Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Climate change impacts on the potential distribution of Vatica lanceaefolia (Roxb.) Blume, a critically endangered tree species of Northeastern India

Anubhav Bhuyan
Tezpur University
Suhasini Hazarika
Tezpur University
Shilpa Baidya
Tezpur University
Kuladip Sarma
Gauhati University
Bijay Thakur
Tezpur University
Amit Prakash
Tezpur University
Ashalata Devi
ashalatakh23@gmail.com

Tezpur University

Research Article

Keywords: Bioclimatic variables, Climate change, Critically endangered, Ensemble model, Species Distribution Model, Vatica lanceaefolia

Posted Date: November 12th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-5360737/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Theoretical and Applied Climatology on February 7th, 2025. See the published version at https://doi.org/10.1007/s00704-025-05369-3.

Abstract

Vatica lanceaefolia, a critically endangered tree species, found in Northeast India, Bangladesh, and Myanmar is being exposed to threat due to significant habitat loss and overexploitation. Therefore, it is imperative to intervene in conservation efforts by effectively identifying and assessing the present and future habitats of the species to take necessary measures for their protection. This study aims to assess the suitable habitat of *V. lanceaefolia* considering both present and future scenarios in Northeast India by employing different species distribution modeling (SDM) techniques. An ensemble approach, incorporating five SDMs- Maximum Entropy (MaxEnt), Random Forest (RF), Generalized Linear Models (GLM), Support Vector Machine (SVM), and Multivariate Adaptive Regression Spline (MARS) is used in this study. Twenty-one independent occurrence points of *V. lanceaefolia* were used to model its current and future distribution, along with 23 environmental variables including bioclimatic factors, topographical factors, evapotranspiration, and soil data were used to predict the potential habitat of *V. lanceaefolia* under both current and future climate scenarios (SSP245 and SSP585) given by global climate model (GCM), HadGEM3-GC31-LL. The evaluation of the ensemble model showed excellent performance based on AUC (0.88) and TSS value (0.65). The findings revealed that 2597.69 km^2 (0.99% of the total geographical area, i.e. 262,184 km^2) was identified as a highly suitable habitat for the species in the current scenario. According to model projections, the highly suitable area is predicted to decrease by more than 59.00% from 2597.69 km^2 to 1065.14 km^2, under the SSP245 scenario in the year 2061–2080 in the future climate. The present findings underscore the urgent need for implementing suitable conservation measures to mitigate the impacts of climate change on the species. To adequately address habitat loss due to deforestation and urban expansion, it is crucial to prioritize habitat restoration, targeted protective measures, and additional species-level ecological research of individual species.

1 Introduction

Assessing suitable habitat is important for threatened plant and animal species to understand the overall future conservation potential, especially for conserving restricted-range species that are vulnerable to specific threats such as climate change. Climate change in Northeastern India is expected to have an adverse effect on most of the endemic and threatened plants. This is primarily attributed to the high levels of fossil fuels combustion, deforestation, and farming livestock in this biodiversity hotspot region, over the past decade. Climate change is one of the significant driving forces for species extinction and loss of biodiversity, together with habitat loss and fragmentation (Banda and Tassie 2018; Duflot et al. 2018; Galán-Acedo et al. 2018; Nüchel et al. 2018; Secretariat of the Convention on Biological Diversity 2020). According to the State of the World's Plants and Fungi Report 2020, it was estimated that approximately 40% of plant species of the world are in danger and going to be at risk of extinction. Additionally, Intergovernmental Panel on Climate Change (IPCC) in 2023, projects that the average global temperature will increase by 1.5 ^{0}C due to continued global warming. Climate is the main factor that influenced species' distribution range. Increasing temperatures will have significant effects on the physiological and ecological characteristics of species, as well as their distribution patterns (Hu et al. 2015). Therefore, assessing the suitable habitats and distributions is crucial to acquire important information about populations. This information is vital for managing threatened plant and animal species and formulating suitable conservation strategies. Furthermore, the identification of environmental factors significant to the species and mapping suitable habitats are equally important (Pearson 2007). Assessing suitable models is a useful tool for forecasting different environmental scenarios as well as for managing threatened species' habitat or population.

Advancement in statistical tools and other mathematical modelling approaches have made if feasible to measure the habitat suitability based on the predicted outputs of such models. However, most of the species distribution

models are based on presence-only data that basically indicates the presence of a species, making the prediction more complex while interpreting the results. Species distribution models are one of the most important approaches in predicting how a species will respond to environmental changes. Their use has significantly enhanced our understanding regarding environmental predictors that influence species' distributions. These models are useful for assessing the responses of the global biota to disturbances, climate change, and other causes (Guisan et al. 2017; Kaky et al. 2020). The models aid in determining a habitat that is essential for the management of a species and provide critical information on the probability of species presence in a specific area (Hazarika et al. 2023). *V. lanceaefolia*, a critically endangered plant species, is a significant source of non-timber forest products (NTFP), firewood, and is used for charcoal production (Shiva and Jantan 1998). The species is typically found in lowland dipterocarp forests and contributes significant role in the process of carbon sequestration. Its distribution includes northeastern India, Bangladesh, and Myanmar. However, effective management efforts are hampered by insufficient data on its population size, species distribution, and suitable habitats, making it challenging to develop comprehensive conservation strategies (Shiva and Jantan 1998; Sharma et al. 2022). The geographical range of *V. lanceaefolia* is presently restricted only to Tinsukia, Dibrugarh, Jorhat and Golaghat districts of Assam (Sharma et al. 2022). This limited distribution underscores its potential vulnerability to habitat loss and environmental changes, highlighting the need for management and conservation efforts. Currently, there has been no research conducted using the Species Distribution Model (SDM) approach to investigate the potential suitable habitat and the possible range of *V. lanceaefolia*. By recognizing the potential habitats of *V. lanceaefolia* under different climatic scenarios, conservationists, scientist and policy makers from NGOs or government agencies can focus on reintroducing the species in these regions and can design suitable conservation measures that combine proper management practices to ensure the species' long-term survival. However, the distribution range of *V. lanceaefolia* has not been adequately established in the IUCN species range to date. As a result, it is essential to gather baseline data on the geographical distribution of *V. lanceaefolia* in Northeast India. The objectives of the present study were to: (1) determine the influential bioclimatic variables associated with the current and future potential habitat of *V. lanceaefolia*, and (2) quantify and map the potential areas with high suitability for the occurrence of *V. lanceaefolia* in Northeast India.

2 Materials and methods

2.1 Study area

The present study was conducted in the North-eastern region of India, which comprises the states of Assam, Arunachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Tripura, and Sikkim (Fig. 1). This region is physio-graphically divided into three different landforms: the Eastern Himalaya, the hilly North-Eastern (NE) region and the flat plains of the Brahmaputra and Barak Valley. This area accounts for one-third of the country's biodiversity and supports a variety of biological niches due to the significant influence of the southwest and northeast monsoons (Roy et al. 2015). Geographically, the region is predominantly characterized by hilly terrain, with interspersed plains and valleys. The elevation of the area ranges from sea level to over 7,000 meters above mean sea level (MSL). The annual mean temperature ranges from 5 to 30°C, while the humidity levels range from 70–85% (Sharma et al. 2009). The average annual rainfall varies between 1577 and 6002 mm in this region (Parthasarathy 1995; Jhajharia et al. 2009; Jain et al. 2013).

2.2 Data

2.2.1 Occurrence data

The geographical locations of *V. lanceaefolia* were obtained from different sources, including both the primary and secondary occurrence points. The primary data were accessed from field surveys, while secondary occurrences were downloaded from the Global Biodiversity Information Facility (GBIF) (https://www.gbif.org), India Biodiversity Portal (IBP) (https://indiabiodiversity.org), iNaturalist (https://www.inaturalist.org) and literature (including scientific data, confirmed reports, and research publications). Prior to retrieving the secondary data, a thorough examination of the sources was conducted, which covered the years 2000 to 2024. To validate secondary occurrence points, a preliminary evaluation was conducted using Google Earth Pro. Subsequently, field observations were followed to explicitly study the basic habitat requirements of *V. lanceaefolia*. Presence only data was used to assess the potential habitat of *V. lanceaefolia* in the study area (Elith et al. 2006), evaluating 21 locations for habitat suitability modeling. To minimize model overfitting, we utilized the "wallace" package (http://wallaceecomod.github.io/wallace/, accessed on 17 January 2024) to thin the data, ensuring that only one distribution point was retained within each 0.5 km × 0.5 km grid (Brown et al. 2017). This approach reduced the bias in the prediction results caused by clustered sampling or data redundancy. Finally, the models were run by 21 occurrence points.

2.2.2 Environmental variables

The selection of environmental variables was based on the data availability and anticipated impact on the distribution of *V. lenceaefolia*. A total of 19 bioclimatic variables (Table 1) representing historical and future scenarios were downloaded from Worldclim database (https://worldclim.org) at a resolution of 30 arc-seconds (1 km^2). Elevation data was accessed from the USGS website (https://www.usgs.gov), aspect and slope were generated from the DEM using the ArcGIS Spatial Analyst tools, while evapotranspiration data was accessed from the MODIS/Terra Net Evapotranspiration database (https://www.lpdaac.usgs.gov). Soil data was obtained from the International Soil Reference and Information Centre (ISRIC) (https://www.isric.org) (Poggio et al. 2021). These variables reflect both historical and future scenarios. The data were clipped to cover only the North East (NE) region of India (Hijmans et al. 2005). To predict current (1971 to 2000) and future climate scenarios of the *V. lenceaefolia* two different Shared Socio-economic Pathways (SSPs), namely SSP245 and SSP585 were employed, for the periods of 2041–2060 and 2061–2080, considered as mid-future (MF) and distant future (DF), respectively. The climate projections used in this study were generated by the Hadley Centre Global Environment Model in the Global Coupled Configuration 3.1 (HadGEM3-GC31) from the Coupled Model Intercomparison Project (CMIP6). The Shared Socioeconomic Pathways (SSPs), are scenarios used in climate change study to explore potential socioeconomic conditions and their influence on greenhouse gas emissions, as well as their broader impacts on the climate in future. This specific model was recognized for its proven effectiveness in South and Southeast Asia.

To detect the multicollinearity among the 23 predictor variables, variance inflation factor (VIF) was applied to select a group of non-colinear predictor variables for both current and future scenarios. A VIF score > 10 indicates a significant multicollinearity within the data set. Therefore, a habitat distribution model for *V. lanceaefolia* was developed by utilizing seven specific bioclimatic variables and nine environmental variables. The multicollinearity among the 23 variables changes between the current scenario and future scenarios, resulting in a change in the selection of important variables (Table 1). The subset of important variables remained same across both time periods under scenarios SSP245 and SSP585.

Table 1
Variance inflation factor (VIF) of selected predictor variables downloaded from the open-source database.

Variables	Description	Current	MF_SSP245	DF_SSP245	MF_SSP585	DF_SSP585
		VIF	VIF	VIF	VIF	VIF
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	2.12	2.32	2.03	2.20	2.05
BIO3	Isothermality (BIO2/BIO7) (×100)	2.86	2.97	2.90	3.03	2.78
BIO5	Max Temperature of Warmest Month	4.15	4.18	4.42	4.10	4.54
BIO14	Precipitation of Driest Month	5.87	6.75	3.64	4.60	3.64
BIO15	Precipitation Seasonality (Coefficient of Variation)	2.92	3.48	3.85	3.49	4.06
BIO17	Precipitation of Driest Quarter	8.66	8.97	6.79	7.17	6.78
BIO18	Precipitation of Warmest Quarter	3.19	3.56	3.63	3.58	3.66
Aspect	Aspect	1.01	1.01	1.01	1.01	1.01
Slope	Slope	2.22	2.14	2.25	2.14	2.22
Evapotranspiration	Annual Evapotranspiration	1.32	1.37	1.33	1.28	1.29
CEC_(0–5) cm	Cation exchange capacity (0–5) cm	4.44	5.07	4.73	4.39	4.62
pH_ (0–5) cm	Soil pH (0–5) cm	6.29	7.28	6.67	6.19	6.53
Sand_ (0–5) cm	Sand (0–5) cm	4.70	4.63	4.35	4.32	4.30
Clay_ (0–5) cm	Clay (5–15) cm	4.17	4.40	4.26	4.17	4.16
Nitro_ (0–5) cm	Nitrogen (5–15) cm	1.83	1.88	1.84	1.74	1.86
SOC_ (5–15) cm	Soil organic carbon (5–15) cm	4.35	4.95	4.65	4.84	4.97

2.2.3 Species distribution modelling

Historically, solitary Species Distribution Modelling (SDM) techniques such as maximum entropy (MaxEnt) were frequently employed. Recent research suggests ensemble models, which combine multiple algorithms as proposed by Araújo and New (2007) may provide more accurate predictions in comparison to single models (Zhang et al. 2015). In this study, an ensemble model is utilized to optimize both the accuracy and robustness of SDM predictions. The approach involves the integration of five models: three machine learning techniques, Maximum entropy

(MaxEnt) (Phillips et al. 2004), random forest (RF) (Ripley 2007) and support vector machine (SVM) (Hearst et al. 1998), and two regression model, generalized linear models (GLM) (Nelder and Wedderburn 1972), multivariate adaptive regression splines (MARS) (Friedman 1991). According to Yebeyen et al. (2022), these models demonstrate high predictive accuracy for species distribution based on presence data only. In order to build and evaluate the model, the dataset was divided into two subsets: a training (75%) and testing (25%) sets, using a 10-percentile threshold. Model performance was assessed using the AUC-ROC curve, a comprehensive metric that considers all potential classification thresholds. For each possible threshold, sensitivity and specificity were calculated to avoid arbitrary choices (Manel et al. 2001; Liu et al. 2005). This approach effectively balances the trade-offs between sensitivity and specificity (Pearce and Ferrier 2000; Shabani et al. 2018). The models were developed using the "SSDM" package within the R statistical software (version 4.1.3) as demonstrated by Naimi and Araújo (2016) and were evaluated using the AUC and TSS metrics. AUC scores range from 0 to 1, where higher values signify superior model performance (Lobo et al. 2008). The True Skill Statistic (TSS) is a metric that is affected by the chosen classification threshold; values approaching + 1 indicate perfect alignment between predictions and observed outcomes, while scores of 0 or below suggests no better than random classification (Allouche et al. 2006). TSS values were classified into categories ranging from poor to perfect based on specific numerical thresholds. Similarly, the collective performance of the ensemble model was categorized into different levels from poor to excellent according to its overall score.

3 Results

3.1 Model performance and major environmental factors

The ensemble models effectively predicted the current potential distribution of *V. lanceaefolia*. The five models had different mean AUC values, the MaxEnt model had a maximum mean AUC value of 0.91, the RF and GLM had 0.90, SVM had 87, and the MARS had a least mean AUC value of 0.84. According to the models' mean AUC, MaxEnt model provided the best predictions compared to the other models. It accurately predicted almost 90% of the area across the 10 replicates of each algorithm were used to assess model performance. However, the Maxent model have a higher sensitivity for (0.90) than that of GLM (0.87). Conversely, the RF has a better specificity (0.91) than SVM (.89). Similarly, models with a TSS value ≥ 0.75 demonstrate high-performance prediction ability. TSS values ≥ 0.75 have the potential to be useful and acceptable for analyzing the results of a niche model. The TSS value varied between 0.75 (RF) to 0.63 (MaxEnt). The ensemble model, consisting of MaxEnt, RF, GLM, SVM and MARS, achieved an overall AUC of 0.88, Sensitivity of 0.85, Specificity of 0.80 and TSS of 0.65 (Table 2).

Table 2
Performance evaluation of SDMs using different statistical parameters for current distribution of *Vatica lanceaefolia* in the northeastern region of India

Current	AUC	Sensitivity	Specificity	TSS
MaxEnt	0.91	0.90	0.73	0.63
RF	0.90	0.84	0.91	0.75
GLM	0.90	0.87	0.84	0.71
SVM	0.87	0.83	0.89	0.72
MARS	0.84	0.80	0.87	0.67
Ensemble	0.88	0.85	0.80	0.65

According to the variable contribution curve of the top four environmental variables (Fig. 2), the most suitable habitat conditions for *V. lanceaefolia* are as follows: Precipitation of Driest Quarter (BIO17) ranges from 1 to 99 mm, Cation exchange capacity [CEC_mean_(0–5)cm] varies between 0 to 378 t/ha, Precipitation Seasonality (Coefficient of Variation) (BIO15) ranges from 63.61 to 141.29 mm and Slope varies between 0 to 88.17 degree. It has been observed that distribution of *V. lanceaefolia* is suitable in moist and humid climate, although it has limited ability to withstand low temperature. The ensemble model revealed that BIO17 had the highest contribution (12.51%) among the selected bioclimatic variables for current scenario, followed by CEC_mean_(0–5)cm (10.47%). Under all SSP projections, BIO17 was found to have the highest values, followed by CEC_mean_(0–5)cm. The projected contributions of BIO17 from 2041 to 2060 are estimated to be 11.61% for SSP245 and 10.86% for SSP585 scenario. Between 2061 to 2080, the contributions were observed to be 15.98% and 12.92% for SSP245 and SSP585 respectively (Fig. 2).

In the mid-future, CEC_mean_(0–5)cm contributes 9.84% and 10.44% to the ensemble models for SSP245 and SSP585, respectively. In the distant future, these contributions for SSP245 and SSP585 are 11.98% and 9.96%, respectively as shown in Fig. 2. The model illustrates that variables associated with precipitation have significant contributions to the habitat suitability model developed for *V. lanceaefolia* compared to the biocimatic variables related to temperature (Fig. 2).

3.2 Potential suitable habitat under present and future climatic scenarios

In the current scenario, an ensemble model revealed that the potential habitat for *V. lanceaefolia* covers an area of 2597.69 km², which accounts for 0.99% of the total 262,184 km² area, which is considered to be highly suitable. Subsequently, there is an area of 15413.75km² (5.88%) categorized as moderate suitable, 37723.38 km² (14.39%) as low suitable, and 206449.18 km² (78.74%) as no suitable (Table 3 and Fig. 3).

For future potential distribution in the future, the model predicts that by 2041–2060, there will be 2317.53 km² (0.88%) of highly suitable area under the SSP245 scenario, and 2328.75 km^2 (0.89%) under the SSP585 scenario. The study also projected 1065.14 km^2 (0.41%) and 2394.56 km^2 (0.91%) of the highly suitable area under SSP245 and SSP585 by 2061–2080, respectively. The scenarios projected the change in suitable areas of *V. lanceaefolia* under different future scenarios are shown in Table 3 and Fig. 3. Between 2041 to 2060, there will be a decrease of suitable area under SSP245 and SSP585 is projected to be 10.78% and 10.35%, respectively. Similarly, during the period from 2061 to 2080, the maximum loss of 59.00% in the highly suitable area under SSP245 and a decrease of 7.82% in the SSP585 scenario.

The current distribution of *V. lanceaefolia* ranges from high suitable area to moderately suitable area (Fig. 3) in Assam, Nagaland, Manipur, Meghalaya and Arunachal Pradesh among the states in the northeast India. Due to habitat loss, the species is confined to different forests in these states. Assam has the largest suitable area for *V. lanceaefolia* in the northeastern region, with 2263.74 km², representing 2.89% as highly suitable area of its total geographical area of 78438.00 km². The species thrives in the districts of Lakhimpur, Charaideo, Sivasagar, Jorhat, Majuli, Golaghat, Karbi Anglong, Dima Hasao, Cachar, Hailakandi, Karimganj, Dibrugarh, Tinsukia, and Dhemaji in Assam. According to the projections for the year 2041–2060 and 2061–2080 under SSP245 and SSP585, there will be a gradual decline in highly suitable area with maximum decline would have in the state Assam. Under SSP245 and SSP585 for the period of 2061–2080, there is an initial increase in moderately suitable areas, but this is followed by a decline in both SSP245 and SSP585. The low suitable area of the region shows an increase in SSP585 for the year 2041–2060 as compared to the current scenario.

The second largest state to hold the highly suitable area of *V. lanceaefolia* is Arunachal Pradesh having 310.11 km² (0.37%). The districts which fall under highly suitable area are Papumpare, Lower Siang, West Siang, Siang, East Siang, Lower Dibang Valley, Lohit, Namsai, Tirap, and Changlang. Under the scenarios SSP245 and SSP585, the highly and low suitable area will increase during 2041–2060 and 2061–2080. However, from 2061–2080, there will be a decline in the coverage of these areas under SSP245. Conversely, under SSP245 and SSP585 scenarios, there will be an increase in moderately suitable areas for *V. lanceaefolia* between the year 2041–2060 and 2061–2080.

Nagaland has 23.84 km² (0.14%) of highly suitable area, covering the districts of Longlang, Mokokchung, Wokha, Dimapur, and Peren. Under future climate change scenarios, there will be a steady decline in highly and moderately suitable areas, whereas low suitable areas will increase in SSP245 between the years 2061–2080. But both SSP245 and SSP585 will estimate a decline in low suitable areas between the years 2041–2060 and 2061–2080. In Manipur, there will be decline in suitable areas of *V. lanceaefolia* under SSP245 and SSP585. While, in Mizoram, under both SSP245 and SSP585 scenarios, there will be an increase in moderately suitable area for both time periods of 2061–2080. The other northeastern states, Meghalaya, Tripura and Sikkim have recorded the maximum areas that are least suitable for *V. lanceaefolia*. Overall, the habitat suitability of *V. lanceaefolia* is projected to decrease significantly across all the northeastern states.

Table 3
Comparison of Vatica lanceaefolia's habitat suitability in the India's Northeast India under current and future climate scenarios

Scenario	Estimated area (km2)	No	Low	Moderate	High
		(% change)	(% change)	(% change)	(% change)
Current		206449.18	37723.38	15413.75	2597.69
(2041–2060)	SSP245	206668.73 (0.11)	37605.66 (-0.31)	15592.07 (1.16)	2317.53 (-10.78)
	SSP585	203332.93 (-1.51)	41402.62 (9.75)	15119.7 (-1.91)	2328.75 (-10.35)
(2061–2080)	SSP245	196475.45 (-4.83)	45419.22 (20.40)	19224.19 (24.72)	1065.14 (-59.00)
	SSP585	203447.71 (-1.45)	39120.53 (3.70)	17221.2 (11.73)	2394.56 (-7.82)

4 Discussion

Our results offer the first overview of the current and future potential distributions of *Vatica lanceaefolia* across Northeast india by using ensemble modeling. The study also highlights key environmental and climatic factors influencing its distribution current and future species distributions. Additionally, we calculated changes in the species' potential area and identified potential refuge areas over temporal changes for *V. lanceaefolia*, a critically endangered tree species facing climate change in the Northeast India. The study highlights a significant decline in habitat areas across different climate change scenarios. Projections for the period of 2041–2061 and 2061–2080, under SSP245 scenario, shows a declining trend in the species distribution. The highly suitable area of *V. lanceaefolia* in the northeastern region is projected to reduce by 280.16 km^2 between 2041–2060 and 1532.55 km^2 between 2061–2080 under SSP245. Thus, these findings underscore that habitat loss and fragmentation of habitats poses high risks to species existence in future, resulting a significant change in their distribution and survival. This is because of the adverse effects of climate change on plant species distribution (Swart et al. 2019; Wu et al. 2019).

In the future climatic scenario, it is anticipated that there would be potential habitat loss and a little longitudinal shift of the species due to climate change. The primary environmental variables that influence a species' distribution revealed that precipitation and temperature are the main environmental variables that influence or affect the potential distribution of *V. lanceaefolia.* From the ensemble model it can be indicated that the current spatial distribution of *V. lanceaefolia* is influenced by the amount of Precipitation of Driest Quarter, CEC_mean_(0–5)cm, Precipitation Seasonality (Coefficient of Variation) and Slope. Environmental factors such as Precipitation of Driest Month and Max Temperature of Warmest Month, Slope, Aspect and other soil data are expected to strongly influence the distribution of the species, in future climatic scenarios. Comparatively, under all scenarios, the relevant factors that emerge as important are Precipitation of Driest Quarter. Bioclimatic variables as well as topographical variables and soil data contribute immensely to the current and futuristic potential distribution model of *V. lanceaefolia* are within a significant range. Apparently BIO2, which represents the Mean Diurnal Range (Mean of monthly (max temp - min temp)), is the least influential factor in determining the current and future distribution of the species. Similarly, it has been recorded that BIO17, CEC_mean_(0–5)cm, BIO15, and Slope are the maximum contributors to the two current and future climate scenarios respectively. Furthermore, the variables pH_mean_(0–5)cm and BIO14 are Soil pH (0–5)cm and Precipitation of Driest Month have the potential to ultimately affect the distribution of a species and it may be associated with precipitation in the futuristic scenario (Baruah et al. 2017; Sarma and Tanti 2022).

The modeling results of the current and future distribution of *V. lanceaefolia* in northeastern India indicate that there are 2597.69 km^2 areas are highly suitable for the species. Among different states of northeastern region, Assam alone represents a total of 2263.74 km^2 representing a highly suitable area for the density and population distribution of *V. lanceaefolia*. Arunachal Pradesh has an area of 310.11 km^2, followed by Nagaland with an area of 23.84 km^2. The distribution of the *V. lanceaefolia* was recorded in different habitats of Assam which include moist tropical forest (Borah and Devi 2014), home gardens of Barak valley (Devi and Das 2017), lowland rain forest of Kakoi reserve forest (Buragohain et al. 2023), upper Brahmaputra valley regions (Dutta et al. 2023), and tropical wet evergreen forests (Gosai, 2015 and Baidya et al. 2024). *V. lanceaefolia* was recorded as the dominant tree species in tropical semi-evergreen forest of Nambor Wildlife Sanctuary (Giri et al. 2019) and co-dominant species in Poba Reserve Forest (Borah et al. 2019). Hence, the species has been observed in diverse ecological settings ranging from disturbed secondary forests to undisturbed primary forests, indicating its adaptability and resilience. However, despite the presence of suitable habitat in Manipur and Meghalaya, the distribution of *V. lanceaefolia* is significantly limited. The findings emphasized the importance of targeted conservation efforts in the region where the species is most prominent to ensure the long-term survival of *V. lanceaefolia* across its potential range. Furthermore, the lack of previous study on distribution of *V. lanceaefolia* makes it challenging to compare it with other researches. Therefore, our study provides a baseline for understanding the habitat preferences and distribution patterns of *V. lanceaefolia*. Overall the present study establishes basic environmental variables and predicts potential habitat where species would thrive under different environmental conditions in different regions of Northeastern India. Also, our study comprehends in understanding the ecological and conservation measures require for the species in relation to ongoing environmental changes.

5 Conclusion

Under future climate scenarios, the study reveals a significant reduction in the habitat suitability of *V. lanceaefolia*. Projections for the period of 2061–2080 indicate that there would be a decrease in the area of highly suitable habitat. Under the SSP245 scenario, there will be a maximum reduction of high suitable area 1532.55 km^2, while under SSP585 scenario, there will be a reduction of 203.13 km^2. These findings highlight the negative impacts of climate change on the distribution of species. The current distribution of *V. lanceaefolia* is strongly influenced by

precipitation patterns and temperature, with the precipitation of the driest month and precipitation seasonality being critical factors. The model forecasts that these variables, along with other climatic factors, will have a similar effect on the future distribution of the species. The study further confirms that the potential habitat of *V. lanceaefolia* is limited to the northeastern states of the Assam, Arunachal Pradesh, Nagaland, and Manipur. Nevertheless, future climate scenarios predict a reduction in suitable habitats across these regions. In order to ensure the long-term survival of *V. lanceaefolia*, it is essential to address the anticipated habitat reductions and adapt conservation strategies to deal with the changing climatic conditions. Engaging policymakers and local communities, raising awareness, and implementing targeted conservation measures are crucial for the protection of the species and its habitat.

Declarations

Acknowledgements: Authors are highly grateful to the Department of Science and Technology (DST), Government of India for the financial support to Centre of Excellence project at Department of Environmental Science, Tezpur University to conduct the current study. The authors are thankful to Principal Chief Conservator of Forest (PCCF) of Assam for necessary permission to work in protected areas, Assam for assistance and supports provided during the field visit. We also thank Head, Department of Environmental Science and Dean, Research and Development of Tezpur University for administrative and logistic support.

Author contributions Anubhav Bhuyan, Shilpa Baidya, and Bijay Thakur did the field work, collected data and wrote the first draft of the manuscript. Kuldip Sarma, Anubhav Bhuyan and Suhasini Hazarika performed the modelling and analysis. Ashalata Devi, and Amit Prakash conceptualized and supervised the study, corrected the manuscript and arranged funding. All authors contributed ideas and finalized the manuscript.

Funding This work was supported by the Department of Science and Technology (DST), Government of India under grant ID DST/CCP/CoE/182/2019(G).

Data availability Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Ethical approval Not applicable

Consent for publication All authors agree to publish.

Competing interests The authors declare no competing interests.

References

1. 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
2. Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends Ecol Evol 22:42–47. https://doi.org/10.1016/j.tree.2006.09.010
3. Baidya S, Sumant S, Bhuyan A, et al (2024) Tree Population Structure and Carbon Stock of Tropical Wet Evergreen Forest in Dehing Patkai National Park, Assam, Northeast India. Int J Ecol Environ Sci 50:303–313
4. Banda LB, Tassie N (2018) Modeling the distribution of four-bird species under climate change in Ethiopia. Ethiop J Biol Sci 17:0–17

5. Baruah PS, Deka K, Sarma B, et al (2017) Assessment of few unexplored RET plant wealth of Assam, India. J Adv Plant Sci 9:10–15
6. Borah M, Boruah HPD, Neog B (2019) Tree biomass and carbon stock assessment of two forest areas of Assam in North East India. *Pleione* 13(1), 19 – 31 https://doi.org/10.26679/Pleione.13.1.2019.019-031
7. Borah M, Devi A (2014) Phenology, growth and survival of Vatica lanceaefolia Bl.: A critically endangered tree species in moist tropical forest of Northeast India. Trop Plant Res 1:1–12
8. Brown JL, Bennett JR, French CM (2017) SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. PeerJ 5:e4095. https://doi.org/10.7717/peerj.4095
9. Buragohain MK, Dar AA, Babu KN, Parthasarathy N (2023) Tree community structure, carbon stocks and regeneration status of disturbed lowland tropical rain forests of Assam, India. Trees, For People 11:100371. https://doi.org/10.1016/j.tfp.2023.100371
10. Devi NL, Das AK (2017) Plant biodiversity conservation through traditional use and cultivation in Meitei home-gardens in Barak valley of Assam, India. https://doi.org/10.26679/Pleione.11.2.2017.277-289
11. Duflot R, Avon C, Roche P, Bergès L (2018) Combining habitat suitability models and spatial graphs for more effective landscape conservation planning: An applied methodological framework and a species case study. J Nat Conserv 46:38–47. https://doi.org/10.1016/j.jnc.2018.08.005
12. Dutta M, Deb P, Das AK (2023) Factors shaping plant diversity in traditional agroforestry system of dominant ethnic communities of upper Brahmaputra valley regions of Northeast India. Agrofor Syst 97:727–738. https://doi.org/10.1007/s10457-023-00823-5
13. Elith J, Graham HC, Anderson PR, et al (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography (Cop) 29:129–151. https:// doi. org/ 10. 1111/j.2006. 0906- 7590. 04596.x
14. Friedman JH (1991) Multivariate adaptive regression splines. Ann Stat 19:1–67. https://doi.org/10.1214/aos/1176347963
15. Galán-Acedo C, Arroyo-Rodríguez V, Estrada A, Ramos-Fernández G (2018) Drivers of the spatial scale that best predict primate responses to landscape structure. Ecography (Cop) 41:2027–2037. https://doi.org/10.1111/ecog.03632
16. Giri K, Buragohain P, Konwar S, et al (2019) Tree diversity and ecosystem carbon stock assessment in Nambor Wildlife Sanctuary, Assam. Proc Natl Acad Sci India Sect B Biol Sci 89:1421–1428. https://doi.org/10.1007/s40011-018-01072-8
17. Gosai K. (2015) Assessment of Seasonal Variation in Canopy Structure and Greenness in Tropical Wet Evergreen Forests of North East India. Master degree Thesis, Indian Institute of Remote Sensing, Indian Space Research Organisation, Department of Space, Govt. of India, Dehradun, Uttarakhand, India.
18. Guisan A, Thuiller W, Zimmermann NE (2017) Habitat suitability and distribution models: with applications in R. Cambridge University Press
19. Hazarika A, Deka JR, Nath PC, et al (2023) Modelling habitat suitability of the critically endangered Agarwood (Aquilaria malaccensis) in the Indian East Himalayan region. Biodivers Conserv 32:4787–4803. https://doi.org/10.1007/s10531-023-02727-3
20. Hearst MA, Dumais ST, Osuna E, et al (1998) Support vector machines. IEEE Intell Syst their Appl 13:18–28. https://doi.org/10.1109/5254.708428
21. Hijmans RJ, Cameron SE, Parra JL, et al (2005) Very high resolution interpolated climate surfaces for global land areas. Int J Climatol A J R Meteorol Soc 25:1965–1978. https://doi.org/10.1002/joc.1276

22. Hu X-G, Jin Y, Wang X-R, et al (2015) Predicting impacts of future climate change on the distribution of the widespread conifer Platycladus orientalis. PLoS One 10:e0132326. https://doi.org/10.1371/journal.pone.0132326
23. IPCC 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, 184 pp., doi: 10.59327/IPCC/AR6-9789291691647.
24. Jain SK, Kumar V, Saharia M (2013) Analysis of rainfall and temperature trends in northeast India. Int J Climatol 33:968–978. https://doi.org/10.1002/joc.3483
25. Jhajharia D, Shrivastava SK, Sarkar D, Sarkar S (2009) Temporal characteristics of pan evaporation trends under the humid conditions of northeast India. Agric For Meteorol 149:763–770. https://doi.org/10.1016/j.agrformet.2008.10.024
26. Kaky E, Nolan V, Alatawi A, Gilbert F (2020) A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. Ecol Inform 60:101150. https://doi.org/10.1016/j.ecoinf.2020.101150
27. Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography (Cop) 28:385–393. https://doi.org/10.1111/j.0906-7590.2005.03957.x
28. Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr 17:145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x
29. Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence–absence models in ecology: the need to account for prevalence. J Appl Ecol 38:921–931. https://doi.org/10.1046/j.1365-2664.2001.00647.x
30. Naimi B, Araújo MB (2016) sdm: a reproducible and extensible R platform for species distribution modelling. Ecography (Cop) 39:368–375. https://doi.org/10.1111/ecog.01881
31. Nelder JA, Wedderburn RWM (1972) Generalized linear models. J R Stat Soc Ser A Stat Soc 135:370–384. https://doi.org/10.2307/2344614
32. Nüchel J, Bøcher PK, Xiao W, et al (2018) Snub-nosed monkeys (Rhinopithecus): potential distribution and its implication for conservation. Biodivers Conserv 27:1517–1538. https://doi.org/10.1007/s10531-018-1507-0
33. Parthasarathy B (1995) Monthly and seasonal rainfall series for all India homogeneous regions and meteorological subdivisions: 1871-1994. Indian Inst Trop Meteorol Res Rep RR-65
34. Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecol Modell 133:225–245. https://doi.org/10.1016/S0304-3800(00)00322-7
35. Pearson RG (2007) Species' distribution modeling for conservation educators and practitioners. Synth Am Museum Nat Hist 50:54–89.
36. Phillips SJ, Dudík M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. In: Proceedings of the twenty-first international conference on Machine learning. p 83. https://doi.org/10.1145/1015330.1015412
37. Poggio L, De Sousa LM, Batjes NH, et al (2021) SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7:217–240. https://doi.org/10.5194/soil-7-217-2021
38. Ripley BD (2007) Pattern recognition and neural networks. Cambridge university press.
39. Roy A, Das SK, Tripathi AK, et al (2015) Biodiversity in North East India and Their Conservation. Progress Agric 15:182. https://doi.org/10.5958/0976-4615.2015.00005.8
40. Sarma B, Tanti B (2022) Ecological niche modeling for reintroduction and conservation of Aristolochia cathcartii Hook. f. & Thomson (Aristolochiaceae), a threatened endemic plant in Assam, India. J Threat Taxa 14:20597–

20605. https://doi.org/10.11609/jott.6999.14.2.20597-20605

41. Secretariat of the Convention on Biological Diversity. 2020. Global Biodiversity Outlook 5. Available at: https://www.cbd.int/gbo5 [Date accessed: 15 August 2022].
42. Shabani F, Kumar L, Ahmadi M (2018) Assessing accuracy methods of species distribution models: AUC, specificity, sensitivity and the true skill statistic. Glob J Human-Social Sci B Geogr Geo-Sciences, Environ Sci Disaster Manag 18(1).
43. Sharma E, Chettri N, Eriksson M, et al (2009) Climate change impacts and vulnerability in the Eastern Himalayas. https://doi.org/10.53055/ICIMOD.497
44. Sharma N, Deka B, Sharma A, et al (2022) Vatica lanceaefolia: A Critically Endangered Tree of Tropical Forests of Upper Brahmaputra Valley.
45. Shiva MP, Jantan I (1998) Non-timber forest products from dipterocarps. A Rev Dipterocarps–Taxonomy, Ecol Silvic (S Appanah JM Turnbull, eds) CIFOR/FRIM Cent Int For Res Bogor, Indones 187–197
46. Swart NC, Cole JNS, Kharin V V, et al (2019) The Canadian earth system model version 5 (CanESM5. 0.3). Geosci Model Dev 12:4823–4873. https://doi.org/10.5194/gmd-12-4823-2019
47. Wu T, Lu Y, Fang Y, et al (2019) The Beijing climate center climate system model (BCC-CSM): The main progress from CMIP5 to CMIP6. Geosci Model Dev 12:1573–1600. https://doi.org/10.5194/gmd-12-1573-2019
48. Yebeyen D, Nemomissa S, Hailu BT, et al (2022) Modeling and Mapping Habitat Suitability of Highland Bamboo under Climate Change in Ethiopia. Forests 13:1–16. https://doi.org/10.3390/f13060859
49. Zhang L, Liu S, Sun P, et al (2015) Consensus forecasting of species distributions: The effects of niche model performance and niche properties. PLoS One 10:1–18. https://doi.org/10.1371/journal.pone.0120056

Figures

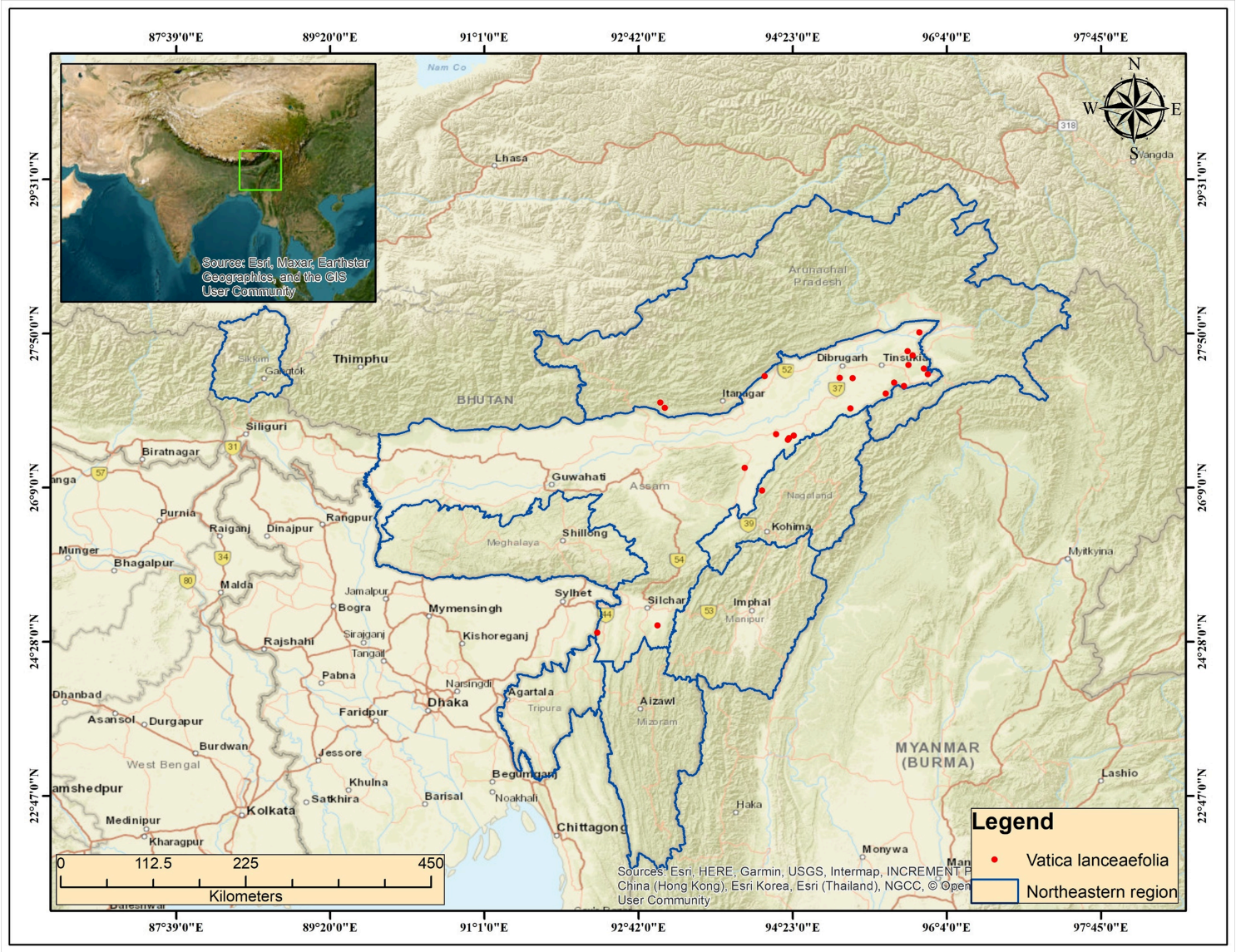


Figure 1

Sampling sites of *Vatica lanceaefolia* in the Northeastern region of India

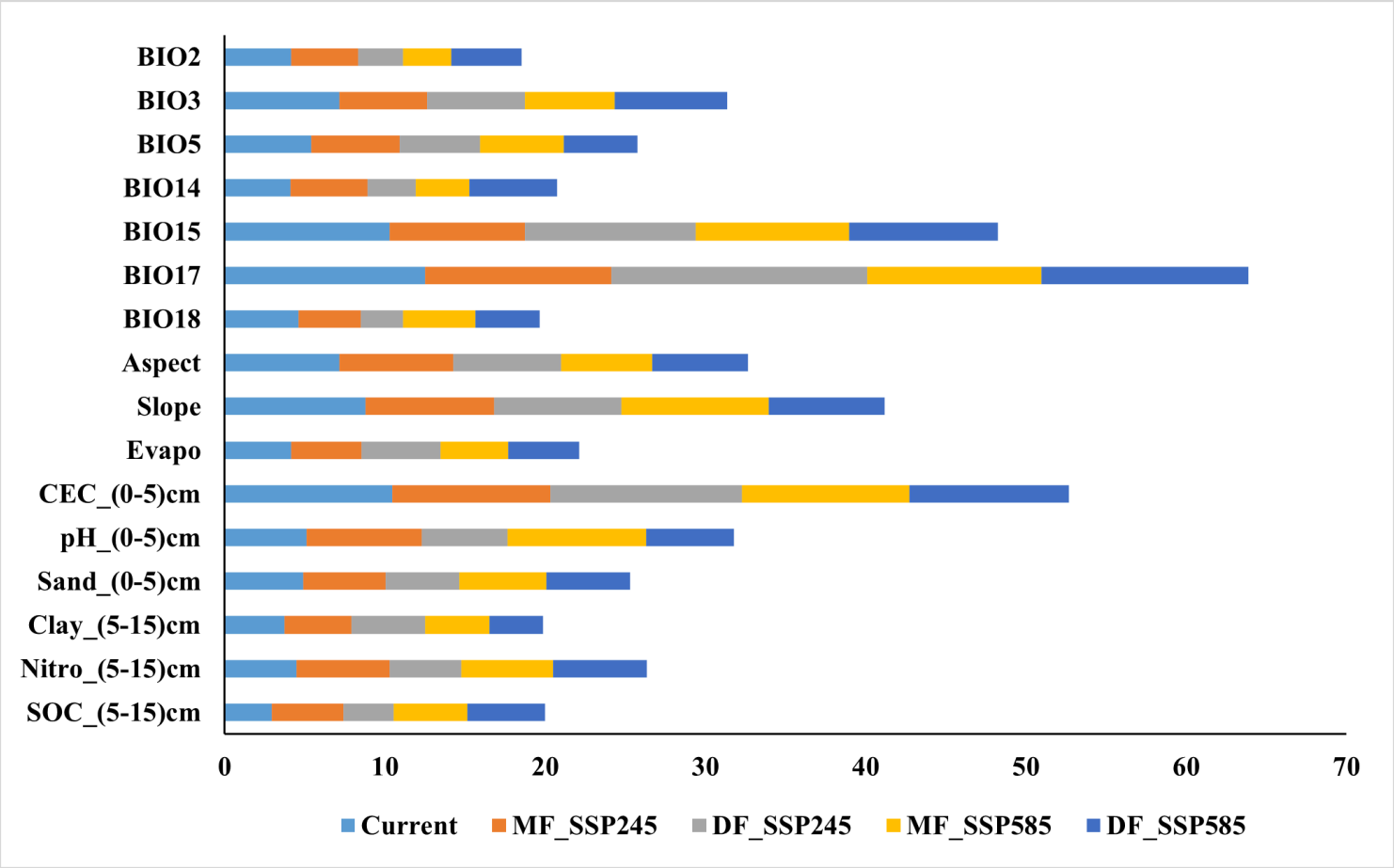


Figure 2

Variables contribution for different scenarios of habitat suitable model, where Distant future (DF) and Mid-future (MF) indicates timeframes i.e. 2061-2080 and 2041-2060

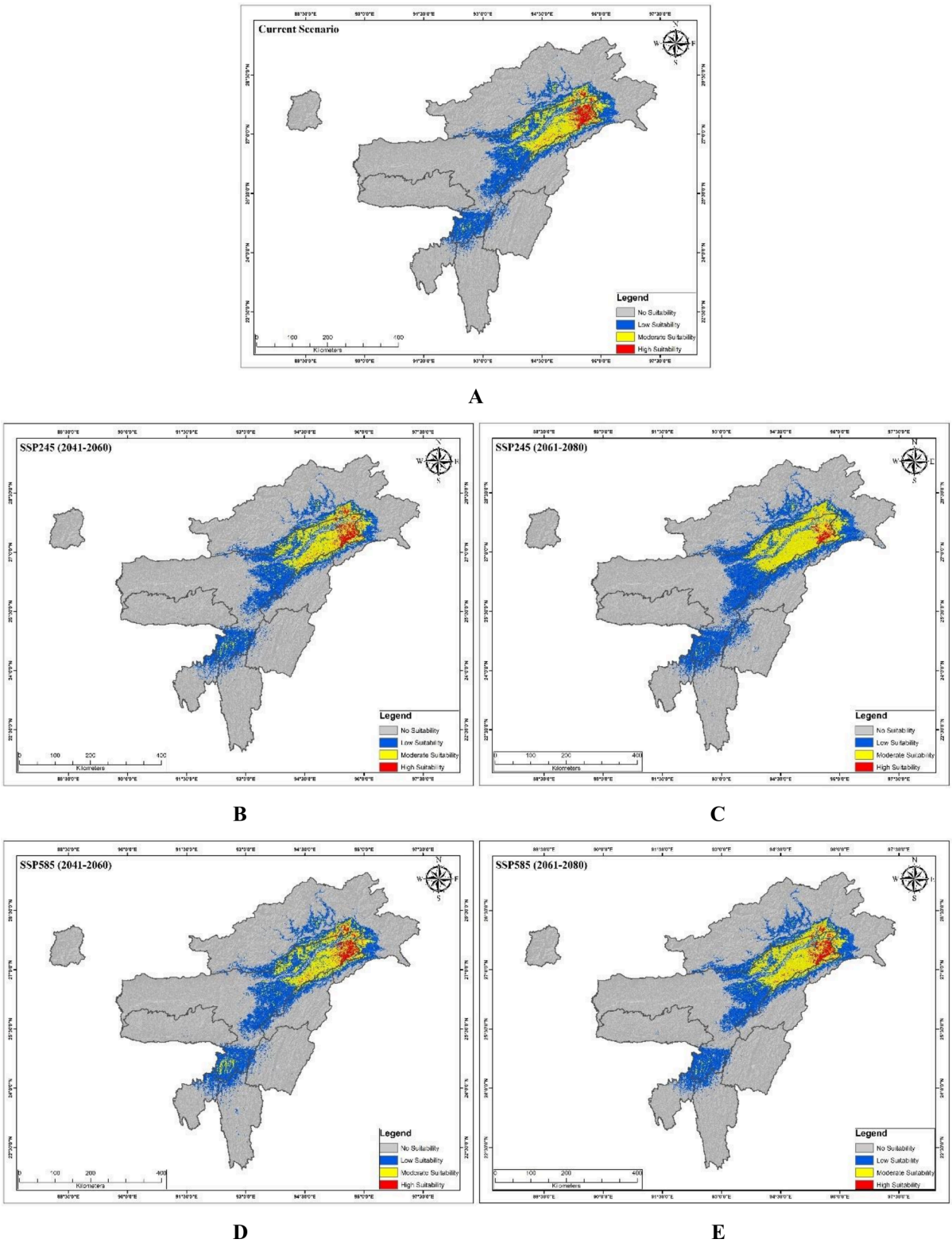


Figure 3

Potential suitable habitat of *Vatica lanceaefolia* in present and future scenarios of SSP245, and SSP585 for the years 2041-2060 and 2061-2080. (A) Current Scenario, (B) SSP245 (2041-2060) Scenario, (C) SSP245 (2061-2080) Scenario, and (D) SSP585 (2041-2060) Scenario, (E) SSP585 (2061-2080) Scenario.