ASSESSING HABITAT SUITABILITY OF PANGOLIN SPECIES (MANIS CRASSICAUDA) IN UTTARAKHAND REGION USING MAXENT MODELLING FOR EFFECTIVE CONSERVATION STRATEGIES

1. ABSTRACT

Biodiversity worldwide is currently confronted with escalating challenges such as habitat degradation, overexploitation, and climate-related concerns. Among the various species, pangolins belonging to the order Pholidota stand out as the sole scaly mammals globally, and unfortunately, they are highly sought-after wildlife commodities, primarily for their scales and meat, making them the most trafficked wild mammals worldwide. The Indian pangolin, specifically the (Manis Crassicaudata) species found in the Uttarakhand region, remains one of the least researched members of this order. Limited knowledge exists regarding their ecological response to environmental and human-induced factors, as well as their spatial distribution patterns in the northern regions. Our study aimed to enhance the ecological understanding of Indian pangolins by investigating the environmental factors associated with the presence of suitable habitats for this species within a protected and favourable area in the lower Himalayan region of Uttarakhand. Through the utilization of primary surveys, GPS data collection, and employing a maximum entropy (MaxEnt) modelling approach, we explored the habitat suitability of Indian pangolins across three districts of Uttarakhand. The results of our model demonstrated a high level of accuracy, as indicated by the average test area under the Receiver Operator Curve metric. Furthermore, the jackknife test highlighted that the slope, normalized difference vegetation index, and elevation played crucial roles as predictors in determining the distribution of pangolin burrows. In conclusion, our study underscores the significance of habitat suitability modelling for comprehending the distribution patterns of pangolins, and it further emphasizes the importance of employing such models in conservation planning and prioritization efforts for this species.

Key Words: Indian Pangolins, Species Distribution Model (SDM), Habitat Suitability, MaxEnt Modelling
Uttarakhand is situated on the southernmost side of the Himalayan Mountain range. The climate and wildlife in this region undergo significant changes as the elevation varies. Due to human activities such as climate change, alteration or loss of habitats, biological invasions, infrastructure development, and various other factors, many wildlife species in Uttarakhand are currently facing the risk of extinction [(Owens & Bennett, 2000) (Thomas et al., 2004) (Sharma et al., 2014)]. The conservation of species relies on our understanding of their ecological requirements, which necessitates knowledge about the distribution of species and their associations with specific habitats (Sands et al., 2012) (Greggor et al., 2016). The identification of the potential distribution of species is a fundamental step in species conservation and can be instrumental in ecological.

Among the eight pangolin species globally, the Indian pangolin (M. crassicaudata) is found in Uttarakhand. They occupy long grasslands and secondary forests, demonstrating adaptability to arid regions and deserts, albeit with a preference for barren, hilly areas. The Indian pangolin thrives in soil conditions characterized by softness and semi-sandy textures, which facilitate its ability to dig burrows (Mahmood et al., 2014) (Latafat et al., n.d.) The Indian pangolin faces threats from poaching due to the high demand for pangolin’s meat and scales, which are consumed and utilized by local communities and also traded in different parts of the world. Various parts of the Indian pangolin are prized for their culinary and medicinal uses. Its scales are coveted for their perceived aphrodisiac qualities and are occasionally fashioned into rings or charms. Additionally, its skin is utilized in crafting leather goods such as boots, ornaments, and shoes (Mahmood et al., 2013). Nomads and skilled local hunters are primarily responsible for carrying out most hunting activities (Mahmood et al., n.d.). Indian pangolin’s skin has been illicitly traded and trafficked for personal use in China since at least the early 2000s. It is noteworthy that pangolins are recognized as the most extensively trafficked protected mammals (Wu, S. B. at al., 2007). The Indian pangolin is officially classified as endangered on the IUCN Red List of Threatened Species and has been listed under Appendix I of CITES since January 2017. It receives protection in all countries within its range (Mahmood et al., n.d.) The primary reason for the threatened status of pangolins is poaching, driven by the illegal trade of their scales and meat (Challender, D., et al., 2014).

Species distribution models are a methodology utilized to identify and characterize potential habitats that are suitable for specific species (Elith et al., 2009). MaxEnt is widely recognized as most used tools for modeling species potential distributions. It employs presence-only occurrence locations along with meaningful environmental variables to construct a favorable habitat map, aiming to provide the most accurate estimation of the species' distribution (S. Phillips et al., 2006). In this study, SDM was employed to examine the potential distribution of the Indian pangolin within the existing land cover types. Maximum entropy modelling (S. J. Phillips et al., 2004) is a widely utilized technique for estimating the distribution of species based on presence-only data. It proves particularly advantageous when modeling species that are under threat and have limited available data (Elith, J. et al., 2011). The objective of this study was to determine the potential habitats of the Indian pangolin within the current land cover types found in Uttarakhand. This analysis aimed to provide guidance for prioritizing future research, monitoring efforts, and conservation initiatives related to the species. Our current research seeks to fill a gap in existing literature concerning illegal pangolin poaching. While extensive studies have been conducted on poaching in regions like Vietnam and China, research specific to the Uttarakhand area remains limited.
3. Study Area

The state of Uttarakhand in India is situated between the latitudes of 28.43°N and 31.27°N and the longitudes of 77.34°E and 81.02°E. Uttarakhand is renowned for its natural environments, encompassing the Himalayas, Bhabar, and Terai regions. It shares borders with China's Tibet Autonomous Region to the north, Nepal to the east, Uttar Pradesh to the south, and Himachal Pradesh to the west and northwest. The state has a total of 13 districts, with Garhwal and Kumaon as its two divisions. Uttarakhand covers a land area of 53,566 km² (20,682 sq. mi), of which ‘86% is mountainous terrain’ and ‘65% is covered by dense forests’ (District Portal, 2012).

The state's rich ecological diversity supports numerous animal species such as bharal, snow leopards, leopards, and tigers, as well as various plants and rare herbs within the Himalayan ecosystem. The Valley of Flowers National Park and Nanda Devi National Park, jointly designated as a UNESCO World Heritage Site, are two notable national parks within the state. In Nainital and Pauri Garhwal District, one can find the renowned Jim Corbett National Park, which happens to be the oldest national park in India. The valley is home to numerous plant species, some of which are exclusive to this region and face global endangerment. These protected areas play a vital role in preserving the exceptional biodiversity found in Uttarakhand (UNESCO, 2012).

4. Methodology
4.1 Species Dataset

The data regarding the presence of pangolins was collected through camera trap surveys and direct observations conducted during three seasons of the year 2022. Three sampling areas were chosen for further monitoring based on the presence data collected in previous years, and they were visited in spring, summer, and autumn of 2022 and 2023. Georeferenced images of pangolins were captured throughout the month. Any presence locations with autocorrelation were eliminated, resulting in a total of 187 distinct pangolin presence locations. Camera trap data and direct observations were combined to achieve broader coverage. The SDM toolbox in ArcGIS Pro 3.0.3 was utilized to spatially filter presence locations within a 1 km radius of each other. To differentiate between Indian and Chinese pangolins and to avoid confusion, only direct sightings of Indian pangolins were considered in areas below 500 m elevation. Additionally, whenever possible, the occurrence of Indian pangolins was confirmed through direct
communication with key informants. In areas with elevations exceeding 500 m, the presence locations of Indian pangolins were determined based on direct sightings and indirect evidence such as burrows, footprints, and scratches.

4.2 Camera Trap Surveys

Non-systematic camera trap surveys were conducted opportunistically within Rajaji National Park during the period from March 2023 to May 2023. This placement approach was intentionally non-random, as the primary objective was to maximize the chances of capturing diverse biodiversity. No specific measures were taken to mitigate potential biases, as the focus was solely on documenting the area's biodiversity using the most efficient methodology available. The camera trap sites were selected based on the expertise of local parabiologists who possessed extensive knowledge of the region from decades of hunting experience. These sites were primarily located within Broadleaved deciduous forests. The distances between the trap sites and the settlements within the forest ranged from one kilometer to ten kilometers.

4.3 Interviews and Secondary Sources

To establish baseline data for a habitat study, a total of 12 semi-structured interview surveys were conducted in March 2023. These surveys were conducted in three tribal villages located within the Dehradun, Nainital, Rajaji National Park, and adjacent areas. The interviews were designed to be conversational in nature, promoting a more relaxed and open discussion rather than rigid question-and-answer sessions. All interviews were conducted in the Garhwali language and later translated into English for analysis purposes. During the interviews, photographs of pangolins were shown to the participants to aid in species identification. Prior to the interview process, informed consent was obtained from every survey participant, with the consent given verbally instead of in written form due to the high levels of illiteracy in the region. Participants were clearly informed that they had the option to decline the interview at any point during the process, and their identities and household data were not collected to ensure anonymity and confidentiality. A comprehensive review of published research papers, books, and articles was conducted to gather background information and establish a robust theoretical framework for the study. Government Data sourced from official government entities such as the Wildlife Institute of India, including statistical reports and publicly available databases, were utilized to analyze trends and patterns in the study.

5. Results

To visually represent their spatial positions, the occurrence points obtained from Google Earth were mapped onto a shapefile of Uttarakhand, along with their corresponding latitude and longitude coordinates.
5.1 Modeling Species Distribution and Ensemble Model

5.1.1 Environmental data

Nineteen climatic factors, obtained from the WorldClim database (http://www.worldclim.org), were used as bioclimatic variables in this study. These variables were downloaded at a geographical resolution of approximately 1 square kilometer and majorly consisting temperature and precipitation data.

5.1.2 Categorical Data

To ensure the most accurate and concise model construction, considering the limited sample size, we incorporated environmental variables based on prior knowledge specific to Indian Pangolins. To successfully execute the model, calculations for Hillshade, Aspect, Slope, and Terrain Ruggedness Index were performed using the Digital Elevation Model (DEM) within ArcGIS Pro Version 3.0.3. Additionally, variables such as ‘Land Cover/Land Use’ (LULC), ‘Normalized Difference Water Index’ (NDWI), ‘Normalized Difference Vegetation Index’ (NDVI), ‘Bare Soil Index’ (BSI), ‘Land Surface Temperature’ (LST), ‘Drainage Density’, ‘Distance to Settlement’, and ‘Distance to Railway’ were computed using data from various sources, which were then processed within ArcGIS Pro Version 3.0.3. To ensure spatial consistency, all raster layers were clipped to the extent of Uttarakhand and resampled to a 1 km resolution to align with the original resolution of the WorldClim data. This approach was adopted to mitigate the risk of overfitting and develop a more parsimonious model given the limited sample size.
Figure 3 Maps showing categorical data used in study
5.1.3 MaxEnt Model

The processed distribution point data along with screened bioclimatic variables, along with all categorical data, were imported into MaxEnt 3.4.1. The bioclimatic variables were evaluated using the Jackknife test. The obtained models were calibrated using 75% of the available records for each species as training (calibration) data, while the remaining 25% were used for model validation as the test dataset. The MaxEnt model was utilized for predicting the potential distributions of Pangolins. The ‘Bootstrap Method’ was applied with 10 repeats, and a maximum of 5000 iterations, along with default selected parameters. MaxEnt is a multivariate technique utilized to estimate the distribution of a species by calculating the probability distribution of maximum entropy. This approach takes into account various constraints that represent our limited understanding of the species' distribution.

![MaxEnt Model](table1_cumulative_threshold_of_species_distribution_model)

![Result produced by MaxEnt showing Suitable Sites for Pangolins habitat](fig4)

<table>
<thead>
<tr>
<th>Cumulative threshold</th>
<th>Omission threshold</th>
<th>Description</th>
<th>Fractional presence area</th>
<th>Training omission rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.000</td>
<td>0.000</td>
<td>Final cumulative value 1</td>
<td>0.122</td>
<td>0.000</td>
</tr>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>Final cumulative value 5</td>
<td>0.062</td>
<td>0.015</td>
</tr>
<tr>
<td>0.000</td>
<td>0.172</td>
<td>Initial testing presence</td>
<td>0.067</td>
<td>0.004</td>
</tr>
<tr>
<td>0.000</td>
<td>0.065</td>
<td>Minimising testing presence</td>
<td>0.074</td>
<td>0.000</td>
</tr>
<tr>
<td>0.050</td>
<td>0.200</td>
<td>Max testing sensitivity</td>
<td>0.089</td>
<td>0.007</td>
</tr>
<tr>
<td>0.050</td>
<td>0.178</td>
<td>Max testing sensitivity plus specificity</td>
<td>0.086</td>
<td>0.001</td>
</tr>
<tr>
<td>0.000</td>
<td>0.065</td>
<td>Minimising testing sensitivity plus specificity</td>
<td>0.074</td>
<td>0.000</td>
</tr>
<tr>
<td>0.000</td>
<td>0.021</td>
<td>Balance testing omission, predicted area and threshold value</td>
<td>0.105</td>
<td>0.000</td>
</tr>
<tr>
<td>0.000</td>
<td>0.188</td>
<td>Relative entropy of threshold and original distribution</td>
<td>0.092</td>
<td>0.000</td>
</tr>
</tbody>
</table>

6. Discussions

In the research, several commonly used thresholds and their corresponding omission rates are provided. When test data is accessible, binomial probabilities are calculated accurately as long as the number of test samples does not exceed 25. Beyond that threshold, a normal approximation to the binomial is utilized. The reported p-values are one-sided and represent the probability that the null hypothesis is true, which states that the test points are not predicted any better than random prediction with the same fractional predicted area. One specific threshold referred to as the "Balance" threshold is determined by minimizing the expression: 6 times the training omission rate plus 0.04 times the cumulative threshold plus 1.6 times the fractional predicted area.

The provided table presents estimations of the relative contributions of environmental variables to the Maxent model. The first estimate is determined by adding the increase in regularized gain to the contribution of the corresponding variable during each iteration of the training algorithm. However, if the change to the absolute value of lambda is negative, it is subtracted instead. The second estimation method involves independently permuting the values of each environmental variable within both the training presence and background data. Subsequently, the model is reassessed using the permuted data, and the resulting decrease in training AUC (Area Under the Curve) is presented in the
table as a percentage. It is essential to exercise caution when interpreting the contributions of variables, particularly when there are correlations between predictor variables. This caution is similar to the approach used in variable jackknife analysis.

The four output formats have a consistent relationship with each other, but they vary in scaling and interpretation. The default output format, cloglog, is the most straightforward to comprehend. It provides an estimation, ranging from 0 to 1, representing the probability of presence. This is a representation of the Maxent model for Pangolin. This figure illustrates the Maxent model for Pangolin, with areas of better predicted conditions represented by warmer colors. The white dots indicate the presence locations that were used for training the model, while the violet dots represent the test locations. The automatic generation threshold for determining Pangolin suitability is set at 0.277. This means that any value between 0 and 0.237 is considered unsuitable for Pangolins. The suitable areas are classified into different grades: areas with low suitability range from 0.38 to 0.54, areas with medium suitability range from 0.54 to 0.77, and areas with high suitability range from 0.77 to 1. The evaluation of model performance involved calculating the Area Under the Receiver Operator Curve (AUC). In our study, models with AUC values exceeding 0.7 were considered satisfactory (S. J. Phillips & Dudík, 2008). MaxEnt, as a multivariate approach, utilizes maximum entropy to estimate the species' distribution by considering constraints that account for our limited knowledge about the distribution. It is widely accepted that an AUC value below 0.7 indicates low model accuracy, while predictions can be relied upon when AUC falls within the range of 0.7 to 0.9. AUC values exceeding 0.9 signify highly accurate predictions suitable for subsequent analysis (S. J. Phillips & Dudík, 2008).

6.1 STATISTICAL ANALYSIS

The provided image depicts the relationship between the cumulative threshold and the omission rate, as well as the predicted area. The omission rate is computed for both the training presence records and, if available, the test records. It is expected that the omission rate closely approximates the predicted omission, given the definition of the cumulative threshold.

The next picture is the receiver operating characteristic (ROC) curve for the same data. Note that the specificity is defined using predicted area, rather than true commission (see the paper by Phillips, Anderson and Schapire cited on the help page for discussion of what this means). This implies that the maximum achievable AUC is less than
1. If test data is drawn from the Maxent distribution itself, then the maximum possible test AUC would be 0.979 rather than 1; in practice the test AUC may exceed this bound.

6.2 RESPONSE CURVES

These curves show how each environmental variable affects the Maxent prediction. The curves show how the predicted probability of presence changes as each environmental variable is varied, keeping all other environmental variables at their average sample value. Note that the curves can be hard to interpret if you have strongly correlated variables, as the model may depend on the correlations in ways that are not evident in the curves. In other words, the curves show the marginal effect of changing exactly one variable, whereas the model may take advantage of sets of variables changing together. In contrast to the previously mentioned marginal response curves, the curves presented below correspond to different models, specifically Maxent models created using only the respective variable. These plots illustrate the relationship between predicted suitability and the individual variable selected, taking into account any dependencies resulting from correlations between the chosen variable and other variables. These curves can be particularly helpful in interpretation when there are notable correlations between variables.

![Response Curves](image)

**Figure 7** Graphs showing individual data layers getting impacted by cumulative threshold

6.3 CLASSIFICATION OF SUITABLE AREAS

Using a MaxEnt model, the projected distribution of various Pangolin species in the Uttarakhand region was determined and subsequently visualized using ArcGIS Pro.
The MaxEnt model’s predictions closely aligned with both the actual distribution area and the Area Under the Curve (AUC) obtained from the most recent climatic data. Probability maps were then categorized using a threshold value, which defines the minimum level below which a specific distribution is deemed unsuitable, allowing for the differentiation between suitable and unsuitable habitats. The low-lying regions of Uttarakhand, namely Dehradun, Haridwar, Pauri Garhwal, and Nainital, demonstrated pixel values ranging from 0.38 to 1, indicating that they are the most favorable areas for the distribution of pangolins. By utilizing the ArcGIS Pro software to extract the most suitable pixels from the site suitability map, it becomes evident that the low-lying regions of Uttarakhand lying in south-west zones, specifically Dehradun, Haridwar, Pauri Garhwal, and Nainital, exhibited pixel values ranging from 0.77 to 1. These values indicate that these areas are highly suitable for the distribution of pangolins.

The presence of pangolins was primarily confirmed based on the existence of burrows, although the exact species of pangolin could not be definitively determined solely from the burrow evidence. Consequently, this study aimed to investigate the distribution of pangolin species and predict potential habitats for them. The MaxEnt modelling approach was used to assess pangolin distribution and identify suitable habitats. The results indicate that approximately only 378 km² of Uttarakhand’s total land area has the potential to serve as suitable habitat for pangolins.

Among the geographic variables considered, elevation emerged as the most influential factor in predicting suitable pangolin habitats, corroborating findings from previous studies (Kavre, 2011) (Shrestha et al., 2015). Uttarakhand’s diverse elevational range contributes to varied weather and climatic conditions, temperature fluctuations, rainfall patterns, and vegetation types, including herbs, shrubs, and trees, which can impact the behavior and habitats of burrowing mammals like pangolins. Therefore, hilly regions with less dense and diverse forests, along with appropriate climate, rainfall, ample food availability, and minimal human disturbance, are more likely to provide
suitable habitats for pangolins. Given these findings, community-based conservation programs should be prioritized to support the preservation of pangolin species.

7. Conclusion

This study, which focused on pangolins in Uttarakhand, represents the first comprehensive research conducted on this endangered species. It involved extensive surveys across various potential areas to gather data on pangolin populations, validating anecdotal information through field surveys. The study successfully compiled a database of pangolin presence records by examining historical reports and gathering first-hand information from previously unrecorded locations. Additionally, the study investigated the key environmental factors associated with pangolin occurrence and distribution. The observed decline in pangolin populations and their absence from areas where they were previously documented highlights the urgent need for conservation efforts. Moreover, the study employed predictive modelling to identify potential pangolin habitats in different regions of the lower Uttarakhand area. These findings are of paramount importance for future research endeavors, the development of conservation and management plans, and ultimately contribute to the long-term preservation of pangolin species, not only in Uttarakhand but also in other regions worldwide.

8. References


