Distribution modelling of *Tor putitora* (Hamilton, 1822), an endangered Cyprinid in the Himalayan river system using MaxEnt

Ranjit Mahato¹, Santoshkumar Abujam², Dhoni Bushi¹, Oyi Dai Nimasow³, Gibji Nimasow^{1*} and Debangshu Narayan Das²

¹Department of Geography, Rajiv Gandhi University, Arunachal Pradesh (INDIA).

²Department of Zoology, Rajiv Gandhi University, Arunachal Pradesh (INDIA).

³Department of Botany, Rajiv Gandhi University, Arunachal Pradesh (INDIA).

*Corresponding Author [Gibji Nimasow] Email: gibji.nimasow@rgu.ac.in

Professor, Department of Geography, Rajiv Gandhi University, Rono Hills, Doimukh – 791112, Arunachal Pradesh (India).

Abstract

Freshwater fishes are threatened by habitat modification, overfishing, pollution, invasive species, and climate change. Hence, habitat assessment using geospatial technology is essential in predicting the distribution of threatened candidates. *Tor putitora*, an endangered game fish is declining due to indiscriminate fishing and habitat destruction. No attempts were made so far to understand habitat predilection of the species. In this study, we used Maximum Entropy (MaxEnt) model with occurrence records and environmental variables. The occurrences and environmental variables were collected from survey and secondary sources and used after removing biases, multicollinearity test and contribution analysis. The model performed well with area under curve (AUC) 0.867 and true skill statistic (TSS) 0.756. The main contributing variables were upstream elevation (35.3%) and flow length (21.4%). The model predicted 10% suitable, 33% slight to moderate suitable and 57% unsuitable. So, modelling is useful in identifying river networks for conservation and eco-restoration of freshwater fishes.

Key words: Freshwater fish; Environmental variables; Habitat predilection; Conservation; Eco-restoration

1. Introduction

Tor putitora (Hamilton 1822) is known popularly as Golden mahseer or Himalayan mahseer and it belongs to the family Cyprinidae under order Cypriniformes. It is one of the largest freshwater fish of the Indian subcontinent inhabiting mainly the rivers of the Himalayan foothill region (Bhatt and Pandit 2015). The species has a restricted area of occupancy in general within India, Afghanistan, Pakistan, Nepal, Bhutan, Myanmar, and Bangladesh (Rajbanshi and Csavas 1982; Talwar et al. 1991; Oo 2002). Behaviourally, being potamodromous (migrating within freshwaters) one, the species mostly uses smaller hill streams as their spawning grounds (generally rapid streams, riverine pools, and lakes characterized by large boulders, pebbles and

gravel) with water temperature range of 11 to 30.5° C, alkaline pH and dissolved oxygen concentration in the range of 6.4–11 mg/l (Joshi 1994; Bhatt et al. 2004). Compared to the other congeners, *T. putitora* is elongated but laterally compressed having longer head and head length is significantly greater than body depth at dorsal fin origin. Further, the head is slightly blunt with a large mouth and prominent lateral line stripes whereas the body is dim greenish dorsally, yellowish and silvery white ventrally. The distal margin of pelvic, anal and caudal fins show tint of reddish-golden colour as well as at maturity the body above its lateral line shows usually golden colour, which might be absent in the juvenile stage (Darshan et al. 2019). According to IUCN (2021), *T. putitora* is enlisted as an endangered (EN) species.

Across the global range, a total of 17 species of *Tor* are considered valid (Fricke et al. 2021). Out of which, eight (8) species have been reported from India namely *Tor barakae* (Arunkumar and Basudha 2003), *Tor khudree* (Sykes 1839), *Tor kulkarnii* (Menon 1992), *Tor malabaricus* (Jerdon 1849), *Tor mosal* (Hamilton 1822), *Tor putitora* (Hamilton 1822), *Tor remadeviae* (Kurup and Radhakrishnan 2011) and *Tor tor* (Hamilton 1822). In the North-eastern states of India, only 3 species of genus *Tor* namely *Tor barakae*, *T. putitora* and *T. tor* have been reported to date. Interestingly, *T. putitora* is the state fish in many states of India viz., Arunachal Pradesh, Himachal Pradesh, Uttarakhand and the union territory of Jammu and Kashmir. The Indus, Ganges and Brahmaputra river basins are the native range of *T. putitora* (Jha et al. 2018). Among the Indian species of the genus, *T. putitora* is the important food and a popular game fish after *T. tor*. It constitutes an amazing fishery in different foothill water bodies of the country and nowadays, many reservoirs are being stocked with the species both for fishery and ecotourism development. However, creation of dams across certain rivers has alarmingly degraded habitat and shrink their natural breeding grounds. Furthermore, high mortality of brood and juvenile fishes often have been witnessed due to the act of indiscriminate overfishing in the rivers and the species has been estimated to decline by more than 50% already (Jha et al. 2018). *T. putitora* is also popular and important high-value indigenous fish species in Nepal (Rai et al. 1997). The species is also reportedly declining from its natural habitat in Nepal (Bista et al. 2007). However, in recent years, the success in artificial breeding at some research stations has provided enthusiasm on developing it for commercial cultivation in natural waters (Rai et al. 2006). In Bhutan, it is commercially important game fish, and has a very high table value. It is locally called as *Serngya* (Yellow fish), and represents the symbol of good luck in the Bhutanese belief system (Ministry of Agriculture and Forests, Bhutan 2020). Unfortunately, the natural stock in the country is also declining due to deterioration of the natural habitat to irreparable extent.

Species Distribution Modelling (SDM) predicts the distribution of a species across geographic space and time using environmental data and occurrence records. Hence, the significance of ecological factors like topographic, climatic, land cover, physico-chemical properties of soil and water, etc. cannot be ignored in determining the distribution pattern of freshwater fishes (Roy et al. 2021). Further, the alteration in thermal regimes, land use and hydrology has been reported to cause fish isolation and fragmentation within the terrestrial landscapes leading to changes in the distributional patterns (Johnson et al. 2012; Greshishchev et al. 2015; Ruiz-Navarro et al. 2016; Hamilton et al. 2018). Therefore, in view of the declining

population of natural stock, estimating the potential distribution of this sought after freshwater fish through ecological modelling is highly indispensable for policy making and planning particularly towards habitat restoration and implementing conservation strategies. So far, hardly any attempts have been made to model the distribution of freshwater fishes, particularly *T. putitora* in the riverine systems of the Himalayas. This research article hitherto revealed that Maximum Entropy (MaxEnt) approach (Phillips et al. 2006) can predict the distribution range of an endangered Cyprinid fish (*T. putitora*) on the backdrop of the occurrence records and environmental factors very precisely and successfully. It is expected that such outcome would be able to provide important information for rational aquatic resource management and *ex-situ* or *in-situ* conservation of the species like golden mahseer in its natural habitat.

2. Materials and methods

2.1. Species occurrence data

As per the available literature, the targeted species is mainly distributed in the rivers of the Himalayan region. Hence, the study area encompasses the union territory of Jammu and Kashmir in the west to the North-East Indian states in the east including the countries of Nepal and Bhutan (Fig. 1). The occurrences of *Tor putitora* in the study area were collected from both primary and secondary sources. The primary occurrence data were recorded during the field survey using handheld Garmin GPSMAP 64csx. The dismo package of R (Version 4.0.3) has been used to access the secondary occurrence data from the Global Biodiversity Information Facility (GBIF 2019) following Hijmans et al. (2011). The search string '*Tor putitora*' in GBIF returned 138 occurrence data and another 42 species occurrences were consulted from different published journal articles. Thus, a total of 180 occurrence records were collected. Out of this, the species records lacking geographic coordinates and falling outside the study area were removed. Further, a snapping tolerance of 1 km was used to move the occurrence points to the closest freshwater pixel and the points falling outside these pixels were removed. The new geographical coordinates (104 Nos.) were used as presence records for testing the geographical biasness (Supplementary - 1). The occurrence records are often reported to be biased due to easy accessibility of an area or near road or areas of high population density that can distort the species' niche resulting into error in estimates of the model performances (Kadmon et al. 2004; Veloz 2009). Hence, the biasness in the occurrence records were removed using spThin package in R, which uses an algorithm to randomize the removal of occurrence locations within a specified distance threshold (Aiello-Lammens et al. 2015). The resulting dataset retained the records of species presence within 10 km spatial grid and returned a dataset of 58 occurrence points for the model (Supplementary - 2).

Figure 1. Map of the study area. Generated by the authors using ArcGIS 10.3, <http://esri.com>

2.2. Environmental variables

For generating the distribution modelling of *T. putitora*, a set of near-global environmental variables with 1 km resolution were downloaded from the website of EarthEnv project (www.earthenv.org) using R. EarthEnv is a collaborative project of biodiversity scientists and remote sensing experts for monitoring and modelling biodiversity, ecosystems, and climate.

The sources of these variables are WorldClim, Consensus Land Cover, HydroSHEDS and World soil information (ISRIC) as explained in detail by [Domisch et al. \(2015\)](#). The environmental variables include 19 layers of hydroclimatic variables, 12 layers of land cover variables, 10 layers of soil variables, 4 layers each of elevation and slope, and 1 layer each of stream length and flow accumulation. Additionally, 4 variables namely dissolved oxygen, total alkalinity, town and road proximity was generated through Inverse Distance Weighting (IDW) interpolation method from the water samples (field survey) and published literatures using ArcMap 10.3 ([Environmental Systems Research Institute \(ESRI\) 2016](#)). The coordinate points were snapped to the closest freshwater pixel with a snapping tolerance of 1 km before the interpolation. All the predictor variables were clipped to the area of interest (AOI) for further processing.

2.3 Selection of best environmental variables

To achieve parsimonious model, all the predictors with insignificant contribution to the model were removed through multicollinearity test and analysis of percent contributions. Most of the statistical methods face the problem of multicollinearity due to high correlation among the predictor variables ([Miller 2010](#)) that can lead to model uncertainties ([Graham 2003](#); [Braunisch et al. 2013](#); [Vatcheva et al. 2016](#)). Hence, [Merow et al. \(2013\)](#) recommended minimizing the highly correlated environmental variables because the complex features created by MaxEnt are often already highly correlated. In this study, we used Variance Inflation Factor (VIF) correlation approach to reduce multicollinearity problems using usdm package in R ([Naimi and Araújo 2016](#)). In general, a correlation coefficient threshold of > 0.7 is considered to be an appropriate indicator of collinearity ([Dormann et al. 2012](#); [Manzoor et al. 2018](#); [Sony et al. 2018](#); [Farrell et al. 2019](#); [Feng et al. 2019](#)) and a VIF greater than 10 is a signal that the model has a collinearity problem ([Chatterjee and Hadi 2006](#)). Accordingly, in this study, out of 55 variables, 19 predictor variables with correlation coefficient threshold values of < 0.7 ([Engler et al. 2013](#); [Zimmermann et al. 2007](#)) and $VIF < 7$ were retained for additional analysis ([Table 1](#)).

Table 1. Post collinearity variables assessed for inclusion in the model

Then, initially we run MaxEnt model with all 19 predictor variables to examine the variable contributions. Based on the results, 12 variables (elv_01, hydroclim_15, flow_acc_01, lc_avg_01, lc_avg_03, lc_avg_05, lc_avg_07, lc_avg_09, lc_avg_11, soil_avg_09, slope_01, and do) with more than 1% percent contribution were retained. Out of which, 6 best environmental variables (elv_01, hydroclim_15, flow_acc_01, lc_avg_03, soil_avg_09, and do) were selected after consulting available literature on the habitat information of the species and removing the persistent variables of land cover. Finally, we performed 5 replicate runs with 58 presence records and the 6 environmental variables to predict the distribution of *T. putitora* in the Himalayan river system ([Table 2](#)).

Table: 2 Variables retained for the final model (**shown in bold**) with percent contribution and permutation importance (**averaged over 5 replicate runs**)

2.4 Model setup and evaluation

MaxEnt (version 3.4.1) was used to model the potential distribution of *T. putitora*. Out of a range of traditional and machine-learning modelling approaches, MaxEnt is reported to perform well (Phillips et al. 2006; Peterson et al. 2007; Phillips and Dudík 2008; Radosavljevic and Anderson 2014; Nimasow et al. 2016) with high predictive accuracy (Moreno et al. 2011; Yang et al. 2013). MaxEnt is a machine-learning algorithm that builds relationships between presence-only records and a set of environmental variables to estimate a target probability distribution of maximum entropy based on the most uniform distribution (Merow et al. 2013). The specific settings used in the MaxEnt model were logistic output format, cross-validate replicate run type, response curves, jackknife measures of variable importance, 10 percentile training presence, and rest kept as default settings. The MaxEnt output in .ASC format was imported to ArcGIS 10.3 and the relative habitat suitability (Merow et al. 2013) ranging from 0 (unsuitable) to 1 (highly suitable) was reclassified into five classes viz. < 0.1 = unsuitable, 0.1 to 0.3 = slightly suitable, 0.3 to 0.5 = moderately suitable, 0.5 to 0.7 = suitable, and > 0.7 = highly suitable. The flowchart of methodology followed in this study is given in Fig. 2.

To test the model predictive accuracy, we used 5-fold cross-validation, where the occurrence data was randomly split into a number of equal-size groups called ‘folds’, and the models are created leaving out each fold in turn. The overall performance of the model was evaluated using the threshold independent, area under curve (AUC) of the receiver operating characteristic (ROC) curve and True Skill Statistic (TSS). The AUC value ranges from 0.5 to 1.0 where a value of 1.0 indicates perfect fitted model while a value near 0.5 indicates a fit not better than a random (Fielding and Bell 1997; Baldwin 2009). The TSS takes into account both omission and commission errors, and success as a result of random guessing, and ranges from -1 to $+1$, where $+1$ indicates perfect agreement and values of zero or less indicate a performance no better than random (Allouche et al. 2006).

Figure 2. Flowchart of methodology followed in the study.

3. Results

3.1 Model evaluation

The final model used six best environmental variables (minimum upstream elevation, precipitation seasonality, deciduous riparian broadleaf trees, flow length, dissolved oxygen, and depth to bedrock) representing topographic, flow accumulation, land cover, hydroclimatic, soil and water quality. These variables are known to play a significant role in the spawning, feeding habits, and overall distribution of *T. putitora*. The averaged model (5 replicate runs) show high training AUC values (> 0.894) across the models and also higher than the test AUC values. The test AUC values, which show the actual model predictive power, were higher than 0.818. The mean AUC of 0.867 with standard deviation of 0.026 and TSS value of 0.756 suggests that the final model is reasonably good. Out of the six environmental variables, minimum upstream elevation (elv_01) showed the highest relative importance of 35.3%, followed by flow length (flow_acc_01) with 21.4%. The importance of rest of the variables ranged between 1.4 to 7.8% only, indicating intermediate to lower contribution in predicting the suitable habitat of *T.*

putitora. The logistic output (response curves) shows a negative relationship of *T. putitora* with high upstream elevation (peak towards lower values 0 to 1000 m) and deciduous broadleaf trees (peaked towards lower values of 5 to 10%). On the other hand, there was a positive association with flow length (peak towards higher values of 8000 to 30000 m), precipitation seasonality (peak towards higher values), dissolved oxygen (peak towards higher values of 5 to 12 mg/l), and depth to bedrock (peaked towards higher values of 220 to 235 cm). The response curves of the environmental variables are shown in Fig. 3.

Figure 3. Species response curve (derived from MaxEnt). Predictor variables are shown in *x*-axis and species response in *y*-axis represented by Minimum Upstream Elevation (elv_01), Flow Length (flow_acc_01), Precipitation Seasonality (hydroclim_15), Deciduous Broadleaf Trees (lc_avg_03), Depth to Bedrock (soil_avg_09), and Dissolved Oxygen (do).

3.2 Distribution modelling of *T. putitora*

The final suitability map was calculated in terms of the total river length. Out of the total river length of 60290.56 km, the model predicted 34265.75 km (56.83%) as unsuitable, followed by 12979.36 km (21.53%) as slightly suitable and 6807.67 km (11.29%) as moderately suitable. The suitable and highly suitable categories constitute 5003.62 km (8.30%) and 1234.16 km (2.05%) respectively (Table 3 & Fig. 4). Countrywise, India represents a total river length of 3652 km, followed by Nepal with 2248 km under suitable and highly suitable categories. Bhutan showed the lowest river length of 338 km under suitable category. Among the Indian states and union territories, the largest river length under suitable and highly suitable categories was found in Arunachal Pradesh with 1626 km, followed by Uttarakhand with 960 km and Himachal Pradesh with 772 km. The union territory of Jammu & Kashmir, Darjeeling district of West Bengal and Sikkim recorded suitable river lengths of 165 km, 69 km and 61 km respectively (Supplementary - 3A & B).

Figure 4. Predictive habitat suitability of *T. putitora* in the Himalayan river system. Generated by the authors using ArcGIS 10.3, <http://esri.com>

Table 3. Suitable categories of *T. putitora* in the Himalayan river system

A visual examination of the final model shows majority of the upper reaches of the Himalayan rivers as unsuitable for *T. putitora* due to higher elevation. On contrary, the relatively low altitude reaches, mainly comprising of the Indian states of Himachal Pradesh, Uttarakhand, West Bengal, Sikkim, and Arunachal Pradesh including the countries of Nepal and Bhutan represents the suitable and highly suitable habitat of *T. putitora* (Fig. 4). Besides, the model also predicted majority of the suitable areas in the river flow length of 8000 to 30000 m. The suitable habitats were mainly predicted in the stems of the major rivers like the Sutlej (Himachal Pradesh), Ganga and Ramganga (Uttarakhand), Trishuli and Kosi (Nepal), Teesta (West Bengal and Sikkim), Puna Tsang Chhu (Bhutan), and Kameng, Subansiri, Siang, Dibang, Lohit, and Noa-Dihing (Arunachal Pradesh). These localities represent the main spawning and feeding grounds of *T. putitora* in the Indian subcontinent due to suitable topographic and hydroclimatic characteristics.

The model predicted only 10.35% of the total river length as suitable habitat of *T. putitora* in the entire Himalayan river system that point towards its limited area of habitation. The model also predicted about 11% of the total river length under moderately suitable that forms the future potential habitats of the species and needs proper exploration and confirmation. In view of the endangered status and reported declining population in wild stocks, the model results would be helpful in identifying the priority areas for conservation and restoration of *T. putitora* in the Himalayan river system.

4. Discussion

The results show the occurrences of *T. putitora* in larger streams and rivers in low minimum upstream elevation and elongated flow lengths. Besides, low percentage of riparian deciduous broadleaf trees, high precipitation seasonality, high dissolved oxygen, and high depth to bedrock also play low to moderate role in predicting the distribution of *T. putitora*. Based on the results, the species did not occur in the absolute smallest streams in conformity with its occurrence reports from the major rivers and tributaries of Himalayan river system (Bhatt and Pandit 2015). The model showed a negative relationship of *T. putitora* with high upstream elevation by restricting the predictions to the elevation range of < 1000 m (with highest possible range up to 2100 m). The results are in agreement with the reported occurrence of the species up to an altitude of 2000 m (Cordington 1946), up to 850 m (Talwar and Jhingran 1991), 600–1200 m (Singh and Kumar 2000), 70 to 1891 m (Nautiyal 2014) in India, and 2100 m in Nepal (Shrestha 1997). The hydroclimatic variables like temperature and precipitation are highly correlated with the elevation. Hence, the high predictive power of the model in the lower altitude areas can also be attributed to the temperature and precipitation requirements of the species. As per previous studies, the spawning grounds of *T. putitora* are characterized by water temperature varying from 11 to 30.5° C (Joshi 1994; Bhatt et al. 2004) and the feeding grounds in the range of 14 to 22° C (Bhatt et al. 2004).

The river habitat is an important determinant of the condition of fish assemblages in rivers (Wang et al. 2006; Fialho et al. 2008) and there is a reciprocal relationship between river habitat and the flow length. In simple words, an elongated flow of river offers the possibilities of more diverse river habitats. Flow length is a linear aspect of drainage network and refers to the total length of stream channels in the drainage basin (Horton 1945). The flow length had a positive relationship, with higher values of flow length predicting a high likelihood of *T. putitora* occurrences. Generally, *Tor* species prefer fast-flowing rivers and streams with rocky/stony substrate (Nautiyal 2014). Hence, the presence of higher pools and ripples in the lower gradients along elongated stretches of the stream offers suitable habitats for the species.

Riparian forests are an important habitat for freshwater fish worldwide, because forests provide fish with both shelter and food including fruits, seeds, and canopy insects (Goulding 1980; Lowe-McConnell 1987; Gerking 1994). Hence, the moderate contributions of deciduous broadleaf trees in the model predictions could be attributed to the feeding requirements of the species. The rivers along the Himalayan foothills, considered to be the actual geographic range of *T. putitora* is mostly characterized by deciduous broadleaf forest. Therefore,

the decayed and semi-decayed plant matter of the riparian forests forms essential food for the species. The feeding habit of *T. putitora* on plant matters has been reported in India (Nautiyal and Lal 1984; Kishor et al. 1998), Bangladesh (Alam et al. 2002) and Nepal (Mahaseth 2015).

We also found a positive association of *T. putitora* with high precipitation seasonality, dissolved oxygen and depth to bedrock. Seasonality plays a vital role in influencing the perseverance of living organisms. Seasonal shifts in climatic conditions influence the availability of resources, which determines the presence or absence of species in an environment at temporal and spatial scale (Dayton 2008). Precipitation seasonality, a measure of the range of annual precipitation plays a key role in breeding and feeding habits of freshwater fishes. Dissolved oxygen is an important parameter in assessing water quality because of its influence on the organisms living within a body of water. According to Bhatt et al. (2004), the dissolved oxygen varies from 5.2 to 12.9 mg/l in feeding grounds while it ranges between 6.4 to 11 mg/l in the spawning grounds of *T. putitora* (Joshi 1994; Bhatt et al. 2004). The preference of depth to bedrock varies among different freshwater fishes. The model prediction towards the higher values of depth to bedrock may be attributed to the habitat preferences of *T. putitora* in large water volumes and river beds mostly covered with sand, silt and small boulders (Bhatt et al. 2004).

The findings of wide distribution and restricted area of occupancy of *T. putitora* is in accordance with the report of IUCN (2021). The results also predicted suitable areas in the rivers of Northeast Himalayas particularly Kameng, Subansiri, Siang, Dibang, Lohit and Noa-Dihing in Arunachal Pradesh that has been reported as unexplored potential habitats of *T. putitora* (Bhatt and Pandit 2015). Hence, these predictions are important in view of the estimated 50% decline in the past and 80% decline in future wild stocks of the species due to regulations in its river habitats (Sharma et al. 2015; IUCN 2021). As per available literature, habitat alteration by dams, barrages, illegal sand/boulder mining, poaching and indiscriminate fishing of brooders are largely responsible for the endangerment of *T. putitora* (Pathani 1994; Lakra et al. 2010; Nautiyal 2011; Pandit and Grumbine 2012; Khajuria et al. 2013; Gupta et al. 2014; Sharma et al. 2015). Besides, the life history and ecological traits like migratory habit, low fecundity and delayed sexual maturity have been also suggested to play a vital role in evaluating the vulnerability of the species (Nautiyal and Singh 1989; Shrestha 1997; Ogle 2002; William et al. 2005). Therefore, conservation and restoration strategies should be focused towards the unexplored areas and the critical suitable habitats. According to experts, *ex-situ* conservation of the species through development of hatcheries would be a sustainable conservation action but less genetic variability has been observed in hatchery stocks compared to its wild populations (Pradhan et al. 2011). So, priority must be given to the potential unexplored habitats by restricting human activities like fishing, extraction of sand, gravels and boulders, and by withdrawing the proposed dams or proper environmental impact assessment before construction of dams for hydropower. As of now, numbers of dams have been proposed in the major rivers of the Himalayan region which may possibly further reduce the already fragmented habitat of *T. putitora*.

5. Conclusion

Our study shows the predictive distribution of *Tor putitora* in the Himalayan river system using MaxEnt modelling technique. We used topographic, hydroclimatic, land cover, soil and water quality variables to evaluate the performance of model. The performance of model was high and reasonably good. The model predicted only 10.35% of the total river length as suitable to highly suitable indicating its limited area of habitation. About 11% of the river length has been predicted as moderately suitable which call for further explorations and confirmation of the species occurrence. Based on the reported declining wild stocks of the species, we recommended to focus conservation activities towards unexplored and critical habitats following the model predictions. We also recommend to lessen anthropogenic activities like indiscriminate fishing, sand/boulder mining, and construction of dams over the suitable river habitats of *T. putitora*. Further, our model appears to be constrained by the vastness of the study area. So, we recommend researchers to carry out distribution modelling at watershed level by integrating the local level environmental factors for more reliable prediction of habitat suitability.

Disclosure statement

The authors declare that they have no potential conflict of interests

Funding

There was no funding for the work.

Data Availability

Data available on request from the authors

Author contributions

Ranjit Mahato and Santoshkumar Abujam prepared the initial draft of the manuscript. Ranjit Mahato prepared the thematic layers, and Dhoni Bushi collected the fish sample. Oyi Dai Nimasow and Gibji Nimasow commented, edited and revised the manuscript by consulting the relevant literature. Debanshu Narayan Das further revised the manuscript after proofreading. Finally, the manuscript was read by all the authors and approved it to submit for possible publication.

Acknowledgments

The authors are thankful to the Soil and Limnological Laboratory, Department of Geography, Rajiv Gandhi University, Rono Hills, Doimukh, Arunachal Pradesh (India) for facilitating necessary infrastructural and laboratory facilities to carry out the work. The authors also acknowledge GBIF, EarthEnv, and the authors of other published articles for making available the data sources for this work.

References

- Aiello-Lammens ME, Boria RA, Radosavljevic A, Vilela B, Anderson RP. 2015. spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* 38:541–545. <https://doi.org/10.1111/ecog.01132>
- Alam MA, Dewan S, Hossain MI, Hussain MG. 2002. Food and feeding habit of *T. putitora* (Hamilton). *Univ J Zool Rajshahi Univ.* 21:21–22.
- Allouche O, Tsoar A, Kadmon R, 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *J Appl Ecol.* 43:1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Arunkumar L, Basudha C. 2003. *Tor barakae*, a new species of mahseer fish (Cyprinidae: Cyprininae) from Manipur, India. *Aquaculture* 4(2):271–276.
- Baldwin R. 2009. Use of maximum entropy modeling in wildlife research. *Entropy* 11:854–866. <https://doi.org/10.3390/e11040854>
- Bhatt JP, Nautiyal P, Singh HR. 2004. Status (1993–1994) of the endangered fish Himalayan mahseer *Tor putitora* (Hamilton) (Cyprinidae) in the mountain reaches of the river Ganga. *Asian Fish Sci.* 17:341–355.
- Bhatt JP, Pandit MK. 2016. Endangered Golden mahseer *Tor putitora* Hamilton: a review of natural history. *Rev Fish Biol Fisheries* 26:25–38. <https://doi.org/10.1007/s11160-015-9409-7>
- Bista JD, Wagle SK, Prasad S, Gurung TB, Rai AK. 2007. Growth and maturity of sahar *Tor putitora* in controlled environment in mid hill Pokhara Valley, Nepal. *Asian Fisheries Society, Indian branch, Kochi, India*, p. 143.
- Braunisch V, Coppes J, Arlettaz R, Suchant R, Schmid H, Bollmann K. 2013. Selecting from correlated climate variables: A major source of uncertainty for predicting species distributions under climate change. *Ecography* 36:971–983. <https://doi.org/10.1111/j.1600-0587.2013.00138.x>
- Chatterjee S, Hadi AS. 2006. Analysis of collinear data. In: Shewhart WA, Wilks SS. (Eds.), *Regression analysis by example*. John Wiley & Sons Inc., Hoboken, NJ. pp. 221–258.
- Cordington KDB. 1946. Notes on Indian Mahseer. *J Bombay Nat History Soc.* 46:334–336.
- Darshan A, Abujam SK, Das DN. 2018. *Biodiversity of Fishes in Arunachal Himalaya*, Academic Press, Elsevier, Massachusetts.
- Dayton GA. 2008. Seasonality. In: Jørgensen SE, Fath BD. (Eds.), *Encyclopedia of Ecology*. Elsevier B.V., Amsterdam, Netherlands, pp. 3168–3171.
- Domisch S, Amatulli G, Jetz W. 2015. Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution. *Sci data* 2(1):1–13. <https://doi.org/10.1038/sdata.2015.73>. Data available online at <http://www.earthenv.org/>

- Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carré G, Marquéz RG, Gruber B, Lafourcade B, Leitão PJ, et al. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1):27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Engler R, Waser LT, Zimmermann NE, Schaub M, Berdos S, Ginzler C, Psomas A. 2013. Combining ensemble modeling and remote sensing for mapping individual tree species at high spatial resolution. *Ecol Manage.* 310:64–73. <https://doi.org/10.1016/j.foreco.2013.07.059>
- Environmental Systems Research Institute 2016. ArcMap 10.3 (Version 10.3) [Computer software]. Environmental Systems Research Institute, Redlands, California, USA.
- Farrell A, Wang G, Rush SA, Martin JA, Belant JL, Butler AB, Godwin D. 2019. Machine learning of large-scale spatial distributions of wild turkeys with high-dimensional environmental data. *Ecol Evol.* 9(10):5938–5949. <https://doi.org/10.1002/ece3.5177>
- Feng X, Park DS, Liang Y, Pandey R, Papeş M. 2019. Collinearity in ecological niche modeling: Confusions and challenges. *Ecol Evol.* 9(18):10365–10376. <https://doi.org/10.1002/ece3.5555>
- Fialho AP, Oliveira LG, Tejerina-Garro FL, de Mérona B. 2008. Fish-habitat relationship in a tropical river under anthropogenic influences. *Hydrobiologia* 598:315–324. <https://doi.org/10.1007/s10750-007-9165-3>
- Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ Conserv.* 24(1):38–49. <https://doi.org/10.1017/S0376892997000088>
- Fricke R, Eschmeyer WN, Van der Laan R. 2021. Eschmeyer's catalog of fishes: genera, species. Online version: <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>. (Accessed 1 March 2021).
- GBIF: The Global Biodiversity Information Facility 2019. What is GBIF? <https://www.gbif.org/what-is-gbif> (Accessed 5 May 2021).
- Gerking SD. 1994. *Feeding Ecology of Fish*. Academic Press, San Diego.
- Goulding M. 1980. *The Fishes and the Forest*. University of California Press, Los Angeles.
- Graham MH. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology* 84(11):2809–2815. <https://doi.org/10.1890/02-3114>
- Greshishchev V, Onikura N, Iyooka H. 2015. Environmental factors influencing fish species distribution in irrigation channels around Ariake Sea, Kyushu, Japan. *Irrigat Drainage Sys Eng.* 4:139. <https://doi.org/10.4172/2168-9768.1000139>
- Gupta, N., Sivakumar, K., Mathur, V.B., Chadwick, M.A., 2014. The ‘tiger of Indian rivers’: stakeholders’ perspectives on the golden mahseer as a flagship fish species. *Area* <https://doi.org/10.1111/area.12124>

- Hamilton F. 1822. An account of the fishes found in the river Ganges and its branches. George Ramsay and Co., Edinburgh.
- Hamilton CM, Batemana BL, Gorzoa JM, Reid B, Thogmartinc WE, Peeryd MZ, Heglunde PJ, Radeloffa VC, Pidgeona AM. 2018. Slow and steady wins the race? Future climate and land use change leaves the imperiled Blanding's turtle (*Emydoidea blandingii*) behind. Biol Conserv. 222:75–85. <https://doi.org/10.1016/j.biocon.2018.03.026>
- Hijmans RJ, Phillips S, Leathwick J, Elith J. 2011. dismo package for R (Version 4.0.3). <https://cran.r-project.org/package=dismo>.
- Horton, R.E. 1945. Erosional development of streams and their drainage basins: Hydro-physical approach to qualitative morphology. Bull. Geol. Soc. Am. 56, 275–370. <https://doi.org/10.1177/030913339501900406>
- IUCN 2021-1. Red List of Threatened Species. <https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T126319882A126322226.en>. (Accessed 1 June 2021).
- Jerdon TC. 1849. On the fresh-water fishes of southern India. Madras J Lit Sci. 15(2):302–346.
- Johnson JA, Parmar R, Ramesh K, Sen S, Murthy RS. 2012. Fish diversity and assemblage structure in Ken River of Panna landscape, central India. JoTT 4:3161–3172. <https://doi.org/10.11609/JoTT.o3024.3161-72>
- Joshi CB. 1994. Alikhad (Sutlej river system): juvenile habitat and breeding grounds. In: Nautiyal P. (Ed.), Mahseer the game fish. Jagdamba Publication, Dehradun, pp. B102–B111.
- Kadmon R, Farber O, Danin, A. 2004. Effect of roadside bias on the accuracy of predictive maps produced by bioclimatic models. Ecol Appl. 14:401–413. <https://doi.org/10.1890/02-5364>
- Khajuria B, Langer S, Tripathi NK. 2013. Status of Golden mahseer (*Tor putitora*) in Jammu region (J&K). Int J Rec Sci Res. 4:1154–1156.
- Kishor B, Bhatt JP, Rawat VS, Nautiyal P. 1998. Variations in food habit of the Himalayan mahseer – *Tor putitora* (Ham.) inhabiting the Ganga river system in Garhwal region. Indian J Fish 45:113–118.
- Kurup BM, Radhakrishnan KV. 2011. *Tor remadevii*, a new species of *Tor* (Gray) from Chinnar Wildlife Sanctuary, Pambar River, Kerala, southern India. J Bombay Nat Hist Soc. 107(3):227–230.
- Lakra WS, Goswami M, Sarkar UK. 2010. Conservation biology of Indian mahseer. Indian J Anim Sci. 80:98–108.
- Lowe-McConnell RH. 1987. Ecological studies in tropical fish communities. Cambridge University Press, Cambridge.
- Mahaseeth VK. 2015. Food and feeding habit of *T. putitora* of Mahakali River, Nepal. BIBECHANA 13:121–131.

- Manzoor SA, Griffiths G, Lukac M. 2018. Species distribution model transferability and model grain size—finer may not always be better. *Sci Rep.* 8(1):7168. <https://doi.org/10.1038/s41598-018-25437-1>
- Menon AGK. 1992. Taxonomy of mahseer fishes of the genus *Tor* Gray with description of a new species from the Deccan. *J Bombay Nat Hist Soc.* 89(2):210–228.
- Merow C, Smith MJ, Silander Jr JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography* 36(10):1058–1069. <https://doi.org/10.1111/j.1600-0587.2013.07872.x>
- Miller J. 2010. Species distribution modeling. *Geogr. Compass* 4(6):490–509. <https://doi.org/10.1111/j.1749-8198.2010.00351.x>
- Ministry of Agriculture and Forests, Bhutan 2020. Bhutan: A pioneer in Mahseer conservation. <http://www.moaf.gov.bt/bhutan-a-pioneer-in-mahseer-conservation/> (Accessed 11 June 2021).
- Moreno R, Zamora R, Molina JR, Vasquez A, Herrera MÁ. 2011. Predictive modeling of microhabitats for endemic birds in South Chilean temperate forests using Maximum entropy (Maxent). *Ecol Inform.* 6:364–370. <https://doi.org/10.1016/j.ecoinf.2011.07.003>
- Naimi B, Araújo MB. 2016. Sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* 39:368–375. <https://doi.org/10.1111/ecog.01881>
- Nautiyal P. 2011. The golden mahseer (a threatened fish of Himalaya). Lambert Academic Publishing, Leipzig.
- Nautiyal P. 2014. Review of the art and sciences of Indian Mahseer (game fish) from nineteenth century: road to extinction or conservation? *Proc Natl Acad Sci India Sect B Biol Sci.* 84(2):215–236. <https://doi.org/10.1007/s40011-013-0233-3>
- Nautiyal P, Lal MS. 1984. Food and feeding habits of fingerlings and juveniles of mahseer (*Tor putitora* Ham) in Nayar river. *J Bombay Nat Hist Soc.* 81:642–646.
- Nautiyal P, Singh HR. 1989. River projects and endangered hill stream mahseer *Tor putitora* in Indian uplands of Garhwal Himalaya, a critical evaluation. *Him J Env Zool.* 3:129–134.
- Nimasow G, Nimasow OD, Rawat JS, Tsering G, Litin T. 2016. Remote sensing and GIS-based suitability modeling of medicinal plant (*Taxus baccata* Linn.) in Tawang district, Arunachal Pradesh, India. *Curr Sci.* 110(2):219–227. <https://doi.org/10.18520/cs/v110/i2/219-227>
- Ogle SN. 2002. Mahseer breeding and conservation and possibilities of commercial culture: the Indian experience. In: Petr T, Swar SB. (Eds.), *Coldwater fisheries in the trans-Himalayan countries*. FAO Fisheries Technical Paper 431, pp. 193–212.
- Oo W. 2002. Inland fisheries of the Union of Myanmar. In: Petr T, Swar SB. (Eds.), *Cold water fisheries in the Trans-Himalayan countries*. FAO Fish. Technical Paper, 431.

- Pandit MK, Grumbine RE. 2012. Potential effects of ongoing and proposed, hydropower development on terrestrial biological diversity in the Indian Himalaya. *Conserv Biol.* 26:1061–1071. <https://doi.org/10.1111/j.1523-1739.2012.01918.x>
- Pathani SS. 1994. Biology of *tor* and *putitor* mahseer. In: Nautiyal P. (Ed.) Mahseer the game fish. Jagdamba Publications, Dehradun, pp. B86–B90.
- Peterson AT, Papeş M, Eaton M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30:550–560. <https://doi.org/10.1111/j.2007.0906-7590.05102.x>
- Phillips S, Dudík M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31:161–175. <https://doi.org/10.1111/j.0906-7590.2008.5203.x>
- Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecol Model.* 190(3-4):231–259. <https://doi.org/10.1016/j.ecolmodel.2005.03.026>
- Radosavljevic A, Anderson RP. 2014. Making better Maxent models of species distributions: complexity, overfitting and evaluation. *J Biogeogr.* 41(4):629–643. <https://doi.org/10.1111/jbi.12227>
- Rai AK, Rana AK, Bastola KB. 1997. Comparative growth study of Indigenous fish species Mahseer, Katle and Asla fed with local feed ingredients. *J Nep Agric Res.* 1:57–61.
- Rai AK, Wagl, SK, Pradhan N, Basnet SR, Baidya AP, Gurung TB, Bista JD, Shrestha MK, Thapa AB. 2006. Status of sahar *Tor putitora* and its prospects in Nepal. First International Symposium of the mahseer, Kualampur, Malasiya from 29–30 March 2006. pp. 1–23.
- Rajbanshi KG, Csavas I. 1982. Aquaculture Development in Bhutan - a report prepared for the establishment of Fish Seed Production Centre Project. FAO FI:DP/BHU80/007, Field Doc. 1, p. 46.
- Roy S, Ray S, Saikia SK. 2021. Indicator environmental variables in regulating the distribution patterns of small freshwater fish *Amblypharyngodon mola* in India and Bangladesh. *Ecol Indic.* 120:106906. <https://doi.org/10.1016/j.ecolind.2020.106906>
- Ruiz-Navarro A, Gillingham PK, Britton JR. 2016. Predicting shifts in the climate space of freshwater fishes in Great Britain due to climate change. *Biol Conserv.* 203:33–42. <https://doi.org/10.1016/j.biocon.2016.08.021>
- Sharma KK, Mohan VC, Kouser U. 2015. Comparative accounts of merestic count and morphometric measurements of Golden mahseer (*Tor putitora*) among Chenani hydroelectric dam, Jhajjar stream and Dansar stream, (J&K) India. *Indian J Appl Res.* 5:772–774.
- Shrestha TK. 1997. The mahseer in the rivers of Nepal disrupted by dams and ranching strategies. R.K. Printers, Teku, Kathmandu.

- Singh HR, Kumar N. 2000. Some aspects of ecology of hillstreams; stream morphology, zonation, characteristics, and adaptive features of ichthyofauna in Garhwal Himalaya. In: Datta Munshi JS. (Ed.), Modern trends in Fish Biology Research, Narendra Publishing House, New Delhi. pp. 1–18.
- Sony RK, Sen S, Kumar S, Sen M, Jayahari KM. 2018. Niche models inform the effects of climate change on the endangered Nilgiri Tahr (*Nilgiritragus hylocrius*) populations in the southern Western Ghats, India. Ecol Eng. 120:355–363. <https://doi.org/10.1016/j.ecoleng.2018.06.017>
- Sykes WH. (1839). On the fishes of the Deccan. Proc Zool Soc Lond. 6:157–165.
- Talwar PK, Jhingran AG. 1991. Inland fishes of India and adjacent countries. Vol. 1. A.A. Balkema, Rotterdam, p. 541.
- Vatcheva KP, Lee M, McCormick JB, Rahbar MH. 2016. Multicollinearity in regression analyses conducted in epidemiologic studies. Epidemiol. 6:227. <https://doi.org/10.4172/2161-1165.1000227>
- Veloz SD. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. J Biogeogr. 36:2290–2299. <https://doi.org/10.1111/j.1365-2699.2009.02174.x>
- Wang L, Seelbach PW, Hughes RM. 2006. Introduction to landscape influences on stream habitats and biological assemblages. In: Hughes RM, Wang L, Seelbach PW. (Eds.), Landscape influences on stream habitats and biological assemblages. Am. Fish Soc. Symp. 48, Bethesda, MD, pp. 1–23.
- Wang L, Seelbach PW, Hughes RM. 2006. Introduction to landscape influences on stream habitats and biological assemblages. Am Fish Soc Symp. 48:1–23.
- William WLC, Tony JP, Daniel P. 2005. A fuzzy logic expert system to estimate intrinsic extinction vulnerabilities of marine fishes to fishing. Biol Conserv. 124:97–111. <https://doi.org/10.1016/j.biocon.2005.01.017>
- Yang X-Q, Kushwaha S, Saran S, Xu J, Roy P. 2013. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in lesser Himalayan foothills. Ecol Eng. 51:83–87. <https://doi.org/10.1016/j.ecoleng.2012.12.004>
- Zimmermann NE, Edwards TC, Moisen GG, Frescino TS, Blackard JA. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. J Appl Ecol. 44(5):1057–1067. <https://doi.org/10.1111/j.1365-2664.2007.01348.x>

Table 1. Post collinearity variables assessed for inclusion in the model

Variable code	Variable name	Unit	Source
elv_01	Upstream elevation (min)	Meter [m]	HydroSHEDS
slope_01	Upstream slope (min)	Degree * 100	HydroSHEDS
flow_acc_01	Flow length	Count	HydroSHEDS
hydroclim_03	Isothermality (BIO2/BIO7) (×100)	* 100	WorldClim
hydroclim_15	Precipitation Seasonality (Coefficient of Variation)	* 100	WorldClim
lc_avg_01	Evergreen/Deciduous Needle leaf Trees	[%]	Consensus land cover (CLC)
lc_avg_03	Deciduous Broadleaf Trees	[%]	Consensus land cover (CLC)
lc_avg_04	Mixed/Other Trees	[%]	Consensus land cover (CLC)
lc_avg_05	Shrubs	[%]	Consensus land cover (CLC)
lc_avg_07	Cultivated and Managed Vegetation	[%]	Consensus land cover (CLC)
lc_avg_09	Urban/Built-up	[%]	Consensus land cover (CLC)
lc_avg_10	Snow/Ice	[%]	Consensus Land Cover
lc_avg_11	Barren	[%]	Consensus land cover (CLC)
lc_avg_12	Open Water	[%]	Consensus land cover (CLC)
soil_avg_04	Silt content mass fraction	[%]	World soil information (ISRIC)
soil_avg_07	Cation exchange capacity	[cmol/kg]	World soil information (ISRIC)
soil_avg_08	Bulk density of the fine earth fraction	[kg / m ³]	World soil information (ISRIC)
soil_avg_09	Depth to bedrock (R horizon) up to max. 240 cm	[cm]	World soil information (ISRIC)
do	Dissolved oxygen	[mg/l]	IDW interpolation in ArcGIS 10.3

Table 2. Variables retained for the final model (**shown in bold**) with percent contribution and permutation importance (**averaged over 5 replicate runs**)

Variable code	Variable name	Contribution (%)	Permutation importance (%)
elv_01	Upstream elevation (min)	35.3	41.7
flow_acc_01	Flow length (m)	21.4	4.4
lc_avg_03	Deciduous broadleaf trees	7.8	2.1
lc_avg_01	Evergreen/Deciduous needle leaf trees	5.4	8.7
lc_avg_05	Shrubs	3.9	0.6
hydroclim_15	Precipitation Seasonality (coefficient of variation)	3.5	9.1
lc_avg_07	Cultivated and managed vegetation	3.5	6.6
do	Dissolved oxygen	2.3	2.2
Slope_01	Upstream slope (min)	2.3	0.7
lc_avg_11	Barren	2.2	5.9
lc_avg_09	Urban/Built-up	2.2	0.4
soil_avg_09	Depth to bedrock (R horizon) up to max. 240 cm	1.4	1.5

Table 3. Suitable categories of *T. putitora* in the Himalayan river system

Categories	Suitability index	River length (km)	Percentage
Unsuitable	< 0.1	34265.75	56.83
Slightly suitable	0.1 – 0.3	12979.36	21.53
Moderately suitable	0.3 – 0.5	6807.67	11.29
Suitable	0.5 – 0.7	5003.62	8.30
Highly suitable	> 0.7	1234.16	2.05
Total		60290.56	100.00

Figure 1. Map of the study area. Generated by the authors using ArcGIS 10.3, <http://esri.com>

Figure 2. Flowchart of methodology followed in the study

Figure 3. Species response curve (derived from MaxEnt). Predictor variables are shown in x -axis and species response in y -axis represented by Minimum Upstream Elevation (elv_01), Flow Length (flow_acc_01), Precipitation Seasonality (hydroclim_15), Deciduous Broadleaf Trees (lc_avg_03), Depth to Bedrock (soil_avg_09), and Dissolved Oxygen (do).

Figure 4. Predictive habitat suitability of *T. putitora* in the Himalayan river system. Generated by the authors using ArcGIS 10.3, <http://esri.com>