P. ISSN: 0971-8184 II E. ISSN: 0976-1926
DOI: 10.61949/0976-1926.2024.v37i02.15

Indian Journal of Plant Genetic Resources

RESEARCH ARTICLE

Comparative Assessment of Different Earth System Models for Habitat Suitability of *Cuminum cyminum* (Linn.) Crop: A Machine Learning Evaluation from Arid and Semi-Arid Hot Areas of the India

Manish Mathur[1] and Preet Mathur[2*]

Abstract

Crop species modeling is limited to prime crops like wheat, rice, maize, soyabean, etc., and few attempts have been made for spies' crops grown in arid and semi-arid regions. This study examines the habitat suitability of *Cuminum cyminum* (cumin), a crop grown in Rajasthan and Gujarat, India. The widely used WorldClim dataset and CMCC-BioClimInd bioclimatic dataset were tested for predictive and elucidative power. This evaluation calculated AUC and omission rate using the MaxEnt entropy method. The WorldClim dataset includes three-time frames and four greenhouse gas scenarios. The CMCC-BioClimInd dataset included five Earth System Models (ESMs): GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, and NorESM1-M, which included two emission scenarios (4.5 and 8.5). The results indicate that both data sets have similar predictive accuracy. However, the optimal predictive areas for this crop differed significantly between the two model types. Annual precipitation, precipitation seasonality, precipitation during the coldest quarter, and potential evapotranspiration Hargreaves are the main factors affecting this crop's growth and expansion into new regions. Our research offers a novel standpoint for the implementation of this crop in potential new areas within Rajasthan and Gujarat.

Keywords Agricultural Production and Innovation, *Cuminum cyminum,* Climate Action, CMCC-BioClimInd, Maxent, Precipitation, Species Distribution Modelling.

[1]ICAR-Central Arid Zone Research Institute, 342003, Jodhpur, India.

[2]Jodhpur Institute of Engineering and Technology, Jodhpur, India.

***Author for correspondence:**

preetm9535@gmail.com

Received: 08/03/2024 **Revised:** 30/05/2024

Accepted: 05/06/2024

How to cite this article: Mathur M and P Mathur (2024) Comparative Assessment of Different Earth System Models for Habitat Suitability of *Cuminum cyminum* (Linn.) Crop: A Machine Learning Evaluation from Arid and Semi-Arid Hot Areas of the India. *Indian J. Plant Genet. Resour.* 37(2): 316-340. **DOI:** 10.61949/0976-1926.2024.v37i02.15

Introduction

India is the world's largest producer, exporter, and consumer of spices, leveraging its rich spice cultivation history. India has about 63 spices, 20 of which are seed spices. Cumin, fenugreek, coriander, and fennel are grown in India. A popular spice is cumin, scientifically known as *Cuminum cyminum* L. and it contributes Rs 2884.80 crore to the annual export value of seed spice crops, which total Rs 19505.81 crore. India satisfactorily meets 50% of global demand, demonstrating its importance in the global market (Sharma *et al.,* 2018; Kumar *et al.,* 2023). Rajasthan and Gujarat produce over 80% of India's cumin. Despite a smaller scale, Uttar Pradesh, Delhi, Uttarakhand, Madhya Pradesh, Chhattisgarh, and West Bengal have also grown this aromatic spice. Over a decade (2011–2020), this crop production in Gujarat and Rajasthan showed significant spatial differences. Cumin cultivation, production, and productivity are highest in Surendranagar, Gujarat. Banaskantha, Patan, Kutch, Rajkot, Junagadh, Porbandar, and Ahmedabad follow. Barmer had the most cumin land in Rajasthan, followed by Jodhpur, Jalore, Nagaur, Jaisalmer, and Ajmer (Kumar *et al.,* 2023).

Cumin, a small annual herbaceous plant in the Apiaceae (Umbelliferae) family (Fig. 1A and B), is susceptible to frost damage,

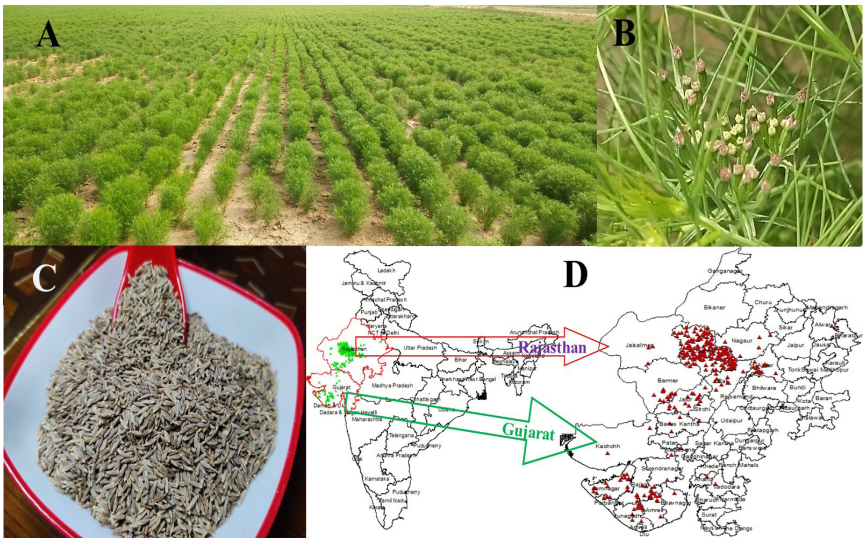


Fig. 1: Crops of *Cuminum cyminum* at vegetative stage (a), *C. cyminum* at fruiting stage (b), *C. cyminum* seeds (c), and spatially thinned sampling locations at Rajasthan and Gujarat states of India (D)

especially during flowering and seed formation. Aroma and pharmaceutical uses are valued in its seeds (Fig. 1C). The cumin plant, known for its resilience in arid climates, thrives in tropical or subtropical climates. The growth phase lasts 100–120 days. This species grows best at 20–30°C and 30–270 cm of annual precipitation. Sowing is best between mid-November and the first week of December. These climatic influences on cumin crops require scientific inventories to determine the reactions of various climatic models that predict the spatial and temporal fluctuations in the suitability of habitats for this crop, which is mostly found in Rajasthan and Gujarat.

Species Distribution Modelling (SDM) predicts and explains the geographical parameters that help species survive in the face of climate change (Mathur and Mathur, 2023). Scholars are using many global bioclimatic index datasets due to climate data accessibility (Morales-Barbero and Vega-Alvarez, 2019). The WorldClim dataset (Fick and Hijmans *et al.,* 2017; Poggio *et al.,* 2017) is a leading global climate dataset used in many fields. It is useful in ecological hydrology, climatology comparisons, carbon stock and temporal variation evaluation, and niche modelling of endangered plant species.

The dataset contains 19 bioclimatic variables related to temperature and precipitation, along with historical, current, 2050, and 2070 temporal frames. Additionally, the dataset includes Representative Concentration Pathways (RCPs) for Green House Gas Scenarios. The WorldClim dataset offers four spatial resolutions from 30 seconds (~1 km^2) to 10 minutes (~340 km^2), making it valuable. However, the limited scope of the 19 bioclimate variables makes it difficult to simulate plant and organism distribution because flora growth and proliferation require considerations beyond precipitation and temperature. As shown, some regions have high evaporation despite high precipitation. Thus, plants transfer water through various mechanisms. Climate change is increasingly projected using Earth system models (ESMs) that combine atmospheric and oceanic dynamics with cryosphere and biosphere models (Asch *et al.,* 2021). Future climate data depend on CO_2 emission scenarios, which are usually represented by representative concentration pathways (RCPs) or shared socioeconomic pathways (SSPs) that represent optimistic and pessimistic global technological, economic, and social development (Mathur and Mathur, 2023).

Noce *et al.* (2020) introduced the CMC-BioClimInd, a 35-variable bio-climatic indicator with a 0.5° grid resolution, for five Earth System Models (ESMs) in CMIP5: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, NorESM1-M. This new data set covered 1960–1999 and 2040–2079 and 2060–2099. The inclusion of spatial information that is missing from the WorldClim dataset allows for the precise assessment of climate change's impact on the protection and management of fauna, flora, and ecological assets across various spatial scales and domains of study. The CMCC-BioClimInd can quickly and easily determine the correlation between the species under investigation and the climatic variables, mitigating fluctuations from the future trajectory and the physical attributes of various indicators. Noce *et al.* (2020) examined the interrelationships between climate variables in two models, CMCC-BioClimInd and WorldClim, and the causes of their differences. CMCC-BioClimInd predicts species distribution and simulates climate change with high accuracy, comprehensiveness, and efficacy.

In SDM analysis, bioclimatic model types like GCMs and ESMs, as well as RCPs and SSPs, cause predictive variabilities in area and production. These factors, alone or together,

shape SDM analysis results through complex interactions. According to our knowledge, CMCC-BioClimInd and WorldClim data have not been used to predict spice crop spatial extents. The current study compares WorldClim and CMCC-BioClimInd's RCPs and time frames on *Cuminum cyminum* crops habitat suitability in India's arid and semi-arid regions. Five ESMs, two time periods (2050 and 2070), and four Green House Gas Scenarios (GHS)—RCP 2.6, 4.5, 6.0, and 8.5 of WorldClim dataset were associated with two horizons. Both models were compared to WorldClim version 2 bio-climatic predictors.

Materials and Methods

Distribution data-set

Distributional records for this species were obtained from data repositories such as GBIF (https://www.gbif.org), published literature like Chandawat *et al.* (2008); Meena *et al.* (2012); Verma and Kumar (2015); Kant *et al.* (2017); Mehriya and Ramesh (2018); Pagaria and Sharma (2019) as well as information collected from Thasildar (revenue) office of Jodhpur (20221) and Jalor (2023), outreach activities carried out by different Kissan Vigyan Kendras (KVKs) like Amereli http://www.jau.in/attachments/SeedVillage/KVK-Amreli.pdf; Khaat http://www.jau.in/attachments/SeedVillage/KVK-Khapat.pdf;

Targhadia http://www.jau.in/attachments/SeedVillage/KVK-Targhadia.pdf and Jamnagar http://www.jau.in/attachments/SeedVillage/KVK-Jamnagar.pdf) of Junagarh Agricultural University worked under Gujarat seed village programe (http://www.jau.in/index.php/extension-40/seed-village-programme) and KVKs of Barmer (https://www.aujodhpur.ac.in/kvk-ma.php?office=9) operated under Agriculture University, Jodhpur. The coordinates of sites/village mentioned in published literature as well as government sources were identified through google earth and projected them with GIS ArcMAP on a WGS84 coordinate system (Coban *et al.,* 2020). In order to lessen the effects of spatial autocorrelation and eliminate redundant entries, we employed the spatial thin window of the "Wallace Software", a Graphical User Interface based on the R programming language (Kass *et al.,* 2018), with a thinning distance of 10 kilometres.

Bio-Climatic Variables

CMCC-BioClimInd is available in full from PANGAEA.1 (https://doi.pangaea.de/10.1594/PANGAEA.904278?format=html#download). There are 805 NetCDF4 files with a 0.5 × 0.5° grid resolution and global coverage (excluding Antarctica) (Noce *et al.,* 2020). As mentioned, we used five ESMs: CMIP5: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, NorESM1-M, and two RCPs: 4.5 and 8.5 for 2040-79 (60) and 2060-99 (80). We used ArcGIS 10.2 spatial analysis function to extract the simulated values of 35 bioclimatic indicators for India and clipped the data set using geoprocessing (Wang *et al.,* 2023). Observational data from WorldClim version 2.0 (Fick and Hijmans, 2017) predicted species distributions. 19 bioclimatic variables (Hijmans *et al.,* 2005) were downloaded at 30 seconds (~1 km^2) and converted to ASCII (or ESRI ASCII) in DIVA-GIS version 7.5 (Coban *et al.,* 2020) for current and two future climatic scenarios (2050- and 2070-time frames, respectively;

Table 1: Variables Importance Values of different bio-climatic variables studied with specific models using Maxent tool

Bioclimatic Models		Bio-Climatic Variables (BC)												
		1	2	3	4	6	7	11	12	15	19	27	29	34
Current		8.9	2.5	0.7	5.1	0.2	0.9	3.9	21	40.2	16.2	-	-	-
2050 RCPs	2.6	9.4	7.4	1.3	1.6	4.5	0	6.8	23	27.7	18.9	-	-	-
	4.5	10	7.5	1.1	1.3	6.9	0.7	1.3	19	41.9	10.8	-	-	-
	6	10.9	7.1	3.7	0.1	6.6	0.8	6.5	16	27.5	20.9	-	-	-
	8.5	8.6	6.3	1.5	2.2	5.7	0.6	4.9	18	36.3	16.4	-	-	-
2070 RCPs	2.6	1.9	5.5	0.1	2.5	3.5	0.4	0.8	9.9	64.2	11.2	-	-	-
	4.5	4.9	6.7	1.3	1.7	8.7	1	3.2	14	38.7	20.1	-	-	-
	6	15	7.9	1.9	1.5	0.9	0	6.2	17	27.3	22.7	-	-	-
	8.5	3.1	8.7	0.1	1.8	9.3	0.6	1.2	14	43.4	17.4	-	-	-
GFDL	45 (60)	8.3	0	0.2	5.5	0.8	0.2	1.1	10	7	41.4	2.1	7	16.3
	45 (80)	4.2	0	0.2	5.7	4.5	0.1	0.5	13	6.6	35.3	2.3	1.1	26.6
	85 (60)	10	0.6	0.4	2.4	0.3	0.9	7.6	20	13.2	21	1.4	1.9	20.4
	85 (80)	12.3	0.3	0.4	5.8	0.1	0.6	4.3	16	1.9	34	0.2	3.4	20.4

HADGAM	45 (60)	1.6	0.4	0	1	0.8	0.9	1	13	58.5	13.2	0.4	1.9	7.6
	45 (80)	4.5	0.4	0	0.7	0.4	0.2	1.6	16	58.9	5.3	1.2	1	10.1
	85 (60)	15.1	0.2	0	1.4	3.5	0.8	1.1	18	41.6	5.6	0.7	0.5	11.1
	85 (80)	3.9	0.3	0.3	0.1	4.7	0.3	0.6	13	21.4	36.1	0.3	3.3	16.3
IPSL	45 (60)	10.2	0	2.5	5.1	0.8	3.3	0.3	14	21.5	30.9	3.7	0.3	7.3
	45 (80)	14.9	1	0.1	4.6	0	0.4	1.3	16	16.8	29.2	0.7	4.5	10.7
	85 (60)	1	0.3	0.4	5.3	3.8	2.5	1.1	20	21.1	10.7	26.1	2.9	4.5
	85 (80)	12.1	0.5	0.9	3.1	1.4	1	17	30	16.9	0.6	3.9	1	11.6
MIRCOS	45 (60)	7.2	0.7	0	0.7	1.1	0.7	1.2	6.1	21.8	37.3	3.1	1.1	19.7
	45 (80)	3.4	0.1	0.1	6.8	1	0.9	1.7	9.5	11.2	38.4	1.2	4.7	21
	85 (60)	7.2	0	0.1	1.1	5	0.2	1.3	7.1	37.5	19.3	0.4	0.5	20.4
	85 (80)	2.7	0.3	0.2	2.4	0.2	0.6	0.7	8.2	32.9	21	0.4	11	19.6
NORSEM	45 (60)	2.1	2.9	0	0.8	2.3	2.1	5.8	3.8	44.7	24.3	0.1	0.8	10.3
	45 (80)	3.5	0.7	0	1	3.7	2.7	5.7	6.3	41.1	22	0.5	2.3	10.5
	85 (60)	3.6	0.2	0.8	0.8	0.5	0.2	1.1	3.3	73	8.8	0.7	2	5.6
	85 (80)	3.6	0.2	0.9	2.4	1.3	0.8	7.6	4.7	54.8	17	0.4	0.5	5.8

BC-1 Annual mean temperature; BC-2 Mean diurnal range; BC-3 Isothermality; BC-4 Temperature seasonality; BC-6 Minimum temperature of coldest month; BC -7 Temperature annual range; BC-11 Mean temperature of coldest quarter; Bc-12 Annual precipitation; BC-15 Precipitation seasonality; BC-19 Precipitation of coldest quarter; BC-27 Simplified continentality index; BC-29 Mean temperature of coldest month and BC-34 Potential Evapotranspiration Hargreaves

Zhang *et al.,* 2022). RCPs 2.6, 4.5, 6.0, and 8.5 are represented by these datasets. The four RCPs range from high (RCP 8.5) to low (RCP 2.6) planned concentrations. In RCP 2.6, aggressive mitigation and low emissions reduce greenhouse gas (GHG) concentrations. The maximum emission scenario is RCP 8.5, while intermediate scenarios are RCP 4.5 and 6.0. In GHG concentration pathways, radioactive forcing (global energy imbalance) stabilizes 2.6, 4.5, 6.0, and 8.5 W/m^2 (Chaturvedi *et al.,*2012). The units and mathematical expressions for each bio-climatic parameter are in supplementary Table 1.

The climate during this crop's cultivation greatly affects its distribution and production. Instead of using a multi-collinearity test, we only consider bio-climatic variables that occur during cultivation, have an annual influence, and follow this rationale. So, 13 variables were chosen. annual mean temperature (BC-1), mean diurnal range (BC-2), isothermality (BC-3), temperature seasonality (BC-4), minimum temperature of coldest month (BC-6), temperature annual range (BC-7), mean temperature of coldest quarter (BC-11), annual precipitation (BC-12), precipitation seasonality (BC-15), precipitation of coldest quarter (BC-19), simplified continentality index (BC-27), mean temperature of coldest month (BC-29), and potential evapotranspiration Hargreaves

Projection transformation

It is critical to standardize the projections of the Bio-Climatic variables obtained from disparate sources and at varying resolutions prior to extracting data and generating forecasts using the MaxEnt model. Our analytical approach included ArcMap and ArcToolbox to follow a predefined methodology. The projection delineation was originally explained in the "projection and transformation" section of the Data Management Tools interface (Jijon *et al.,* 2021). We used the World Geodetic System 1984 EASE Geographic Coordinate System (GCS) for this. We converted the habitat class raster file projections to WGS 1984 web Mercator (auxiliary sphere-3857) in order to use Arc Map's "calculate geometry" window to quantify area under each habitat suitability class (see below).

Species Distribution Modelling

The present study used Maxent 3.4.1 (http://www.cs.princeton.edu/schapire/Maxent/) to simulate and predict *C. cyminum* plausible geographic distribution likelihood using bio-climatic models of different timeframes and RCPs. This tool's discrete execution with each predictor in isolation allows us to accurately measure their impact on the species' distributional pattern. During modeling, 70% of this crop's spatially thinned distribution data were randomly selected as training data and 30% as testing data. The random background point count was 10,000 (Zhang *et al.,* 2021). We set the regularization multiplier to 0.1 to avoid test data overfitting. (Phillips *et al.,* 2006). Linear, quadratic, and hinge properties were used. The environment parameters were configured using the Jackknife method, while the rest were left at their software defaults. threshold-independent receiver-operating characteristic (ROC) analyses and an

area under the receiver operating curve (AUC) were used to calibrate and validate Maxent model evaluation and estimate model prediction accuracy (Elith *et al.,* 2006). The performance of the model was classified as failing (0.5-0.6), poor (0.6-0.7), fair (0.7-0.8), good (0.8-0.9), or excellent (0.9-1). The closer the AUC value to 1, the farther away from the random distribution, the greater the correlation between environmental variables and the predicted geographical distribution of species, and the more accurate the performance of the model (Mathur *et al.,* 2023).

The effects of bio-climatic variables on species distribution were assessed using Variable Importance values and response curves (Mathur and Mathur, 2023). We then used ArcGIS to convert the Maxent output ASCII file into raster format and used this to categorize the habitat areas as optimal (1.0 to 0.80), moderate (0.80 to 0.60), marginal (0.60 to 0.40), low (0.40 and 0.20), and absent (< 0.20) for this species.

Area -Yield Estimation

A simple regression analysis with Minitab (2016) software was used to compare the projected areas and their effects on cumin yield under the optimum habitat type. The relationships between village/site area (ha.) and cumin average yield (kg/ha.) were examined. For both parameters, Rajasthan and Gujarat government offices provided 225 site data sets. Raster file transformation into Keyhole Markup Language (KML) configuration allowed comparative visualization of projected changes in the optimum habitat type. Current and specific bio-climatic model KML files were superimposed, transformed in ArcMap, and saved as tiff files (Mathur and Mathur, 2023).

We used WorldClim dataset with two RCPs, 4.5 and 8.5, with 2050- and 2070-time frames and five ESMs with both RCPs and time frames to study the impact of model types and individual use of WorldClim and ESMs on area under different habitat suitability classes.

Results

Models performances

Through a comprehensive analysis of diverse sources, we found 387 species occurrence sites. After using Wallace Software's spatial thin window feature (Kass *et al.,* 2018), we eliminated all but one instance of a record within a specific region using a 10-kilometre thinning distance. SDM development was completed by including 300 spatially autocorrelation-free *C. cyminum* presence points (Fig. 1D).

Omission-commission plots show cumulative threshold selection's effect on anticipated area and sample point autocorrelation (test and training). The test sample omission rate should match the projected rate (Mathur *et al.,* 2023). With most predictors, test omission lines are well matched or above with projected omission (Supplementary Figs 1 to 7), indicating no sampling autocorrelation and no sampling bias in our model attributes. The area under the receiver operating curve (AUC) was used to evaluate the Maxent model's performance for predicting distribution of this species. Result revealed that that this machine learning method performs excellently (AUC > 0.94) with all the studied predictors and their AUC curves are depicted in Supplementary Figs 8 to 14.

Variable Importance and Response curves

Table 1 shows the variable importance value of bio-climatic variables with different climatic model datasets and RCPs, and Supplementary Figs 15 to 29 show the response curves of the most influential two variables. WorldClim models for current, 2050, and 2070 and future RCPs scenarios showed that BC-12 (annual precipitation), BC-15 (precipitation seasonality), and BC-19 (coldest quarter precipitation) were the most important variables for this species' habitat suitability, with precipitation seasonality being the most dominant. In this study, the response curve for precipitation seasonality showed that cumin presence increases up to 150 mm, so this species will survive better in intermediate precipitation conditions. Annual precipitation (BC-12) response curves with current, 2050 (RCP 2.5, 4.5, 8.5) also indicated that the area receiving 450 to 550 mm rainfall had the highest probability of its occurrence. Precipitation during the coldest quarter (BC-19), the second most influential factor for this crop in 2050 (RCP 6.0) and 2070 (RCP 2.6, 4.5,6.0, and 8.5), and their response curves show that this species prefers 5-8 mm of rainfall.

With other ESMs, like NORSEM, HADGAM [45(60), 45 (80), 85(60)] and MIRCOS [(85(60) and 85(80)], precipitation seasonality was also recorded as most influential factor. Among these ESM, NORSEM and MIRCOS showed the identical response curves for these variables as with WorldClim data set, however occurrence probability of cumin is lesser than that of with WorldClim data set. Interestingly, HADGAM 45 (60) and 45 (80) exhibited very less range for peak occurrence probability as compared to other models. Similarly, precipitation of the coldest quarter (BC-19) was found to be most influential variable with GFDL-ESM as well as with IPSL [45(60), 45(80) and 85(60)] and with MIRCOS-ESM particularly with 45(60) and 45 (80) and their response curve range are similar to WorldClim data set, however, the peak of occurrence probability is smaller with these ESMs. With GFDL, MIRCOS, NORESM and HADGEM 45 (80) and 85 (80), potential evapotranspiration Hargreaves was also recorded as influential factor for occurrence probabilities of this species.

Area (km^2) of different habitat suitability class with various bio-climatic models.

Spatial extents of different habitat suitability classes identified for *C. cyminum* with help of MaxENT modelling

are depicted in Fig. 2 (with current and 2050 time-frames its four RCPs), Fig. 3 (2070 time-frames its four RCPs), Fig. 4 (GFDL ESM), Fig. 5. (HADGAM ESM); Fig. 6 (IPSL ESM); Fig. 7 (MIRCOS ESM) and Fig. 8 (NORSEM ESM). Based on cell values in raster output processed with ArcMap, we can categorize four types of suitability classes with 0.20-point break and these were designated as optimum, moderate, marginal and low (Mathur and Mathur 2023) and the area under these classes (km^2) are depicted in Table 2.

Graphical representation of spatial extent of optimum suitable habitats was found to be higher in Rajasthan state compared to Gujrat state under current and 2050 and 2070 WorldClim data set. Within these data-set, single continuous larger area was observed with current climatic conditions which is de-fragmented into several small populations extending at both the states (Fig. 2 and 3) with future time-frames. Among the such data-set, higher area under optimum (89623.33 km^2; 16.64 Per cent optimum cumin area with reference to total geographical areas of Rajasthan and Gujarat), and moderate (104171.7 km^2) suitability classes were recorded with 2050 RCP 8.5 and 4.5, respectively. While lower areas under these suitability classes were also recorded with same time-frame (2050) but with RCP 6.0 and 2.6, respectively. Interestingly, results revealed the area under the optimum class are higher with RCP 8.5 with both the future time-frame in comparison to other RCPs scenarios (Table 2) and with such result it can be said that this crop will show good performance against higher greenhouse gas conditions. Higher areas under marginal (108433.9 km^2) and low (91999.91 km^2) class were recorded with current bio-climatic conditions (Fig. 2, Table 2).

Similarly, with ESMs, higher area under optimum (246962.1 km^2; 45% of total geographical study areas) and moderate (186496.4 km^2) habitat classes were recorded with MIRCOS45(60) ESM (Fig. 7), while the lowest areas under these classes (134068.9 km^2 and 76626 km^2) were recorded with GFDL85(60; Fig. 4) and HADGSM 85(80; Fig. 5), respectively. Higher (125005.7 km^2) as well as lower (54959.87 km^2) areas under marginal class were recorded with NORESM85(60) and NORESM 45 (80) ESM, respectively (Fig. 8). This analysis indicated the decrease of non-suitable areas for this crop with other ESM compared to WorldClim data set encompassing both current and two future time-frames and four respective RCPS. Further, range of per cent geographical areas of cumin under optimum habitat was recorded higher (24.90 to 45.86) with five studied earth system models, in comparison to WorldClim (10.38-16.64) data sets (Table 2).

Area -Yield Estimation

Regression analysis yield significant linear relationships between between area of villages/sites (Ha.) and cumuin average yield (kg/ha.) and this relaitonships (Supplementary figuere 30) can be equate as

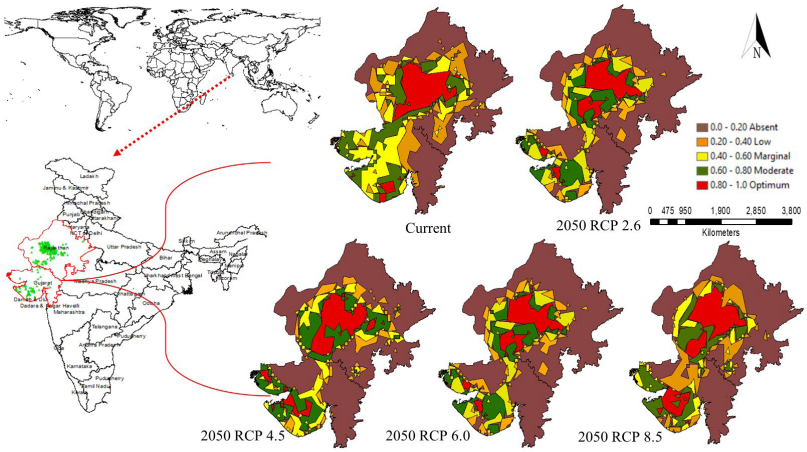


Fig. 2: Habitat suitability of *C. cyminum* with current and 2050 time-frames its four RCPs

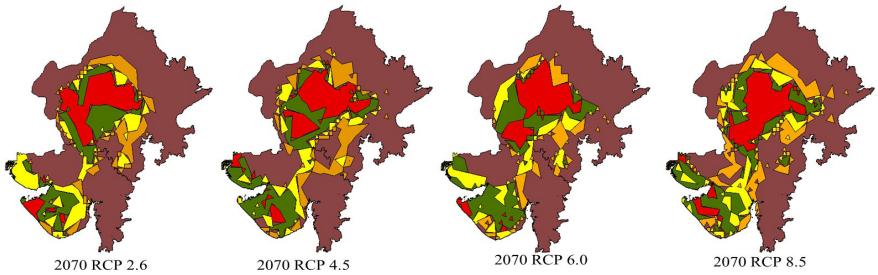


Fig. 3: Habitat suitability of *C. cyminum* with current and 2070 time-frames its four RCPs

Table 2: Area (km^2) under different habitat suitability classes calculated with different bio- climatic models. Total geographical area of Rajasthan (342239) and Gujarat (196244) is 538483 Km^2

Bio-Climatic Models	*Optimum*	*Moderate*	*Marginal*	*Low*	*Per cent of optimum cumin area with reference to total geographical area*
WorldClim Current	72827.01	97553.31	108433.9	91999.91	13.52
2050 RCP 2.6	58283.43	76823.04	72802.89	54599.65	10.82
2050RCP 4.5	77621.71	104171.7	65469.76	47075.38	14.41
2050 RCP 6.0	55911.12	102284.6	76407.01	77864.04	10.38
2050 RCP 8.5	89623.33	86964.15	51211.94	83803.57	16.64
2070RCP2.6	74349.43	77433.97	69356.89	53908.02	13.81
2070 RCP4.5	74145.89	94801.41	48913.03	63967.24	13.77
2070 RCP6.0	74500	94360.5	60744.76	45558.55	13.84
2070 RCP 8.5	84050.36	89353.67	59695.58	78311.14	15.61
GFDL45(60)	173676	109573.2	87584.69	145716.1	32.25
GFDL45(80)	178782.5	97022.43	69942	185083.4	33.20
GFDL85(60)	134068.9	173401.4	76822.86	130838	24.90
GFLD85(80)	170104.7	153232.6	79503.79	94690.18	31.59
HADGSM 45(60)	155075.8	144953.3	71598.04	119254.3	28.80
HADGSM 45(80)	141438.4	123043.7	83335.18	130652.2	26.27
HADGSM 85(60)	176950.4	133939.3	85391.74	73551.69	32.86
HADGSM 85(80)	198980.2	76626	97975.61	121631.6	36.95
IPSL45(60)	184419.9	102888.2	95343.54	103708	34.25
IPSL45(80)	203160	106134.1	90498.76	100448.1	37.73
IPSL85(60)	160706.6	146080.3	95123.81	95123.81	29.84
IPSL85(80)	219438.6	97785.15	97856.31	121072.6	40.75
MIRCOS45(60)	246962.1	186496.4	112656.9	56140.29	45.86
MIRCOS45(80)	197463.4	108102.4	89920.51	115849.4	36.67
MIRCOS85(60)	183633.3	109833.8	87886.21	111462.4	34.10
MIRCOS85(80)	164490.5	110256.4	118375.8	143026.3	30.55
NORESM45(60)	217094.2	149136.3	114966.2	79519.77	40.32
NORESM45(80)	209111.3	148829.2	54959.87	106591.8	38.83
NORESM85(60)	155563.1	109752.6	125005.7	102631.7	28.89
NORESM85(80)	188765.7	141083	86779.29	56786.85	35.06

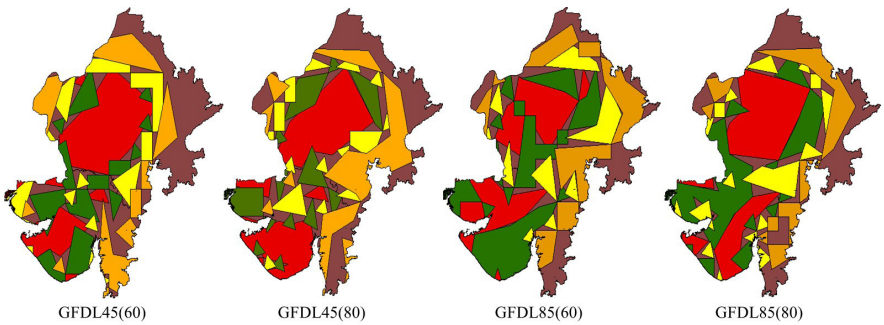


Fig. 4: Habitat suitability of *C. cyminum* with GFDL earth system model its four RCPs

Table 3: Two-way Analysis of Variance for area (km^2) of cumin with different model number settings

ESMs models and their RCPs	*ANOVA Variables*	*Area*	
		Model	*Habitat types*
Both Model	F calculated	2.94*	26.54*
	F critical	1.62	2.71
CMCC-BioClimInd only	F calculated	0.29^{NS}	36.77*
	F critical	1.77	2.76
WorldClim only	F calculated	1.42^{NS}	6.5*
	F critical	2.35	3

*and Bold number represents the significant values at 5% level; NS = non-significant

Table 4: Two-way Analysis of Variance for area (km^2) for examining the effect of EMS and RCP 4.5 and 8.5)

Habitat Suitability	*Area (Calculated F values)*	
	ESMs	*RCP (4.5 and 8.5)*
	F calculated	*F calculated*
Optimum	9.9*	5.05*
Moderate	2.39^{NS}	4.77^{NS}
Marginal	1.79^{NS}	1.0^{NS}
Low	3*	3.54^{NS}

F critical for EMS = 2.81 and 4.84 for RCP; *and Bold number represents the significant values at 5% level; NS = non-significant

$$Area\ of\ villages\ or\ sites\ (ha) = 515.4 + 0.026 \times cumin\ average\ yield\ (Kg\ per\ ha)\ R^2 = 0.75^{*}\ n = 225$$

Comparative results are depicted in Figs 9 to 15 pertaining to 2050 RCPs, 2070 RCPs, GFDL, HADGEM, IPSL, MIRCOS and NORSEM ESMs, respectively. With current scenario, maximum optimum areas come under Jodhpur, Pali, Ajmer, Jalor, Nagaur and Parts of Barmer districts of Rajasthan and Amreli, Junagadh and parts of Jamnagar and Bhavnagar (Fig. 9). With 2050 RCP 2.6, we found -19.57 per cent reduction in area and -15.71 in yield as compared to current. With this specific time frame and RCP, reduction was mostly observed at Jodhpur, Pali and Amreli and Junagadh areas. However, some new areas of optimum habitat are occurred at Kachchh and Jamnagar areas of Gujarat and Barmer and Jalor and Bikaner areas of Rajasthan (Fig. 9a). With RCP 4.5 of the same time-frame, 6.57 and 5.18 per cent increase in area and yield were recorded. Major gain was recorded at Kachchh, Jamnagar, Rajkot, Surendra Nagar of Gujarat and Barmer and Jalor areas of Rajasthan (Fig. 9b). Similar to RCP 2.6 of this time-frame, -23.23 and -18.27 per cent reduction were recorded in area and yield, respectively, with RCP 6.0. Specifically, with this time frame, highest area and yield gain +23.06 and +18.14, respectively were recorded with RCP 8.5 (Fig. 9d) and new areas were majorly occurred at Rajkot, Jamnagar, Surendranagar of Gujarat state and Bikaner, Jaisalmer and Jalor districts of Rajasthan.

Opposite to 2050, current study revealed the gain in area and yield with all four RCPs belongs to 2070 time-frame (Fig. 10 a

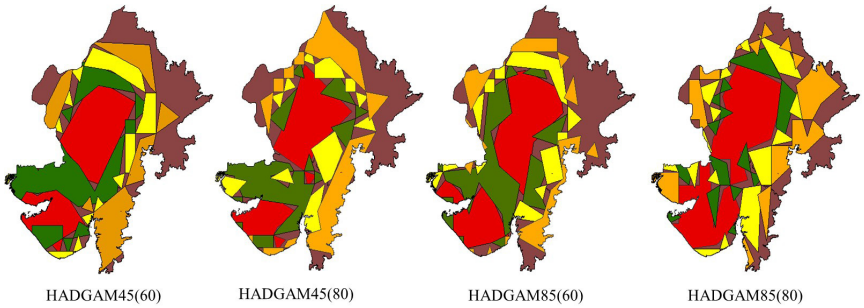


Fig. 5: Habitat suitability of *C. cyminum* with HADGAM earth system model its four RCPs

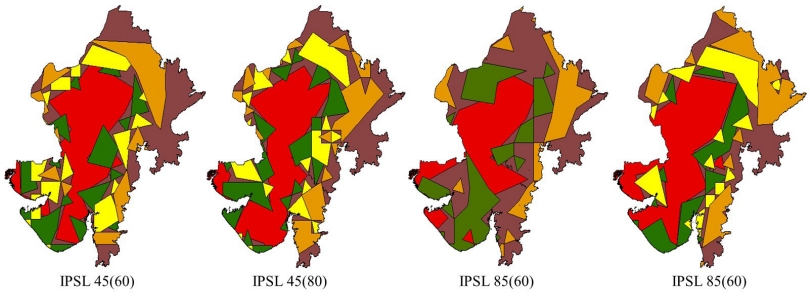


Fig. 6: Habitat suitability of *C. cyminum* with IPSL earth system model its four RCPs

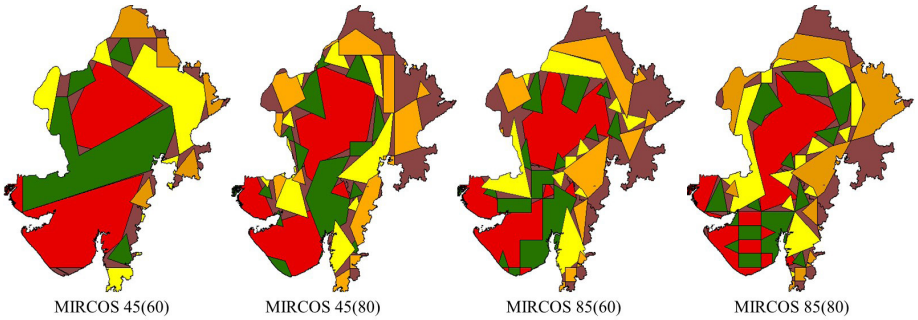


Fig. 7: Habitat suitability of *C. cyminum* with MIRCOS earth system model its four RCPs

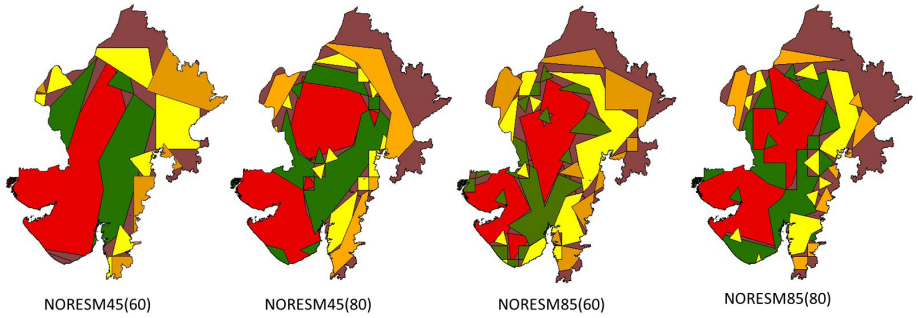


Fig. 8: Habitat suitability of *C. cyminum* with NORSEM earth system model its four RCPs

to d) and the maximum + 11.29 and + 8.86 gain in area and yield, respectively was recorded with RCP 8.5, and such gains were highlighted at Jamnagar, Rajkot, Kachchh and Banas Kantha districts of Gujarat and Barmer and Jalor district of Rajasthan. With GFDL-ESM, except 85 (60) more than twice the area gain was recorded (Fig. 11 a to d) and among them, maximum area +145.49 and excepted yield +114.45 were recorded with RCP 4.5 (Fig. 11b). Such new areas are covering most parts of the Gujarat covering entire Jamnagar, Rajkot, Amrelli, Bhavnagar, parts of Surendranagar and Sabar Kantha, while at Rajasthan it covering most parts of Barmer, some parts of Bikaner, Nagaur, Ajmer and Pali districts. Similar higher gain +173.22, +136.27 in area and yield, respectively was recorded with HADGEM 85 (80 Fig. 12d). However, compared to GFDL 45 (80), additional extension of optimum habitat was recorded at Kachchh, Banas Kantha areas of Gujarat and Bikaner, Jaisalmer, Sirohi and Rajsamand districts of Rajasthan. With IPSL 85(80) +201.31 % increase in area and +158.37 % increase in yield (Fig. 13d) were recorded and such area increase were observed both at Gujarat (Jamnagar, Rajkot, Kachchh, Surendranagar, Patan, Banas Kantha and Gandhinagar) and Rajasthan (entire Barmer and Jalor, Pali and Jodhpur district and some parts of Jaisalmer and Sirohi). Among all studied ESM, 239.11% area gain and 188.10% gain in yield were recorded with MIRCOS45(60) covering entire areas of Jamnagar, Junagadh, Amreli, Bhavnagar, Rajkot, Surendranagar, Anand, Gandhinagar, Sagar Kantha, Kheda, and some parts of Patan and Bharuch in Gujarat state (Fig. 14a) and similarly areas of Jodhpur, and parts of Barmer, Nagaur, Pali, Ajmer and Jaisalmer. Similarly, results of NORSEM ESM are depicted in Fig. 15 a to d.

ANOVA analysis suggested significant variabilities in areas brought by use of combination of both model types only, while their individual use were non-significant for areas. In comparison to this, impacts of habitat types with different setting of model numbers were statistically significant with all variables (Table 3). Similar statistical analysis was also carried out to examine the impact of ESMs and two RCPs namely 4.5 and 8.5 on area under different habitat suitability classes. ANOVA result revealed the statistically significant result for area under optimum habitat with both types of climatic data-sets and their RCPs, while area under moderate and marginal habitats these predictors were non-significant (Table 4).

Discussion

Climate-based geographic crop modeling is essential for predicting the best cultivation sites and yield. This step helps sustain food production (Alsafadi *et al.,* 2023). This study examined how climatic projections affect cumin's current and future geographical distribution in India's arid and semi-arid regions. This study compares four greenhouse gas scenarios (RCP 2.6, 4.5, 6.0, and 8.5) using the WorldClim data set for two future time frames and five ESMs, CMIP5: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, and NorESM1-M, from the recently

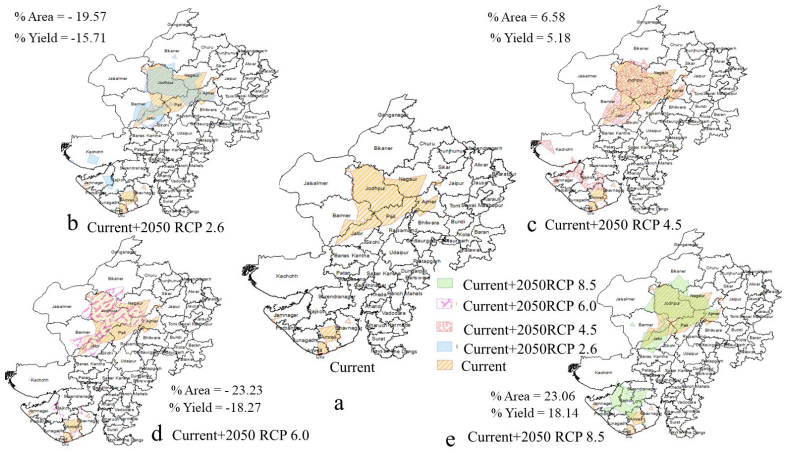


Fig. 9: Superimposition of current optimum suitability sites with different RCPs of 2050 time-frames

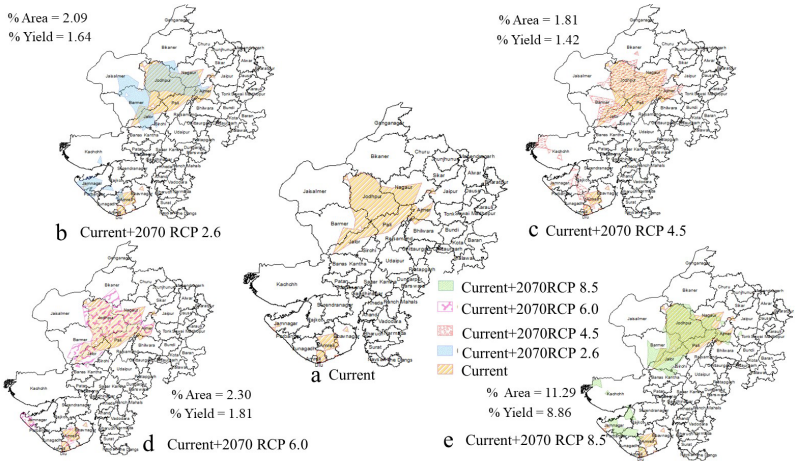


Fig. 10: Superimposition of current optimum suitability sites with different RCPs of 2070 time-frames

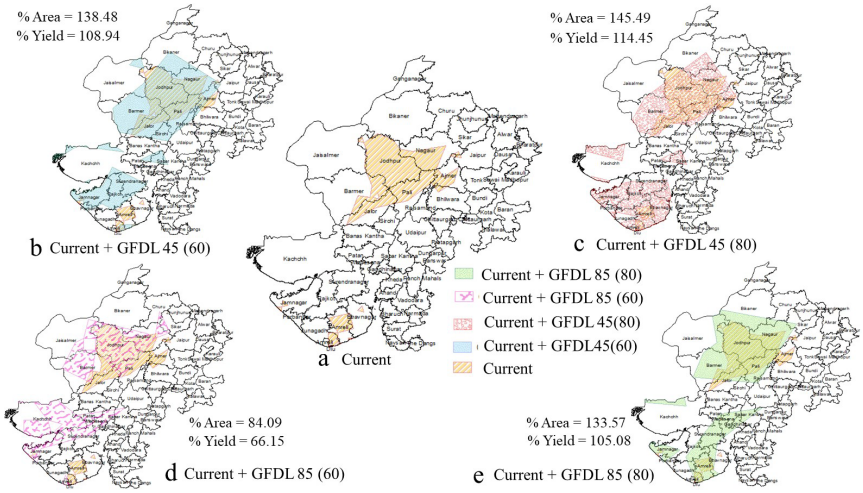


Fig. 11: Superimposition of current optimum suitability sites with GFDL ESM

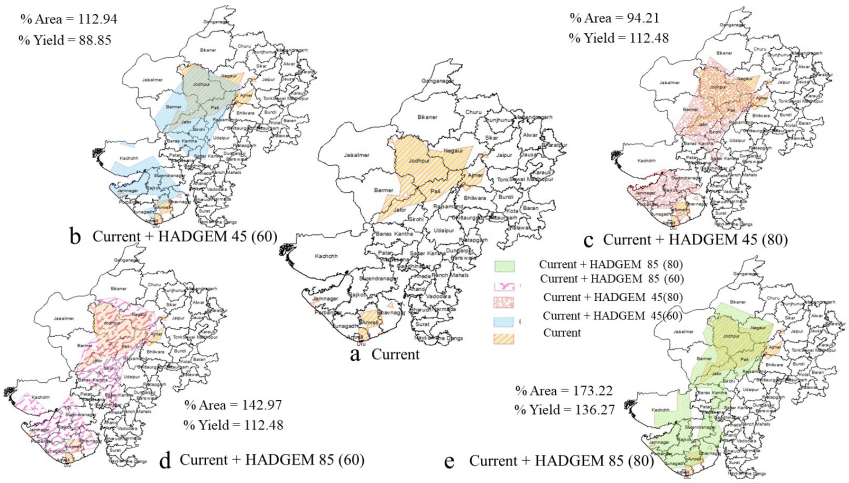


Fig. 12: Superimposition of current optimum suitability sites with HADGEM ESM

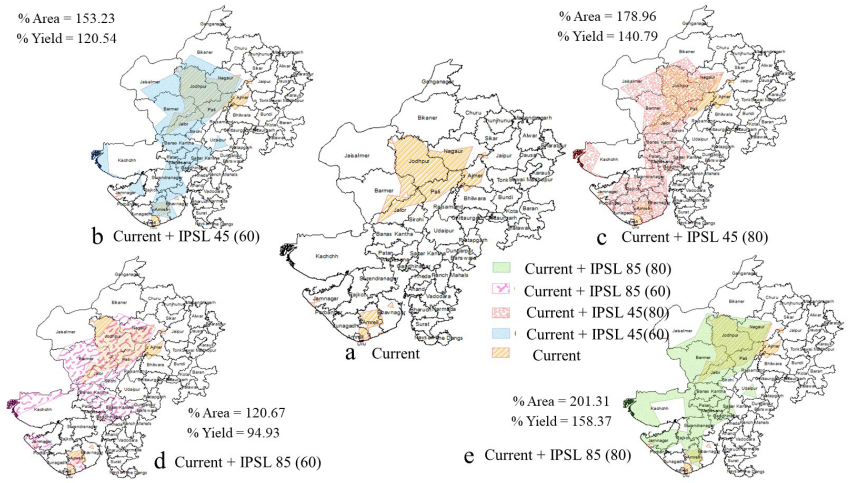


Fig. 13: Superimposition of current optimum suitability sites with IPSL ESM

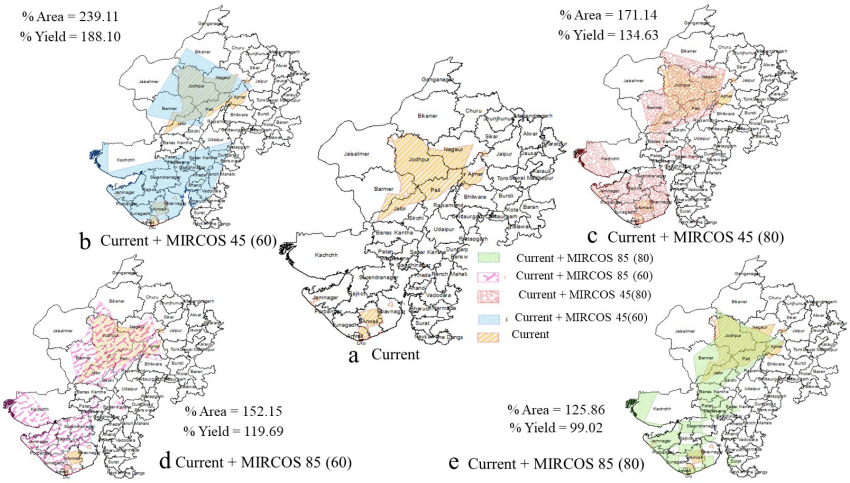


Fig. 14: Superimposition of current optimum suitability sites with MIRCOS ESM

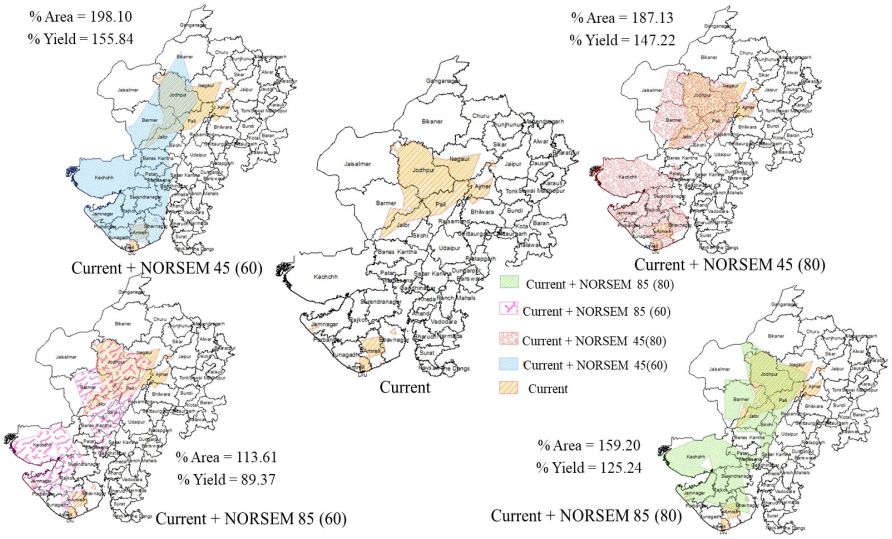


Fig. 15: Superimposition of current optimum suitability sites with NORSEM ESM

developed BioClimInd dataset for two RCPs, 4.5 and 8.5, of two-time frames, 2040-79 (60) and 2060-99 (80). Kogo *et al.* (2019a) used HAdGEM2-ES and CCM4 GCM models with RCP 4.5 and 8.5 to compare the current and future maize distribution in Kenya. CCSM4 showed minor differences. Zhang *et al.* (2017), Xu *et al.* (2021) examined how different climatic models affect maize in China, eastern Africa, and respectively. Kogo *et al.* (2019b) examined how climate change affects maize growth and productivity across regions. They found several effective general circulation models (GCMs): ECHAM5, CCSM, HadCM3, CSIRO-MK3, CGCM3.1, UKLO, and MIROC3.2. They also recommended multi-model ensemble GCMs for crop modeling to improve accuracy and reliability. In contrast, Alsafadi *et al.* (2023) used three models from the Coupled Model Intercomparison Project Phase 6 (CMIP6): BCC-CS2-MR (Beijing), CanESM5 (Canadian model), and IPSL-CM6A_LR (France) to simulate the effects of predicted climate change on wheat cultivation in Syria's semi-arid southwest. Their study found that the Beijing model was more accurate than the others, especially for wheat production in the examined area. Koo *et al.* (2017) recommend climatic multi-models with diverse temporal frameworks to reduce climate pattern ambiguity. Liu *et al.* (2023) used five GCMs and three CO_2 levels for *Calligonum mongolicum*, an arid shrub. GCMs and emission scenarios introduce uncertainty, which they discuss.

The dataset WorldClim 2.1 is widely cited in academic literature. This dataset contains 19 bio-climatic variables with varying resolutions. These variables cover recent climate conditions from 1970 to 2000 and projected future climate scenarios for 2021-2040, 2041-2060, 2061-2080, and 2081-2100. WorldClim uses 9 Generative Circulation and Earth System Models to project future climate. They are part of the Coupled Model Intercomparison Project Phase 6 (CMIP6), which compares and evaluates climate models globally. Note that WorldClim has limitations. A major drawback is the lack of spatial details needed by models to accurately assess climate change impacts. This limitation limits the dataset's ability to predict and account for climate patterns and processes' spatial heterogeneity (Zhang *et al.,* 2023a). When considering spatial heterogeneity, a medium-resolution grid cell can encompass a climate environment with hundreds of meters of elevation difference (Poggio *et al.,* 2017). The complex process of plant growth requires considerations beyond precipitation and temperature, which may limit the 19 bioclimate variables' ability to simulate plant and organism distribution (Mathur and Mathur, 2023). As shown, evaporation increases in some regions despite heavy precipitation.

The CMCC-BioClimInd dataset improves community-valued spatial information accessibility in two ways. First, it includes 35 historical and future bioclimatic indicators. Second, it uses models and other analytical methods to ensure climate change impact assessments are reliable and accurate, taking into account uncertainties.

In this study, under most predictors and according to the interpretation criterion, our omission lines are well matched or above with projected omission, indicating no sampling autocorrelation and no sampling bias in our model attributes. Perhaps this is because we cleaned the spatial locations of 387 cumin cultivation sites across the study area using a spatial auto-correlation filter before performing our SDM on a subset of 300 of these sites.

Receiver operating characteristic (ROC) test is threshold-independent and does not require thresholds for presence versus absence predictions (Peterson *et al.,* 2007). The area under the curve (AUC) of ROC measures a model's ability to distinguish between species-rich and species-poor sites

(Elith *et al.,* 2006). The machine learning method excels in this study (AUC > 0.94; Mean 0.97±0.01) for all studied predictors. Chao-Yun *et al.* (2012) reported similar SDM results for Asian Piper nigrum spice. Zhang *et al.* (2023b) used the AUC tool to compare the WorldClim and CMCC-BioClimInd datasets efficacy. The researcher classified their study using numerical bio-climatic variables. These variables included 19 WorldClim variables, CMCC-BioClimInd (1–19 bio-climatic predictors), CMCC-BioClimInd (1–35 bio-climatic predictors), and CMCC. 11 global invasive species were studied using these classifications, yielding average AUC values of 0.948, 0.951, 0.963, and 0.969. They found that the simplified continentality index (the divergence between the mean values of the most hot and cold months across a given span of years) and modified Kira warmth index (energy inputs in the warmest portion of the year), derived from CMCC-BioClimInd, were effective explanatory tools.

Response curves have helped quantify the logistic probability of species presence and environmental factors. This has greatly improved our understanding of the species' complex ecological niche (Mathur *et al.,* 2023). This study found that WorldClim dataset annual precipitation (BC-12), seasonal precipitation (BC-15), and coldest quarter precipitation (BC-19) are key factors in cumin species habitat suitability. This analysis included the present, 2050-, and 2070-time frames and four RCP scenarios. Precipitation seasonality had the greatest impact on cumin species habitat suitability.

Precipitation seasonality measures and analyzes monthly precipitation variability over a year. The index is the ratio of monthly aggregate precipitation standard deviation to mean precipitation (Noce *et al.,* 2020). In fragile ecological systems like arid and semi-arid regions, crop productivity depends on precipitation timing and quantity during the growth cycle. This study stressed the importance of precipitation during the growth cycle. Numerous researchers emphasize the importance of precipitation in rainfed wheat cultivation (Aixia, 2022; Zhang, 2022). Murugan *et al.* (2022) examined how precipitation seasonality affects cardamom distribution in southern India under RCP 4.5 and 8.5. Their findings show that a significant change in precipitation seasonality negatively affects spice cultivation, maturation, and productivity. Buckland *et al.* (2022) reported similar effects of precipitation seasonality on Garcinia indica distribution in India, *Dalbergia latifolia* in Nepal, and *Opuntia ficus-indica* and *Euphorbia tirucalli* worldwide.

Response curves for three bio-climatic variables help us identify cumin's optimal growing conditions. These areas have a precipitation seasonality threshold of 150 mm (BC-15), an annual precipitation range of 450 to 550 mm (BC-12), and 5 to 8 mm of rainfall during the coldest quarter (BC-19). Precipitation seasonality was also the most influential factor for NORSEM, HADGAM [45(60), 45 (80), 85(60)], and MIRCOS [(85(60) and 85(80)], whose response curves were identical to WorldClim but had lower peak occurrence probabilities. GFDL, MIRCOS, NORESM, and HADGEM 45 (80) and 85 (80) also showed that potential evapotranspiration Hargreaves affected species occurrence probabilities. The temperature-based Hargreaves equation calculates potential evapotranspiration (Hargreaves *et al.,* 1985). Mean monthly maximum temperature - mean monthly minimum temperature for the month of interest and mean air temperature were included.

The cumin plant's physiological processes and secondary metabolite production are affected by many environmental stresses. These responses help the plant develop tolerance and survive harsh conditions (Pandey *et al.,* 2015). The physiological responses of *C. cyminum* to water stress include the synthesis of proline, soluble sugars, and essential oils (Kazemi *et al.,* 2018). Water-related variables in cumin cultivation have been extensively studied. Under experimental control conditions (Mehriya *et al.,* 2020), irrigation water and crop evapotranspiration (IW/Etc) increased plant height, branch count, and photosynthetic rate. Cumin seed yield increased significantly due to these improvements. Cumin crop yield attributes and increase are due to optimal precipitation and soil moisture levels. These conditions promote efficient nutrient uptake and a good soil-water air relationship with higher root zone oxygen (Mathur, 2013). The root zone's optimal moisture levels boost plant physiological activity and dry matter accumulation. The correlation between a consistent precipitation pattern throughout the year and cumin development during the colder season is linked to many complex physiological mechanisms. These mechanisms reduce leaching, increase photosynthetic rates, and help plants absorb essential nutrients. Due to these factors, cumin seed yield increases. Water-related climatic variables boost root development and soil moisture extraction. This phenomenon may promote abundant vegetative growth in water-rich conditions (Mathur and Sundaramoorthy, 2013). Based on the data, water-related factors rather than temperature-related factors determine the likelihood of this important crop in the studied reg

MaxENT modelling has revealed the spatial distribution of *C. cyminum* habitat suitability classes. Using the WorldClim dataset, the optimal class had greater spatial coverage under the RCP 8.5 scenario compared to other RCP scenarios and across future time frames. C. cyminum may benefit from the "CO_2 fertilization effect" and increase crop production (e.g., wheat; Alsafadi *et al.,* 2023) under high CO_2 conditions.

Our comparison of optimum habitat spatial extension and yield provided useful crop information. Currently, Jodhpur, Pali, Ajmer, Jalor, Nagaur, and parts of Barmer in Rajasthan and Amreli, Junagadh, and parts of Jamnagar and Bhavnagar are the best areas. These findings are supported

by Singh and Kumar (2016); Sharma *et al,* (2018);; Pagaria and Sharma (2020).

With comparative analysis, areas and yield fall under gain (+) or loss (-) with different RCPs and datasets could also be marked. Area and yield only decreased with 2050 RCP 2.6 and 6.0. MIRCOS45(60) again recorded maximum area and yield in Jamnagar, Junagadh, Amreli, Bhavnagar, Rajkot, Surendranagar, Anand, Gandhinagar, Sagar Kantha, Kheda, and some parts of Patan and Bharuch in Gujarat, and most of western Rajasthan, including Jodhpur, Barmer, Nagaur, Pali, and Jaisalmer.

ANOVA analysis of areas affected by models (both ESMs and WorldClim, ESMs only, and WorldClim only) and their RCPs and habitat types (optimum, moderate, marginal, and low) showed significant variabilities in areas caused by both model types, but not by their individual use. In contrast, habitat types with different model numbers affected all variables statistically (Table 3). The effects of ESMs and RCPs 4.5 and 8.5 on habitat suitability class area were also examined. ANOVA showed that ESMs and RCPs were significant for optimum habitat but not moderate or marginal habitats (Table 4).

This species cannot be grown in Rajasthan's irrigated northwest plain (Sriganganagar and Hanumangarh), sub-humid southern plain (Bhilwara, Udaipur, and Chittorgarh), humid southern plain (Dungarpur, Banswara, and Chittorgarh), or humid southeastern plain. The agro-climatic zonation of Gujarat shows which areas are best for cumin cultivation. Salve *et al.* (2017) reported rising cumin area and yield in Banaskantha district, Gujarat.

WorldClim and CMCC-BioClimInd models had excellent model accuracy, but their optimum areas were statistically different. Two-way ANOVA analysis is used in previous studies to identify the major cause of area uncertainty. Real *et al.* (2010) suggested that GCMs, not CO_2 emission scenarios, cause climate change modeling uncertainty. In their modeling, Hamit *et al.* (2018), and Zhu (2019) found that RCPs or SSPs caused such uncertainty, while GCMs/ESMs were ignored. Thus, climate change studies must assess the relative contributions of ESMs and RCPs/SSPs. Liu *et al.,* (2023) used two-way ANOVA to examine the significance of 5 GCMs and 3 SSPs for range dynamics of an arid shrub *Calligonum mongolicum*. Their analysis showed that GCMs had a significant impact on gain area, loss area, and total area, indicating that selecting these GCMs is uncertain for all aspects of range change. SSPs scenarios only affected gain and loss areas, not total area. It shows that SSP selection did not affect total area simulation but did affect gain and loss area estimation.

The current, 2050, and 2070 WorldClim data sets showed that Rajasthan had more optimally suitable habitats than Gujarat. Under current climate, a single continuous larger area was observed in these datasets, but it is expected to break up into several small populations across both states (Figs 2 and 3). The aforementioned analysis found fewer unsuitable regions for growing this crop using alternative Earth System Models (ESMs) than the WorldClim dataset. Two-way ANOVA analysis showed that using both model types (WorldClim + five EMS of CMCC-BioClimInd) together caused statistically significant area differences, while using either model type alone did not. However, different habitat types had statistically significant effects under different model numbers. Statistics were also used to examine how ESMs and two RCPs, 4.5 and 8.5, affected habitat suitability classes. In an ANOVA, both types of climatic data-sets and their RCPs showed statistically significant results for the area under optimum habitat, but not moderate or marginal habitats (Table 4).

Conclusion

With this research, we simulated the habitat suitability and estimated climate thresholds for *C. cyminum* (cumin) crop in arid and semi-arid parts of the India. We also projected changes in the suitable habitat (particularly optimum) under various climate change model and emission scenarios. The obtained climatically suitable habitat map and habitat change maps could be guiding us to introduce this species in new areas of Rajasthan and Gujarat. The obtained climate thresholds possess the potential to be utilized by policymakers in deducing the appropriateness of the habitat for cumin, thereby fortifying the management of said regions.

Supplementary Materials: Table 1 and figs 1 to 30

Acknowledgements

MM thankful to the Director, ICAR-CAZRI for giving approval to him for attending training on R-Programming that enhance his working capacity using ENM modelling techniques. PM (Jodhpur Institute of Engineering and Technology, Jodhpur, India) thankful to their Director for extending their academic help.

Competing interests

The authors declare that they have no competing interests

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data that support the findings of this study specifically geo-coordinates of the species are available on request from the corresponding author, [Preet Mathur]. The data are not publicly available due to avoid the duplication of the work within the same geographical area

Funding

The authors declare no specific funding for this work

References

Aixia R, Z Weifeng, S Anwar, L Wen, D Pengcheng, H Ruixuan and S Min (2022) Effects of tillage and seasonal variation of rainfall on soil water content and root growth distribution of winter wheat under rainfed conditions of the Loess Plateau China. *Agric. Water Manag.* 268: 107533

Alsafadi K, S Bi, HG Abdo, H Almohamad, B Alatrach, AK Srivastava, M Al-Mutiry, MAS Chandra and S Mohammed (2023) Modeling the impacts of projected climate change on wheat crop suitability in semi-arid regions using the AHP-based weighted climatic suitability index and CMIP6. *Geosci. Lett.* 10: 1-21

Asch RG, JM Holding, DJ Pilcher, S Rivero-Calle and KA Rose (2021) Ecological Applications of Earth System Models and Regional Climate Models. *Front. Mar. Sci.* 8: 1-5.

Buckland CE, AJAC Smith and DSG Thomas (2022) A comparison in species distribution model performance of succulents using key species and subsets of environmental predictors. *Eco. Eval.*12: 1-16.

Chandawat MS, MS Chouhan and RF Thakor (2008) Adoption of improved cumin production technology by the growers of Jalore district. *Guj. J. Ex. Edu.* 39: 80-82.

Chao-Yun H, F Rui, M Ribeiro, T Le-he, W Huo-song, Y Jian-feng, Z Wei-quan and Y Huan (2011) Modelling the potential geographic distribution of Black Pepper (*Piper nigrum*) in Asia using GIS Tools. *J. Integr. Agric.* 11 (4): 593-599.

Chaturvedi RK, J Joshi, M Jayaraman, G Bala and NH Ravindranath (2012) Multi-model climate change projections for India under representative concentration pathways. *Curr. Sci.* 103(7): 791-802.

Elith J, HC Graham, PR Anderson, M Dudik, S Ferrier, A Guisan, JR Hijmans, F Huettmann, R Leathwick and A Lehmann (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-151.

Fick SE and RJ Hijmans (2017) WorldClim 2: New 1–km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37: 4302–4315.

GBIF.org (Accessed on 2 February 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.t5duy6

Hamit S, N Abdushalih, X Li, H Shao, A Jiesisi and A Guli (2018) Effects of Climate Change and Human Activities on the Distribution Pattern of *Calligonum mongolicum* Turcz. *Arid Zone Res.* 35: 1450–1458.

Jijon JD, KH Gaudry, J Constante and C Valencia (2021) Augmenting the spatial resolution of climate-change temperature projections for city planners and local decision makers. *Environ. Res. Lett.* 16: 1-12.

Kant K, B Singh, YK Sharma, SR Meena, RK Solanki, BK Mishra, NK Meena and AU Siddiqui (2017) Diversity of nematodes population in cumin (Cuminum cyminum) crop in semi-arid regions of Rajasthan. *Indian J. Agri. Sci.* 87 (12): 80-83.

Kass JM, B Vilela, ME Aiello-Lammens, R Muscarella, C Merow and RP Anderson (2018) *Wallace*: A flexible platform for reproducible modelling of species niches and distributions built for community expansion. *Methods Ecol. Evol.* 9: 1151–1156.

Kazemi H, SMM Mortazavian and MG Javid (2118) Physiological responses of cumin (*Cuminum cyminum*) to water stress conditions. *Iranian J. Field Crop Sci.* 48(4): 1099-1113.

Kogo BK, L Kumar, R Koech and CS Kariyawasam (2019a) Modelling climate suitability for rainfed maize cultivation in Kenya using a maximum entropy (MaxENT) approach. *Agronomy* 9: 1-18.

Kogo BK, L Kumar, R Koech and P Langat (2019b) Modelling Impacts of Climate Change on Maize (Zea mays L.) Growth and Productivity: A Review of Models, Outputs and Limitations. *J. Geosci. Environ. Prot.* 7: 76-95.

Koo KA, SU Park, WS Kong, S Hong, I Jang and C Seo (2017) Potential climate change effects on tree distributions in the Korean peninsula: Understanding model and climate uncertainties. *Ecol. Modell.* 353: 17–27.

Kumar M, PK Sahoo, DK Kushwaha, S Gudi and G Singh (2023) Tackling the Constraints of Cumin Cultivation and Management Practices. *Ann. Agric. Sci.* 8(3): 1-14.

Kumar S, T Roshni, E Kahya and MA Ghorbani (2020) Climate change projections of rainfall and its impact on the cropland suitability for rice and wheat crops in the Sone river command, Bihar. *Theor. Appl. Climatol.* 142(1): 433–451.

Liu G, Y Zhang, Q Lu, K An, Y Li, D Xiong, G Li and S Du (2023). Uncertainties of Climate Change on Suitable Habitat of *Calligonum mongolicum* Turcz. (Polygonaceae) in Desert Areas of Central Asia. *Forests* 14: 1-14.

Mathur M, P Mathur and H Purohit (2023) Ecological niche modelling of a critically endangered species *Commiphora wightii* (Arn.) Bhandari using bioclimatic and non-bioclimatic variables. *Ecol. Process.* 12: 1-21.

Mathur M (2013) Kinetic of nutrient uptake and their utilization efficiency in a seratonious plant-*Blepharis sindica*. *Asian J. Bio. Sci.* 8 (1): 94 -106.

Mathur M and P Mathur (2023). Prediction of global distribution of *Ganoderma lucidum* (Leys.) Karsten: a machine learning maxent analysis for a commercially important plant fungus. *Indian J. Ecol.* 50 (2): 289-305.

Mathur M and S Sundaramoorthy (2013) Mineral concentration and their deviation from optimum percentage in *Tribulus terrestris*. *Indian J. Ecol.* 40(1): 19-23.

Mathur P and M Mathur (2023) Machine learning ensemble species distribution modelling of an endangered arid land tree *Tecomella undulata:* a global appraisal. *Arab. J Geosci.* 165: 131.

Meena ML, D Singh and NK Sharma (2012) Impact of front-line demonstrations on yield enhancement of cumin: a case in arid zone of Rajasthan. *Indian J. Ext. Edu.* 48 (1&2): 103-105.

Mehriya ML and Ramesh (2018) Impact of front-line demonstration (FLD) on cumin in Jodhpur and Nagaur district. *Chem. Sci. Rev. Lett.* 7 (26): 449-453.

Mehriya ML, N Geat, SH Sarita, MA Matttar and HO Elansary (2020) Response of drip irrigation and fertigation on cumin yield, quality, and water-use efficiency grown under arid climatic conditions. *Agronomy* 10: 1-17.

Minitab (2016) Statistical Software, Version 16. Minitab Incorporation, State College.

Morales-Barbero J and J Vega-Álvarez (2019) Input matters matter: Bioclimatic consistency to map more reliable species distribution models. *Methods Ecol. Evol.* 10: 212–224.

Murugan M, M Alagupalamuthirsolai, K Ashokkumar, A Anandhi, R Ravi, J Rajangam, MK Dhanya and KS Krishnamurthy (2022) Climate change scenarios, their impacts and implications on

Indian cardamom-coffee hot spots; one of the two in the world. *Front. Sustain. Food Sys.* 6: 1-16.

Noce S, L Caporaso and M Santini (2020) A new global dataset of bioclimatic indicators. *Sci. Data* 7:1-12

Pagaria P and S Sharma (2019) Production and marketing constraints for cumin see in Barmer district. *Int. J. Curr. Microbiol. App. Sci.* 8(3): 1828-1832

Pagaria P and S Sharma (2020) Trends in area, production and yield of cumin crops in Barmer district of Rajasthan. *Rashtriya Krishi* 15: 23-26.

Pandey S, MK Patel, A Mishra and B Jha (2015) Physio-Biochemical Composition and Untargeted Metabolomics of Cumin (*Cuminum cyminum* L.) Make It Promising Functional Food and Help in Mitigating Salinity Stress. *PLoS ONE* 10(12): 1-25

Peterson AT, M Papes and M Eaton (2007) Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography* 30: 550-560.

Phillips SJ, RP Anderson and RE Schapire (2006) Maximum entropy modeling of species geographic distributions. *Ecol. Modell.* 190 (3-4): 231–259.

Poggio L, E Simonetti and A Gimona (2017) Enhancing the WorldClim data set for national and regional applications. *Sci. Total Environ.* 625: 1628-1643.

Real R, AL Marquez, J Olivero and A Estrada (2010) Species distribution models in climate change scenarios are still not useful for informing policy planning: An uncertainty assessment using fuzzy logic. *Ecography* 33: 304–314.

Salve PD, RR Patel, RM Patel and AS Patel (2017) Cultivation and production cost of cumin in Banaskantha district of north Gujarat. *Int. Res. J. Agri. Eco. Stat.* 8 (1): 138- 142

Sharma SS, SS Rao, RS Singh, RP Sharma and PN Dubey (2018) Evaluation of potential cumin growing area in hot arid region of Jaisalmer district. *Int. J. Seed Spices* 8(1): 50-55.

Singh RB and A Kumar (2016) Agriculture dynamics in response to climate change in Rajasthan. *The Delhi Uni. J. Humanities Social Sci.* 3: 115-138.

Thasildar office Jodhpur (Rajasthan, Directorate of Economics and Statistics/TRS/DES/2021/657-636 dated 12.04.2020)

Thasildar office Jodhpur (Rajasthan, Directorate of Economics and Statistics/TRS/2022-23/3936-3945 dated 30.01.2023)

Verma VK and Kumar P (2015) Marketing behaviour of cumin in Jodhpur district of Rajasthan. *Eco. Aff.* 137-142.

Wang M, Z Hu, Y Wang and W Zhao (2023) Spatial Distribution Characteristics of Suitable Planting Areas for *Pyrus* Species under Climate Change in China. *Plants* 12: 1559.

Xu F, B Wang, C He, DL, Liu, P Feng, N Yao, R Zhang, S Xu, J Xue and H Feng (2021) Optimizing Sowing Date and Planting Density Can Mitigate the Impacts of Future Climate on Maize Yield: A Case Study in the Guanzhong Plain of China. *Agronomy* 11: 1-18.

Zhang L, F Wang, H Song, T Zhang, D Wang, H Xia and R Min (2022) Effects of projected climate change on winter wheat yield in Henan China. *J. Clean Prod.* 379: 134734

Zhang Y, J Tang, G Ren, K Zhaod and X Wang (2021) Global potential distribution prediction of *Xanthium italicum* based on Maxent Model. *Sci. Rep.* 1-11

Zhang F, C Wang, C Zhang and J Wan (2023a) Comparing the Performance of CMCC-BioClimInd and WorldClim Datasets in Predicting Global Invasive Plant Distributions. *Biology* 12: 1-16.

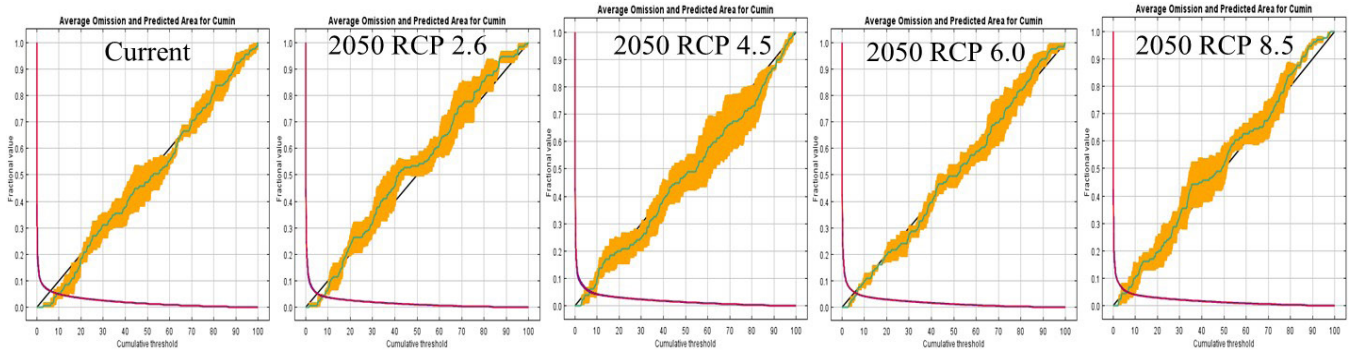
Zhang FX, CJ Wang, CH Zhang and JZ Wan (2023b) Using a new dataset of bioclimatic variables to predict invasive plant species distribution. *Biology* 12: 1-16

Zhang Y, Y Wang and H Niu (2017) Spatio-temporal variations in the areas suitable for the cultivation of rice and maize in China under future climate scenarios. *Sci. Total Environ.* 601: 518–531.

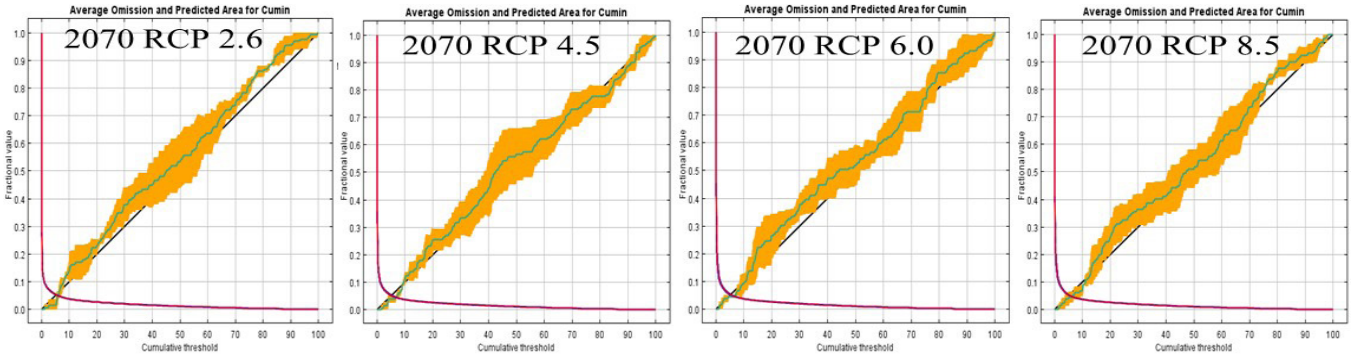
Zhu N (2019) Predicting the Geographic Distribution of *Calligonum mongolicum* under Climate Change. *J. Distribution Res.* 39: 136–144.

Supplementary Table 1: Details of 35 bio-climatic variables opted from five earth system models and WorldClim dataset. (Details of each bio-climatic variables can be found in Noce, S., Caporaso, L. and Santini, M. 2020. A new global dataset of bioclimatic indicators. Scientific Data, 7: 398 https://doi.org/10.1038/s41597-020-00726-5)

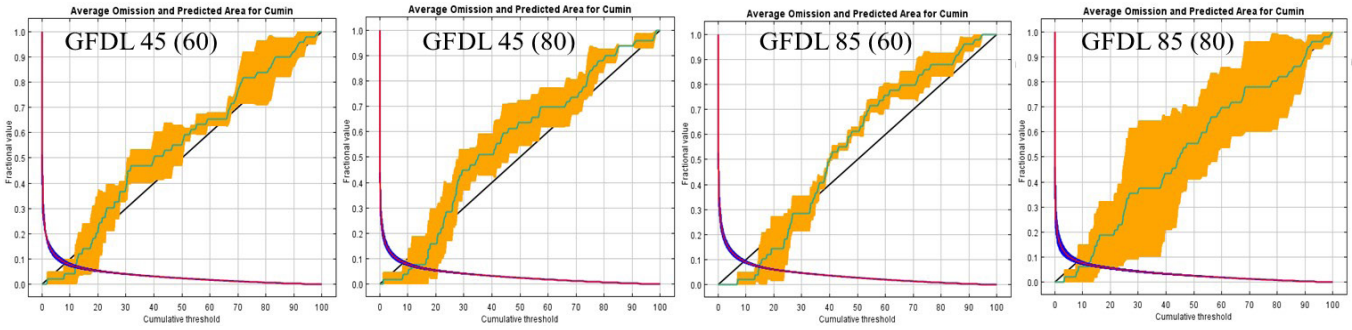
Code	*Name*	*Unit*
Bio1	Annual mean temperature	°C
Bio2	Mean diurnal range	°C
Bio3	Isothermality	%
Bio4	Temperature seasonality	°C
Bio5	Maximum temperature of warmest month	°C
Bio6	Minimum temperature of coldest month	°C
Bio7	Temperature annual range	°C
Bio8	Mean temperature of wettest quarter	°C
Bio9	Mean temperature of driest quarter	°C
Bio10	Mean temperature of warmest quarter	°C
Bio11	Mean temperature of coldest quarter	°C
Bio12	Annual precipitation	mm
Bio13	Precipitation of wettest month	mm
Bio14	Precipitation of driest month	mm
Bio15	Precipitation seasonality	%
Bio16	Precipitation of wettest quarter	mm
Bio17	Precipitation of driest quarter	mm
Bio18	Precipitation of warmest quarter	mm
Bio19	Precipitation of coldest quarter	mm
Bio20	Ellenberg quotient	°C/mm
Bio21	Yearly positive temperature	°C
Bio22	Sum of annual temperature	°C
Bio23	Ombrotermic index	mm/°C
Bio24	Yearly positive precipitation	mm
Bio25	Modified Kira coldness index	°C
Bio26	Modified Kira warmth index	°C
Bio27	Simplified continentality index	°C
Bio28	Mean temperature of warmest month	°C
Bio29	Mean temperature of coldest month	°C
Bio30	Mean temperature of driest month	°C
Bio31	Mean temperature of wettest month	°C
Bio32	Modified Thermicity index	°C
Bio33	Ombrothermic index of summer and the previous month	mm/°C
Bio34	Potential Evapotranspiration Hargreaves	mm
Bio35	Potential Evapotranspiration Thornthwaite	mm



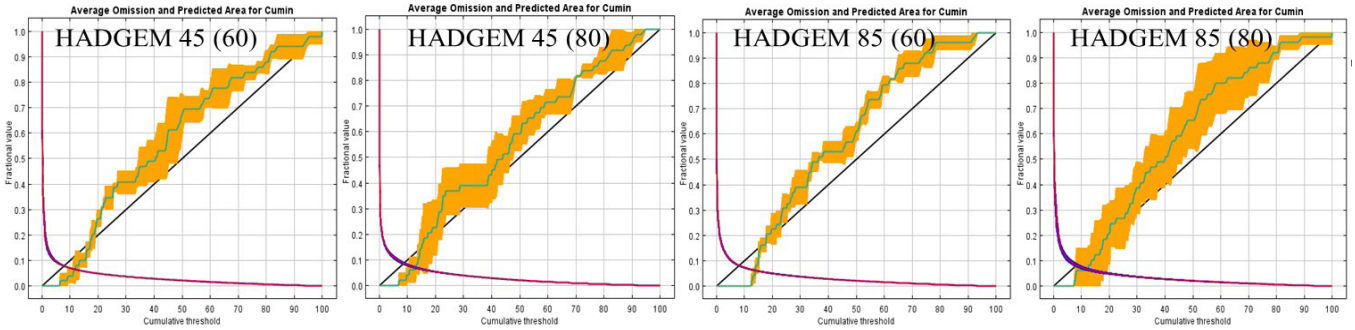
Supp. Fig. 1: Analysis of omission/commission with current and 2050 bio-climatic timeframe and its RCPs projection



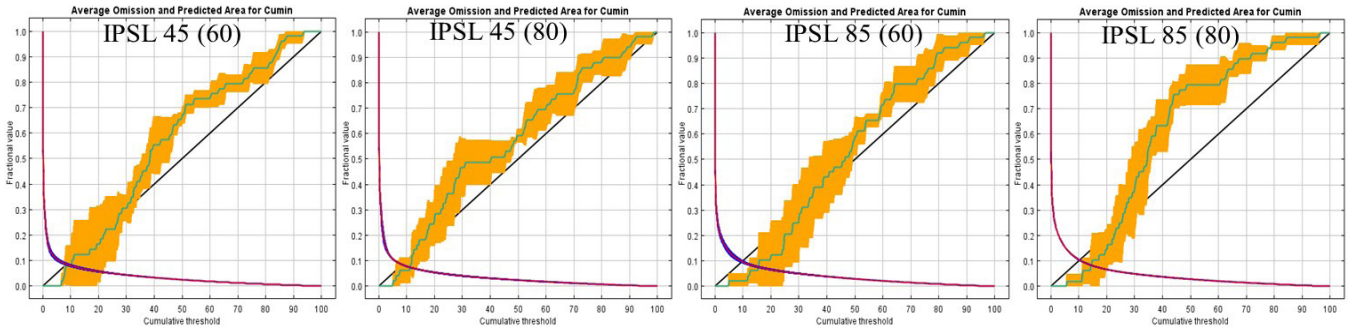
Supp. Fig. 2: Analysis of omission/commission with current and 2070 bio-climatic timeframe and its RCPs projection



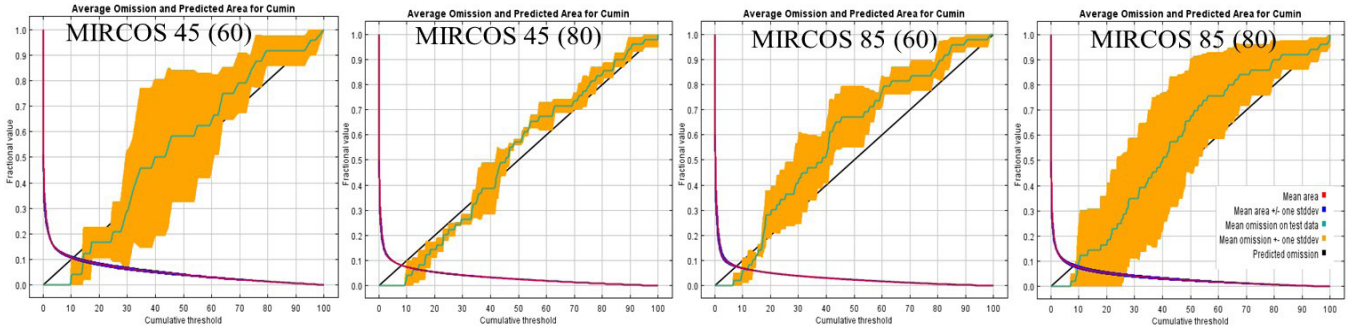
Supp. Fig. 3: Analysis of omission/commission with GFDL earth system model with two time frames and its projections



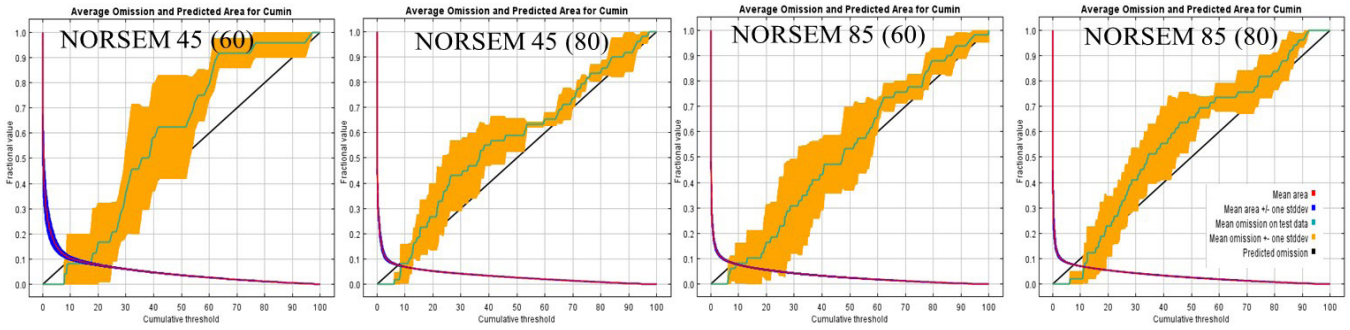
Supp. Fig. 4: Analysis of omission/commission with HADGEM earth system model with two time-frames and its projections



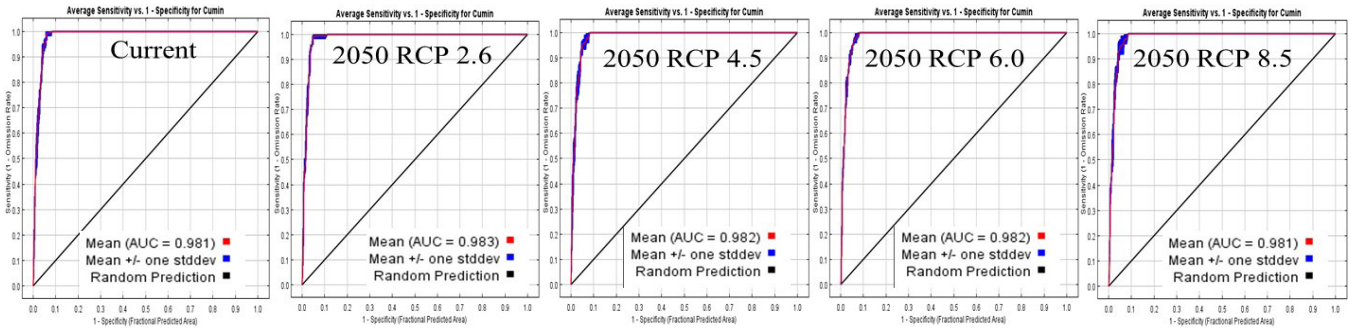
Supp. Fig. 5: Analysis of omission/commission with IPSL earth system model with two time-frames and its projections



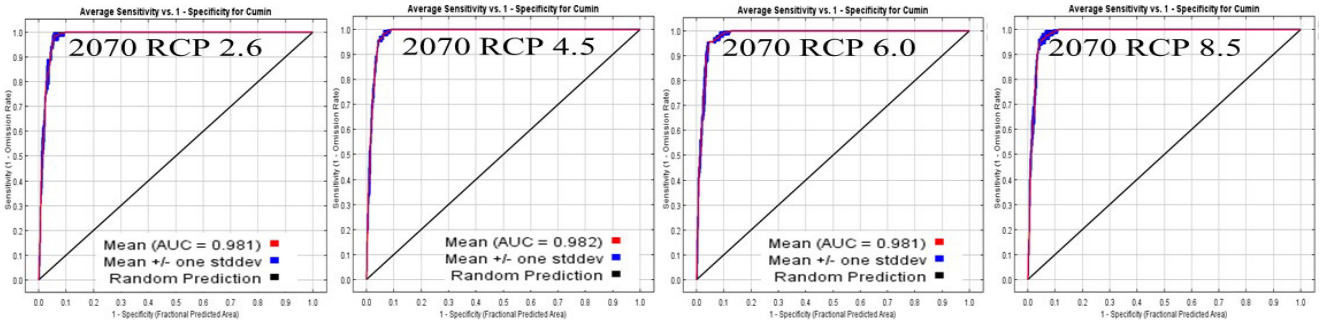
Supp. Fig. 6: Analysis of omission/commission with MIRCOS earth system model with two time-frames and its projections



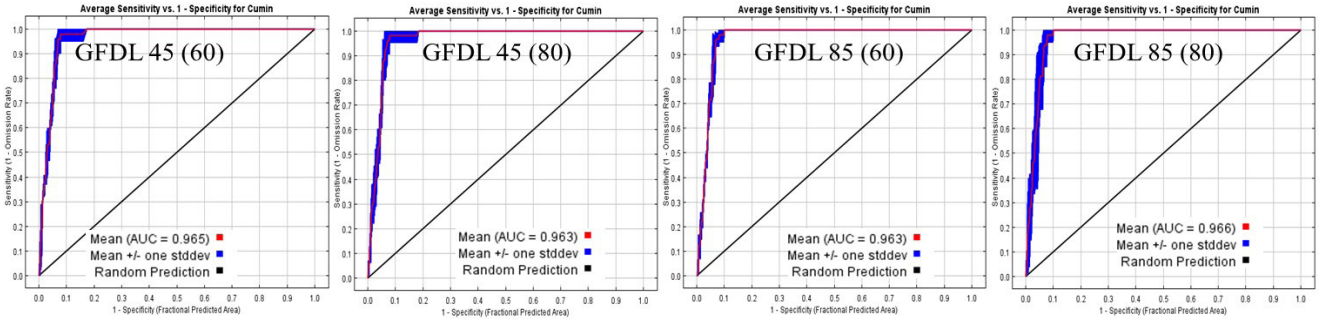
Supp. Fig. 7: Analysis of omission/commission with NORSEM earth system model with two time-frames and its projections



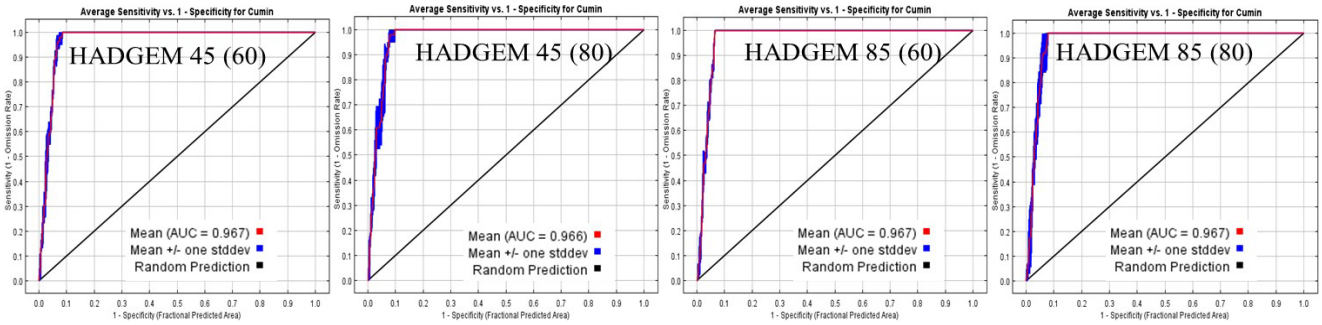
Supp. Fig. 8: Receiver Operating Characteristic (ROC) curves: with current and 2050 bio-climatic timeframes and RCPs projection



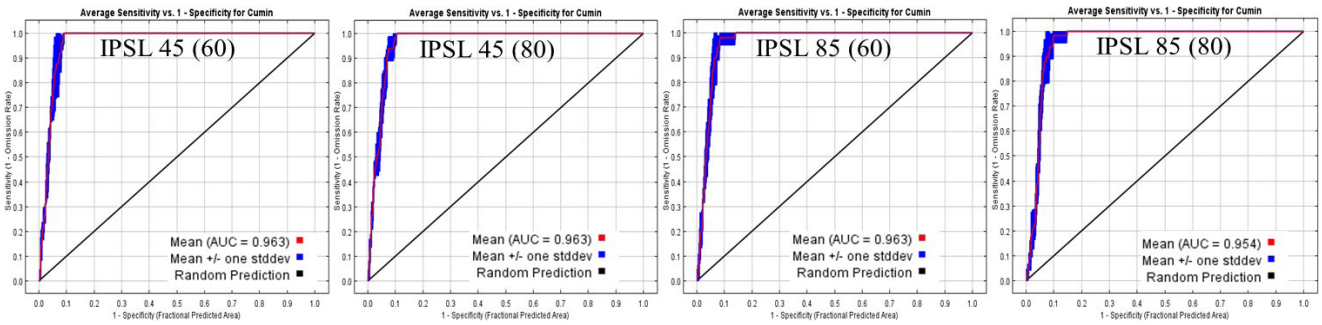
Supp. Fig. 9: Receiver Operating Characteristic (ROC) curves: with 2070 bio-climatic timeframes and RCPs projection



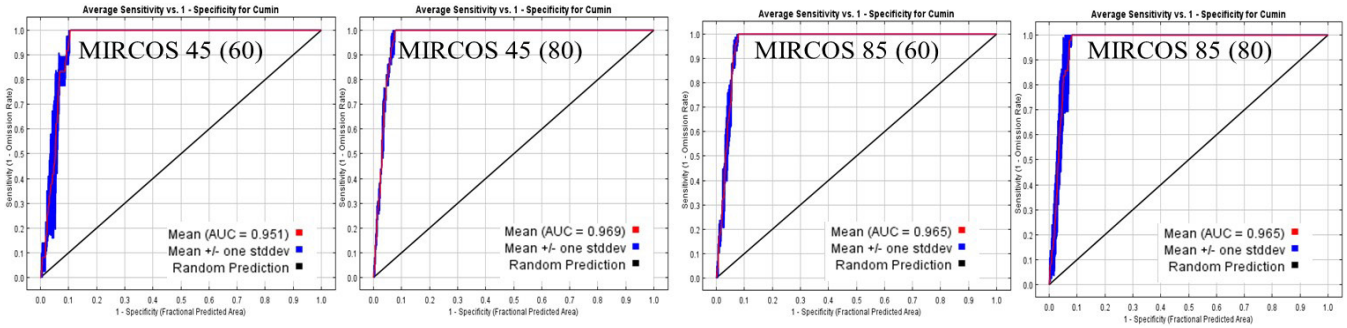
Supp. Fig. 10: Receiver Operating Characteristic (ROC) curves: with GFDL earth system model using two time frames



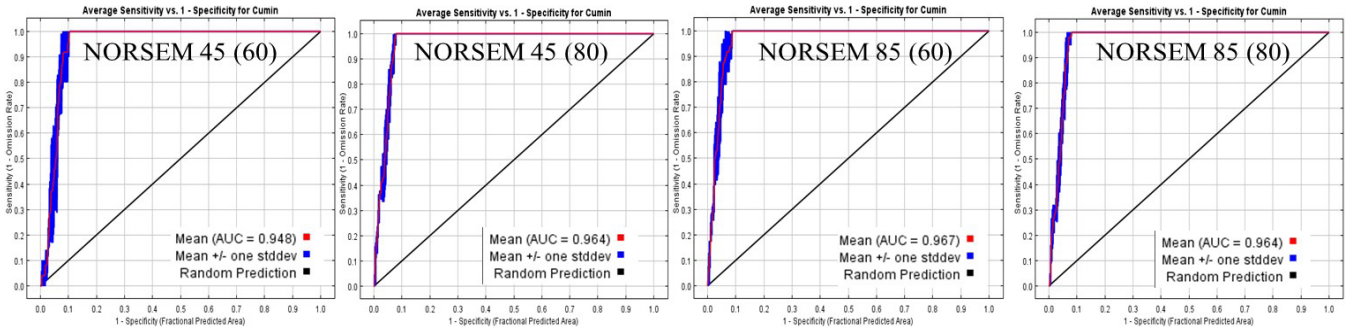
Supp. Fig. 11: Receiver Operating Characteristic (ROC) curves: with HADGEM earth system model using two time frames



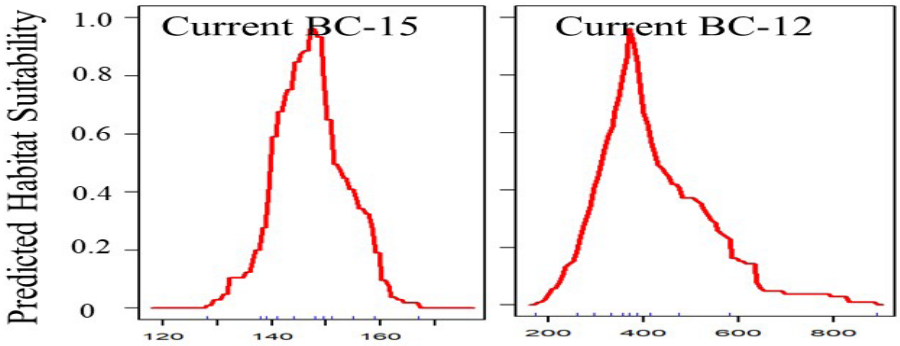
Supp. Fig. 12: Receiver Operating Characteristic (ROC) curves: with IPSL earth system model using two time frames



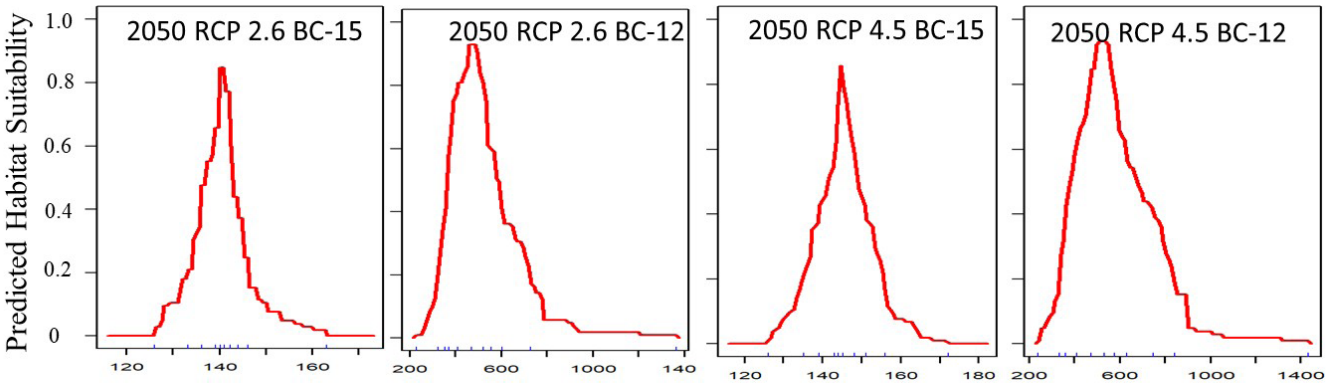
Supp. Fig. 13: Receiver Operating Characteristic (ROC) curves: with MIRCOS earth system model using two time frames



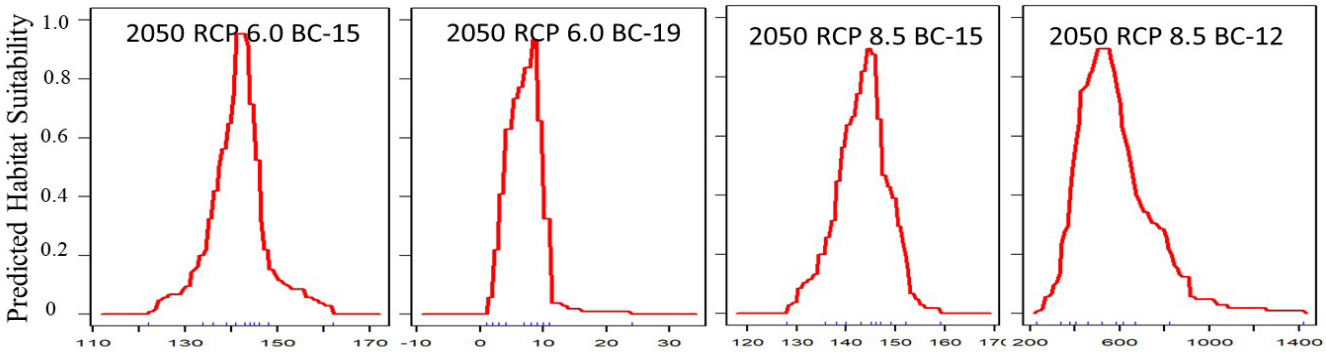
Supp. Fig. 14: Receiver Operating Characteristic (ROC) curves: with NORSEM earth system model using two time frames



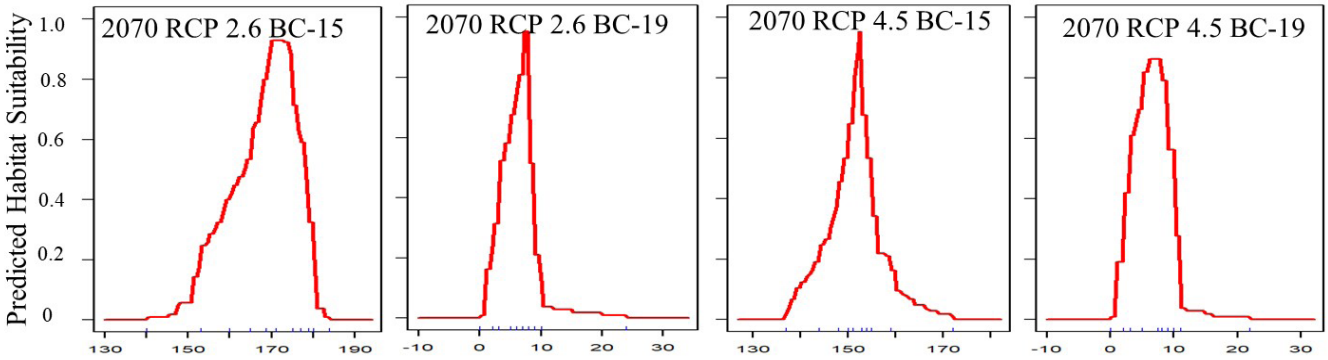
Supp. Fig. 15: Response curves with current-climatic parameters illustrating the likelihood of *Cuminum cyminum* habitat suitability



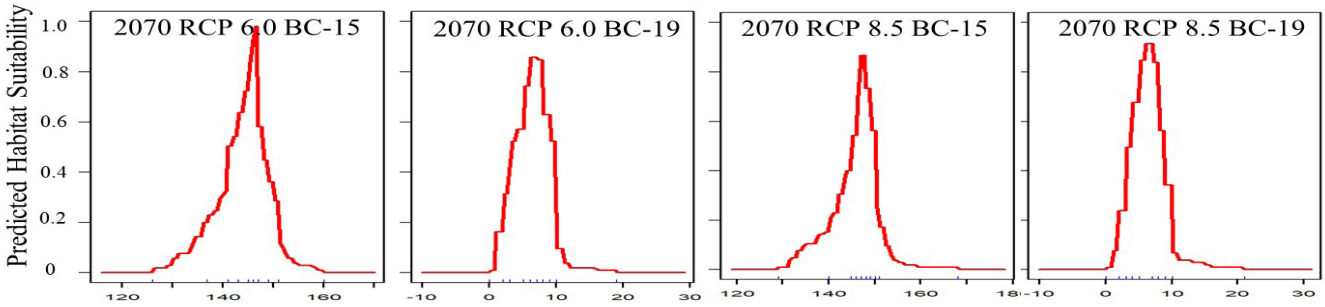
Supp. Fig. 16: Response curves with RCP 2.6 and 4.5 of 2050 bio-climatic time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



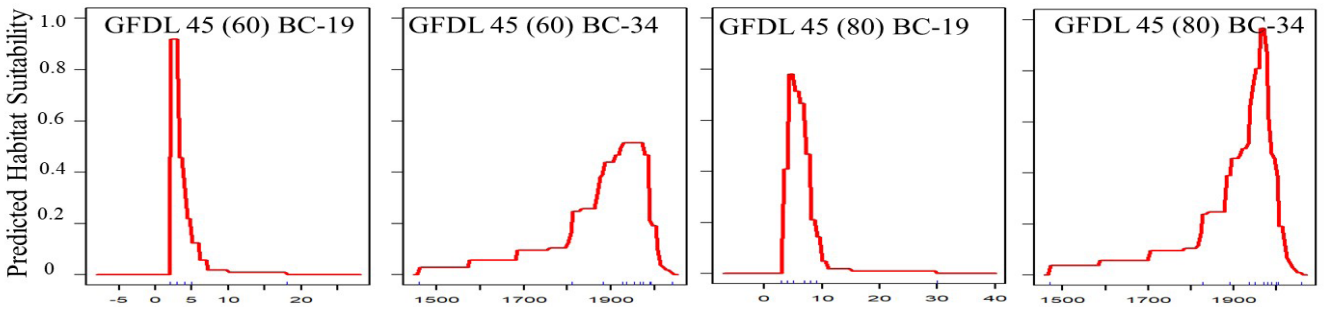
Supp. Fig. 17: Response curves with RCP 6.0 and 8.5 of 2050 bio-climatic time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability.



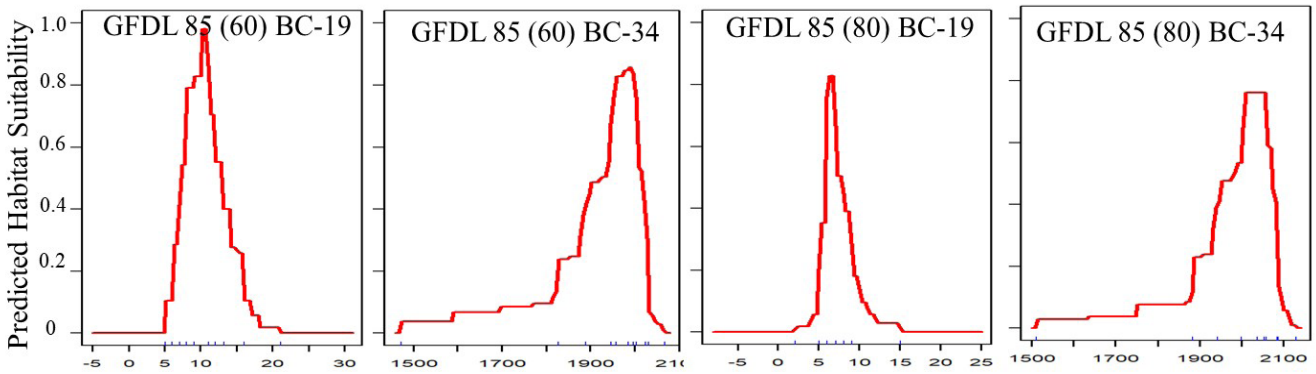
Supp. Fig. 18: Response curves with RCP 2.6 and 4.5 of 2070 bio-climatic time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



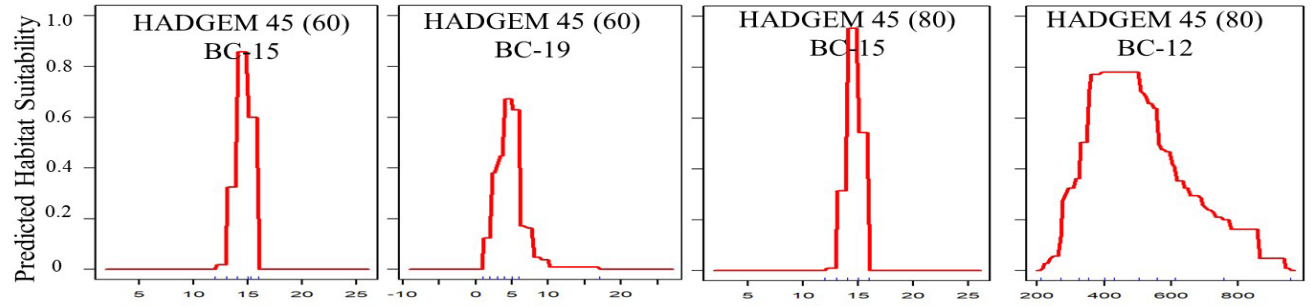
Supp. Fig. 19: Response curves with RCP 6.0 and 8.5 of 2070 bio-climatic time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



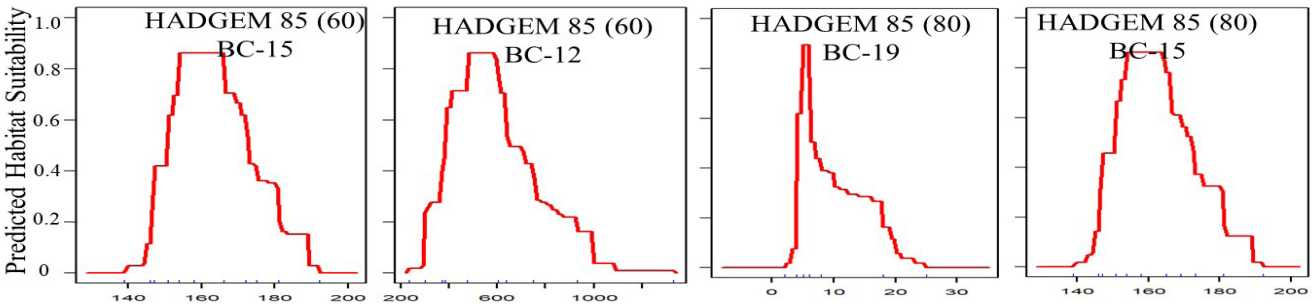
Supp. Fig. 20: Response curves with GFDL earth system model with 45 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



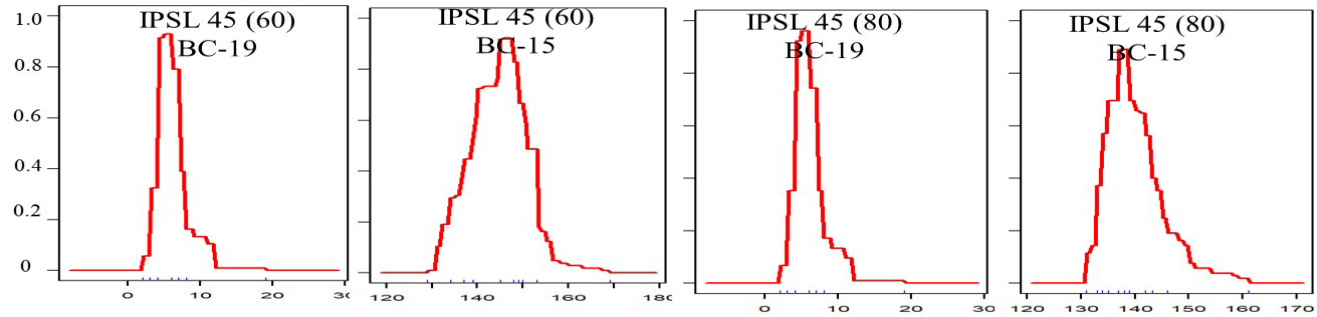
Supp. Fig. 21: Response curves with GFDL earth system model with 85 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



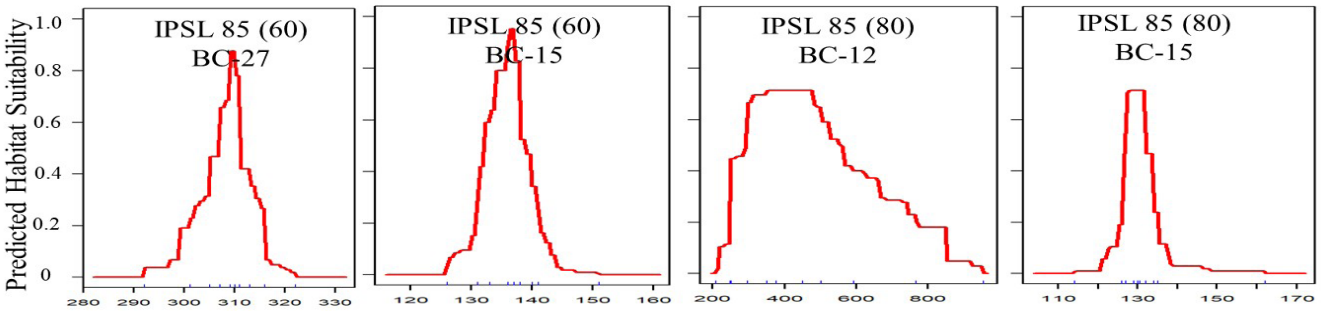
Supp. Fig. 22: Response curves with HADGEM earth system model with 45 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



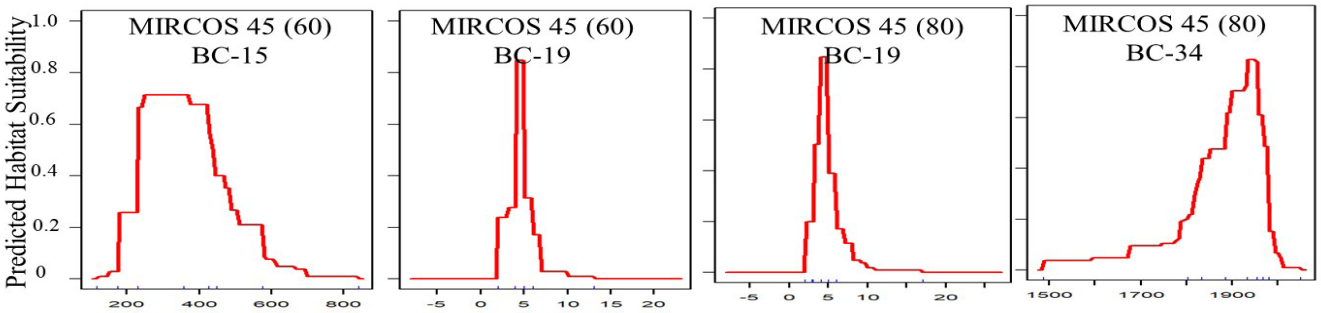
Supp. Fig. 23: Response curves with HADGEM earth system model with 85 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



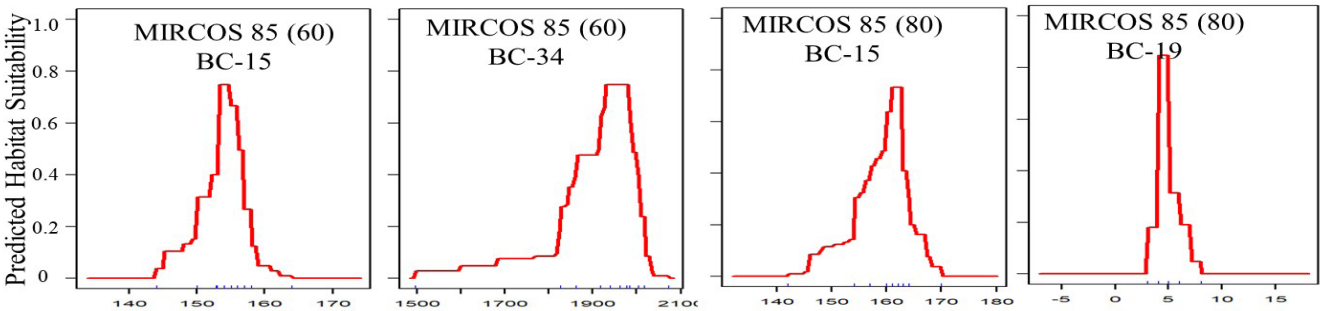
Supp. Fig. 24: Response curves with IPSL earth system model with 45 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



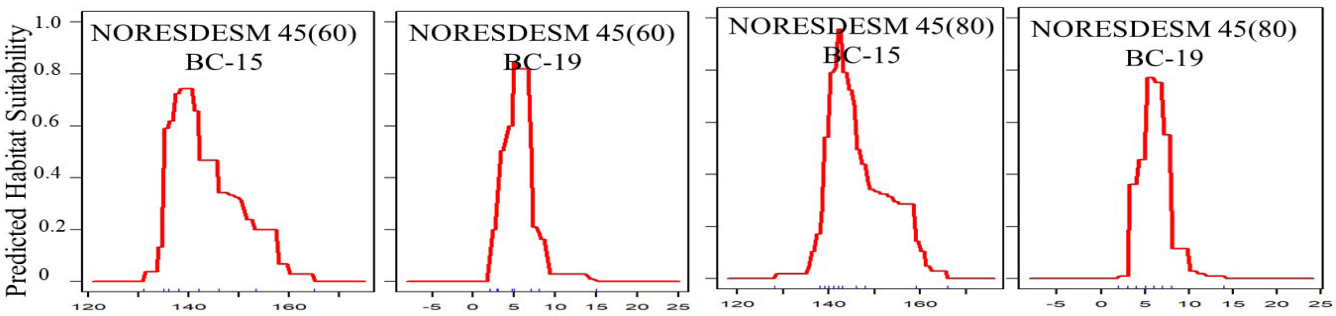
Supp. Fig. 25: Response curves with IPSL earth system model with 85 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



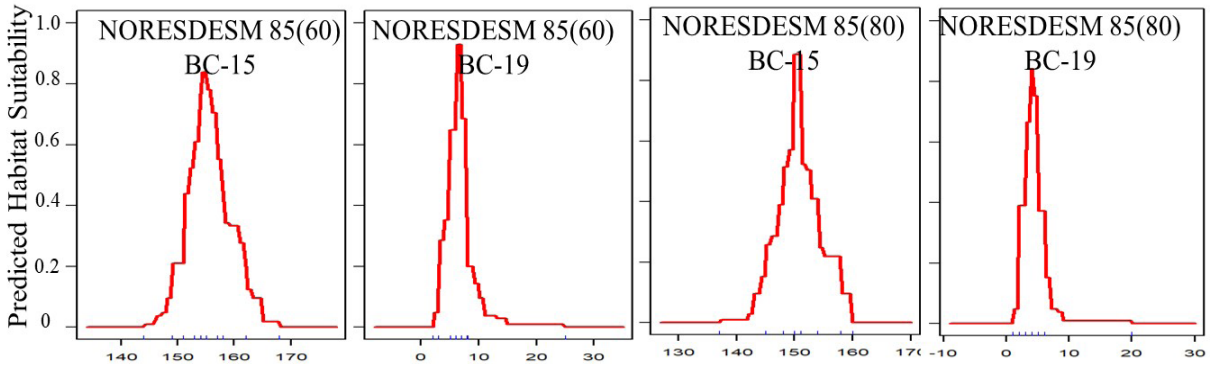
Supp. Fig. 26: Response curves with MIRCOS earth system model with 45 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



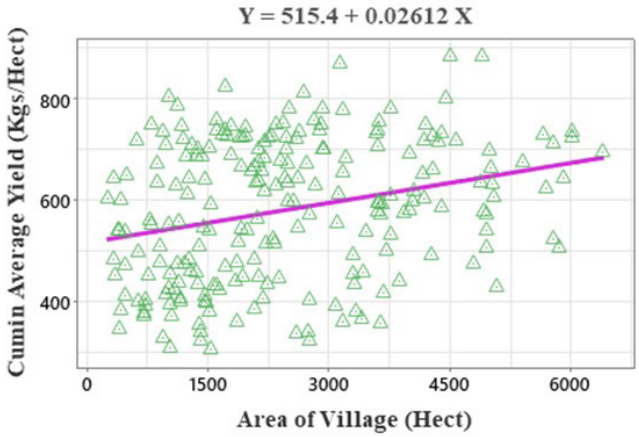
Supp. Fig. 27: Response curves with MIRCOS earth system model with 85 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



Supp. Fig. 28: Response curves with NORESDESM earth system model with 45 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



Supp. Fig. 29: Response curves with NORESDESM earth system model with 85 scenario and two time-frame illustrating the likelihood of *Cuminum cyminum* habitat suitability



Supplementary Fig. 30: Simple regression analysis between area of village (Ha.) and *Cuminum cyminum* average yield (Kg/ha.) n = 225; R^2 0.75*