PREDICTIVE HABITAT SUITABILITY MODELLING OF *AXIS PORCINUS* (HOG DEER) UNDER CURRENT AND FUTURE CLIMATE CHANGE SCENARIOS IN PUNJAB, PAKISTAN

AZEEM, A.[1,2*] – AHMED, S. R.[1] – QADIR, A.[1] – HUSSAINY, A. S.[2]

[1]*College of Earth & Environmental Sciences, University of the Punjab, Lahore, Pakistan*
(phone: +92-312-980-1510)

[2]*The Urban Unit, 503 Shaheen Complex, Egerton Road, Lahore, Pakistan*
(phone: + 92-315-410-5894)

**Corresponding author*
e-mail: aneeqa.azeem@gmail.com; phone: +92-315-410-5894: fax: +92-37-504-356

(Received 5th Mar 2021; accepted 14th May 2021)

Abstract. It is anticipated that climate change will cause biodiversity loss by altering natural habitats and species distribution. The main purpose of the study was to model the current and future distribution of *Axis porcinus* to predict the changes in their habitat in Punjab, Pakistan. A Total of 32 variables including bioclimatic, natural, topographical and human impact variables were prepared. Dimension reduction was done by three methods, namely the Pearson's correlation, multi-collinearity analysis and principal component analysis (PCA) to achieve the appropriate number of predictors. The study predicted the potential distribution of species by the 2050s and 2070s under representative concentrative pathways (RCPs) RCP 4.5 and RCP 8.5 climate change scenarios using earth observations and maximum entropy (MaxEnt) machine learning model. Results revealed that highly suitable areas for current distribution of *Axis porcinus* cover 451.3 Km^2. According to future projections suitable habitat will face an 18.7% decline by 2050s according to RCP 4.5 scenario or 52.8% based on the RCP 8.5 scenario which is alarming and protection measures are crucial. Based on the current and future distribution of the species, three priority conservation areas covering 632 Km^2 for *Axis porcinus* are identified. This study supports the formulation of current conservation policies and strategies for protection of the species keeping in view the impact of climate change scenarios.
Keywords: *species distribution, geospatial big data, modelling, maxEnt, machine learning*

Introduction

Habitat is a place which is very important for the existence, reproduction and population development of organisms. The habitat can effect directly the distribution, quality, and quantity and survival rate of the organism (Zhang et al., 2019). Currently the habitat loss or disintegration of habitat is the major factor which is threatening the survival of organisms (Brooks et al., 2002; Haddad et al., 2015). It is anticipated that climate change will cause biodiversity loss by altering the natural habitat and distribution of species. The likely impacts of climate change can be mitigated by effectively targeting conservation strategies to important habitats and sensitive ecosystems (Qin et al., 2017). This is possible through modeling the future distributions of threatened species and identifying their likely geographic range and habitats (Qin et al., 2017). Species distribution modeling can be carried out through various methods though quality of data limit choices (Engler et al., 2004; Kwon et al., 2016). MaxEnt (maximum entropy) is a powerful machine learning model which uses the probability distribution of maximum entropy to estimate target probability distribution (Phillips et

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

al., 2006). This framework performs well even with scarce presence data and narrow-range species (Phillips et al., 2006; Qin et al., 2017).

The *Axis porcinus* (Hog Deer) was an ‘endemic’ species, geographically confined to South and Southeast Asia. It was once widespread, but the population has declined rapidly across its geographic range. It is estimated that the global decline rate over the last 21 years is 50% and that the species has declined by more than 90% within its Southeast Asian range (Angom et al., 2020). In India, the species is distributed throughout the northern plains and in the Northeast region. On the basis of these estimates (the average for the species in three generations), the species has been listed as “Endangered” in the international union for conservation of nature (IUCN) Red List (IUCN, 2019). Hunting, habitat loss and habitat degradation have been the major drivers of the decline (Angom et al., 2020).

In Punjab, it is distributed in riverine forest. Its wild populations suffer from immense hunting pressures and habitat destruction due to some agricultural activities (Timmins et al., 2015). Much of its natural habitat is drying out due to anthropogenic control by local communities. The population density is approximately 11.1animals/km^2 in riverine forest of Taunsa barrage and 1.2 animals/km^2 in riverine forest of Narowal in Punjab (Arshad et al., 2012; Iqbal et al., 2013).

The extant population is now patchily distributed and highly fragmented. In view of the aforementioned, this research was carried out to identify habitat suitability of *Axis porcinus* (Hog Deer) in Punjab Pakistan as aid in conservation plan.

An increase in average global temperature up to 5.8 °C has been predicted by The International Panel of Climate Change (IPCC) (Griggs and Noguer, 2002). As a few species have responded to the increase of 0.6 °C in temperature during the previous century, substantial change in species’ distribution and natural ecosystems are expected in future (Root et al., 2003). As, individual species respond differently to change in environment, identifying and prioritizing species at risk is imperative for effective mitigation and conservation (Jones et al., 2013). Shifts in geographic range of species under various climate change scenarios can be predicted by incorporating variables developed from climate change projections. The magnitude of climate change in future depends on the pathways chosen today regarding greenhouse gas emissions, concentrations and radiative forcing. Representative concentration pathways (RCPs) are four trajectories developed by IPCC which represent different scenarios (of emissions, concentrations and radiative forcing) and consequential future climate (Van Vuuren et al., 2011). The RCPs are named according to range of radiative forcing values in year 2100. The four scenarios based on radiative forcing values and the equivalent CO_2 concentrations are summarized in *Table 1*. In the current study RCP 4.5 and RCP 8.5 have been tested.

Table 1. *Summary of RCPs*

Scenario	Description	CO2 equivalent (ppm)
RCP 2.6	Peak in radiative forcing at approx. 2.6 W/m2 before 2100 and decline	490
RCP 4.5	Stabilization without overshoot pathway to 4.5 W/m2 before 2100	630
RCP 6.0	Stabilization without overshoot pathway to 6 W/m2 before 2100	800
RCP 8.5	Rising radiative forcing pathway leading to 8.5 W/m2 in 2100	1313

APPLIED ECOLOGY AND ENVIRONMENTAL RESEARCH 19(4):3181-3201.
http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

In order to reduce the threat to various species, we have to understand the suitability of habitat for each species and to eliminate the influencing factor which causes loss or destruction of habitat (Austin, 2002). From an ecological point of view, the environmental analyst can predict and effect the species distribution and their suitable habitat (Wiens, 2011).

The maximum entropy is basically used to find out the probability distribution for occurrence of an event with greatest uncertainty and subject to some constraint that statistical distribution moment match with the sample moments of observations. MaxEnt can be used for presence-only (PO) data which is equivalent to Poisson point process model, a spatial statistical model for counted data. For habitat suitability mapping we use a large number of landscape variables in order to predict habitat suitability. Many of those landscape variable are highly correlated to one another, which leads to multi-collinearity in habitat and resource selection models (Farrel et al., 2019).

Species distribution models (SDMs) are of great importance in informing conservation planning of species for global climate change. We embolden the conservation community to clasp a coherent use of SDMs as mean of participation of stakeholders in consultations of future scenarios and decisions required to meet the desired outcome of conservation management. The escalating developments in geographical information system (GIS) and remote sensing have made it possible to integrate it with statistical models like species distribution models (Rahman et al., 2019). Maximum entropy (MaxEnt) is a powerful machine learning model which uses the probability distribution of maximum entropy to estimate target probability distribution (Phillips et al., 2006). Its simplicity of use made it one of the commonly used algorithm of specie distribution modelling (Rahman et al., 2019). This framework performs well even with scarce presence data and narrow-range species (Phillips et al., 2006; Qin et al., 2017). In this paper maximum entropy method through MaxEnt (version 3.4.1) was employed for current and future distribution modelling of *Axis porcinus* in Punjab province Pakistan using climate change scenarios.

Materials and methods

Study area

The study is carried out for Punjab, the land of five rivers, is amongst the most heavily irrigated landscapes on earth with a canal system spread all over the province. Its location map has been shown as *Figure 1*. Approximately 80 mammals, 10 amphibians, 85 reptiles, and 500 birds have been reported from Punjab (Ali, 2008). This vast diversity of flora and fauna in Punjab is attributed to its geographical position, topography, and climate.

Methodology

The detailed habitat suitability model with overall structure is described as *Figure 2*. Occurrence data, climatic, natural land feature, topographical feature and human impact feature are collected initially to configure independent variables of our model. We constructed the multiple MaxEnt models depending upon different predictors sets. In order to assess the impact of climate change on distribution of *Axis porcinus,* Hadley Centre Global Environment Model version 2 (HadGEMV2-AO) is employed in

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

distribution modelling. Finally results and model evaluation was done using area under the Curve (AUC) deployed for models performance assessment.

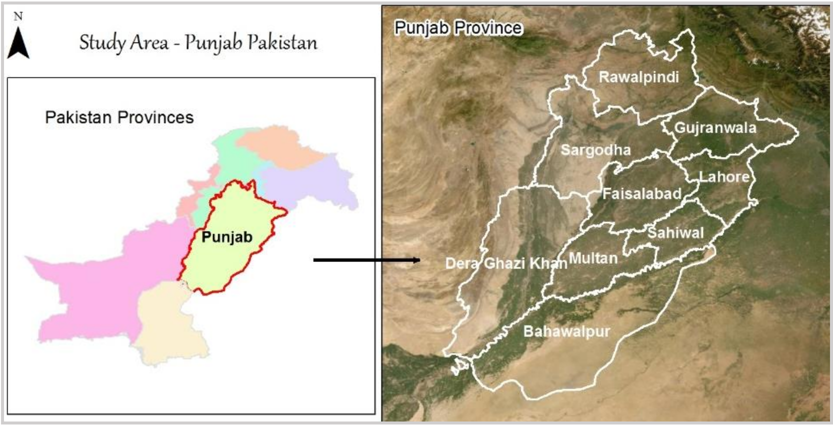

***Figure 1.** Study area - Punjab Pakistan*

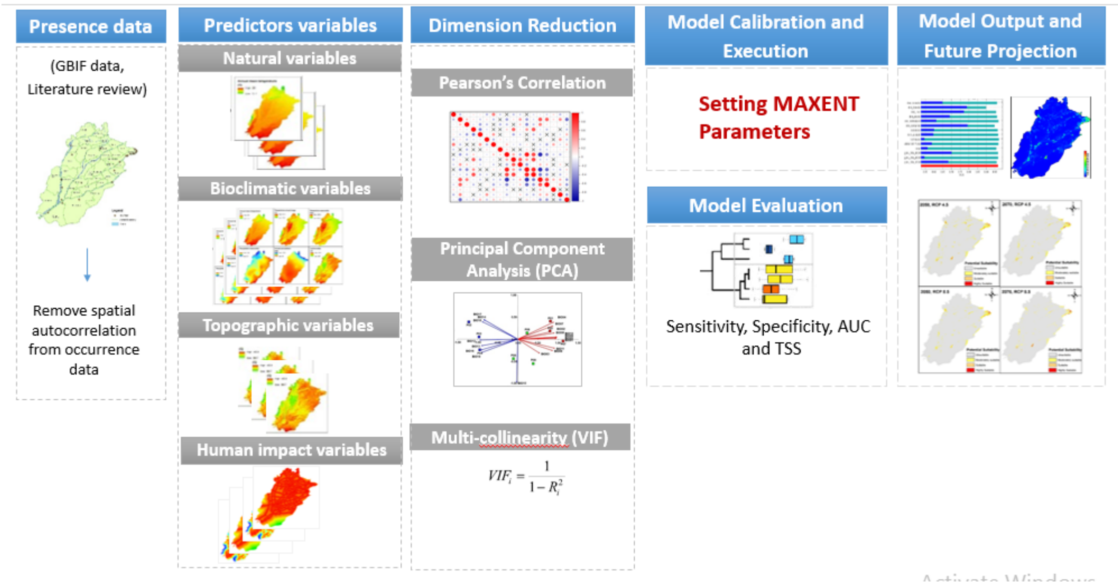

***Figure 2.** Overall Methodological process of habitat suitability model*

Dataset construction

The performance of any model generally depends on the quality and size of the datasets used for training. To build our datasets we reviewed various databases that have observation of various species globally. We gathered and compiled the occurrence data for *Axis porcinus* from three main sources. The sources include (a) GBIF (Global Biodiversity Information Facility) (https://www.gbif.org), (b) published literature and (c) expert knowledge. The model accuracy would be affected if the points are too close from each other therefore to reduce the inherent spatial bias presence records were screened in ArcGIS (version 10.4) to eliminate spatial autocorrelation and guarantee independence; nearest neighborhood analysis was used for this purpose (Bosso et al., 2016; Kwon et al., 2016) using SDM Toolbox in ArcGIS 10.3. Out of total 38 occurrence record, 26 spatially rarified presence points were reserved and used in final modelling.

Species habitats are closely related to climatic and land conditions. The selection of predictor environmental variables considers their restraining impact on specie distribution and spatial correlation among these variables.

Therefore bio-climatic (Bio 01 - Bio 19) variables were collected, obtained from Worldclim website at 30 arc sec resolution (https://www.worldclim.org/). Climatic variables are frequently used for modeling as they have direct effect on distribution modeling (Guisan and Zimmermann, 2000). Thirteen other incorporated variables includes; topography, natural and human impact variables that may have direct or indirect relation with *Axis porcinus* habitat. In total 32 predictor's variables were used for our habitat modelling. The description of the habitat suitability predictor variables along with code and measurement units are given in *Table 2*. To estimate the impact of future climate on *Axis porcinus*, general circulation model (GCM) Hadley Centre Global Environment Model version 2 (HadGEM2-AO) was used to test habitat suitability in the 2050s (2041-2060) and 2070s (2061-2080), RCPs 4.5 and RCP 8.5. HadGEM2-AO played an important role in assessing future climate at national level.

Preparing predictors variables and preprocessing

Conducting habitat suitability requires preprocessing of the collected predictors variables. As The datasets of predictors variables were collected from various sources including environmental earth data. Species distribution models that employed geospatial earth data are frequently used to predict the spatial patterns of species. Yet there are ample mismatches in the spatial and temporal resolution of these datasets (Yang, 2019). Our study area Punjab province consist of huge area of 205344 Km^2 therefore big geospatial datasets were involved in preparing predictors variables.

Big geospatial data and predictor's variable

The earth observation data and derived products or information are vital to understand, model and predicting natural processes as well as the current and future state of human- Earth system (Sudmanns, 2020).

Big geospatial data is a new data rigorous method comprises of large volumes of data with spatial information, containing data related to land, environment, oceans, the atmosphere and human activities. Big spatial data is created by different earth observing satellites. The general properties of big data includes; large volumes, multitemporal, and multisource datasets. In addition to this it has physical correlation with earth observation, communication, and computation and network technologies at its fundamental. Google Earth Engine (GEE) provides terabytes of satellite-based data which is the main support for national and regional scale level analysis without downloading the huge size of datasets which consume lots of time in processing and analysis.

In some cases utilizing environmental data is as straight forward as directly detecting species for mitigation of an invasive species. Like reflectance properties of vegetation which is the foundation of mapping plants using various kinds of indices e.g. normalized difference vegetation index (NDVI) or enhanced vegetation index (EVI), can be derived from remote sensing data. SDMs uses environmental data which can be categorized as bio physical or climatic data. Both of them can be measured by proximal

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

sensing and derived from geospatial big data (Yang, 2019). A list of environmental or predictor's variables that can be derived from big geospatial data is provided as *Table 3*. Information on spatial and temporal resolution, geographical coverage and extent of the data is also given.

Table 2. *List of predictor variables and datasets*

Sr. No.	Description	Code	Unit/scale range
	Bioclimatic variables		
1	Annual mean temperature	Bio_01	°C
2	Mean diurnal range	Bio_02	°C
3	Isothermality	Bio_03	%
4	temperature seasonality	Bio_04	%
5	Max. temp of warmest month	Bio_05	°C
6	Min/temp of coldest month	Bio_06	°C
7	Temperature annual range	Bio_07	°C
8	Mean temperature of wettest quarter	Bio_08	°C
9	Mean temperature of driest quarter	Bio_09	°C
10	Mean temperature of warmest quarter	Bio_10	°C
11	Mean temperature of coldest quarter	Bio_11	°C
12	Annual precipitation	Bio_12	mm
13	Precipitation of wettest month	Bio_13	mm
14	Precipitation of driest month	Bio_14	mm
15	Precipitation seasonality	Bio_15	%
16	Precipitation of wettest quarter	Bio_16	mm
17	Precipitation of driest quarter	Bio_17	mm
18	Precipitation of warmest quarter	Bio_18	mm
19	Precipitation of coldest quarter	Bio_19	mm
	Natural variables		
20	Normalized difference vegetation index	NDVI	Scale -1 to + 1
21	Distance to wetlands	Den_Wetland	Meters
22	Distance to rivers	Dist_ rivers	Meters
23	Distance to forest	Dist_ forest	Meters
	Topography		
24	Elevation	Elevation	Ranges 2247 to 47
25	Eastness	Eastness	Scale -1 to + 1
26	Slope	Slope_rad	Scale 0 to 1
27	Northness	Northness	Scale -1 to + 1
	Human impact		
28	Human population density	Population	Map units (m2)
29	Distance to settlements	Dist_Settl	Meters
30	Distance to roads	Dist_rd	Meters
31	Density of distributary canals	Den_ Distributary	Map units (m2)
32	Density of industries	Den_ Industries	Map units (m2)

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Table 3. *Environmental or predictor's data used in specie distribution modelling along with variables that can be derived from big geospatial data*

Mission/sensor	**Predictor variables**	**Spatial resolution**	**Temporal resolution**	**Extent and coverage**
NASA-MODIS	Vegetation and water indices (e.g., NDVI, EVI, SAVI, NDWI) Land Surface Temperature (LST), Land cover	0.25–1 km	4 times/day	2001–present, Global
USGS Landsat series	Vegetation and water indices (e.g., NDVI, EVI, SAVI, NDWI) Land cover	30 m	16 days	1972–present, Global
ESA SENTINEL missions	Vegetation and water indices (e.g., NDVI, EVI, SAVI, NDWI) Land Surface Temperature, Land cover	10–300 m	3–10 days	2015–present, Global
NOAA VIIRS	NDVI, NDWI, LST imagery, human settlements	375–750 m	1 day–monthly	2015–present, Global
Global Precipitation Measurement Mission (GPMM)	Precipitation	11 km	2–3 h	2015–present, Global
Tropical Rainfall Measuring Mission (TRMM)	Rainfall	28 km	3 h–7 d	1998–2015, Tropical and subtropical regions
Data sets	**Predictor variables**	**Spatial resolution**	**Temporal resolution**	**Extent and coverage**
WorldClim BIO Variables	2 m air temperature and precipitation	~1 km	LTA	1950–2000, Global
MODIS Land Cover Type/Dynamics	Landcover	0.5–1 km	Yearly/twice a year	2001–present, Global
Copernicus Land Cover	Landcover	100 m	Multiyear	2015–present, Global
ASTER-GDEM V2	Digital Elevation Model, Elevation related variables (slope, aspect, hillshade, curvature etc)	30 m		Global

All predictors variable were grouped into four categories; included climatic, topographic, natural features and human impact features for the sake of understanding. Proximity distances layers from land features are regarded as critical for modelling because species traits and their habitat are closely related. The performance of habitat suitability models have improved in various studies by considering the distance between environmental layer and species occurrence. That is why we engaged proximity distances as input variables. NDVI, distance to forests, distance to rivers, distance to wetlands were categorized as natural variables. NDVI was computed from Sentinel 2A using formula NDVI = B8-B4 / B8-B4, for year 2020 with <10% of cloud cover, in Google Earth Engine (GEE). The output was used to resample at the resolution of ~ 808 m. Rivers were extracted from the Landsat OLI (operational land imager) data classification, using maximum likelihood classification technique in Google Earth

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Engine. Training data for this generated on google earth as polygons. 70% of the data was used for training and 30% for validation purpose. Wetlands were digitized on Google Earth and converted to shapefile. Forest dataset was collected from the urban unit, Govt. of Punjab, Pk as shapefile.

Elevation, slop, eastness, northness were classified as topographic variables. Topographical features were extracted from ASTER-GDEM V2 digital elevation model having 30 m resolution (https://asterweb.jpl.nasa.gov/gdem.asp).

Population density, distance to roads, distance to settlements, density of canals, density of industries were grouped as human impact variables. LandScan™ (https://landscan.ornl.gov/) population count data with 1 km m resolution was used as raster. Settlements data was extracted from the Landsat OLI data classification using maximum likelihood classification technique in Google Earth Engine. Data of roads and industries location was obtained from the Urban Unit government of Punjab, Pakistan, as shapefile format. Data on canals/distributaries were collected from the Irrigation department government of the Punjab as vector file.

Moreover distance to features for all layers were constructed using Euclidian distance function in Spatial Analyst toolbar in ArcGIS 10.4. A Euclidian distance output raster holds the measured distance form each cell to the nearest source. Kernel density function was performed to create density rasters for all layers. This kernel density calculated the density (magnitude per unit area) from features using kernel function. All dataset and predictor variables layers were prepared at a single cell size (808 m), geographic extent and projection. These ASCII files were employed in the modelling analysis. The pre-processing process is described in *Figure 3*.

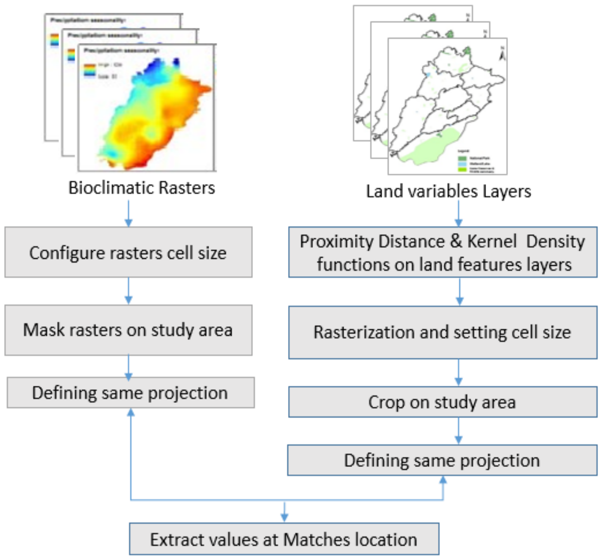

Figure 3. *The pre-processing process*

Dimension reduction

Too many variable may over fit the model and affect model accuracy. The over-fitting of the model prompted by correlation among predictor variables was avoided using three methods of dimension reduction. Two widely used statistical approaches were used in hierarchical way to reduce multi-collinearity among variables. First a

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

priori selection was done based on Pearson's Correlation to remove highly correlated bioclimatic variables with r > 0.75 as cutoff value. Only the uncorrelated and relevant bioclimatic variables were retained. Then principal component analysis (PCA) of environmental variable were performed. Pearson's correlation test gave 3 bio-cliamtic variables based on r = 0.75 encompassing bio 1, 4, 7 and 12 which were also retained because of their importance according to experts and also to assess their contribution. Correlation analysis among uncorrelated environmental variables is shown in *Figure 4*. This gave a set of 18 predictor's variables.

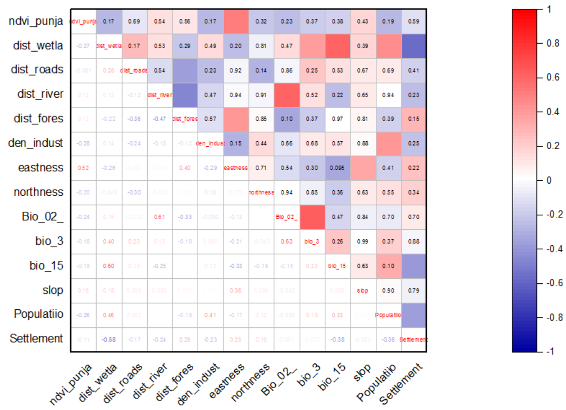

Figure 4. *Pearson's correlation among the predictor's variables*

The PCA gave 15 variables. The scree plot of all principal components is shown as *Figure 5*. The first 9 principal components (PCs) explained 91% of variability in the environmental variables. A 3D-biplot of first three principal components are show in *Figure 6*.

Further the multicollinearity analysis gave 13 environmental variable having variance inflation factor (VIF) less than 5, that shows there is no multicollinearity among them (*Table 4*).

Table 4. *Multicollinearity analysis results*

Variables	Tolerance	VIF
slop	50.3%	1.99
Bio 02	50.1%	2.00
Bio 03	21.9%	4.52
Bio 15	23.2%	4.31
NDVI	30.0%	3.33
dist_wetland	29.7%	3.37
dist_roads	35.1%	2.85
dist_river	41.8%	2.39
dist_forest	35.9%	2.78
den_industry	49.7%	2.01
den_canals	22.3%	4.49
eastness	26.1%	3.84
northness	54.4%	1.84

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

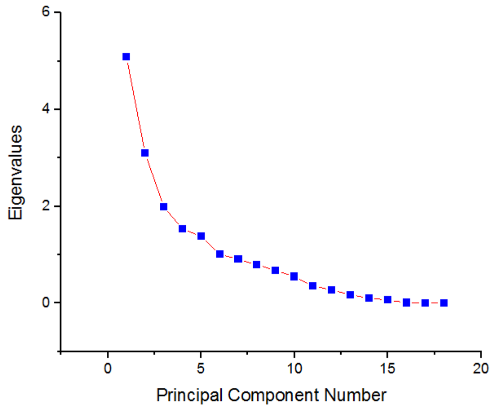

Figure 5. *Scree plot of all principal components*

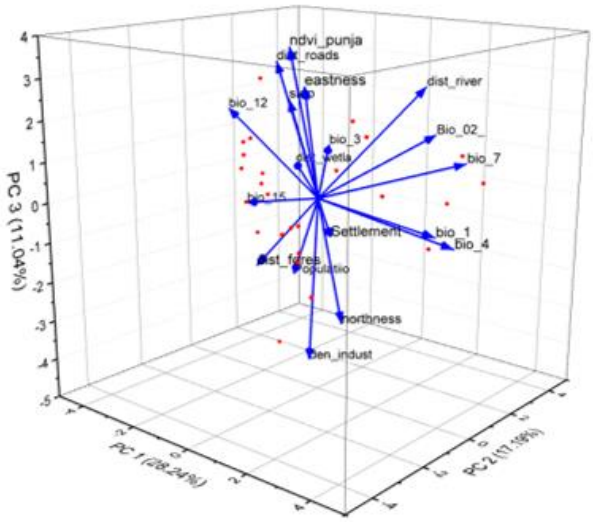

Figure 6. *3D Bi-plots of first 3 principal components*

On the basis of the three dimension reduction methods, three models were build, first with predictors set given by Pearson's correlation (P1), second with the predictors set obtained from PCA and third with variables got from VIF analysis (P3). Influence of each environmental variable on the presence probability of species was visualized using the response curves. The model will test these three predictors sets (P1, P2, P3) with different number of variables, to compare the results and augmenting model accuracy. The predictor's set along with variables used in these models are described in *Table 5*.

MaxEnt modelling

Maximum entropy method was selected from among SDMs for its superior accuracy at a number of presence-only records between 15 and 100 (Hernandez et al., 2008) as well as its capacity to handle categorical and continuous predictors interactively (van Gils et al., 2014).

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Table 5. *Predictor's sets used in model*

Sr.	Predictor sets	Variables
P1	Predictor set 1	den_industry, dist_forest, dist_river, dist_roads, dist_settlement, dist_wetlands, eastness, ndvi, northness, population, slop, Bio01, Bio02, Bio03, Bio04, Bio07, Bio12, Bio15
P2	Predictor set 2	dist_forest dist_river, dist_roads, dist_settlement, dist_wetlands, eastness, elevation ndvi, northness, population, slop, Bio02, Bio03, Bio07, Bio15
P3	Predictor set 3	Bio02, Bio03, Bio15, slop, NDVI, dist_wetland, dist_roads, dist_river, dist_forest, den_industry, den_canal, eastness, northness

Setting model parameter

The MaxEnt model was used to predict the potential suitable distribution of *Axis porcinus* in three time periods (current, 2050s, and 2070s). MaxEnt can give higher quality results based on the model settings. In this study, auto features were used to optimize model complexity and over-fitting was controlled by default regularization multipliers of 1. Parameterization was performed using 1000 maximum iterations along with 10 percentile training presence threshold and bootstrapping analysis for model validation. Therefore, species occurrence information for model calibration was divided into a training set (75% of total occurrence records) and test set (25% of total occurrence records) for design assessment.

These settings are good enough for allowing algorithms to give close to optimum performance (Phillips et al., 2017). Random seed bootstrapped runs were set to 10 empirically to create average SDM.

Jackknife sensitivity analysis was done for the estimation of each variable's contribution in the models. Influence of each environmental variable on the presence probability of species was visualized using the response curves. Jackknife sensitivity analysis was done for the estimation of each variable's contribution.

Evaluation matrices

The performance of our models was evaluated on the basis of four matrices: sensitivity, specificity, AUC and TSS. These metrics are being used to access the specie distribution modelling performance. AUC (area under the receiver operator curve) was used as it is widely accepted in SDM studies (Merow et al., 2013; Kane et al., 2017).

Presence–absence models are normally assessed by comparing a set of validation locations with the predictions by constructing a confusion matrix which observe the number of true positive, false positive, false negative and true positive cases predicted by the model (*Table 6*).

Table 6. *Confusion matrix for model evaluation*

Predicted	Observed	
	Presence	Absence
Presence	a	b
Absence	b	d

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Overall accuracy is described as the rate of correctly classified cells. Sensitivity is defined as the probability of correctly predicted a presence whereas the specificity is the probability of correctly predicted an absence. TSS normalize the overall accuracy by the accuracy that might have occurred by chance. TSS is not affected by prevalence or the size of the validation set. Its value ranges from -1 to + 1 where value close to + 1 is optimal (Shanks, 2019). Sensitivity, specificity, and TSS were calculated using *Equations 1–3.*

$$\text{Sensitivity} = \frac{a}{a+c} \tag{Eq.1}$$

$$\text{Specificity} = \frac{b}{b+d} \tag{Eq.2}$$

$$\text{TSS} = \text{sensitivity} + \text{specificity} - 1 \tag{Eq.3}$$

We used these four mercies to evaluate the performance of habitat suitability models as in general, sensitivity, specificity and TSS together is used for most of the ecological modelling researches (Rew et al., 2020).

Results and discussion

Models evaluation and its variables importance under current climate

The average test AUC for 10 replicate runs for model 1 is 0.9105, model 2 is 0.956 and model 3 is 0.954 with predictors sets P1, P2 and P3 respectively which are above 0.9, its mean that our models had high discrimination ability as evident from test AUC values which is indication of excellent models (Pearce and Ferrier, 2000; Manel et al., 2001). According to AUC model assessment criteria, 0.9 to 1.0 is excellent, 0.8 to 0.9 is good, 0.7 to 0.8 is general and 0.6 to 0.7 is poor (Swets, 1998). AUC poor indicated the performance that is no better than the random expectation whereas 1 represents perfect discrimination (Thuiller et al., 2005).

The Jackknife test results of models indicated that the three factors that contributed most to habitat suitability of *Axis porcinus* for all model were distance to forest, distance to rivers and distance to wetlands, and bio 15. The percentage contribution of these three variables were 66.9%, 68.2% and 60.5% for predictor set P1, P2, and 3 respectively. From the variables, six are common to each model set: distance to forest, distance to rivers and distance to wetlands, bio 15, density of industries and NDVI.

Performance comparison of three models are shown in *Table 7*.

Table 7. *Performance comparison of models*

Evaluation matrix (Avg.)	Sensitivity	Specificity	AUC	TSS
Model 1 (P1)	**0.7429**	**0.9848**	0.941	**0.7169**
Model 2 (P2)	0.6333	0.974	**0.956**	0.6181
Model 3 (P3)	0.672	0.9803	0.950	0.6523

The evaluation criteria of AUC, sensitivity, specificity and TSS are described in *Table 8* in order to assess the model results.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Table 8. *Evaluation criteria of AUC and TSS*

Model	AUC	TSS
Excellent	≥0.9	≥0.8
Good	0.8-0.9	0.6-0.8
Average	0.6-0.8	0.4-0.6
Poor	≤0.6	≤0.4

P1 was the best performing model in terms of sensitivity, specificity and TSS, whereas in terms of AUC model 2 performs the best as shown in *Table 7*. Its shows all the models perform reasonably well in predicting the habitat suitability of *Axis porcinus*. On whole Model 1 (P1) perform better than others and used further for analysis of current distribution of *Axis porcinus*. *Figure 7* shows ROC curve and sensitivity vs 1-specificity graphics along with Jackknife regularized training gain of best performing model. The relative contribution and permutation importance of predictor's variables to the MaxEnt model is described in *Table 9*.

Table 9. *Variable's contribution and permutation importance*

Variable	Percent contribution	Permutation importance
dist_wetlands	19.3	3.1
dist_river	16.6	9.3
dist_forests	16.1	24.1
Bio 15	14.9	17
den_industry	8.4	13.5
ndvi_punjab	6.6	6.6
northness	3.8	2.8
dist_roads	3.2	3
eastness	2.2	3.5
Population	2.2	7.1
Bio 12	1.5	1.4
Bio 03	1.5	1.1
slop	1.1	5.1
Bio 07	0.7	0.4
Bio 02	0.6	0.8
dist_settlement	0.5	0.5
Bio 04	0.4	0.1
Bio 01	0.3	0.4

Habitat suitability map

Distribution map was prepared using the logistic output of best maxEnt model in ArcGIS 10.4, with values ranges from lowest (0) to highest (1). Habitat suitability map was generated by classifying the raster in four categories according to expert experience method: 0-0.2 is unsuitable; 0.2-0.4 is low; 0.4-0.6 is moderate; and 0.6-1 is high (Ansari and Ghoddousi, 2018; Yang et al., 2013). Areas for the highly suitable habitat were calculated. *Figure 8* shows the spatial distribution best performing model of

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

habitat suitability model for *Axis porcinus*. Areas categorized as highly suitable are optimal habitats for *Axis porcinus*.

Current potential distribution of Axis porcinus

We developed the spatial distribution map from output probability raster by creating high (>0.6), moderate (0.4-0.6), low (0.2-0.4) and unsuitable (<0.2) occurrence classes. The area of predicted suitable habitat categorized in four classes for best performed model is presented in *Table 10*.

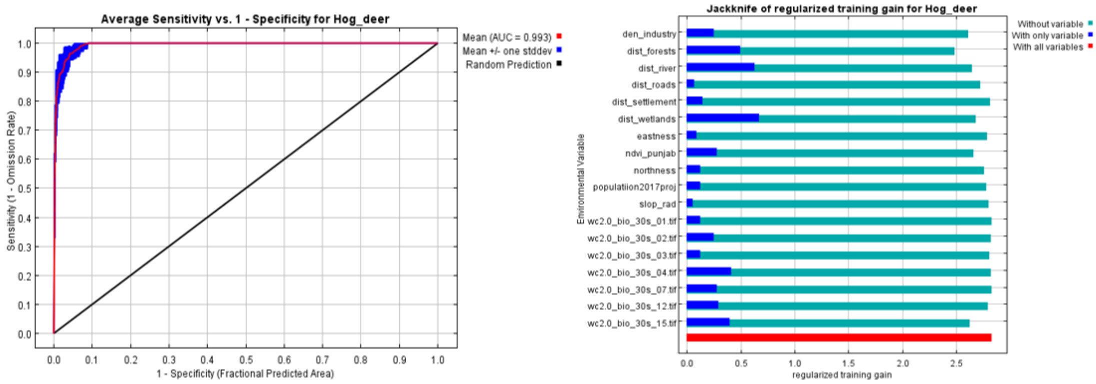

Figure 7. *Sensitivity vs 1-specificity curve and Jackknife regularized training gain of Axis porcinus*

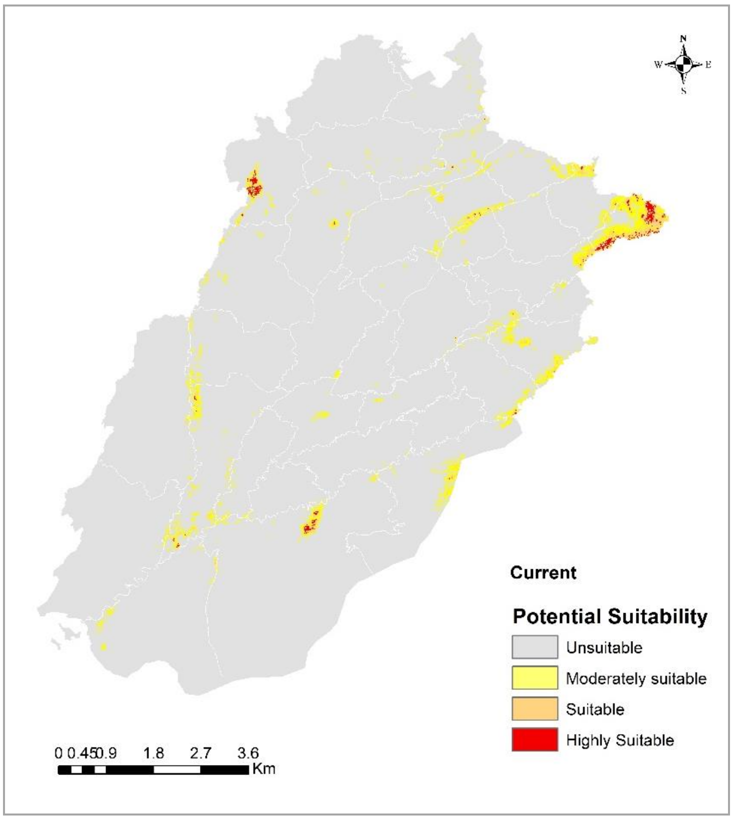

Figure 8. *Habitat suitability map*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Table 10. *Hog deer habitat suitability distribution in Punjab*

Sr.	Classes	Area km^2	Proportion
1	Unsuitable	201194	98.0
2	Low	2878	1.40
3	Moderate	845	0.41
4	High	309	0.15

According to its spatial distribution the highly suitable and suitable areas for Axis porcinus are in Narowal district (Shakargarh and Narowal tehsils) in the riverine forest, Chasma barrage site which is also a Ramsar site and in Lal Suhanra Bio-reserve in Cholistan, Bahawalpur. The low suitability occurrence zones includes Muzaffargh, Kot addu and Alipur tehsils along the river Chenab and river Indus. In total 0.15% area in Punjab is highly suitable for Axis porcinus conservation areas, while 0.41% is categories as moderately suitable areas. The correctness of modelled environment niche varies by variables but the overall there is a decent match among the predicted and observed data values. The histogram of data of each variable at predicted highly suitable habitat areas by MaxEnt and the histogram of data values at presence locations are very well conformed to each other (*Fig. 9*).

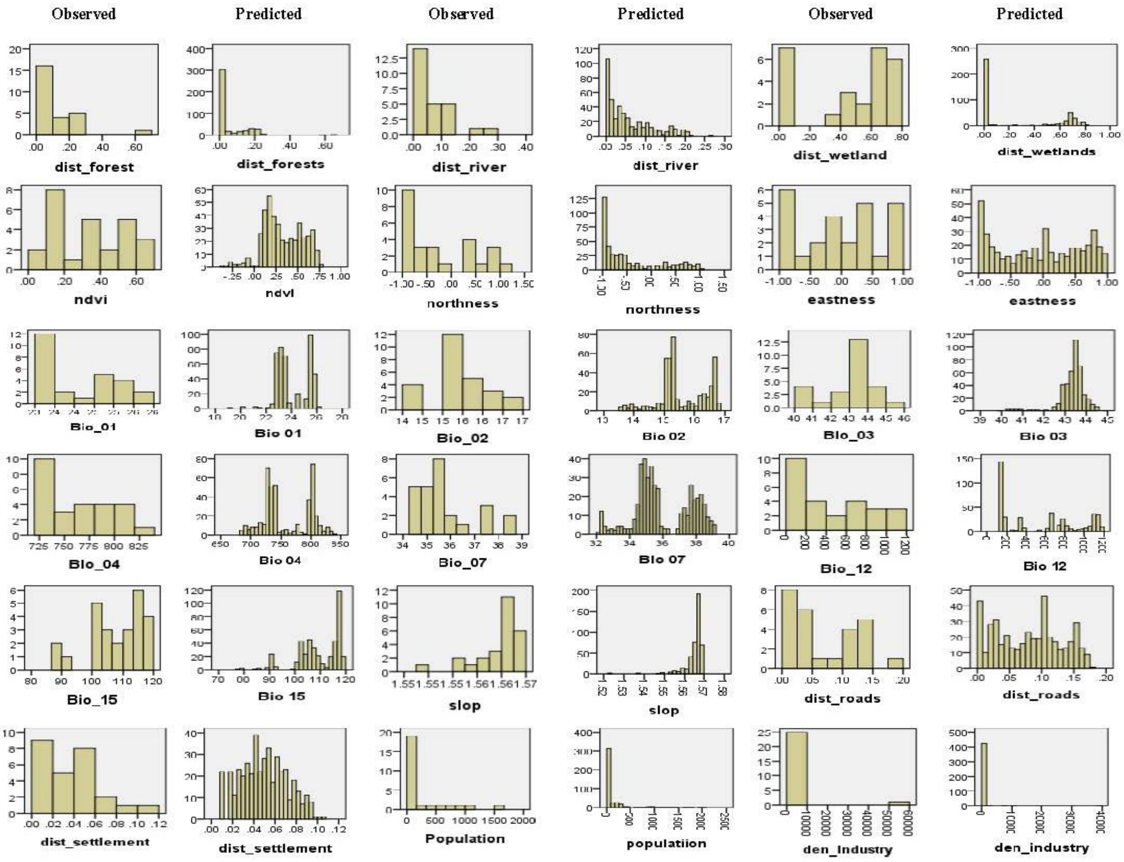

Figure 9. *Histograms of predicted values for highly suitable habitat versus observed values*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Models evaluation and its variables importance for future scenarios

The AUC values for 2050s RCP 4.5, 2070s RCP 4.5, 2050s RCP 8.5 and 2070s RCP 8.5 were 0.955, 0.921, 0.936 and 0.896 respectively, indicates that the outputs were very accurate. The jackknife analysis shows that the distribution of *Axis porcinus* was mainly influenced by the Bio 12 (annual precipitation) in all predictions with 49.3%, 48.3%, 53% and 44.2% contribution respectively for 2050s RCP 4.5, 2070s RCP 4.5, 2050s RCP 8.5 and 2070s RCP 8.5 (*Fig. 10*).

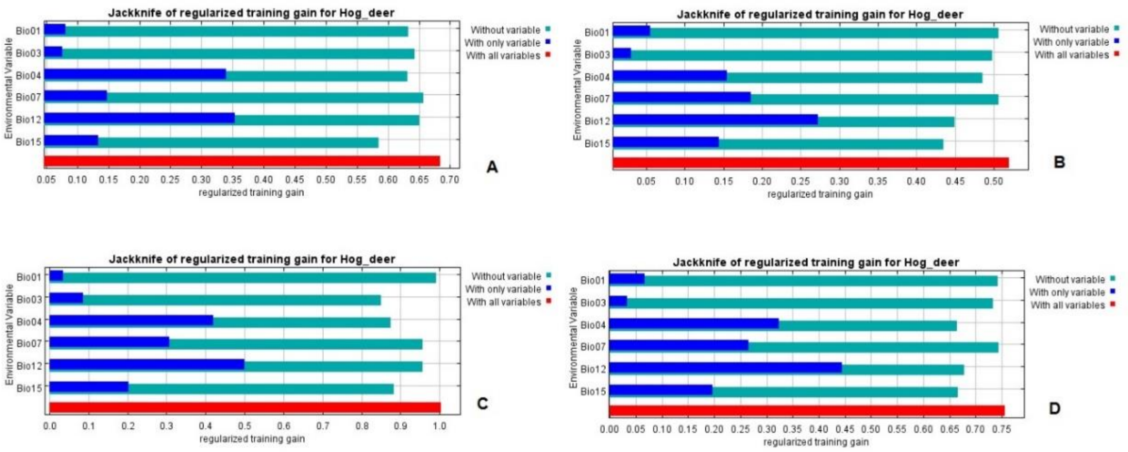

Figure 10. *Jackknife of regularized training for (a) 2050s RCP 4.5, (b) 2070s RCP 4.5, (c) 2050s RCP 8.5 and (d) 2070s RCP 8.5*

Future predicted distribution of Axis porcinus

The species distribution maps for 2050 and 2070 (HadGEM2-AO emission scenario) revealed a reduction in area of highly suitable habitat for *Axis porcinus* calculated for each scenario (*Fig. 11; Table 11*). For RCP 4.5, there is an 18.7% reduction in suitable habitat for *Axis porcinus* in 2050, and 1.0% reduction in the suitable habitat for 2070. For RCP 8.5, 52.7% reduction is predicted for 2050 and 15.5% reduction is predicted by year 2070. Suitable and highly suitable areas particularly the habitat associated with river Indus is predicted to shrink due to fluctuations in precipitation seasonality and annual temperature range.

Priority conservation areas

Based on habitat suitability maps derived from current and potential future distribution of *Axis porcinus*, we identified three priority areas for *Axis porcinus* conservation covering total area of 632 km^2 (*Fig. 12*). Highly suitable areas under current and future scenarios were merged to quantify areas. The first priority area is in Narowal District in riverine forest belt with area of approx. 386 km^2. This entire area is human subdued and is not protected area. It has two major chunks. The second priority conservation is the areas comprised of Chasma barrage which is a wildlife sanctuary and a protected area need effective management plan (cover 110 km^2). Third conservation area is in Bahawalpur district with an area of 121 Km^2 and is a bio reserve called Lal Suhanra Bio reserve which should have a comprehensive management plan to foster *Axis porcinus* population.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

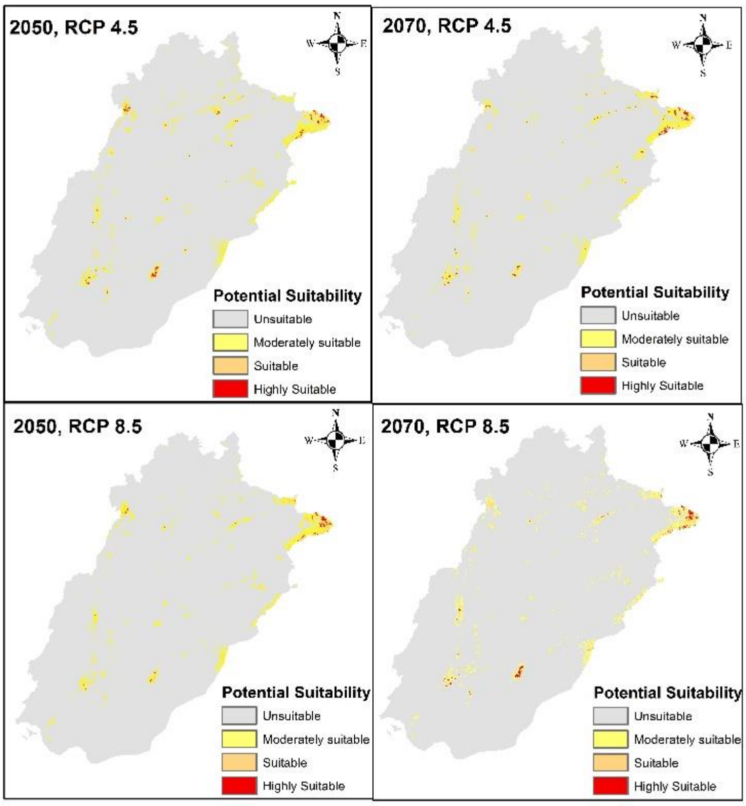

***Figure 11.** Habitat suitability for Axis porcinus 2050s RCP 4.5, RCP 8.5 and 2070s RCP 4.5, RCP 8.5*

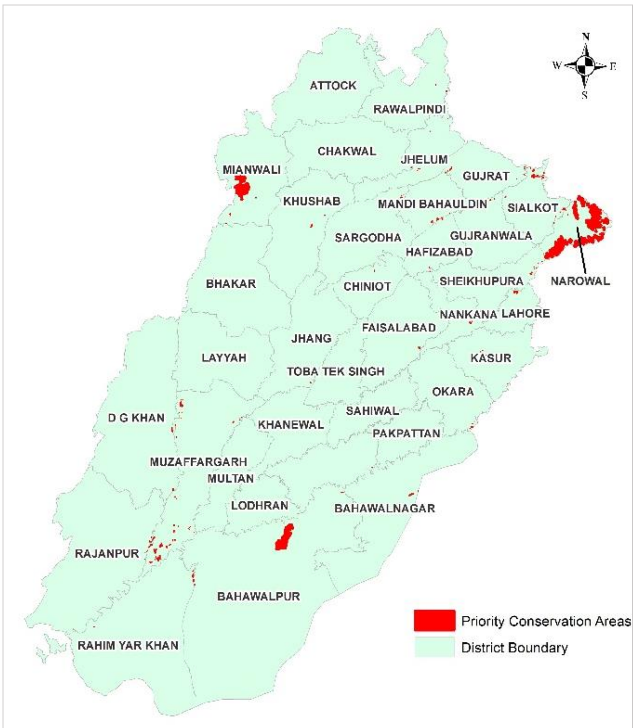

***Figure 12.** Priority conservation areas for Axis porcinus in Punjab*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

Table 11. *Predicted suitable areas for current and future condition*

Class	Current (Km^2)	RCP 4.5 (Km^2)		RCP 8.5 (Km^2)	
		2050	**2070**	**2050**	**2070**
Not suitable	197714.1	197601.9	198295.0	198659.5	197037.7
Moderately suitable	4991.2	5361.2	4542.0	4541.0	5745.3
Suitable	1472.5	1295.8	1350.3	1215.3	1458.1
Highly suitable	451.3	366.9	447.5	215.2	381.2

Conclusion and recommendations

It is concluded that MaxEnt which a machine learning method was successful in producing habitat suitability maps using current and future distribution of *Axis porcinus* utilizing geospatial big data and its derived products. The current and future distribution maps depicts that the combination of limited numbers of good quality presence data, a wide-ranging geospatial datasets of bioclimatic factors, environmental and anthropogenic predictors, constitute a very good habitat suitability model.

The current distribution of *Axis porcinus* shows that the highly suitable habitat areas are very limited and are primarily located in Narowal district, Chasma barrage site and in Lal Suhanra Bioreserve. Therefore, change predicted in these areas according to climate change scenarios should not be ignorable given that these areas constitute only a little percentage of total area of Punjab province. The future model projections from 2050s and 2070s for IPCC climate change scenarios indicated that species distribution would be affected significantly by climate change. According to all RCPs scenarios the suitable habitat of *Axis porcinus* is predicted to shrink. Detail conservation management plans should be prepared for these areas which are prone to climate change to counter the climate change impacts. This study suggests to incorporate future climate change scenarios in formulating current conservation policies and strategies for protection of species.

Although the predictive ability of our models were very good but there are some sources of uncertainty in identifying potential habitat due to lack of information on threat factors of many endangered wildlife species. Therefore, further improvement in accuracy of predictive models should be considered. In future work to ensure the reliability of suitability model, we plan to lighten the possible biases of non-surveyed occurrence data.

Acknowledgements. This paper and research behind it is technically supported by our colleagues at the Urban Unit Pakistan, who provided insight and skills. We would also like to show gratitude to Dr. Muhammad Ali Nawaz, PhD (Ecology and Natural Resource Management), Norwegian University of Life Sciences, Norway. The generosity and expertise of one and all have improved this study in innumerable ways and saved us from many errors; those that inevitably remain are entirely our own responsibility.

REFERENCES

[1] Ali, M. A. (2008): Studies on calving related disorders (dystocia, uterine prolapse and retention of fetal membranes) of the river buffalo (Bubalus bubalis) in different agro ecological zones of Punjab Province Pakistan. – Doctor of Philosophy Thesis in Theriogenology, University of Agriculture, Faisalabad.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

[2] Angom, S., Tuboi, C., Ghazi, M. G. U., Badola, R., Hussain, S. A. (2020): Demographic and genetic structure of a severely fragmented population of the endangered hog deer (Axis porcinus) in the Indo-Burma biodiversity hotspot. – PLoS ONE 15(2): e0210382. https://doi.org/10.1371/journal.pone.0210382.
[3] Ansari, M., Ghoddousi, A. (2018): Water availability limits brown bear distribution at the southern edge of its global range. – Ursus 29(1): 13-24.
[4] Austin, M. (2002): Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. – Ecological Modelling 157(2-3): 101-118. https://doi.org/10.1016/S0304-3800(02)00205-3.
[5] Arshad, M., Ullah, I., Chaudhry, M. J. I., Khan, N. U. H. (2012): Estimating hog deer *Axis porcinus* population in the riverine forest of Taunsa Barrage Wildlife Sanctuary, Punjab, Pakistan. – Records: Zoological Survey of Pakistan 21: 25-28.
[6] Bosso, L., di Febbraro, M., Cristinzio, G., Zoina, A., D. (2016): Shedding light on the effects of climate change on the potential distribution of Xylella fastidiosa in the Mediterranean Basin. – Biological Invasions 18(6): 1759-1768. DOI: https://doi.org/10.1007/s10530-016-1118-1.
[7] Brooks, T. M., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., Rylands, A. B., Konstant, W. R., … Magin, G. (2002): Habitat loss and extinction in the hotspots of biodiversity. – Conservation Biology 16(4): 909-923. https://doi.org/10.1046/j.1523-1739.2002.00530.x.
[8] Engler, R., Guisan, A., Rechsteiner, L. (2004): An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. – Journal of Applied Ecology 41(2): 263-274. DOI: https://doi.org/10.1111/j.0021-8901.2004.00881.x.
[9] Farrel, A., Wang, G., Rush, S. A., Martin, J. A., Belant, J. L., Butler, A. B., Godwin, D. (2019): Machine learning of large-scale spatial distributions of wild turkeys with high-dimensional environmental data. – Ecology and Evolution 9(10): 5938-5949.
[10] Griggs, D. J., Noguer, M. (2002): Climate change 2001: the scientific basis. Contribution of working group I to the third assessment report of the intergovernmental panel on climate change. – Weather 57(8): 267-269.
[11] Guisan, A., Zimmermann, N. (2000): Predictive habitat distribution models in ecology. – Ecological Modelling 135(2-3): 147-186. DOI: https://doi.org/10.1016/S0304-3800(00)00354-9.
[12] Haddad, N. M., Brudvig, L. A., Clobert, J., Davies, K. F., Gonzalez, A., Holt, R. D., Collins, C. D. (2015): Habitat fragmentation and its lasting impact on Earth's ecosystems. – Science Advances 1(2): 1-9. https://doi.org/10.1126/sciadv.150005210.1126/sciadv.1500052.
[13] Hernandez, P. A., Franke, I., Herzog, S. K., Pacheco, V., Paniagua, L., Quintana, H. L., Vargas, J. (2008): Predicting species distributions in poorly-studied landscapes. – Biodiversity and Conservation 17(6): 1353-1366.
[14] Iqbal, M., Prince, A., Khan, M. A., Nayyer, A. Q., Akhtar, M. (2013): Ecology and population status of hog deer from Narowal, Pakistan. – International Research Journal of Biological Sciences 2(7): 19-24.
[15] IUCN (2019): The IUCN Red List of Threatened Species. – https://www.iucnredlist.org/en (accessed on 15 October 2019).
[16] Jones, M. C., Dye, S. R., Fernandes, J. A., Frölicher, T. L., Pinnegar, J. K., Warren, R., Cheung, W. W. (2013): Predicting the impact of climate change on threatened species in UK waters. – PloS ONE 8(1): e54216.
[17] Kane, K., Debinski, D. M., Anderson, C., Scasta, J. D., Engle, D. M., Miller, J. R. (2017): Using regional climate projections to guide grassland community restoration in the face of climate change. – Front. Plant Sci. 8: 730. DOI: 10.3389/fpls.2017.00730.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201

[18] Kwon, H.-S., Kim, B.-J., Jang, G.-S. (2016): Modelling the spatial distribution of wildlife animals using presence and absence data. – Contemporary Problems of Ecology 9(5): 515-528. DOI: https://doi.org/10.1134/S1995425516050085.
[19] Manel, S., Williams, H. C., Ormerod, S. J. (2001): Evaluating presence–absence models in ecology: the need to account for prevalence. – Journal of Applied Ecology 38(5): 921-931.
[20] Merow, C., Smith, M. J., Silander, J. A. (2013): A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – Ecography 36: 1058-1069. DOI: 10.1111/j.1600-0587.2013.07872.x.
[21] Pearce, J., Ferrier, S. (2000): An evaluation of alternative algorithms for fitting species distribution models using logistic regression. – Ecological Modelling 128(2-3): 127-147.
[22] Phillips, S. J., Anderson, R. P., Schapire, R. E. (2006): Maximum entropy modeling of species geographic distributions. – Ecological Modelling 190(3-4): 231-259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.
[23] Phillips, S., Anderson, R., DudõÂk, M., Schapire, R., Blair, M. (2017): Opening the black box: an open-source release of Maxent. – Ecography 40(7): 887-893. DOI: https://doi.org/10.1111/ecog.03049.
[24] Qin, A., Liu, B., Guo, Q., Bussmann, R. W., Ma, F., Jian, Z., Pei, S. (2017): Maxent modeling for predicting impacts of climate change on the potential distribution of Thuja sutchuenensis Franch., an extremely endangered conifer from southwestern China. – Global Ecology and Conservation 10: 139-146.
[25] Rahman, A. A., Mohamed, M., Tokiman, L., Sanget, M. S. M. (2019): Species distribution modelling to assist biodiversity and conservation management in Malaysia. In IOP Conference Series. – Earth and Environmental Science 269(1): 012041.
[26] Rew, J., Cho, Y., Moon, J., Hwang, E. (2020): Habitat suitability estimation using a two-stage ensemble approach. – Remote Sensing 12(9): 1475.
[27] Shanks, R. E. (2019): Assessing the transferability of a species distribution model for predicting the distribution of invasive cogongrass in Alabama. – A thesis presented to the Faculty of the USC Graduate School University of Southern California in partial fulfillment of the requirements for the degree master of science.
[28] Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., Blaschke, T. (2020): Big Earth data: disruptive changes in Earth observation data management and analysis? – International Journal of Digital Earth 13(7): 832-850.
[29] Roberts, T. J. (1997): The Mammals of Pakistan. – Oxford University Press, .
[30] Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C., Pounds, J. A. (2003): Fingerprints of global warming on wild animals and plants. – Nature 421(6918): 57.
[31] Swets, J. A. (1998): Measuring the accuracy of diagnostic systems. – Science 240: 1285-1293.
[32] Van Vuuren, D. P., Edmonds, J. A., Kainuma, M., Riahi, K., Weyant, J. (2011): A special issue on the RCPs. – Climatic Change 109(1-2): 1.
[33] Thuiller, W., Richardson, D. M., Pyšek, P., Midgley, G. F., Hughes, G. O., Rouget, M. (2005): Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. – Global Change Biology 11(12): 2234-2250.
[34] van Gils, H., Westinga, E., Carafa, M., Antonucci, A., Ciaschetti, G. (2014): Where the bears roam in Majella National Park, Italy. – Journal for Nature Conservation 22(1): 23-34.
[35] Wiens, J. J. (2011): The niche, biogeography and species interactions. – Philosophical Transactions of the Royal Society B: Biological Sciences 366(1576): 2336-2350. https://doi.org/10.1098/rstb.2011.0059.
[36] Yang, X. Q., Kushwaha, S. P. S., Saran, S., Xu, J., Roy, P. S. (2013): Maxent modeling for predicting the potential distribution of medicinal plant, Justicia adhatoda L. in Lesser Himalayan foothills. – Ecological Engineering 51: 83-87.

[37] Yang, C., Yu, M., Li, Y., Hu, F., Jiang, Y., Liu, Q., Gu, J. (2019): Big Earth data analytics: a survey. – Big Earth Data 3(2): 83-107.

[38] Zhang, J., Jiang, F., Li, G., Qin, W., Li, S., Gao, H., Zhang, T. (2019): Maxent modeling for predicting the spatial distribution of three raptors in the Sanjiangyuan National Park, China. – Ecology and Evolution 9(11): 6643-6654.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/1904_31813201