Ensemble species distribution model of threatened Cycas L. species of Kannur district and Kerala, India

Kannankodantavida Manjusha
Central University of Kerala
Kavya Jeevan
Central University of Kerala
Shalu George
Central University of Kerala
Nadirsha Puthiyasurambi Nawab
Central University of Kerala
Mukesh Lal Das
Central University of Kerala
Anbazhagi Muthukumar
Central University of Kerala
Muthukumar Muthuchamy
mmuthukumar@cukerala.ac.in
Central University of Kerala

Research Article

Keywords: Cycas, ensemble, modeling, niche breadth, niche overlap, threatened

Posted Date: June 18th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4399560/v1

Additional Declarations: No competing interests reported.

Abstract

The single cycad genus *Cycas* L. of the family Cycadaceae requires special care as they come under the threatened category. The current study focussed on predicting the current, future and paleoclimatic distribution of potentially suitable habitats of *Cycas circinalis* and *Cycas nathorstii* in Kerala state and its second northernmost Kannur district. Ensemble function in “sdm” package used to combine five modeling algorithms, namely Generalized Linear Model (GLM), Generalized Additive Model (GAM), Random Forest (RF), Bioclim, Boosted Regression Tree (BRT), and Maximum Entropy (MaxEnt). Individual model validation used the area under the curve and true skill statistics value. Annual mean precipitation was the most contributed predictor for the current suitability model of both species, followed by mean annual temperature and precipitation of the driest month. *C. circinalis* exhibited a high frequency of least suitable regions, and *C. nathorstii* showed higher moderately suitable areas in Kerala and suitable regions in the Kannur district. The paleoclimatic suitability of both species in Kerala and Kannur district requires further supporting evidence. The distribution model of *C. circinalis* to future SSP 245 and SSP 585 scenarios showed a slight increase in suitability from 2021–2040 to 2081–2100 for both extents. Similarly, it favours slightly the suitability of *C. nathorstii* in all four periods. The niche breadth metric resulted in the habitat specificity for both species in Kerala and less specificity of habitats in the Kannur district, along with niche overlap among them. The study will aid in exploring the new populations in the area based on the model and develop conservation efforts.

Introduction

The identification of ecological niches, comparisons of niche similarity between species pairs in geographic and environmental space (Warren et al., 2021), prediction of species distribution under past, current, and future climate scenarios (Franklin, 2023), risk of species invasion (Briscoe Runquist et al., 2019), and disease prevalence (Dicko et al., 2014) are among the many research areas where SDMs are applicable. To predict potentially suitable habitats across a landscape or region, habitat suitability models use actual relationships between a species’ distribution and environmental variables like climatic, edaphic and topographical parameters (Bradley et al., 2012; Zimmerman et al., 2010).

There are several models available for the prediction of habitat suitability. They include Genetic Algorithm for Rule-set Prediction (GARP) (Stockwell, 1999), BIOCLIM (Busby, 1991), Random Forest (RF) (Breiman, 2001), Generalised Additive Model (GAM) and Generalised Linear Models (GLM) (Guisan et al., 2002), Generalized Regression Analysis and Spatial Prediction (GRASP) (Lehmann et al., 2003), Classification and Regression Trees (CART) (Breiman et al., 1984), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), Support Vector Machine (SVM) (Cortes & Vapnik, 1995) and Maximum Entropy (MaxEnt) (Phillips et al., 2006).

The concept of ensemble forecasting was first introduced in the pivotal work “The Combination of Forecasts” by Bates and Granger (1969). They reported a lower mean error for the combined predictions than any component individual forecasts (Araujo & New, 2007). Ensembling points to the average of the

individual models used and could be a mix of machine-learning and regression-based techniques (Hysen, 2022). An appropriate source of inspiration could be found on open-source platforms, like the R project for statistical computations and programming (Araujo & New, 2007). Weightage to the models used will be the same or based on the model that performed better (Thuiller, 2004). For the Weighted Averaging (WA) method, individual model outputs are chosen using interpolative metrics such as the True Skill Statistic (TSS) and Area Under the Curve (AUC) (Allouche et al., 2006) and the individual models are then aggregated using an averaging function. Other methods for processing individual forecasts to ensemble include mean, median and Principal Component Analysis median (PCAm) (Araujo & New, 2007). By combining four modeling algorithms, namely the generalized boosting model (GBM), generalized linear model (GLM), multiple adaptive regression splines model (MARS) and random forest model (RF) using the Biomod2 package, Xie et al. (2022) predicted the distribution of endemic conifers in China. Unique R packages, like 'sdm' (Naimi & Araujo, 2016), offer consistent and cohesive frameworks for managing several SDM approaches on a single interface. By using this package, models can be compared, tuned and ensembled.

The distribution prediction of five Boswellia species in Africa's drylands used the ensemble approach in the 'sdm' package (Esser & Weldemariam, 2023). Many researchers have reported the superiority of ensemble models over individual models (Araujo & New, 2007; Hao et al., 2019; Mohammadi et al., 2022). Potential habitats of *Citron medica* in south, east and southeast Asia and Australia were predicted using the ensemble function in the "sdm" package (Naimi & Araujo, 2016) for current and future Representative Concentration Pathways (RCP) scenarios (Maurya et al., 2023).

Cycas L. is a threatened solitary cycad genus of the family Cycadaceae worldwide and requires conservation practices (Srivastava, 2021). Researchers used many modeling methods to predict the habitat suitability of threatened cycad species and conserve them. Logistic regression models were utilized to know the environmental variables affecting the distribution of *Macrozamia macdonnellii* (F.Muell. ex Miq.) A. DC., of Central Australia and found the influence of microclimate on the growth (Preece et al., 2007). The MaxEnt output of habitat suitability predicted the contraction and population shift upslope from the protected areas with a rise in mean temperature for five allopatric *Macrozamia* species of Australia (Laidlaw & Forster, 2012). Current and future distribution patterns of *Cycas pectinata* in Southeast Asia under climate change scenarios were modelled using the MaxEnt algorithm. They noted the population reduction and difficulty adapting to the changing climate (Pradhan & Chettri, 2017).

Mudannayake et al. (2019) predicted the distribution of *C. zeylanica* in Sri Lanka using the MaxEnt modeling approach and revealed its existence in South East Asia and Southern Western Ghats, India. Carvajal-Hernandez et al. (2020), while evaluating the IUCN status of a micro endemic species of Mexico *Ceratozamia miqueliana*, predicted the potential distribution using MaxEnt modeling. Bashyal et al. (2020) attempted to determine the distribution status and suitability of *C. pectinata* in the Chure range of Nepal by integrating ArcGIS 10.2.2 and the analytical hierarchy process. They included topographic and edaphic variables to model the distribution.

The elucidation of the habitat suitability of *Cycas* species in Kerala (KE) and this state's second northernmost Kannur district (KD) will pave the way for conserving the endangered genus of the area. Hence, the current study designed to predict the potential distribution of two species, *C. circinalis* L. and *C. nathorstii* J. Schust. and both species are coming under the IUCN Red List of threatened species. The study compares the paleoclimatic, current, and future distribution of the two species, evaluates niche breadth, and determines niche overlap between them. The modeling of the current and future distribution of the species will aid in describing new populations in the area and in adopting conservation strategies.

Materials and methods

Species distribution modeling

The extent of the study area for modeling the distribution of two species and the occurrence points used for modeling shown in Fig. 1.

Occurrence data

For the present study, the GPS coordinates of the individuals were recorded from the field surveys. For *C. circinalis* and *C. nathorstii*, the potential distribution for KE and KD was checked. Presence records for *C. circinalis* and *C. nathorstii* included the primary and secondary occurrences. Secondary occurrences were downloaded from the Global Biodiversity Information Facility (GBIF) for *C. circinalis* and from the India Biodiversity Portal (IBP) for *C. nathorstii*, and data cleaning was done in RStudio. This data was merged with the primary occurrence data. Both occurrence records underwent spatial thinning using the "spThin" package. From 50 GPS coordinates, 23 occurrence records (8 primary and 15 secondary) remained for *C. circinalis* for the entire KE by manually removing the similar coordinates (Feng et al., 2019) and spatial thinning and 18 occurrence points (14 primary and 4 secondary) for *C. nathorstii* from 22 data points along South Indian states and Sri Lanka. The number of background points was tuned to reduce the data imbalance; for *C. circinalis*, it is 200 and for *C. nathorstii, it* is 100. Both species' datasets were split into training (80 per cent) and test (20 per cent) datasets for model run and validation.

Environmental variables

The chosen environmental or Bioclimatic variables were downloaded from PaleoClim (Brown et al., 2018) and WorldClim with 30s (0.93 x 0.93 = 0.86 km^2) resolution. Current (1979–2013) (Karger et al., 2017) Bioclimatic variables, climate data for the Last Glacial Maximum (LGM ~ 20,000 years ago) (Karger et al., 2021) and Mid-Holocene (MH = 6000 years ago) (Fordham et al., 2017) were downloaded from PaleoClim. Future climatic projections of Coupled Model Intercomparison Project 6 (CMIP6) based on Shared Socio-economic Pathways (SSP) scenarios (SSP 245 and SSP 585) were downloaded from WorldClim 2.0. The global climate model – the Model for Interdisciplinary Research on Climate version 6.0 (MIROC6) for SSP245 (Intermediate scenario) and SSP585 (Extreme scenario) was used for future distribution prediction. The period for projections were 2021–2040, 2041–2060, 2061–2080, and 2081–2100.

As the Bioclimatic variables were derived from temperature and precipitation, collinearity existed between them. Hence, firstly, six Bioclimatic variables were chosen, namely Annual mean temperature (Bio 1), Maximum temperature of the warmest month (Bio 5), Minimum temperature of the coldest month (Bio 6), Annual mean precipitation (Bio 12), Precipitation of the wettest month (Bio 13) and Precipitation of driest month (Bio 14). Pearson's correlation matrix checked the multicollinearity among the chosen variables, and variables with correlation coefficients of less than 0.7 were retained for further modeling. Hence, the remaining variables for modeling included Bio 1, Bio 12, and Bio 14. Each variable's contribution to the model was calculated using the AUC metric.

Ensemble species distribution modeling using sdm package and validation

Ensemble approach using the "sdm package" (Naimi & Araujo, 2016) in RStudio modelled the current distribution of the two *Cycas* L. species, namely *C. circinalis* L and *C. nathorstii* and predicted the paleoclimatic and future distribution of the species. Models were run using the "sdm" package in RStudio. The averaging utilized six models, namely Generalized Linear Model (GLM), Generalized Additive Model (GAM), Random Forest (RF), Boosted Regression Tree (BRT), and MaxEnt, out of which GLM and GAM are regression-based algorithms, and RF, BRT and MaxEnt are machine-learning techniques (Hijmans & Elith, 2013). Bioclim is the first developed species distribution model, a profiling technique (Booth et al., 2014; Hijmans & Elith, 2013). Ten replications were used for each model using the bootstrapping method. Individual models were validated using AUC and TSS (True Skill Statistic) metrics. The models were combined using the ensemble function in the package, and AUC criteria were used to weigh the models.

The generated raster predictions were made in the habitat suitability map with five suitability classes (Mahatara et al., 2021) in ArcMap 10.6 by clipping the extent to KE state and KD. The suitability classes include 0-0.2 (Not suitable), 0.21–0.4 (Least suitable), 0.41–0.6 (Moderately suitable), 0.61–0.8 (Suitable), and 0.81- 1 (Highly suitable). The area of each class was calculated using DIVA-GIS and plotted as a graph using Origin 9.0.

Estimation of niche breadth and niche overlap

Mean niche breadth (Levin, 1968) was calculated using the "dismo" package in Rstudio from the ensemble models for all the periods for both species and plotted using OriginPro 9.0. The value ranges from 0–1 indicates narrow and broad niche breadth or environmental tolerance. The predicted niche occupancy (PNO) profiles for each environmental variable were derived using the 'pno' function in the "phyloclim" package for niche overlap or similarity estimation. These PNOs calculated niche overlap as Schoener's D statistic (Schoener, 1968). The value ranges from 0 (no overlap) to 1 (complete overlap).

Results and discussion

Model validation

The validation statistics of the modeling outputs for *C. circinalis* and *C. nathorstii* are shown in Table 1. Model validation utilized two statistics, AUC and TSS values, out of which AUC ranges from 0 to 1 and values near to 1 denote the model's better performance (Philips et al., 2006). AUC value > 0.75 is better, according to Elith (2002). TSS values range from − 1 to + 1, and less than zero indicates poor performance (Allouche et al., 2006).

Table 1
Validation statistics for the suitability models of *C. circinalis* and *C. nathorstii*

Modelling algorithm	AUC		TSS	
	C. circinalis	*C. nathorstii*	*C. circinalis*	*C. nathorstii*
GLM	0.84	0.9	0.72	0.82
GAM	0.84	0.94	0.69	0.89
RF	0.86	0.98	0.73	0.93
BRT	0.86	0.97	0.73	0.91
MaxEnt	0.87	0.98	0.75	0.93

The five modeling algorithms used in predicting the habitat suitability of *C. circinalis* exhibited an AUC value greater than 0.8 and a TSS of 0.7 and above. The maxEnt model ranks first with an AUC value of 0.87 and a TSS of 0.75. All five models of *C. nathorstii* have an AUC value ≥ 0.9 and a TSS above 0.8. MaxEnt and RF models exhibited higher AUC (0.98) and TSS (0.93). The MaxEnt model worked better for both species, again supporting the algorithm's effectiveness and popularity (Wisz et al., 2008). The results indicate that the better model performance for both species and predicted the habitat suitability for the extent of KE and KD.

Contribution of Bioclimatic variables

The importance of variables for each model is combined and averaged based on the AUC metric to get the per cent contribution of the Bioclimatic variables on the current distribution of both species. It is found that for *C. circinalis*, Bio 12 (37.8%) is the highly contributing variable, followed by Bio 1 (24.6%) and Bio 14 (8.5%). Similarly, for *C. nathorstii*, Bio 12 contributes 68.7%, followed by Bio 1 (7.5%) and Bio 14 (6%). Hence, this indicates that annual precipitation (Bio 12) favours the existence of both species in the KE and KD. According to Swart et al. (2018) annual mean precipitation contributes secondly to the suitability prediction of *Encephalartos latifrons*, a South African cycad after geological features. The contribution of variables is shown in Table 2 for *C. circinalis* and *C. nathorstii*.

Table 2
Percentage contribution of predictor variables to the five modelling algorithms

Predictor variables	Percentage contribution of *C. circinalis*	Percentage contribution of *C. nathorstii*
Bio 12	37.8	68.7
Bio 1	24.6	7.5
Bio 14	8.5	6

Moreover, Bio 12 immensely favouring *C. nathorstii* over *C. circinalis* itself, showing that in KE and KD, suitability is higher for *C. nathorstii*. On the other hand, annual mean temperature (Bio 1) contributes more to the distribution of *C. circinalis* than *C. nathorstii*. The present-day climate is actually characterized by much seasonal rainfall patterns where the evolutionary cycad relicts survive (Nagalingum et al., 2011; Yessoufou et al., 2014). They had occupied from the Miocene and survived the xeric habitats, but some cycads adapted to mesic habitats with sufficient moisture (Gutiérrez-Ortega et al., 2017). Hence, the contribution of Bio 12 and Bio 1 for the existence of *C. circinalis* and *C. nathorstii* points to the mesic nature of the species. At the same time mean temperature also contributes less to the suitability of both species. When compared to *C. nathorstii* Bio 1 is contributing more to *C. circinalis*.

Current distribution of the species

Along KE, *C. circinalis* exhibited not suitable, least suitable, and moderately suitable areas, and it is devoid of suitable and highly suitable areas. The predicted suitability map is displayed in Fig. 2a. The scenario is similar in KD; for both extents, the least suitable areas were higher, with an area of 20967.66 sq. km for KE and 1957.36 sq. km for KD. This is followed by not suitable areas (16075.12 sq. km) in KE and moderately suitable areas (868.6 sq. km) in the KD. On the other hand, the ensemble model of *C. nathorstii*, shown in Fig. 2b, exhibited all five suitability areas in KE. Except for suitable areas, all other suitability classes were in the KD. The projected model highlighted that suitable and highly suitable areas were present in the northernmost Malabar region of KE. In KE, moderately suitable areas were higher, with an area of about 17426.87 sq. km. In the KD, suitable areas were higher, counting to 1780.83 sq. km, followed by highly suitable areas of 775.82 sq. km. The predicted suitability maps of both species show that KE and KD favours *C. nathorstii* when compared to *C. circinalis*.

Lindstrom and Hill (2007) recorded the survival of *C. circinalis* in Kerala, which ranges from hilly to coastal regions. From Kalliasseri panchayat, KD, Sreedharan (2004) reported the presence of *C. circinalis* as part of biodiversity checklist. The existence of *C. nathorstii* a Sri Lankan native species, reported from Kerala for the first time at Palaruvi Hills, Kollam (Antony et al., 2010) and the present model shows least and moderately suitable areas.

Field exploration at KD has identified both species through morphological characters with the help of expertise as part of the delimitation of quantitative characters using clustering (Unpublished data). Hence, the existing studies support the predicted models in this study.

Reconstruction of paleoclimatic distribution of the species

The paleoclimatic distribution of *C. circinalis* is shown in Fig. 3. It depicts suitable, moderately suitable, and least suitable areas during the LGM period in KE. KE has mostly favoured moderately suitable areas. In the KD, only moderately suitable regions were present in both LGM (3029.78 sq. km) and MH (3028.06 sq. km) periods. The MH period favoured mostly moderately suitable areas (36399.5 sq. km) in KE, with few regions being suitable (2304.8 sq. km) and very few regions being least suitable (21.5 sq. km). The comparison graph of changes in area for *C. circinalis* with current and paleoclimatic period is displayed in Fig. 4a-b for KE and KD.

In the case of *C. nathorstii* during LGM and MH in KE, both moderately suitable and suitable areas were available, and the proportions of each class opposed each other. The reconstructed paleoclimatic habitat suitability in KE and KD for both LGM and MH are shown in Fig. 5. In LGM, few least suitable regions (119.36 sq. km) were also present. The area change compared to the KE's and KD's current scenario with LGM and MH suitability is shown in Fig. 6 and Fig. 7.

Figure 7.**tif** Comparison of area change between current scenario and two paleoclimatic period of *C. nathorstii* in KD

In Kerala, compared to current suitability, not suitable and least suitable areas were absent and characterized by moderately suitable regions. However, in LGM, the number of moderately suitable areas was higher, and vice versa, in the MH period. In the KD, LGM favoured moderately suitable regions (3041.85 sq. km), and MH favoured suitable areas (3045.31 sq. km).

According to Wanner et al. (2008), the climate systems' boundary conditions of the MH period have not much change when compared to the current climate systems. The temperature of land and air showed a decline during the MH period on entire globe, but tropical and subtropical land surface experienced varying or mixed climatic conditions (Steig, 1999). Hence, the suitability of the two species in MH may be interpreted when compared to the current distribution, as there is less variation in the climate systems as well as the mixed or mesic conditions in the tropics. Further explanation requires paleontological shreds of evidence and evolutionary studies as it is unfamiliar.

Future projection of the species

The projection of the distribution model of *C. circinalis* to future SSP 245 and SSP 585 scenarios showed a slight increase in suitability from 2021–2040 to 2081–2100 in KE and KD. Figure 8a-b, and Fig. 9a-b depicts the habitat suitability maps of *C. circinalis* in future scenarios for both KE and KD. Towards the end of the 21st century, the moderately suitable areas are showing a mild increase, and not suitable areas are reducing. It is evident that in KE not suitable areas are decreasing in both climate scenarios of four time periods. When comparing it with current suitability in KE, not suitable areas decrease with increasing moderately suitable regions.

The least suitable regions also showed a slight decrease, as they may be converted to moderately suitable pixels in future periods. Figure 10a Fig. 10b shows the trend of the suitability of *C. circinalis* in the KE and KD by comparing it with the current suitability. The KD also showed a similar trend to KE, with an increase in moderately suitable areas and a decrease in not suitable and least suitable regions.

Habitat suitability maps of *C. nathorstii* in future scenarios for all four time periods are depicted in Fig. 11a-b and Fig. 12a-b. Future climate is favouring the distribution of *C. nathorstii* to a certain extent. While comparing both climate SSP scenarios, there is not much variation between them. All four time periods in both scenarios show a slight increase in suitable areas in KE and KD. Highly suitable regions of the KD will completely disappear in both scenarios when the time moves to 2081–2100. Hence, there is a slight disfavour of future climate for the distribution of *C. nathorstii* in the KD. Figure 13a-b shows the Comparison of area change in current and future scenarios of *C. nathorstii* in KE and KD.

Compared to the current scenario, suitable areas were reduced for KE in SSP 245 and SSP 585, moving from 2021–2040 to 2081–2100. Moreover, the least and highly suitable areas were diminishing and the moderately suitable and suitable areas were increasing in both scenarios in KE. The KD has had no notable change in the least suitable areas. Compared to the current distribution, moderately suitable and highly suitable regions are becoming the least. In both scenarios for all the periods, suitable areas were high.

The current climate of *C. circinalis* in KE and KD is not very favourable as there are no suitable and highly suitable areas in the state. However, while comparing it with future climate scenarios, the percentage of increase in moderately suitable areas shows a slightly positive response of the species to climate change. The current suitability model significantly contributed to mean annual precipitation and annual mean temperature. In the case of *C. nathorstii*, current climatic conditions favour highly suitable areas, which disappear in 2081–2100 for both scenarios. Opposingly, future scenarios favour suitable regions of both KE and KD. In many parts of India daily mean rainfall and maximum daily rainfall have an increase in all future scenarios (Vinod & Agilan, 2022). Hence, the presence of suitable regions in all future scenarios for both *C. circinalis* and *C. nathorstii* may be connected to the availability of rainfall.

Comparison of niche breadth of both species

Niche breadth summarizes the environmental tolerance of a species to its surroundings. It ranges from 0 to 1, where 0 indicates that the species are specific to particular habitats than others, and 1 indicates equal suitability of all habitats (Banta et al., 2012). Niche breadth was calculated as mean separately for KE and Kannur for both species. The mean niche breadth of *C. circinalis* in KE was 0.13 and 0.28 in KD for the current climate scenario. The result designates that the environmental tolerance level of *C. ciricinalis* to different environments is too low; hence, it is specific to particular habitats.

During LGM and MH, the mean niche breadth was 0.46 and 0.47, respectively, which is comparatively higher than the current specificity to habitat in KE. The *C. circinalis* in KD had a niche breadth of 0.5 and 0.54 for LGM and MH, respectively. It points to adaptive and suitable habitats for *C. circinalis* in past

climatic conditions. As the paleoclimatic interpretation requires further explorations in the area, it isn't easy to draw an apt conclusion about the suitability of this species in the past. The intermediate scenario (SSP 245) for the four time periods in the future gives out a niche breadth, increasing from 0.17–0.21, a low environmental tolerance level in KE.

Likewise, in KE, SSP 585, the extreme future climate scenario exhibits a niche breadth from 0.17–0.26 from 2021–2040 to 2081–2100. In the SSP 245 scenario, the niche breadth has a value of 0.33–0.37, and in SSP 585, it is from 0.33–0.42, which is a higher niche breadth for *C. circinalis* in the KD compared to the future scenarios in KE. Among future scenarios, SSP 585 during 2081–2100 showed a high value of niche breadth, 0.26 in KE and 0.42 in KD, but still a narrower niche breadth. The mean niche breadth of the species for all the scenarios in KE and KD is depicted in Fig. 14a-b.

From this, the tolerance level of the species from the paleoclimatic period to the current shows a severe decline, which tends to rise during both future scenarios along the four periods in KE and KD.

Figure 15a-b shows the mean niche breadth of *C. nathorstii* in KE and KD for all the scenarios. In the case of *C. nathorstii, the* current climatic scenario has a niche breadth value of 0.27 in KE and 0.61 in KD.

Compared with *C. circinalis*, *C. nathorstii* shows higher niche breadth in the KD, indicating the establishment and adaptation of *C. nathorstii* to varied habitats than *C. circinalis*. During the past climate scenario, the LGM and MH, *C. nathorstii*, showed a niche breadth of 0.54 and 0.53, respectively, in KE. On the other hand, in KD, it is 0.57 for LGM and 0.64 for MH. During SSP 245, *C. nathorstii* shows a niche breadth range of 0.28–0.30 and SSP 585 of 0.28–0.33 in KE. In contrast, KD has a niche breadth of 0.58–0.61 for both SSP scenarios. It again shows that *C. nathorstii* is more suitable to the climatic conditions of KD than KE in future scenarios.

Altogether, both species of *Cycas* have a certain level of habitat specificity and environmental requirements for their healthy existence in KE and KD. The specificity makes them vulnerable to climate change and lessens the habitat suitability of both species. The less specificity of habitats for both species to KD may be explained as the adaptation of the individuals in the area which is again supported by the existence of individuals throughout the district (field observations of author) during the study. Exploration of *Cycas* populations in many unexplored parts of KE will help to strengthen the diversity and distribution aspects, and the study of physiological responses of the species at varied habitats can determine the effect of changing climatic parameters.

Niche overlap of both species in KE

The quantification of niche overlap from SDMs is mainly based on a geographical context. Schoener's D statistic explained the overlap of SDMs in the case of the three Bioclimatic variables used for modeling for both species in KE. The result of niche overlap between both species is shown in Fig. 16.

Interestingly, the current distributional range of the species showed a statistic of 0.85 and 0.84 for Bio 1 and Bio 12, respectively. In Bio 1, Bio 12 and Bio 14, the niche overlap in the paleoclimatic period is above

0.9. On the other hand, in future scenarios, all three climatic requirements will also have a statistic above 0.75. Though complete overlap is not there, the value is near 1, which indicates the similar niche sharing between both species in the case of Bio 1, Bio 12 and Bio14. The limitation of this study in quantifying the niche overlap is that it is devoid of other components which determine the species distribution range.

Beyond these climatic variables, many other topographic, edaphic, and Biotic components influence the species' distribution limits (Guisan & Thuiller, 2005). The niche overlap among the species in this study for the three crucial variables may be the existence of competition or lack of competition due to the oversupply (Colwell & Futuyma, 2013). Hence, the study needs further exploration by including the other factors to develop an SDM and to identify the factors affecting niche overlap.

Conclusions

Cycas species require particular concern for their existence as they come under the threatened category. The leading cause of the threat is habitat degradation due to human interference. The development of SDMs in time will help to understand the species distribution range in current habitats and future scenarios. In turn, this will aid in developing conservation strategies for species' existence. The ensemble model, superior to the individual models, predicted the potential distribution of the two *Cycas* species in KE state and its one district, Kannur. The current study proved the suitability of *C. nathorstii* in KE rather than *C. circinalis*, which means that the Bioclimatic variables favour the existence of *C. nathorstii* rather than *C. circinalis*. All three Bioclimatic variables in the study contributed to their living, though annual mean precipitation (Bio 12) contributed higher. The physiological response and need for this requirement want to be studied further, as Cycads are generally categorized as xerophytes. Though the future scenarios affect the distribution of both species, compared to the current habitat suitability, there is not much of a profound impact on the distribution. The niche breadth or environmental tolerance of *C. circinalis* shows habitat preference and specificity higher than that of *C. nathorstii*. A niche similarity or overlap exists among the species, which points to the competition or the surplus of the Bioclimatic variables. If competition exists among the species, this can be linked to the low suitability of *C. circinalis* to *C. nathorstii* in KE and the KD. Gathering evidence for the reasons for overlap requires further research. Future research can also be focussed on the new population discovery in the unexplored parts of KE, the physiological changes of the species in the changing climate and its environmental tolerance, and the effect of other topographical, edaphic and Biotic factors on its distribution range.

Declarations

Acknowledgements

I, Kannankodantavida Manjusha (first author), express gratitude to the Council of Scientific and Industrial Research (CSIR) for awarding me a Senior Research Fellowship (SRF) to conduct the research.

Disclosure of potential conflicts of interest

The authors declare that they have no conflicts of interest.

Research involving human participants and/or animals

This statement is not applicable.

Informed consent

This statement is not applicable.

Funding

The authors did not receive funding support from any organization for the submitted work.

Financial interests

The authors declare they have no financial interests.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article. The first author received a Senior Research Fellowship from the Council of Scientific and Industrial Research (CSIR), file no – 09/1108(0031)/2018-EMR-1, to conduct the research.

Authors' contribution

Kannankodantavida Manjusha conceptualized the idea, applied the methodology, completed the analysis, wrote the original manuscript, and reviewed it. Kavya Jeevan was involved in framing the R code, methodology, and manuscript editing. Shalu George and Nadirsha Puthiyasurambi Nawab provided suggestions in the analysis and reviewed and edited the manuscript. Mukesh Lal Das edited the manuscript. Anbazhagi Muthukumar reviewed and edited the manuscript. Muthukumar Muthuchamy supervised the research, proposed suggestions and reviewed the manuscript.

Data/code availability

The datasets analyzed during the current study and the R code used are available from the corresponding author upon reasonable request.

References

1. Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology*, 43(6), 1223-1232. https://doi.org/10.1111/j.1365-2664.2006.01214.x
2. Antony, R., Shareef, S.M., & Mohanan, N. (2010). Cycas nathorstii J. Schust (Cycadaceae), a new record for Kerala state, India. *Encephalartos*, 102: 17-18

3. Araujo, M. B., & New, M. (2007). Ensemble forecasting of species distributions. *Trends in Ecology and Evolution*, 22,42–47. https://doi.org/10.1016/j.tree.2006.09.010
4. Banta, J. A., Ehrenreich, I. M., Gerard, S., Chou, L., Wilczek, A., Schmitt, J., Kover PX & Purugganan, M. D. (2012). Climate envelope modelling reveals intraspecific relationships among flowering phenology, niche breadth and potential range size in Arabidopsis thaliana. *Ecology letters*, 15(8), 769-777. https://doi.org/10.1111/j.1461-0248.2012.01796.x
5. Bashyal, A., Bhattarai, S., Gautam, J., Tamang, R., & Bhatta, B. (2019). Status, Distribution and Habitat Suitability Mapping of Cycas pectinata in Chure Range of Makawanpur, Central Nepal. *Journal of Plant Resources*, 18(1), 183
6. Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the operational research society*, 20(4), 451-468. https://doi.org/10.1057/jors.1969.103
7. Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modeling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*, 20(1), 1-9. https://doi.org/10.1111/ddi.12144
8. Bradley, B. A., Olsson, A. D., Wang, O., Dickson, B. G., Pelech, L., Sesnie, S. E., & Zachmann, L. J. (2012). Species detection vs. habitat suitability: are we biasing habitat suitability models with remotely sensed data?. *Ecological Modelling*, 244, 57-64. https://doi.org/10.1016/j.ecolmodel.2012.06.019
9. Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32. https://doi.org/10.1023/A:1010933404324
10. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. New York: Chapman & Hall/CRC. https://doi.org/10.1002/widm.8
11. Briscoe Runquist, R. D., Lake, T., Tiffin, P., & Moeller, D. A. (2019). Species distribution models throughout the invasion history of Palmer amaranth predict regions at risk of future invasion and reveal challenges with modelling rapidly shifting geographic ranges. *Scientific Reports*, 9(1), 2426. https://doi.org/10.1038/s41598-018-38054-9
12. Brown, Hill, Dolan, Carnaval, & Haywood (2018). PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. *Nature – Scientific Data*. 5:180254. https://doi.org/10.1038/sdata.2018.254
13. Busby, J. R. (1991). BIOCLIM - a Bioclimate analysis and prediction system. - In: Margules, C. R. & Austin, M. P. (eds), *Nature conservation: cost effective Biological surveys and data analysis*. CSIRO, pp. 64-68
14. Carvajal-Hernández, C. I., Juárez-Fragoso, M. A., Armenta-Montero, S., Demeneghi-Calatayud, A. P., & Vázquez-Torres, M. (2020). Different criteria for evaluation of the conservation status of Ceratozamia miqueliana (Zamiaceae). *Tropical Conservation Science*, 13, 1940082920952682. https://doi.org/10.1177/1940082920952682
15. Colwell, R. K., & Futuyma, D. J. (1971). On the measurement of niche breadth and overlap. *Ecology*, 52(4), 567-576. https://doi.org/10.2307/1934144

16. Cortes, C., & Vapnik, V. N. 1995. Support Vector Networks. *Machine Learning*, 20, 1-25. https://doi.org/10.1007/BF00994018
17. Dicko, A. H., Lancelot, R., Seck, M. T., Guerrini, L., Sall, B., Lo, M., Vreysen, M.J., Lefrançois, T., Fonta, W.M., Peck, S.L., & Bouyer, J. (2014). Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in Senegal. *Proceedings of the National Academy of Sciences*, 111(28), 10149-10154. https://doi.org/10.1073/pnas.1407773111
18. Elith, J. (2002). Quantitative methods for modeling species habitat: comparative performance and an application to Australian plants. In: Quantitative Methods for Conservation Biology (eds Ferson, S. & Burgman, M.). Springer, New York, pp. 39–58. https://doi.org/10.1007/0-387-22648-6_4
19. Esser, L. F., & Weldemariam, E. C. (2023). Climate changes could jeopardize a main source of livelihood in Africa's drylands. *Journal for Nature Conservation*, 71, 126319. https://doi.org/10.1016/j.jnc.2022.126319
20. Fordham, D. A., Saltré, F., Haythorne, S., Wigley, T. M., Otto-Bliesner, B. L., Chan, K. C., & Brook, B. W. (2017). PaleoView: a tool for generating continuous climate projections spanning the last 21000 years at regional and global scales. *Ecography*, 40(11), 1348-1358. https://doi.org/10.1111/ecog.03031
21. Franklin, J. (2023). Species distribution modeling supports the study of past, present and future Biogeographies. *Journal of Biogeography*. 50:1533–1545. https://doi.org/10.1111/jbi.14617
22. Friedman, J. H. (1991). Estimating functions of mixed ordinal and categorical variables using adaptive splines. Department of Statistics, Stanford University, 1-42.
23. Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters*, 8(9), 993-1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x
24. Guisan, A., Edwards, T.C. Jr., & Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological modelling*,157:89-100. https://doi.org/10.1016/S0304-3800(02)00204-1
25. Gutierrez-Ortega, J. S., Yamamoto, T., Vovides, A.P., Pérez-Farrera, M.A., Martínez, J.F., Molina-Freaner, F., Yasuyuki Watano, Y., and Kajita, T. (2017). Aridifcation as a driver of Biodiversity: a case study for the cycad genus Dioon (Zamiaceae). *Annals of Botany*, 00:1–15 https://doi.org/10.1093/aob/mcx201
26. Hao, T., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2020). Testing whether ensemble modelling is advantageous for maximizing predictive performance of species distribution models. *Ecography*, 43(4), 549-558. https://doi.org/10.1111/ecog.04890
27. Hijmans, R. J., & Elith, J. (2013). Species distribution modeling with R. R Cran Project.
28. Hysen, L., Nayeri, D., Cushman, S., & Wan, H. Y. (2022). Background sampling for multi-scale ensemble habitat selection modelling: Does the number of points matter?. *Ecological Informatics*, 72, 101914. https://doi.org/10.1016/j.ecoinf.2022.101914
29. Karger, D. N., Nobis, M. P., Normand, S., Graham, C. H., & Zimmermann, N. E. (2021). CHELSA-TraCE21k v1. 0. Downscaled transient temperature and precipitation data since the last glacial

maximum. *Climate of the Past Discussions*, 1-27. (chelsa lgm). https://doi.org/10.5194/cp-19-439-2023

30. Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, P., & Kessler, M. (2017). Climatologies at high resolution for the Earth land surface areas. *Scientific Data*. 4 170122. https://doi.org/10.1038/sdata.2017.122.
31. Laidlaw, M. J., & Forster, P. I. (2012). Climate predictions accelerate decline for threatened Macrozamia cycads from Queensland, Australia. *Biology*, 1(3), 880-894. https://doi.org/10.3390/biology1030880
32. Lehman, A., Overton, J.M., & Leathwick, J.R. (2003). GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling* 160:165–183. https://doi.org/10.1016/S0304-3800(02)00195-3
33. Levins, R. (1968). Evolution in changing environments: some theoretical explorations. Princeton Univ. Press, Princeton, NJ.
34. Lindstrom, A. J., and Hill, K. D. (2007). The genus Cycas (Cycadaceae) in India. *Telopea*, 11(4), 463-488.
35. Mahatara, D., Acharya, A. K., Dhakal, B. P., Sharma, D. K., Ulak, S., & Paudel, P. (2021). Maxent modeling for habitat suitability of vulnerable tree Dalbergia latifolia in Nepal. *Silva Fennica*, 55(4), 1-17. https://doi.org/10.14214/sf.10441
36. Maurya, A., Semwal, M., Mishra, B. P., Mohan, R., Rana, T. S., & Nair, N. K. (2023). Distribution modeling for predicting habitat suitability for citron (Citrus medica L.) under climate change scenarios. *Flora*, 304, 152298. https://doi.org/10.1016/j.flora.2023.152298
37. Mohammadi, A., Almasieh, K., Nayeri, D., Adibi, M.A., & Wan, H.Y., (2022). Comparison of habitat suitability and connectivity modeling for three carnivores of conservation concern in an Iranian montane landscape. *Landscape Ecology*, 37 (2), 411–430. https:// doi.org/10.1007/s10980-021-01386-5.
38. Mudannayake, A., Ranaweera, L., Samaraweera, P., Sooriyapathirana, S., & Perera, A. (2019). The morpho-genetic and ecological niche analyses reveal the existence of climatically restricted Cycas zeylanica complex in Sri Lanka. *Scientific Reports*, 9(1), 16807. https://doi.org/10.1038/s41598-019-53011-w
39. Nagalingum, N.S., Marshall, C.R., Quental, T.B., Rai, H.S., Little, D.P., & Mathews, S., (2011). Recent synchronous radiation of a living fossil. *Science*, 334, 796–799. https://doi.org/10.1126/science.1209926.
40. Naimi, B., & Araújo, M.B., (2016). sdm: a reproducible and extensible R platform for species distribution modeling. *Ecography*, 39, 368–375. https://doi.org/10.1111/ecog.01881
41. Phillips, S.J., Anderson, R.P., & Schapire, R.E., (2006). Maximum entropy modelling of species geographic distributions. *Ecological Modelling*, 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026. https://doi.org/10.1016/j.ecolmodel.2005.03.026

42. Pradhan, A., & Chettri, A. (2017). Identifying Protected Areas Suitable for Conservation of Cycas pectinata Buch. Ham. in Southeast Asia Under Climate Change Scenario*. International Journal of Ecology and Environmental Sciences*, 43(2), 129-135.
43. Preece, L. D., Duguid, A. W., & Albrecht, D. E. (2007). Environmental determinants of a restricted cycad in central Australia, Macrozamia macdonnellii. *Australian Journal of Botany*, 55(6), 601-607. https://doi.org/10.1071/BT06122
44. Schoener, T.W., (1968). Anolis lizards in Bimini: resource partitioning in a complex fauna. *Ecology*, 49:704-726. https://doi.org/10.2307/1935534
45. Sreedharan, T.P., (2004). Biological Diversity of Kerala: A survey of Kalliasseri panchayat, Kannur district. Kerala Research Programme on Local Level Development, Centre for Development Studies.
46. Srivastava, R. C. (2021). Living gymnosperms in India. *Ravi Srivastava*. ISBN: 9798526023504.
47. Steig, E. J. (1999). Mid-Holocene climate change. *Science*, 286(5444), 1485-1487. https://doi.org/10.1126/science.286.5444.1485
48. Stockwell, D. (1999). The GARP modelling system: problems and solutions to automated spatial prediction. *International journal of geographical information science*, 13(2), 143-158. https://doi.org/10.1080/136588199241391
49. Swart, C., Donaldson, J., & Barker, N. (2018). Predicting the distribution of Encephalartos latifrons, a critically endangered cycad in South Africa. *Biodiversity and Conservation*, 27, 1961-1980. https://doi.org/10.1007/s10531-018-1519-9
50. Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10, 2020–2027. https://doi.org/10.1111/j.1365-2486.2004.00859.x
51. Vinod, D., Agilan, V. (2022). Impact of Climate Change on Precipitation Over India Using CMIP-6 Climate Models. In: Dikshit, A.K., Narasimhan, B., Kumar, B., Patel, A.K. (eds) Innovative Trends in Hydrological and Environmental Systems. *Lecture Notes in Civil Engineering*, vol 234. Springer, Singapore. https://doi.org/10.1007/978-981-19-0304-5_13
52. Wanner, H., Beer, J., Bütikofer, J., Crowley, T. J., Cubasch, U., Flückiger, J., ... & Widmann, M. (2008). Mid-to Late Holocene climate change: an overview. *Quaternary Science Reviews*, 27(19-20), 1791-1828. https://doi.org/10.1016/j.quascirev.2008.06.013
53. Warren, D. L., Matzke, N. J., Cardillo, M., Baumgartner, J. B., Beaumont, L. J., Turelli, M., Glor, R.E., Huron, N.A., Simões, M., Iglesias, T.L., Piquet, J.C. & Dinnage, R. (2021). ENMTools 1.0: An R package for comparative ecological Biogeography. *Ecography*, 44(4), 504-511. https://doi.org/10.1111/ecog.05485
54. Wisz, M.S., Hijmans, R.J., Li, J., Peterson, A.T., Graham, C.H., Guisan, A. & NCEAS. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distribution*, 14, 763–773. https://doi.org/10.1111/j.1472-4642.2008.00482.x
55. Xie, D., Du, H., Xu, W. H., Ran, J. H., & Wang, X. Q. (2022). Effects of climate change on richness distribution patterns of threatened conifers endemic to China. *Ecological Indicators*, 136, 108594. https://doi.org/10.1016/j.ecolind.2022.108594

56. Yessoufou, K., Bamigboye, S. O., Daru, B.H., & van der Bank, M. (2014). Evidence of constant diversification punctuated by a mass extinction in the African cycads. *Ecology and Evolution*, 4:50–58. https://doi.org/10.1002/ece3.880

57. Zimmerman, N., Edwards, T., Graham, C., Pearman, P., & Svenning, J. (2010). Species distribution modelling. *Ecography*, 33(6), 985-1092. https:// doi.org/10.1111/j.1600-0587.2010.06953.x

Figures

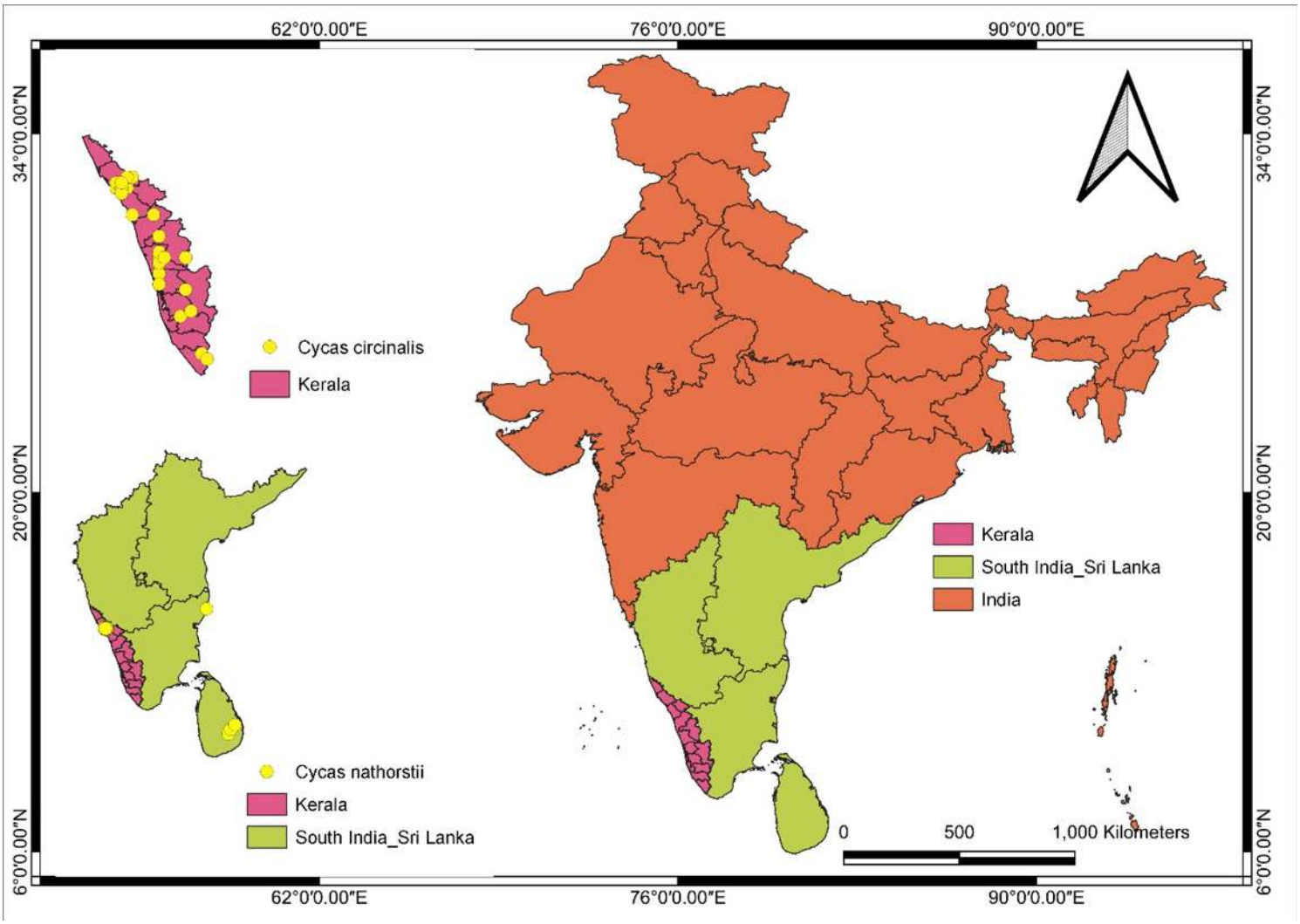

Figure 1

Occurrence points of *C. circinalis* and *C. nathorstii* used for modeling

Environmental variables

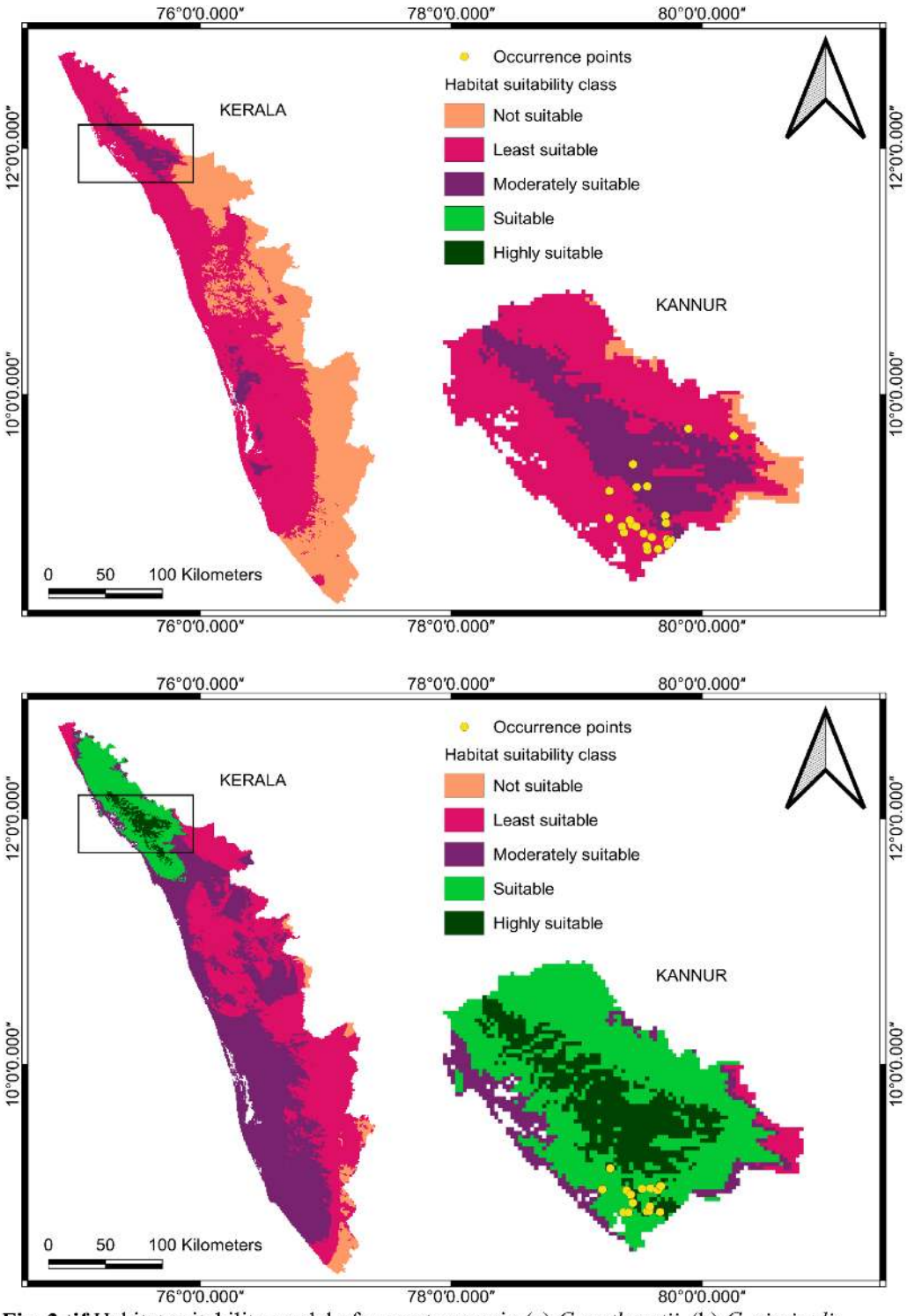

Fig. 2.tif Habitat suitability model of current scenario (a) *C. nathorstii*, (b) *C. circinalis*

Figure 2

Habitat suitability model of current scenario (a) *C. nathorstii*, (b) *C. circinalis*

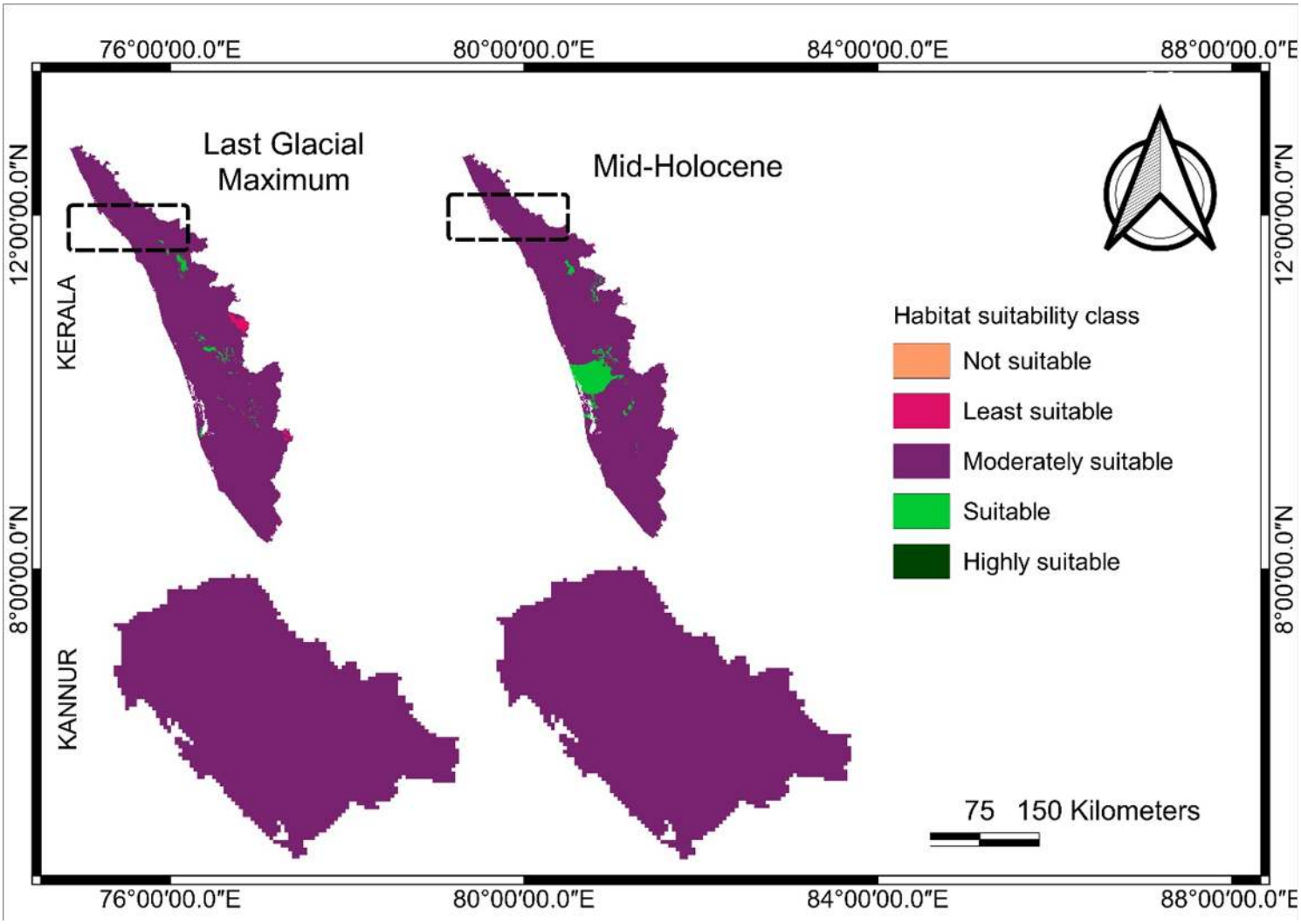

Figure 3

Paleoclimatic distribution model of *C. circinalis*

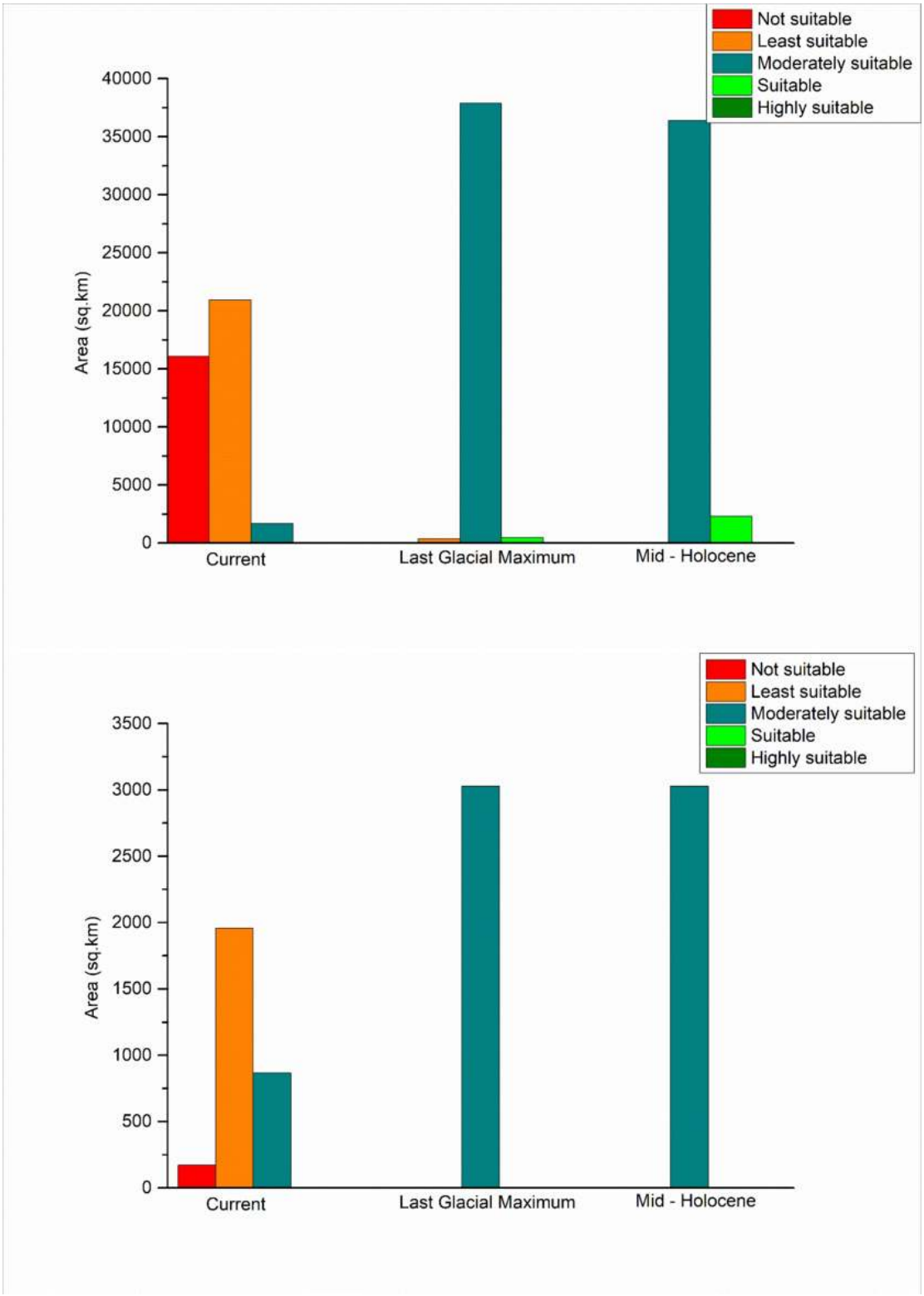

Figure 4

Comparison of area change between current scenario and two paleoclimatic period of *C. circinalis* in (a) KE, (b) KD

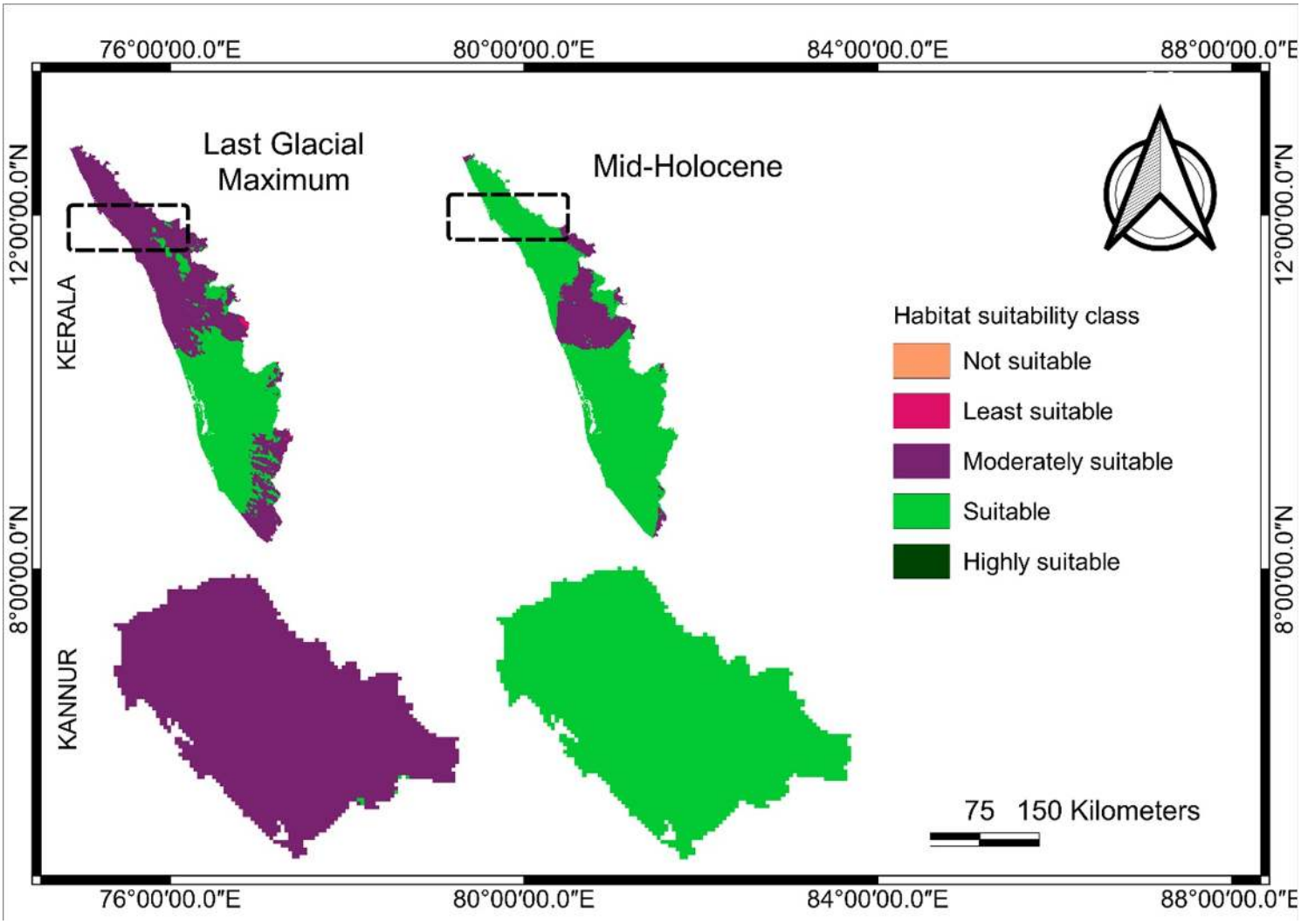

Figure 5

Paleoclimatic distribution model of *C. nathorstii*

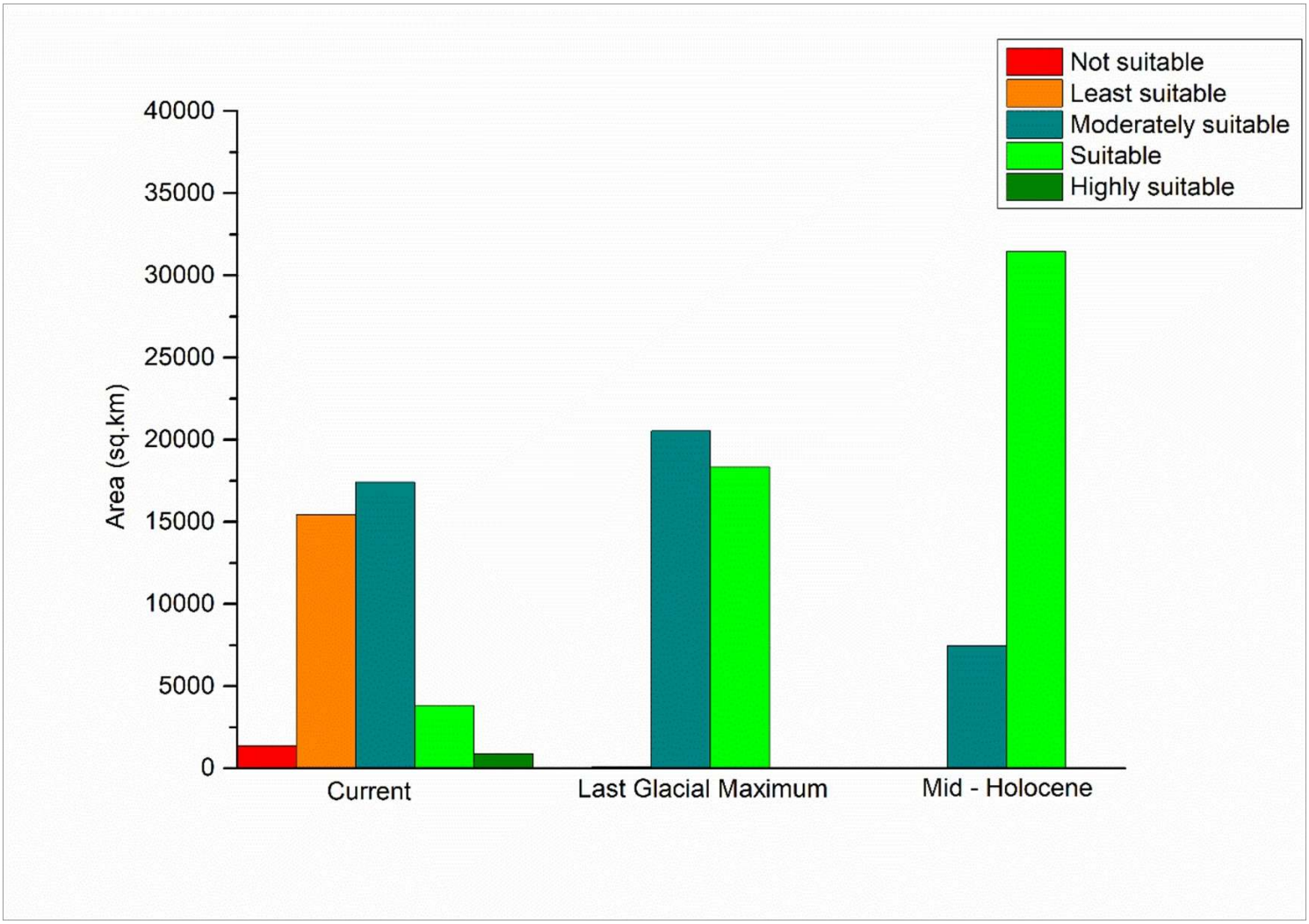

Figure 6

Comparison of area change between current scenario and two paleoclimatic period of *C. nathorstii* in KE

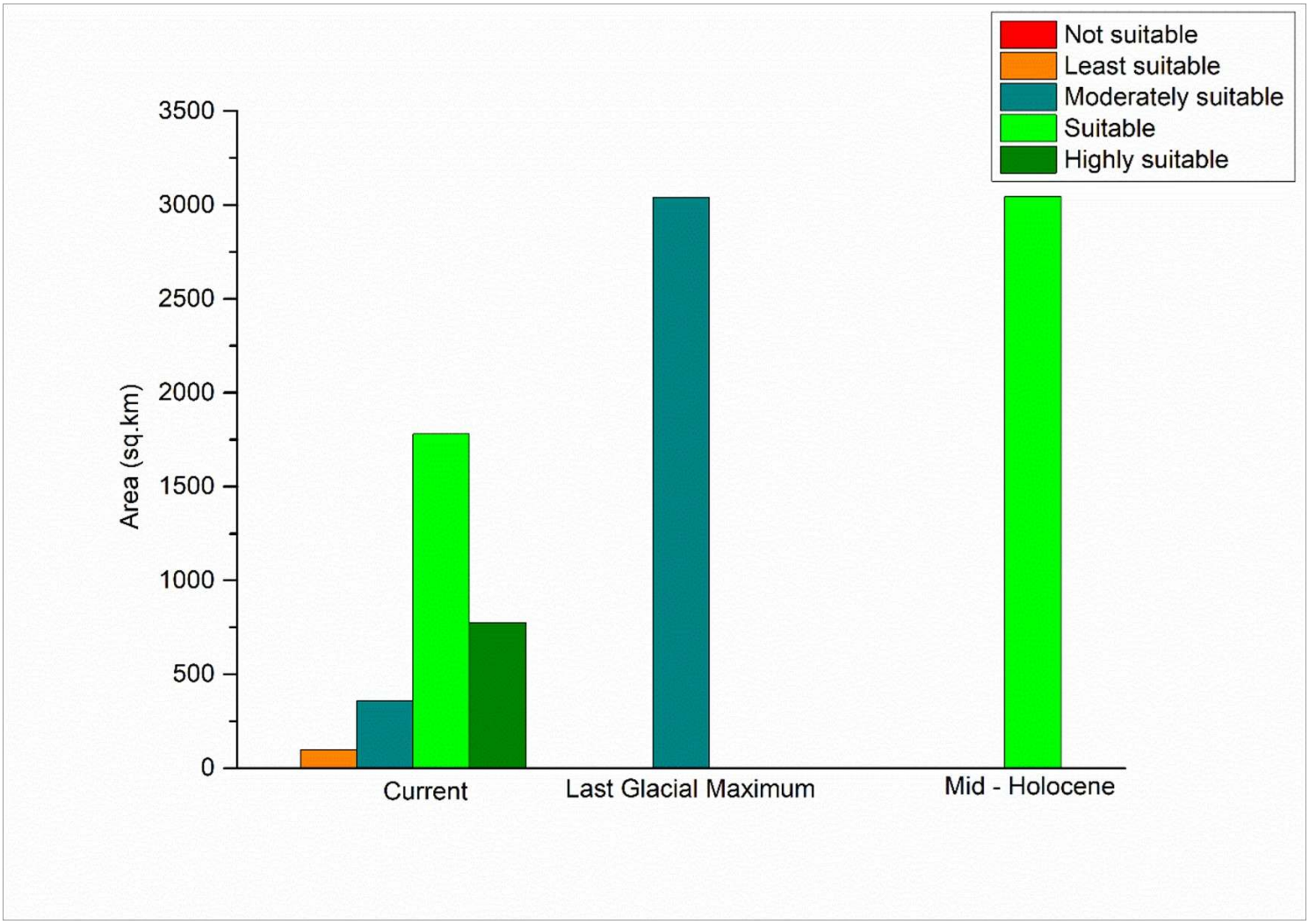

Figure 7

Comparison of area change between current scenario and two paleoclimatic period of *C. nathorstii* in KD

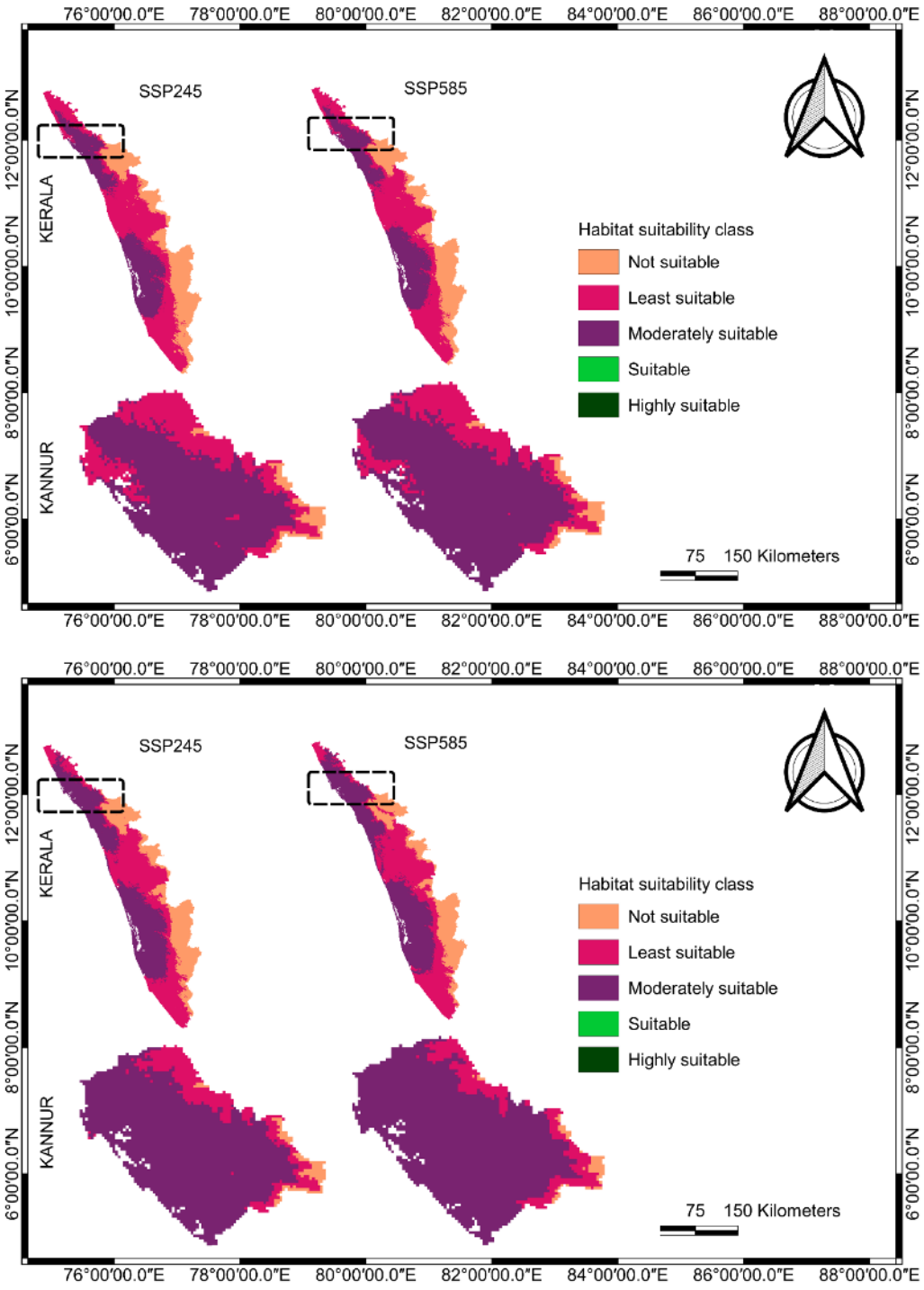

Fig. 8.tif The ensemble model of *C. circinalis* in future scenarios during (a) 2021-2040, (b) 2041-2060

Figure 8

The ensemble model of *C. circinalis* in future scenarios during (a) 2021-2040, (b) 2041-2060

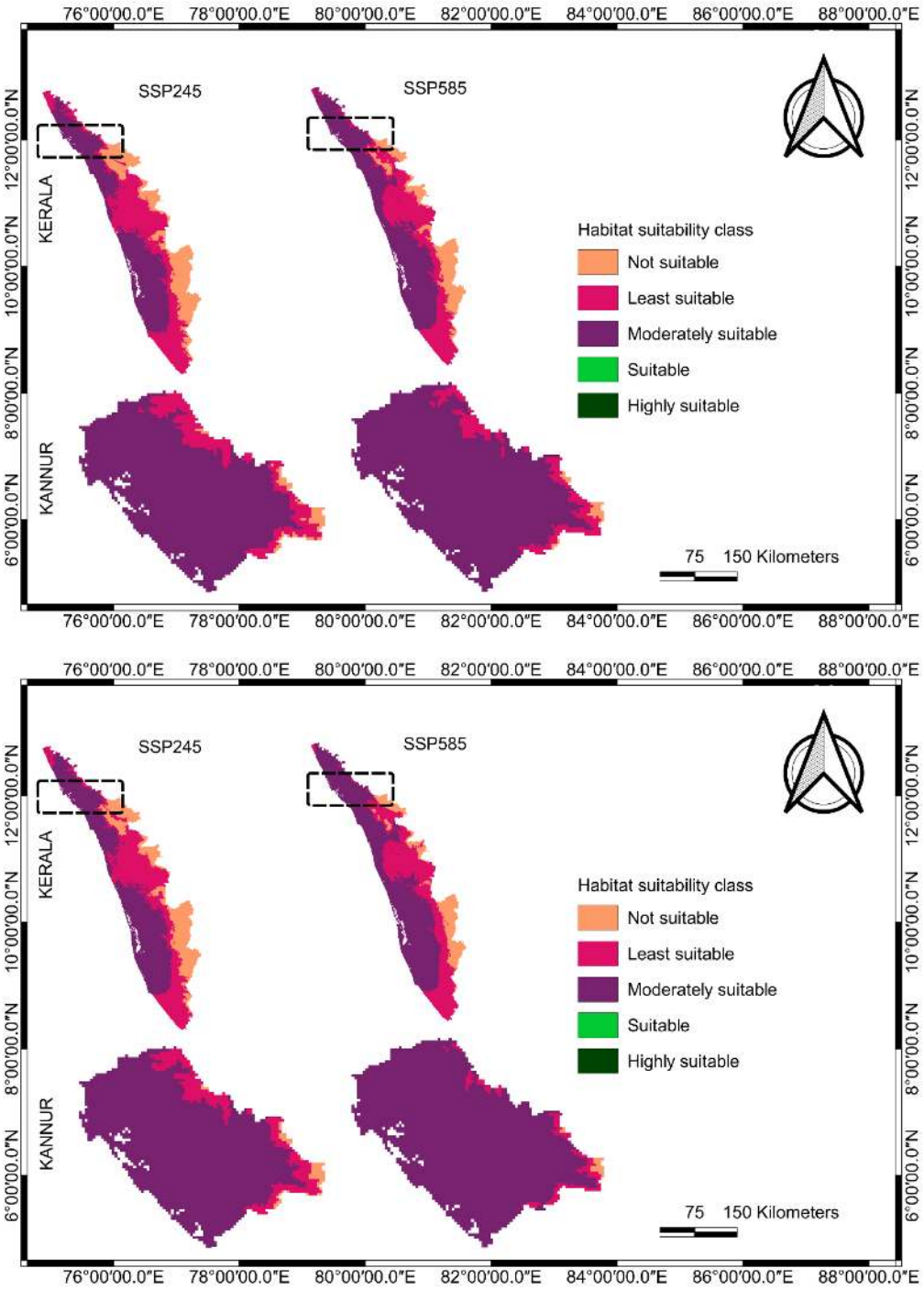

Fig. 9.tif The ensemble model of *C. circinalis* in future scenarios during (a) 2061-2080, (b) 2081-2100

Figure 9

The ensemble model of *C. circinalis* in future scenarios during (a) 2061-2080, (b) 2081-2100

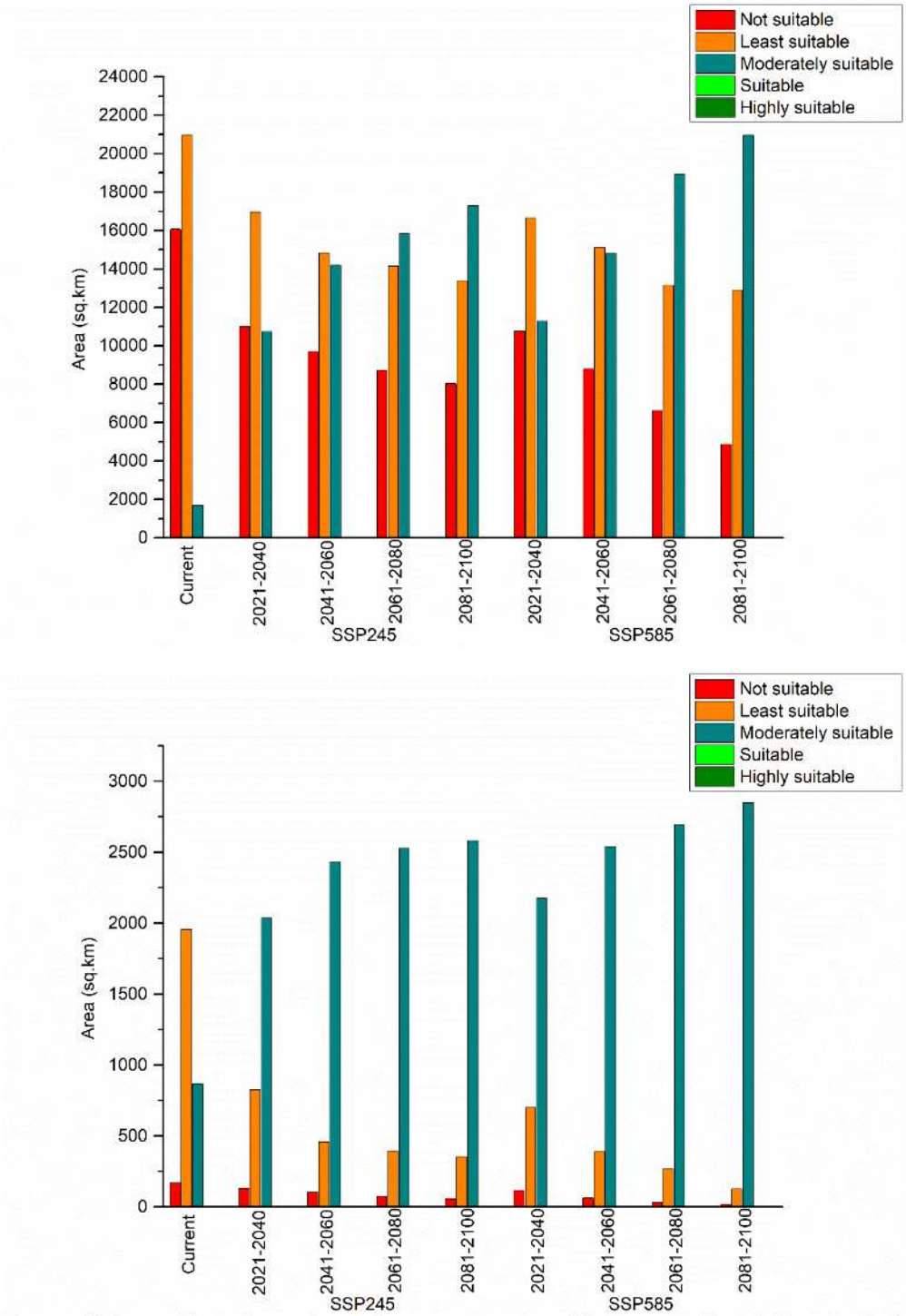

Fig. 10.tif Comparison of area change between current and future scenarios of *C. circinalis* in (a) KE, (b) KD

Figure 10

Comparison of area change between current and future scenarios of *C. circinalis* in (a) KE, (b) KD

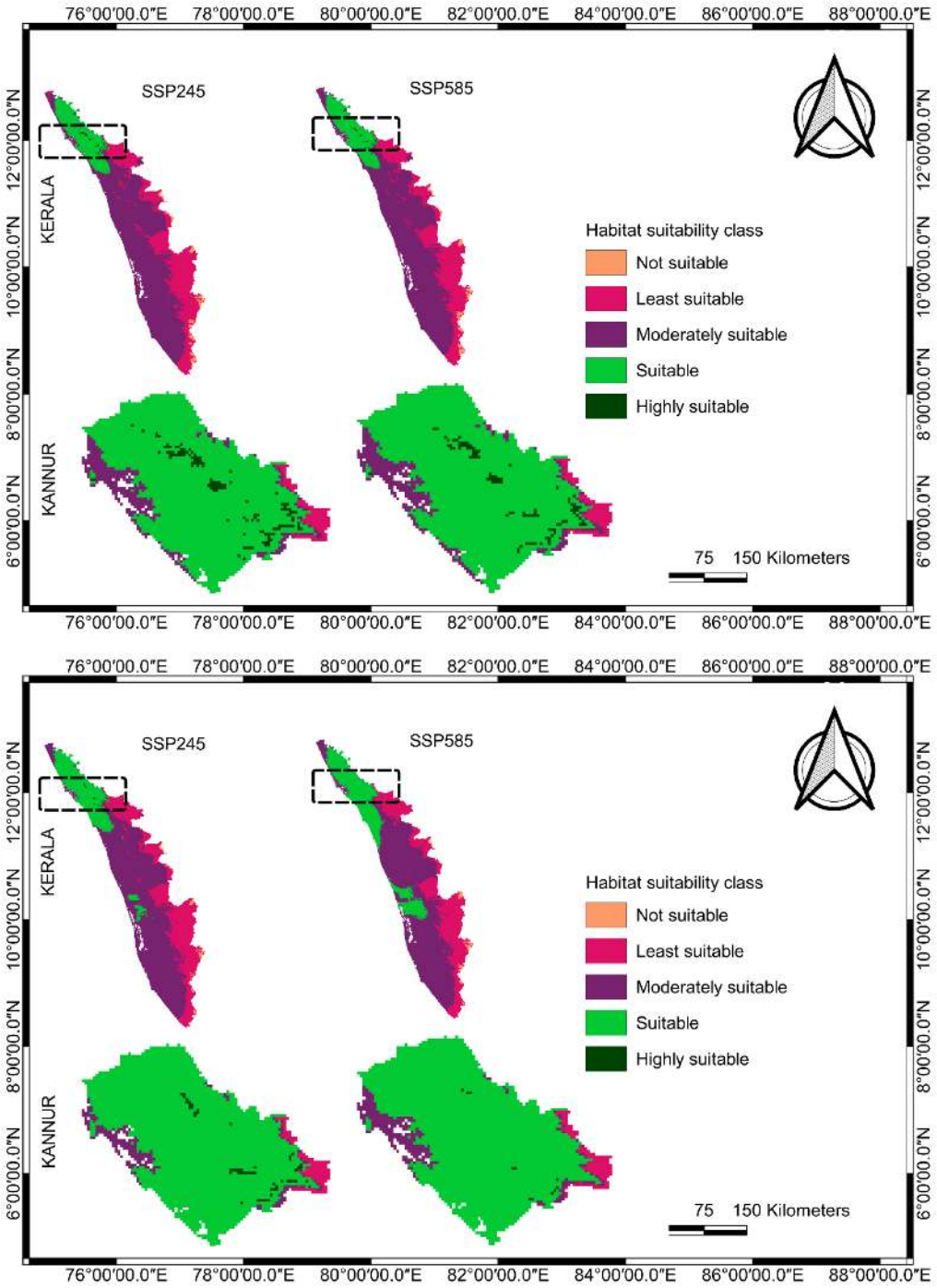

Fig. 11.tif The ensemble model of *C. nathorstii* in future scenarios (a) 2021-2040, (b) 2041-2060

Figure 11

The ensemble model of *C. nathorstii* in future scenarios (a) 2021-2040, (b) 2041-2060

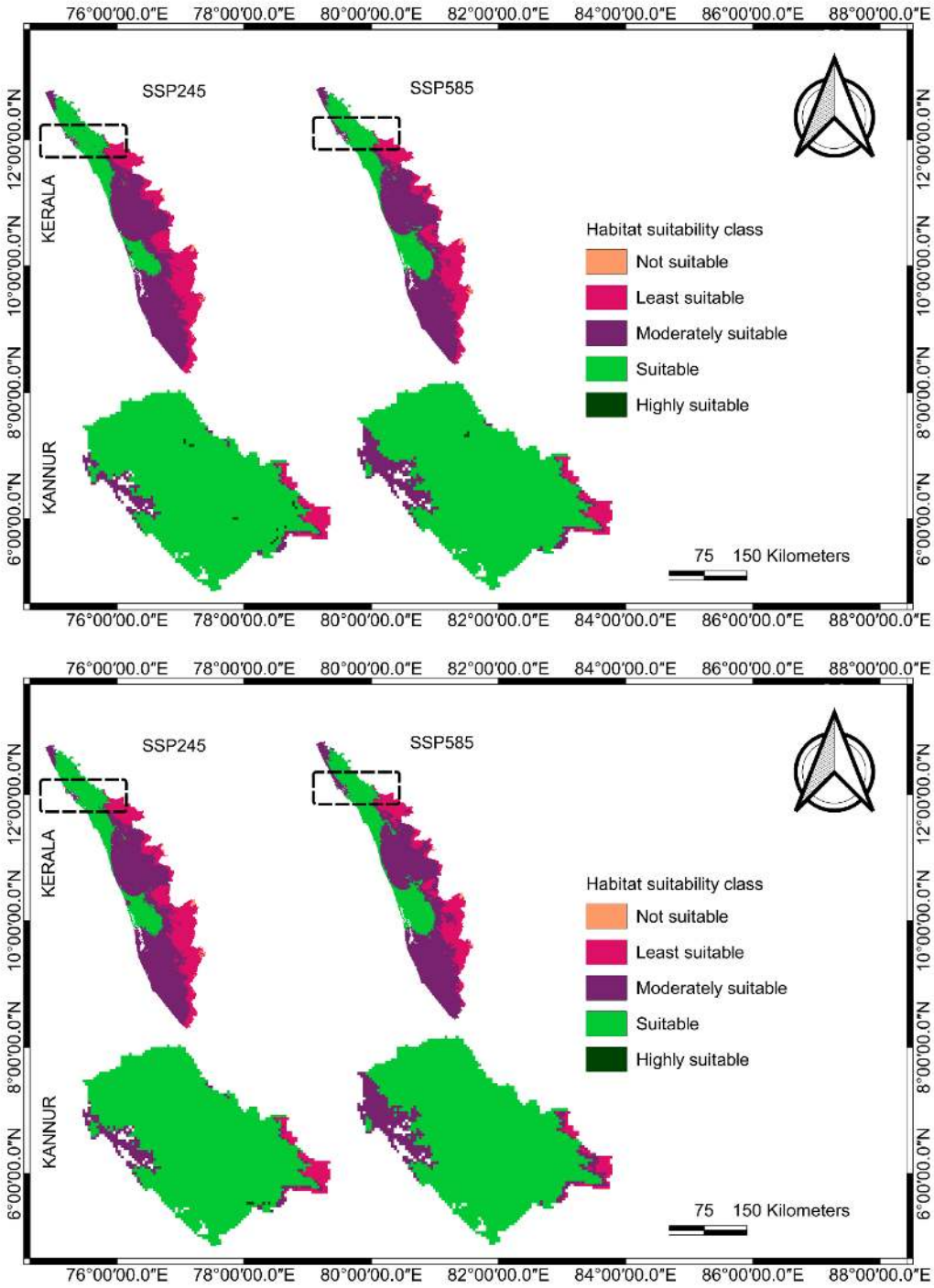

Fig. 12.tif The ensemble model of *C. nathorstii* in future scenarios during (a) 2061-2080, (b) 2081-2100

Figure 12

The ensemble model of *C. nathorstii* in future scenarios during (a) 2061-2080, (b) 2081-2100

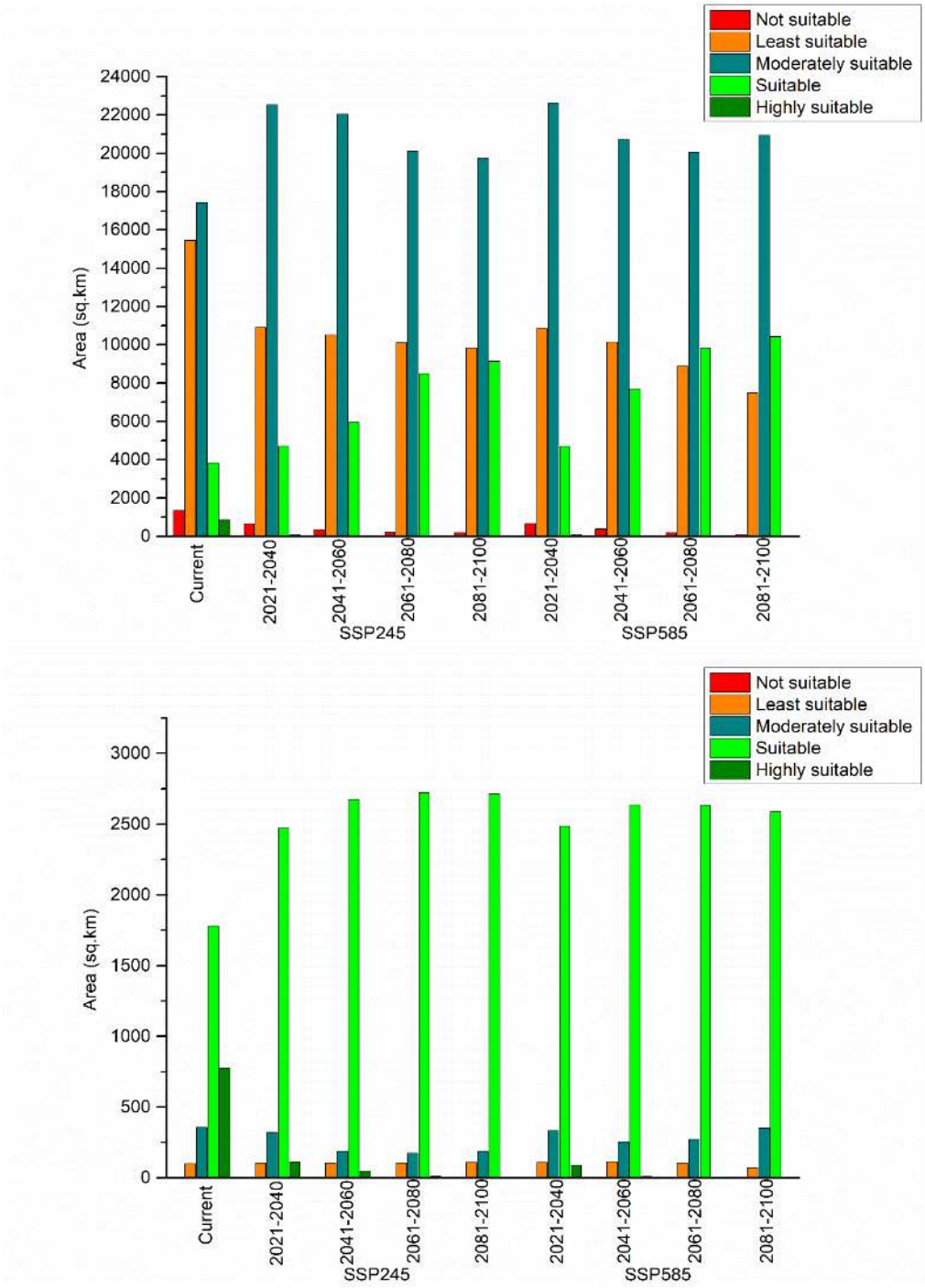

Fig. 13.tif Comparison of area change between current and future scenarios of *C. nathorstii* in (a) KE, (b) KD

Figure 13

Comparison of area change between current and future scenarios of *C. nathorstii* in (a) KE, (b) KD

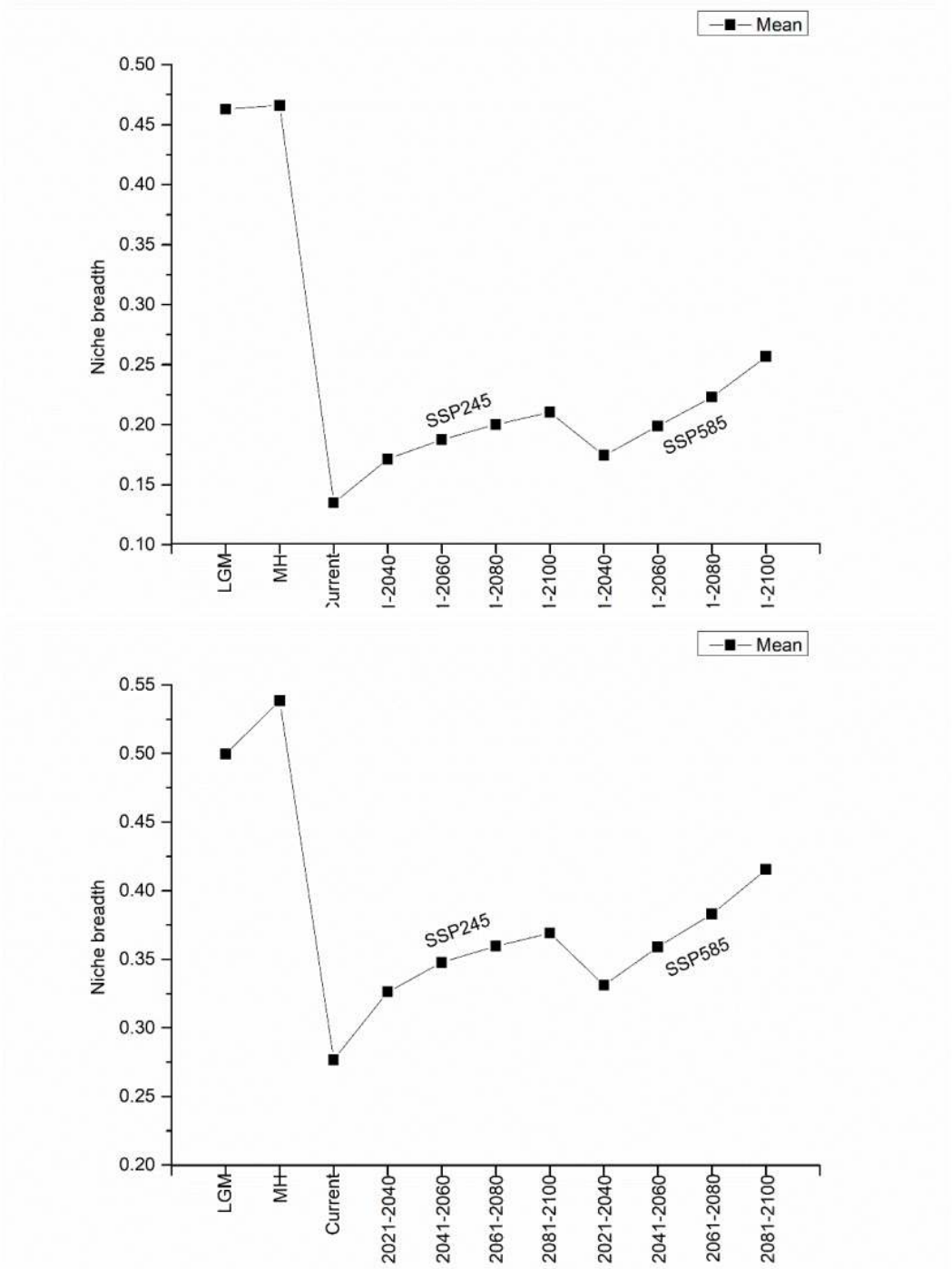

Fig. 14.tif Niche breadth in paleo, current and future scenarios of *C. circinalis* in (a) KE, (b) KD

Figure 14

Niche breadth in paleo, current and future scenarios of *C. circinalis* in (a) KE, (b) KD

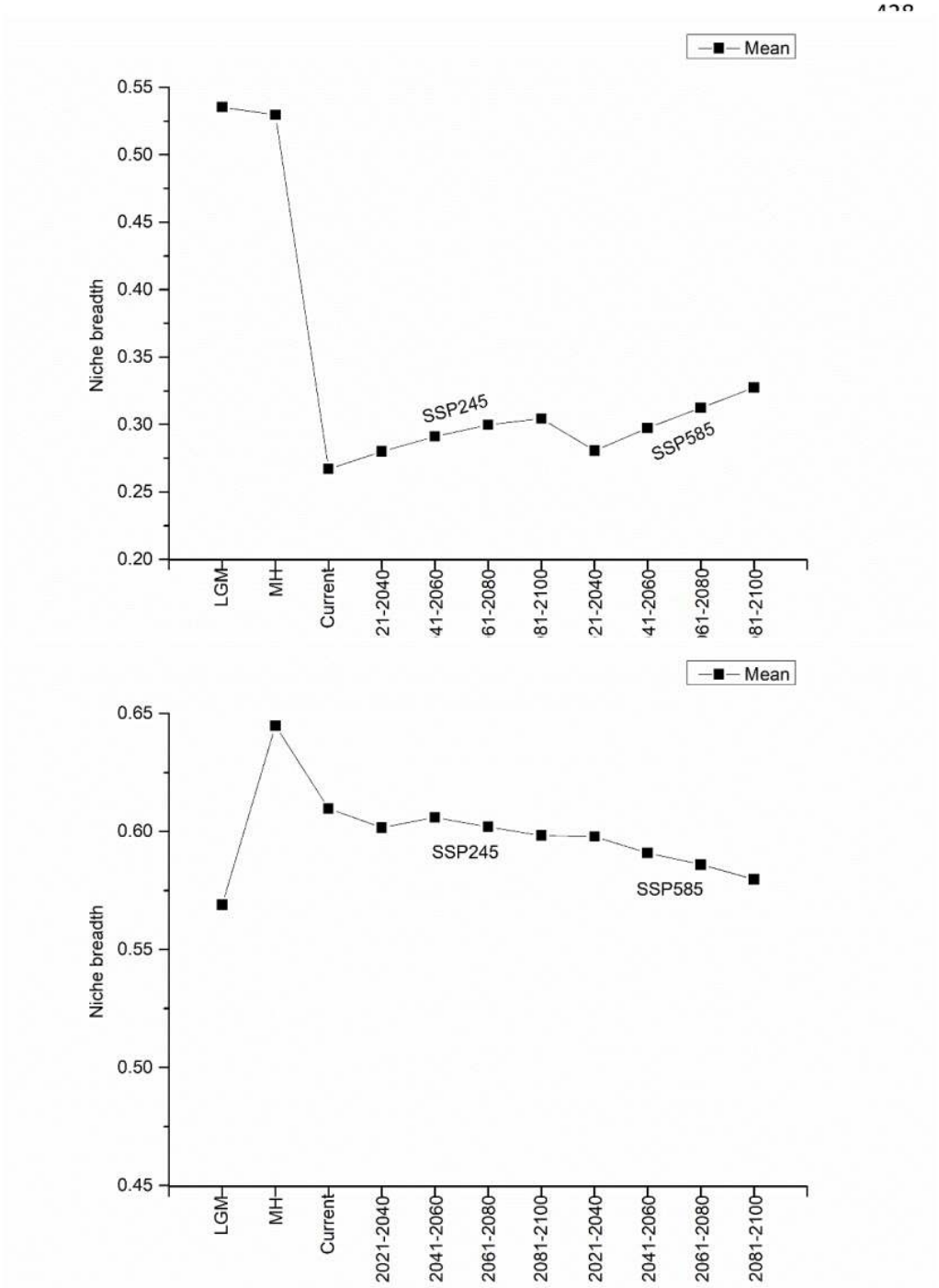

Fig. 15.tif Niche breadth in paleo, current and future scenarios of *C. nathorstii* in (a) KE, (b) KD

Figure 15

Niche breadth in paleo, current and future scenarios of *C. nathorstii* in (a) KE, (b) KD

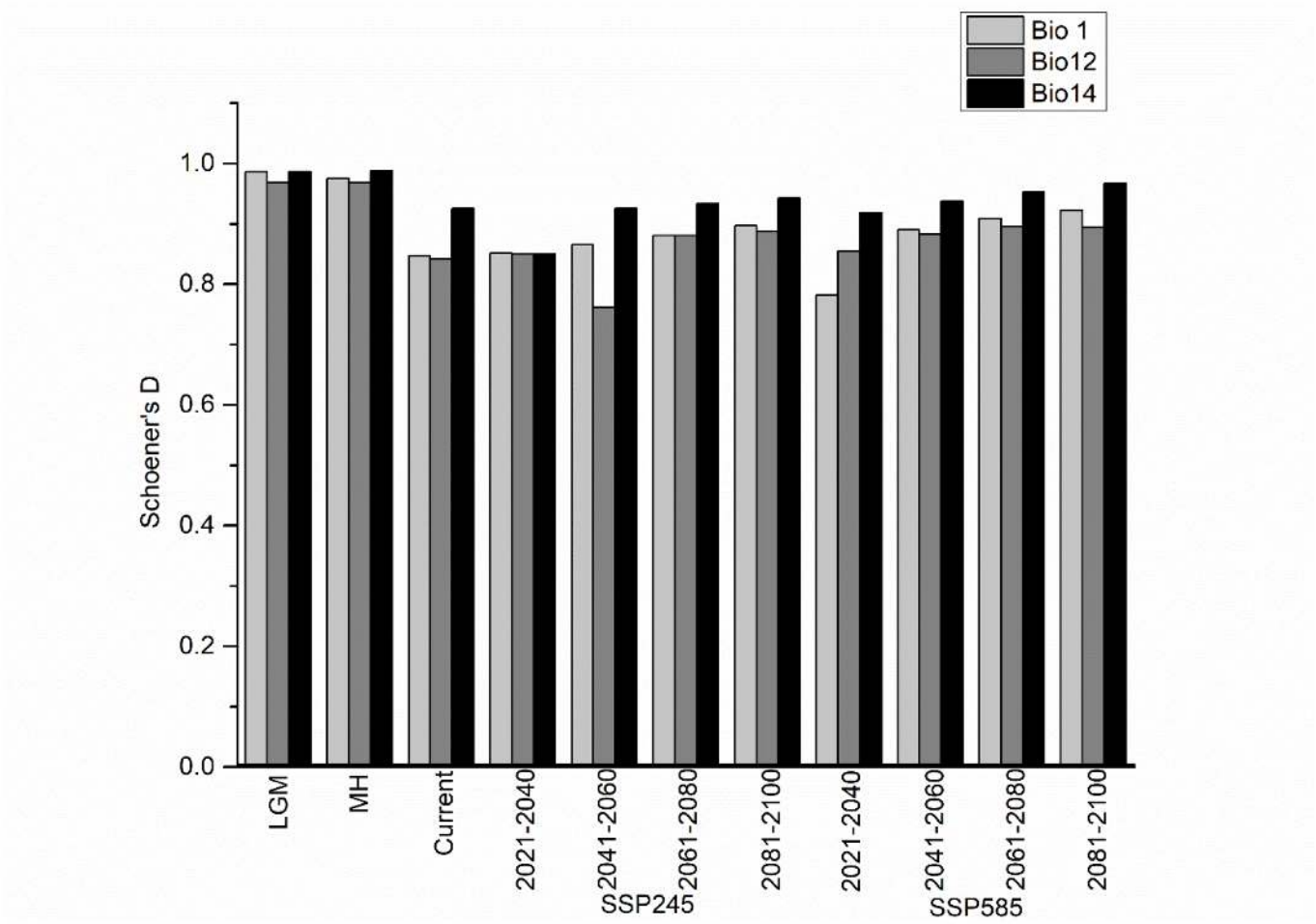

Fig. 16.tif Niche overlap between both species in KE

Figure 16

Niche overlap between both species in KE