pradesh Forest Department official website Available online: https://mpforest.gov.in/ (accessed on Dec 18, 2020).

2. Meiyappan, P.; Roy, P.S.; Soliman, A.; Li, T.; Mondal, P.; Wang, S.; Jain, A.K. India Village-Level Geospatial Socio-Economic Data Set: 1991, 2001 2018.

3. Majumder, A. *Status Report: Tigers, co-predators and prey in Protected Areas of Madhya Pradesh, 2016.*; 2016;

4. Sharma, K.; Rahmani, A.R.; Singh Chundawat, R. Natural History Observations of the four-horned antelope Tetracerus quadricornis. *Journal of the Bombay Natural History Society* **2009**, *106*, 72.

|  |
| --- |
|  |

**Figure S1:** Map showing the forest areas from where secondary data has been obtained (Surveyed forests).