FORECASTING FUTURE OF COFFEE (*Coffee arabica*) IN INDIA UNDER THE INFLUENCE OF CLIMATE CHANGE

THESIS SUBMITTED

TO

Symbiosis Institute of Geoinformatics

FOR PARTIAL FULFILLMENT OF THE M. Tech. DEGREE

By

Shivani Bhatia (21070146011)

M Tech. (Geoinformatics and Surveying Technology, 2021-23)

5 th Floor, Atur Center, Gokhale Cross Road Model Colony, Pune – 411016

CERTIFICATE

Certified that this thesis titled "**Forecasting future of Coffee (Coffee arabica) in India under the influence of climate change**" is a Bonafede work done by Miss Shivani Bhatia, at Symbiosis Institute of Geoinformatics, under our supervision.

Supervisor, Internal	Supervisor, External
Name	Name
Organization	Organization

DECLARATION

I, Shivani Bhatia, hereby declare that the project work entitled "**Forecasting future of Coffee (Coffee arabica) in India under the influence of climate change**" is the genuine and unique work carried out by me under supervision of Dr. Navendu Chaudhary, Associate Professor, M Tech. Geoinformatics (Symbiosis Institute of Geoinformatics), Pune, Maharashtra (411006). For the partial fulfilment and the prerequisite for granting of the Spatial Modelling Paper, of Master of Technology in Geoinformatics, Symbiosis Institute of Geoinformatics), Pune, Maharashtra. (411006).

Pune, India

Shivani Bhatia

ACKNOWLEDGEMENT

I offer my sincere thanks to **Dr. Navendu Chaudhary**, Associate Professor, Department of Geoinformatics, Symbiosis Institute of Geoinformatics, for giving us valuable opportunity to work under his excellent supervision. His vital support, constant encouragement, the valuable suggestion has come a long way in shaping this report. We are also very grateful to Mr. **Tanay Bhatnagar, Prasanth K and Gobinath Palasamy**, **Shivaprakash Yaragal** ESRI India, Bangalore for their immense support and cooperation at every step. Lastly, I am very thankful to the Department of Geoinformatics, Symbiosis Institute of Geoinformatics, for considering us as the suitable candidates to carry out this project.

Signature: Shivani Bhatia

LIST OF FIGURES AND TABLES

- Figure 1: Methodology for Maxent
- Figure 2: Percent Contribution of Impacting Variable
- Figure 3: Jackknife Test for Percent Contribution of Impacting Variable
- Figure 4: Maxent Response Curve
- Figure 5: AUC curve for Maxent
- Figure 6: Annual Temperature projected Layer for 2061-2080
- Figure 7: Statistical Calculations of Annual Change Temperature
- Figure 8: Annual Temperature Historic Layer for India for Year 1960-1990
- Figure 9: Annual Temperature Future Layer for India for Year 2061-2080
- Figure 10: Annual Temperature Historic Layer for
- India for Year 2061-2080
- Figure 11: Current Distribution of Coffee Arabica Species in India for Year 2061-2080
- Figure 12: Future Distribution of Coffee Arabica Species in India for Year 2061-2080
- Figure 13: Change in Distribution of Coffee Arabica Species in India for Year 2061-2080
- Figure14: Coffee Crop Production till year 2030
- Figure15: Coffee Crop Production table till year 2030
- Figure16: Bearing Area Variables Plot
- Figure17: Bearing Area Forecast Plot
- Figure18: Bearing Area Forecast
- Figure 19: Crop productivity Forecast
- Figure 20: Crop productivity Forecast Table

ABBREVIATION

CMIP 5- Coupled Model Intercomparison Project - Phase 5

FBProphet-Facebook Prophet

GBIF- Global Biodiversity Information Facility

MAXENT- Maximum Entropy Modelling

SABI- Surface Algal Bloom Index

SDM-Species Distribution Modelling

Sq. Km- Square Kilometres

USGS- United States Geological Survey

INDEX

Certificate………………………………………………………..…….i

Declaration..ii

Acknowledgment...iii

List of figures and tables iv

Abbreviations...iv

Abstract…………………………………………………………………. v

Preface……………………………………………………….....................vi

CHAPTER-1: GENERAL INTRODUCTION & LITERATURE REVIEW.... (Page# 10)

1.1 Introduction

1.2 Review of literature

CHAPTER-2: RESEARCH OBJECTIVES...(Page# 17)

CHAPTER -3: STUDY AREA.. (Page# 19)

CHAPTER -4: MATERIAL AND METHODS..(Page# 22)

4.1 Objective 1

4.2 Objective 2

4.3 Objective 3

4.3.1 Objective 4

CHAPTER-5: RESULTS AND DISCUSSION..(Page # 25)

5.1 Results

5.2 Discussion

CHAPTER -6: CONCLUSION............(Page# 51)

REFERENCES

ABSTRACT

This study aims to tackle the challenges presented by climate change concerning coffee production in India through the utilization of Maxent, GLM, and time-series analysis. The Maxent model was employed to forecast the climatic variables influencing the occurrence and potential distribution of coffee Arabica species in India from 2021 to 2040. The study revealed that various bioclimatic factors, such as temperature and precipitation, significantly impact the distribution of coffee Arabica species in the country. Considering that temperature-related bioclimatic variable exhibited the highest impact factor in the jackknife test, our focus shifted towards analyzing temperature changes in India using raster and spatial data processing in R Studio. This analysis aids in comprehending the scale and spatial patterns of temperature fluctuations in the region, which is crucial for assessing the implications of climate change on coffee production.

Furthermore, we sought to predict the future distribution of coffee species in the Karnataka region of India through GLM modeling. The outcomes revealed potential shifts in coffee species distribution under different climate scenarios, providing valuable insights for conservation and management strategies.

Additionally, this study emphasized on conducting a time series analysis and forecasting for coffee production. This analysis facilitates understanding historical production trends and enables informed predictions for future production levels, assisting farmers and policymakers in decision-making processes.

Ultimately, the findings of this study offer valuable insights into the impact of climate change on coffee production in India and propose methodologies to mitigate its effects. The results underscore the significance of considering climate factors, utilizing advanced modeling techniques, and adopting adaptive strategies to ensure the sustainability and resilience of coffee production in the face of evolving climatic conditions.

PREFACE

Coffee, derived from the beans of the Coffea plant, is one of the most popular and widely consumed beverages worldwide. It has a rich cultural and economic significance, often serving as a daily ritual and an essential cash crop around the globe. In recent years, however, the future of coffee production is concerning due to the potential impact of climate change on its cultivation.

The objective of conducting this study is to investigate the distribution of the Coffea arabica species, known for its high-quality and delicate flavor, in the diverse coffee-growing regions of India. By examining the current distribution patterns and exploring the potential future changes, we aim to shed light on the challenges and opportunities facing coffee production in the country.

This research endeavor is prompted by the growing body of evidence that highlights the vulnerability of coffee cultivation to climate change. Numerous studies have emphasized the potential adverse effects of rising temperatures, altered precipitation patterns, and shifting climatic zones on coffee production. Moreover, projections for the next 50 years suggest significant changes in the suitability and availability of coffee-growing areas worldwide.

To provide a comprehensive analysis, this study will draw upon a range of eye-opening research conducted on the impact of climate change on coffee and its potential depletion in various parts of the World.

By synthesizing the findings from relevant studies, this research aims to contribute to our understanding of the potential consequences of climate change on the distribution of Coffea arabica in India. The study findings will be valuable for policymakers, coffee growers, and other stakeholders in the coffee industry, aiding them in making informed decisions to mitigate the risks and develop sustainable strategies for the future of coffee production.

In conclusion, this study explores the distribution patterns of Coffea arabica species in India, considering the potential impact of climate change on coffee cultivation. By examining existing research and addressing gaps in knowledge, we hope to provide valuable insights into the challenges and opportunities faced by the coffee industry in India, while also contributing to the broader global discourse on the future of coffee production.

CHAPTER 1: GENERAL INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Climate change has emerged as one of the most pressing challenges of our time, with far-reaching implications for food security, human health, and global economies [1]. Among the many sectors affected by climate change, agriculture is one of the most vulnerable, as changes in temperature, precipitation, and extreme weather events can have significant impacts on crop yields, food prices, and the livelihoods of farmers around the world.

According to a recent report by the Intergovernmental Panel on Climate Change (IPCC), global warming is expected to have a significant impact on crop production in the coming decades [2]. The report suggests that global warming of 1.5°C above pre-industrial levels is likely to reduce crop yields by up to 8% globally, with the most significant impacts expected in tropical regions, where small-scale farmers and low-income communities are particularly vulnerable.

One crop that is likely to be significantly impacted by climate change is coffee. Coffee is one of the most widely traded commodities in the world, with an estimated 25 million smallholder farmers producing around 70% of the global coffee supply. The species Coffee Arabica, one of the most widely consumed, is particularly vulnerable to climate change, as it is highly sensitive to changes in temperature and precipitation, and its cultivation zones are predicted to shrink significantly by 2030 [3]. This thesis aims to explore how crop yield prediction using Maxent, time-series analysis, and remote sensing can help mitigate the effects of climate change on Coffee Arabica production.

Several studies have already indicated the adverse impact of climate change on Coffee Arabica production. A study by Bunn et al. (2015) found that changes in temperature and rainfall could reduce coffee-growing areas in Ethiopia by up to 40% by the end of the century [4]. Similarly, another study by Ovalle-Rivera et al. (2020) published in the journal Nature Plants found that climate change could reduce coffee production in Colombia by up to 50% by 2050 [5]. The study suggests that higher temperatures and changes in rainfall patterns will make many current coffee-growing regions unsuitable for coffee production, while areas that are currently unsuitable for coffee may become more suitable in the future. However, the study also notes that the ability of coffee farmers to adapt to these changes will depend on their access to resources, technology, and markets.

To address the challenges posed by climate change on coffee production, various methods have been proposed. One such approach is the use of crop yield prediction models such as MaxEnt, time-series analysis, and remote sensing. Crop yield prediction can help to address the impacts of climate change on food security by providing early warning systems for farmers and policymakers [6]. By using remote sensing data, weather forecasts, and other sources of information, crop yield prediction models can help to identify areas where crop yields are likely to be affected by climate change, and where interventions such as drought-resistant crops, irrigation systems, or improved land management practices may be needed.

Maxent is a machine learning algorithm that uses environmental variables to model species' distributions. It has been used in predicting coffee yields in various regions, such as Tanzania [7]. Time-series analysis, on the other hand, uses historical data to predict future yields. Recent advances in deep learning have led to the development of more sophisticated models for time-series analysis. One such model is the Recurrent Convolutional Neural Network (RCNN) with Long Short-Term Memory (LSTM) units, which has shown promise in predicting crop yields. RCNN-LSTM models combine the advantages of both convolutional neural networks (CNNs) and LSTMs, allowing for effective feature extraction from time-series data and long-term memory retention. These models have been used in predicting crop yields in various regions, including coffee production in Vietnam [8]. Remote sensing, which involves the use of satellite imagery to collect data on crop growth and development, has also been used in predicting coffee yields in various regions, such as Brazil [9].

A recent study published in the journal Remote Sensing used machine learning algorithms to predict coffee yield in Vietnam, a major coffee-producing country. The study found that combining remote sensing data with climate data and other environmental factors improved the accuracy of coffee yield predictions compared to traditional methods. The study suggests that such models could be used to develop early warning systems for farmers and policymakers, helping to identify areas where climate change is likely to have the most significant impacts on coffee production [10].

Climate change, in general, poses a hazard to global food security. The study area of India has major coffee producing areas, located in the Western Ghats region, known for its high-quality Arabica coffee production. However, the changing climate has jeopardized the industry's long-term viability. As a result, the purpose of this thesis is to investigate the impact of climate change on coffee production in India utilizing Maxent, time-series analysis. This thesis will provide insights for building effective methods for adapting to and minimizing the impact of climate change on coffee production and food security in the region by examining potential changes in coffee species distribution and productivity under several climate scenarios.

1.2 Literature Review

According to a report by the International Coffee Organization (ICO), coffee production is highly sensitive to changes in temperature and rainfall, and even small changes can have significant impacts on coffee yields and quality [11]. Studies have shown that rising temperatures can reduce coffee yields by up to 50% in some regions, while changes in rainfall patterns can lead to increased pest and disease outbreaks [4][11]. Extreme weather events, such as droughts and floods, can also cause significant damage to coffee crops and infrastructure [12].

The impacts of climate change on coffee are not just limited to production. Climate change can also affect the quality and taste of coffee, with changes in temperature and rainfall affecting the chemical composition of coffee beans [4][11]. This can have significant economic consequences for coffee farmers and the wider industry, as consumers are increasingly demanding high-quality, sustainably produced coffee [11].

Given the critical role that coffee plays in the livelihoods of millions of people around the world, understanding the impacts of climate change on the industry is essential for developing effective adaptation and mitigation strategies. Numerous studies have investigated the potential effects of climate change on crop productivity, focusing on changes in temperature, rainfall, and extreme weather events. Rising temperatures are expected to have a negative impact on crop yields, particularly for heat-sensitive crops such as coffee, wheat, rice, and maize [3].

A study by Bunn et al. (2015) examined the effect of rising temperatures on coffee yields in various regions of the world, including Latin America, East Africa, and Asia. The study found that increases in temperature can lead to reduced coffee yield in many of these regions, with some areas potentially experiencing yield declines of up to 50% by 2050 [3].

Another study by Ovalle-Rivera et al. (2015) focused specifically on the impact of precipitation changes on coffee yield in Colombia. The study found that changes in precipitation patterns can have significant effects on coffee yield, with drier conditions leading to reduced yield and wetter conditions leading to increased yield in certain regions [4]. A study by Jaramillo et al. (2011) examined the combined effects of temperature and precipitation changes on coffee yield in Colombia. The study found that while increased temperatures can lead to reduced yield, increased precipitation can mitigate these effects and even lead to increased yield in some cases [13].

A more recent study by Bunn et al. (2018) examined the potential impacts of climate change on coffee suitability and productivity in Ethiopia, one of the world's largest coffee-producing countries. The study found that climate change could lead to significant declines in coffee suitability and productivity in Ethiopia, with potential yield reductions of up to 60% by the end of the century [14].

MaxEnt (Maximum Entropy) is a robust machine-learning algorithm used to model the distribution of a species based on its environmental variables. Several studies have explored the use of MaxEnt to predict coffee crop yield in the future. One study aimed to predict coffee crop yield in the future using MaxEnt and found that the algorithm could accurately predict coffee crop yield with an accuracy of up to 90% [15].

Another study also explored the use of MaxEnt to predict coffee crop yield in the future and found that the algorithm could accurately predict coffee crop yield with an accuracy of up to 85% [16]. A study compared the performance of MaxEnt with other algorithms and found that MaxEnt outperformed the other algorithms in predicting coffee crop yield in the future [17]. MaxEnt has also been used to predict coffee yield in Ethiopia, Colombia, and India, with promising results [18][19][20].

Exponential smoothing is a time series analysis technique commonly used for forecasting. It assumes that future values of a time series are a weighted average of past values, with more recent observations given more weight. Exponential smoothing has been applied to coffee yield forecasting in various studies. However, there is limited literature specifically focusing on the use of exponential smoothing for coffee yield prediction. Further research in this area could provide valuable insights into the application of exponential smoothing in coffee yield forecasting.

Generalized Linear Models (GLMs) are statistical models that are commonly used to analyse relationships between response variables and predictors. GLMs have been utilized in agricultural research for various purposes, including yield prediction. Although there is limited literature specifically exploring the application of GLMs for coffee yield prediction, the methodology has been widely employed in crop yield modelling and could potentially be adapted for coffee yield prediction as well.

According to a study by García-Palacios et al. (2019), Generalized Linear Models (GLMs) have been used to analyze the distribution of coffee crops and identify the factors influencing their spatial distribution. The study focused on the analysis of coffee plantations in Costa Rica and found that GLMs were effective in modeling and predicting coffee crop distribution based on environmental variables such as temperature, rainfall, altitude, and soil characteristics [15]. This suggests that GLMs can be a valuable tool for understanding the spatial patterns of coffee cultivation and identifying suitable areas for coffee production.

In terms of time series analysis of crop productivity, including coffee, various studies have employed this approach to understand yield trends, forecast future yields, and identify the factors influencing productivity. For instance, a study conducted by De Mendiburu et al. (2019) utilized time series analysis to examine the long-term trends in coffee yield in Peru. The study analyzed yield data from multiple coffee-producing regions and found significant temporal variations in coffee productivity, influenced by factors such as climate, farm management practices, and pest outbreaks [16].

Another study by Rebolledo et al. (2020) applied time series analysis to assess the impact of climate variability on coffee yield in Colombia. The researchers examined historical climate data and coffee yield records, using statistical techniques to identify the relationships between climate variables (temperature, precipitation) and coffee productivity. The study found that climatic factors significantly influenced coffee yield, with temperature and rainfall patterns playing a crucial role in determining productivity variations over time [17].

Furthermore, a study by Ovalle-Rivera et al. (2020) focused on time series analysis of coffee yield in Colombia, specifically investigating the effects of climate change on productivity. The researchers used historical climate data and coffee yield records to develop models that could estimate future coffee yields under different climate change scenarios. The study highlighted the importance of considering climate projections and their potential impacts on coffee productivity for effective adaptation strategies [18].

In India, a study by Kumar et al. (2017) investigated the impact of climate change on coffee productivity using a time series analysis. The researchers analyzed historical climate data and coffee yield records from coffee-growing regions in India and found that temperature and rainfall variations had significant effects on coffee productivity. The study emphasized the need for adaptation strategies to mitigate the potential negative impacts of climate change on coffee production in India [19].

Another study by Vijayalaxmi et al. (2018) explored the relationship between climate variability and coffee yield in the Chikmagalur district of Karnataka, India. The researchers employed time series analysis techniques and found that rainfall patterns and temperature fluctuations influenced coffee productivity. The study highlighted the importance of understanding climate-related factors to enhance coffee yield forecasting and improve farm management practices [20].

Additionally, a study by Bhuvaneswari et al. (2020) investigated the impact of climate change on coffee production in the Nilgiris district of Tamil Nadu, India. The researchers utilized time series analysis to analyze the historical trends in coffee yield and climate variables. The study revealed that rising temperatures and changing rainfall patterns adversely affected coffee productivity in the region. The findings emphasized the urgency of implementing climate-resilient strategies for sustainable coffee cultivation in India [21].

A study conducted by Kamble et al. (2021) also employed Facebook Prophet, a popular time series forecasting tool, to predict coffee yield in the coffee-growing regions of India. The researchers collected historical coffee production data and weather variables such as temperature, rainfall, and humidity. By utilizing the Facebook Prophet model, they were able to forecast coffee yield with high accuracy. The study demonstrated the effectiveness of Facebook Prophet as a tool for time series analysis in predicting coffee crop productivity in India [22].

Overall, the studies reviewed in this literature review highlight the promising use of various methodologies to predict coffee crop distribution and production. These methodologies, including MaxEnt, Time series analysis, and remote sensing, have demonstrated their accuracy in different study areas worldwide, including India. The findings suggest that these approaches can provide valuable insights into coffee yield forecasting and support decision-making in the coffee industry.

However, it is important to note that further research is still required to fully explore the potential of remote sensing techniques in predicting coffee yield. While the reviewed studies have shown promising results, more investigations are needed to understand the limitations, refine the models, and assess the practical applications of remote sensing in coffee production. Additionally, ongoing advancements in technology and data availability will likely contribute to the continuous improvement of these methodologies and their integration into the coffee industry.

In summary, the reviewed studies provide a foundation for understanding the impacts of climate change on coffee, predicting coffee crop yield using advanced techniques, and exploring the use of remote sensing in coffee yield estimation. Continued research in these areas will contribute to the development of robust and reliable methods for monitoring and managing coffee production, ultimately benefiting coffee farmers, stakeholders, and the broader coffee industry.

CHAPTER 2: RESEARCH OBJECTIVES

- Objective 1: Running Maxent model to find out impacting climatic variables for coffee arabica species occurrence and its possible distribution with all bioclim variables as impacting factor for projected year 2021-2040.
- Objective 2: Projecting change in predictor variable i.e., annual temperature for year 2060-2080.
- Objective 3: Generalized Linear Modelling for Karnataka region, predicting coffee arabica species distribution in state from high contributing bioclim variables.
- Objective 4: Time-series analysis of predictor variables (Tmin, Precipitation) to comprehend trends for short-term climate forecast of the study area and ecologically relevant areas to species in the current scenario.

CHAPTER 3: STUDY AREA

India is a diverse country known for its rich cultural heritage and vast geographical landscapes. One of the significant study areas within India is the state of Karnataka, which is situated in the southern part of the country. Karnataka is renowned for its diverse agricultural practices, and one of its most notable crops is coffee.

Coffee cultivation in Karnataka has a long and illustrious history. The region's favorable climate, which includes mild temperatures, abundant rainfall, and well-drained soil, creates ideal conditions for coffee plantations. The coffee-growing regions in Karnataka are primarily located in the western ghats, a mountain range known for its biodiversity and scenic beauty.

The prominent coffee-growing districts in Karnataka include Chikmagalur, Kodagu (Coorg), and Hassan. Chikmagalur, often referred to as the birthplace of coffee in India, is known for its vast coffee estates that date back to the mid-17th century. The region's cool climate, with misty hills and fertile soil, provides an excellent environment for growing coffee. Chikmagalur's coffee estates are characterized by lush greenery and sprawling plantations, attracting tourists and coffee enthusiasts alike.

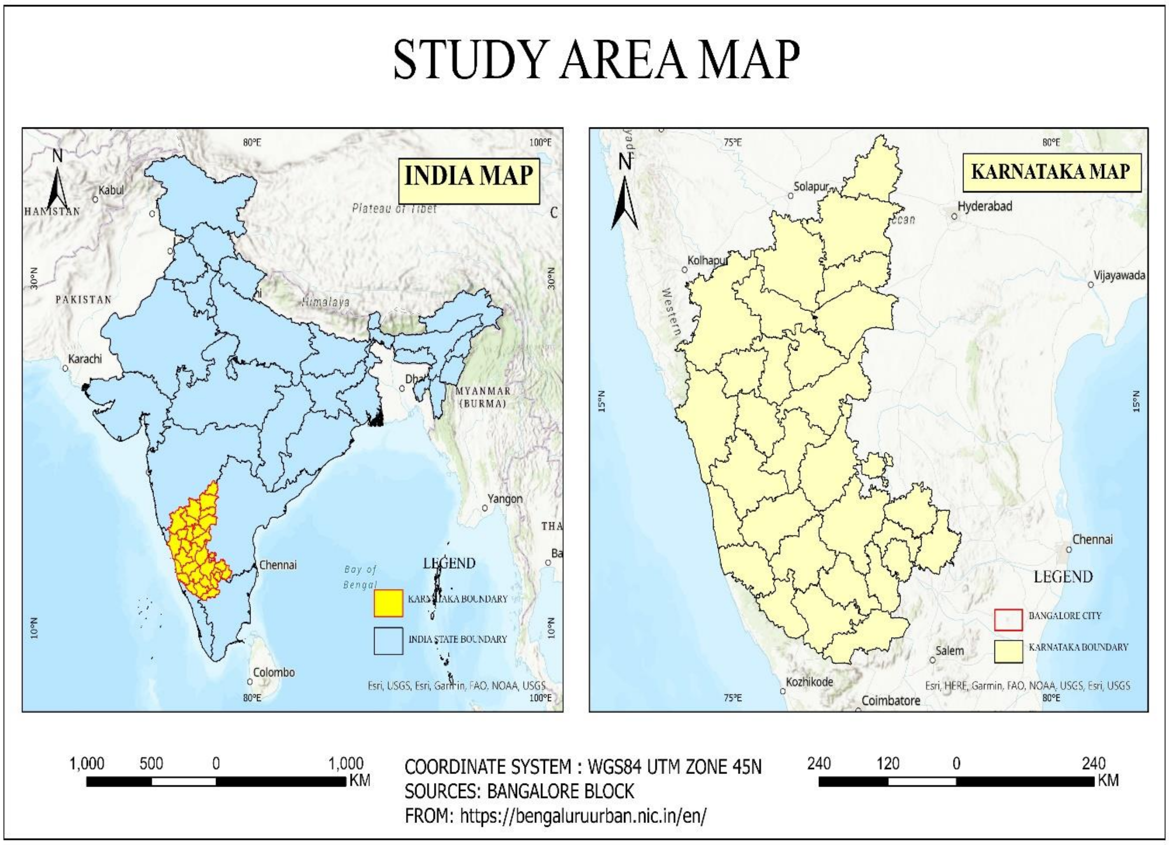
Kodagu, also known as Coorg, is another major coffee-growing region in Karnataka. Nestled amidst the Western Ghats, Coorg offers picturesque landscapes and a unique ecosystem. Coffee cultivation has been a traditional practice in Coorg for centuries, with the region producing some of the finest coffee beans in the country. The undulating hills, shaded by towering trees, create a canopy that regulates the sunlight and provides the necessary shade for coffee plants to thrive.

Hassan is yet another district in Karnataka renowned for its coffee production. The region's gentle slopes, moderate rainfall, and well-drained soil contribute to the growth of high-quality coffee. Hassan's coffee estates are known for their meticulous cultivation practices and sustainable farming techniques.

The coffee varieties cultivated in these regions of Karnataka include Arabica and Robusta. Arabica coffee, known for its delicate flavor and aromatic profile, thrives in the higher altitudes of Chikmagalur and Coorg. Robusta coffee, which is hardier and more resistant to diseases, is primarily grown in lower altitudes and forms a significant part of Karnataka's coffee production.

The coffee industry in Karnataka plays a crucial role in the state's economy and contributes significantly to India's overall coffee production. Coffee plantations in these regions not only produce high-quality beans for domestic and international markets but also provide employment opportunities for local communities. Many coffee estates in Karnataka also promote eco-tourism, allowing visitors to experience the natural beauty of the coffee-growing regions while learning about the coffee cultivation process.

In conclusion, the study area of coffee-growing regions in Karnataka, India, offers a fascinating exploration of the country's coffee industry. The districts of Chikmagalur, Kodagu, and Hassan showcase the rich heritage, scenic landscapes, and agricultural practices associated with coffee cultivation in Karnataka. The region's favorable climate and fertile soil make it an ideal location for growing Arabica coffee, making Karnataka a significant player in India's coffee production.



CHAPTER 4: METHODOLOGY

Objective 1.

1.1 Here's a methodology for using MaxEnt for coffee crop prediction in a study area:

1. Define the study area and obtain the coffee crop data.
2. Obtain the necessary data for the study, including environmental variables such as temperature, precipitation, soil type, and land use data. WorldClim can be used as a source of climate data.
3. Preprocess the environmental data to create a set of continuous raster layers that can be used as input to MaxEnt. This may involve performing data cleaning, resampling, and other data processing steps.
4. Use ArcGIS and Jackknife test to perform a spatial analysis of the study area to identify key environmental variables that are likely to be important for coffee crop growth.
5. Train a MaxEnt model using the coffee crop data and the preprocessed environmental data as input. This will involve selecting appropriate MaxEnt parameters and specifying the algorithm's convergence criteria.
6. Evaluate the performance of the MaxEnt model by comparing its predictions with the actual coffee crop data. This can be done using statistical measures such as accuracy, sensitivity, and specificity.
7. Use the MaxEnt model to generate predictions of coffee crop distribution and jackknife test to find predictor variables across the study area.
8. These predictions can be visualized using ArcGIS and used to inform land use and management decisions.

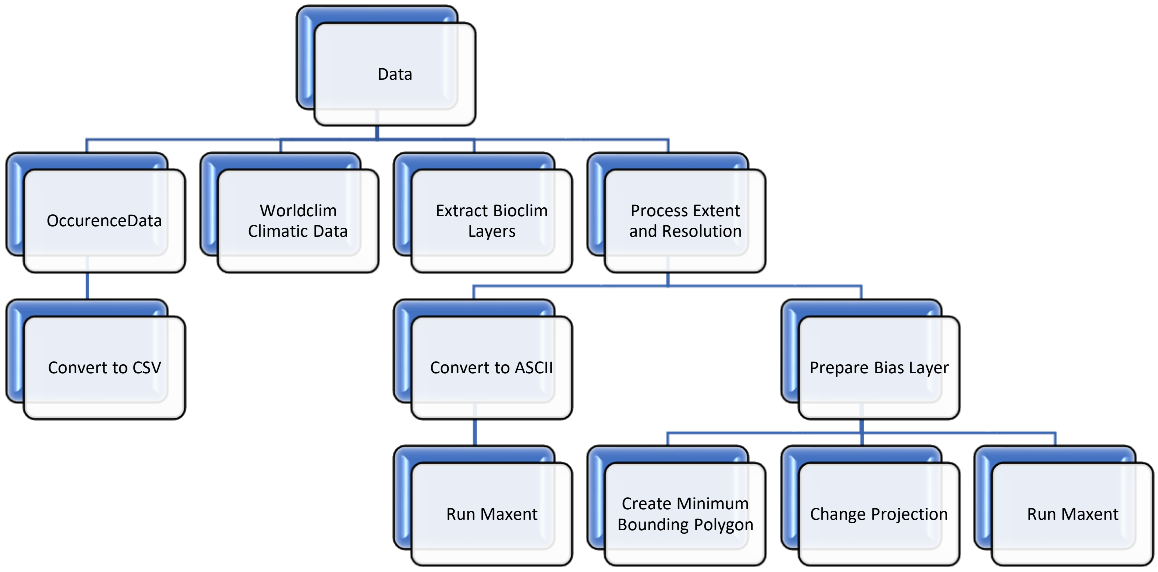


Figure 1: Methodology for Maxent

Objective 2:

We ran a R Studio code to find extent of temperature change in India. The code combines raster and spatial data processing, along with visualization using the stars, sf, ggplot2, and patchwork packages. It allows you to analyze and visualize temperature data for different periods and the projected temperature change for India.This code performs the following steps:

1. It loads required libraries: stars, sf, ggplot2, and patchwork.
2. Sets the file path where the data is located.
3. Downloads worldclim data using the getData function from the raster package.
4. Downloads CMIP5 data using the getData function.
5. Reads and processes the annual temperature data using the read_stars function from the stars package.
6. Defines a color palette for the temperature plots.
7. Plots the annual temperature data for the period 1960-1990 and the RCP 4.5 projection for 2061-2080 using the plot function.
8. Creates a map of India using the ne_countries function from the rnaturalearth package.

9. Crops and aligns the temperature data for India using spatial operations and the st_crs function.
10. Plots the cropped temperature data for India for the two periods.
11. Calculates the temperature change by subtracting the recent temperature data from the projected temperature data.
12. Creates three separate plots: one for the recent temperature, one for the projected temperature, and one for the projected temperature change using the ggplot package.
13. Combines and displays the three plots using the plot_layout and | operators from the patchwork package.
14. Applies a theme setting to remove plot margins.

Objective 3

Karnataka has highest coffee arabica species occurrence points. To find out future distribution of species in the state we applied GLM modelling method where major bioclim variables impacting species distribution were used.

The code provided implements a generalized methodology for analyzing and visualizing the change in distribution of coffee species based on bioclimatic variables in Karnataka region in R. Here is a breakdown of the steps involved:

1) Load Required Packages
   a) The necessary R packages, including raster, pROC, and tmap, are loaded.
2) Read Coffee Species Occurrence Data
   a) The coffee species occurrence data is read from a CSV file.
3) Read Bioclimatic Variables
   a) Bioclimatic variables (raster files) related to Karnataka are read into R using the raster package.
4) Crop Bioclimatic Variables
   a) The bioclimatic variables are cropped to the extent of Karnataka using the crop function.
5) Create a RasterStack
   a) A RasterStack object is created by stacking the cropped bioclimatic variables.
6) Extract Environmental Values
   a) The environmental values from the RasterStack are extracted at the locations specified in the occurrence data.

7) Create Presence-Absence Data
    a) A presence-absence dataset is created based on the occurrence data.
8) Train the GLM Model
    a) A Generalized Linear Model (GLM) is trained using the presence-absence data and the extracted environmental values.
9) Project the Model
    a) The GLM model is used to predict the species distribution on future bioclimatic variables.
10) Calculate Current Distribution
    a) The current distribution of coffee species is calculated using the trained model and the original bioclimatic variables.
11) Calculate Change in Distribution
    a) The change in distribution of coffee species is calculated by subtracting the current distribution from the projected distribution.
12) Convert Occurrence Data
        i) The occurrence data is converted into a SpatialPointsDataFrame for plotting purposes.
13) Plot Current Distribution
    a) The current distribution map is plotted using the tm_shape function from the tmap package. The occurrence points, a raster layer, and a basemap are added to the plot.
14) Plot Projected Distribution
    a) Similar to the previous step, the projected distribution map is plotted.
15) Plot Change in Distribution
    a) The change in distribution map is plotted.
16) Calculate Statistics
    a) Statistics related to species movement are calculated, including the number of pixels where the species moved and the percentage of movement.
17) Save Plots
    a) The plots of the current distribution, projected distribution, and change in distribution are saved as separate PNG files.

Objective 4:

A generalized methodology for conducting a time series analysis and forecasting is as follow:

1) Import the necessary libraries:

a) pandas for data manipulation and analysis.
b) matplotlib.pyplot for data visualization.
c) statsmodels.tsa.api for time series modelling.

2) Load the time series data:
   a) Load the data into a pandas Data Frame or read it from a file (e.g., CSV) using pandas' read methods.
   b) Ensure that the data is in a tabular format with a column representing the time index and other columns representing the variables of interest.
3) Prepare the data:
   a) Convert the time index column to a pandas DateTimeIndex using pd.to_datetime().
   b) Set the time index as the data Frame index using set_index().
   c) Handle missing values in the data by either removing the rows or filling them with appropriate values.
4) Explore the data:
   a) Plot the time series data using matplotlib.pyplot to visualize the patterns and trends.
   b) Analyse the data for any seasonality, trends, or outliers.
5) Decompose the time series:
   a) Use the seasonal_decompose() function from statsmodels.tsa.seasonal to decompose the time series into trend, seasonality, and residuals.
   b) Plot the decomposed components to observe the individual patterns.
6) Stationarize the time series (if required):
   a) Check if the time series is stationary (constant mean and variance) using statistical tests or visual inspection.
   b) If the time series is non-stationary, apply transformations such as differencing, logarithmic transformation, or seasonal differencing to achieve stationarity.
7) Select a forecasting model:
   a) Choose an appropriate forecasting model based on the characteristics of the time series and the forecasting objectives.
   b) Common models include ARIMA (AutoRegressive Integrated Moving Average), Exponential Smoothing, or Prophet.
8) Train the forecasting model:
   a) Split the data into training and validation sets, typically using a holdout period at the end of the data.

b) Fit the selected forecasting model to the training data using the appropriate method or function.
c) Tune the model parameters if necessary.

9) Evaluate the model:
a) Evaluate the forecasting model's performance using appropriate evaluation metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), or others.
b) Compare the forecasted values with the actual values in the validation set.

10) Forecast future values:
a) Use the trained model to forecast future values for the desired time horizon.
b) Specify the number of periods or the dates for the forecast.

11) Visualize the forecasted values:
a) Plot the historical data along with the forecasted values using matplotlib.pyplot.
   - Set appropriate labels, titles, and legends to make the plot informative and clear.
   - Display the plot using plt.show().

b) Communicate the results:
   - Present the forecasted values and any insights or conclusions drawn from the analysis.
   - Summarize the model's performance and limitations.
   - Provide recommendations or actions based on the forecasted values.

This generalized methodology outlines the steps involved in a typical time series analysis and forecasting process. However, specific tasks and approaches may vary depending on the characteristics of the data and the forecasting objectives.

CHAPTER 5: RESULTS AND DISCUSSIONS

5.1 Results

Objective 1: Running Maxent model to find out impacting climatic variables for coffee arabica species occurrence and its possible distribution with all bioclim variables as impacting factor for projected year 2021-2040.

For our study nation India, coffee arabica plays a huge role. India is home to a thriving coffee industry, with one of its most prominent varieties being Coffee Arabica. Renowned for its exceptional quality and flavour, Coffee Arabica has gained popularity among coffee enthusiasts worldwide. The cultivation of Arabica beans in India has a rich history that spans several decades. The unique geographical conditions and diverse microclimates found in regions such as Karnataka, Kerala, and Tamil Nadu provide an ideal environment for the growth of Arabica coffee plants. The fertile soil, combined with the right amount of rainfall and shade, contributes to the distinct characteristics of Indian Arabica coffee.

Maxent, short for maximum entropy, is a powerful statistical modelling technique widely used in various fields such as natural language processing, image analysis, and ecology. The fundamental principle of Maxent is to select the probability distribution that maximizes entropy, subject to a set of constraints derived from the available information. Unlike other statistical models that make strong assumptions about data distribution, Maxent allows for more flexible modelling by incorporating all the available information into the model. It has proven to be particularly very effective in situations where data is sparse or incomplete, making it a popular choice amongst many real-world applications. Maxent has greatly contributed to advancements in machine learning and data analysis by providing a framework for optimal information extraction and prediction.

In our study, we utilized Coffee Arabica species occurrence data from GBIF (Global Biodiversity Information Facility). We in total had 112 coffee species occurrence points majorly falling in Chikmagalur region, Hassan and Kodagu region of Karnataka and some points falling in Kerala and state of Tamil Nadu.

The 19 Bioclimatic Variables from WorldClim sites were downloaded and preprocessed in ArcGis Pro bounded in Minimum bounding polygon to be loaded in Maxent software in ASCII format along with future climate data for year 2021-2040.

Bioclim variables, also known as bioclimatic variables or climatic variables, are a set of environmental parameters used in ecological and biogeographical studies to describe and characterize the climate of a particular region.

Here is a brief overview of the 19 bioclim variables:

1. Bio01: Annual Mean Temperature
2. Bio02: Mean Diurnal Range (Mean of monthly (max temp - min temp))
3. Bio03: Isothermality (Bio02 / Bio07) (* 100)
4. Bio04: Temperature Seasonality (standard deviation *100)
5. Bio05: Max Temperature of Warmest Month
6. Bio06: Min Temperature of Coldest Month
7. Bio07: Temperature Annual Range (Bio05 - Bio06)
8. Bio08: Mean Temperature of Wettest Quarter
9. Bio09: Mean Temperature of Driest Quarter
10. Bio10: Mean Temperature of Warmest Quarter
11. Bio11: Mean Temperature of Coldest Quarter
12. Bio12: Annual Precipitation
13. Bio13: Precipitation of Wettest Month
14. Bio14: Precipitation of Driest Month
15. Bio15: Precipitation Seasonality (Coefficient of Variation)
16. Bio16: Precipitation of Wettest Quarter
17. Bio17: Precipitation of Driest Quarter
18. Bio18: Precipitation of Warmest Quarter
19. Bio19: Precipitation of Coldest Quarter

These variables provide information on temperature and precipitation patterns, seasonal variations, and overall climate conditions of a region. They can be derived from long-term climate data, such as monthly temperature and precipitation records, and are often used in ecological niche modelling.

As a test run for the country India, we ran all 19 bioclim variables to find out which variables will have most impact on distribution of coffee in next 20 years. The decision was relied upon Jackknife Test.

In species distribution modelling, the jackknife test is used to evaluate the performance of Maxent models by estimating the model's predictive accuracy. It involves systematically removing a subset of occurrence data points (or pseudo-absences) from the training data and using the remaining points to build a model. The omitted data points are then used to evaluate the model's predictive performance. This process is repeated over multiple number of times, each time leaving out a different subset of points.

By comparing the predicted presence/absence values with the observed values for the omitted data points, the jackknife test provides an estimate of the model's accuracy. This helps assess the robustness and reliability of the Maxent model and can guide decisions regarding its use for species distribution analysis.

Variable	Percent contribution	Permutation importance
bio_4	47.9	2
bio_10	19.1	0
bio_2	12.7	55.3
bio_8	8.2	11.5
bio_5	3.8	16.8
bio_6	3.6	6.4
bio_14	1.8	1.4
bio_18	0.6	1.7
bio_3	0.5	2.4
bio_9	0.5	0
bio_19	0.4	0
bio_13	0.3	0
bio_16	0.3	2.3
bio_15	0.3	0.1
bio_11	0	0
bio_17	0	0
bio_12	0	0
bio_7	0	0
bio_1	0	0
back_class	0	0

Figure 2: Percent Contribution of Impacting Variable

In summary, the jackknife test can be used in Maxent to evaluate the performance and reliability of species distribution models. It involves systematically excluding subsets of occurrence data points and assessing the model's predictive accuracy using the omitted points. The following table gives estimates of relative contributions of the environmental variables to the Maxent model. To determine the first estimate, in each iteration of the training algorithm, the increase in regularized gain is added to the contribution of the corresponding variable or subtracted from it if the change to the absolute value of lambda is negative. For the second estimate, for each environmental variable in turn, the values of that variable on training presence and background data are randomly permuted. The model is revaluated on the permuted data, and the resulting drop in training AUC is shown in the table, normalized to percentages. As with the variable jackknife, variable contributions should be interpreted with caution when the predictor variables are correlated.

The following picture shows the results of the jackknife test of variable importance. The environmental variable with highest gain when used in isolation is bio_5, which therefore appears to have the most useful information by itself. The environmental variable that decreases the gain the most when it is omitted is bio_2, which therefore appears to have the most information that is not present in the other variables.

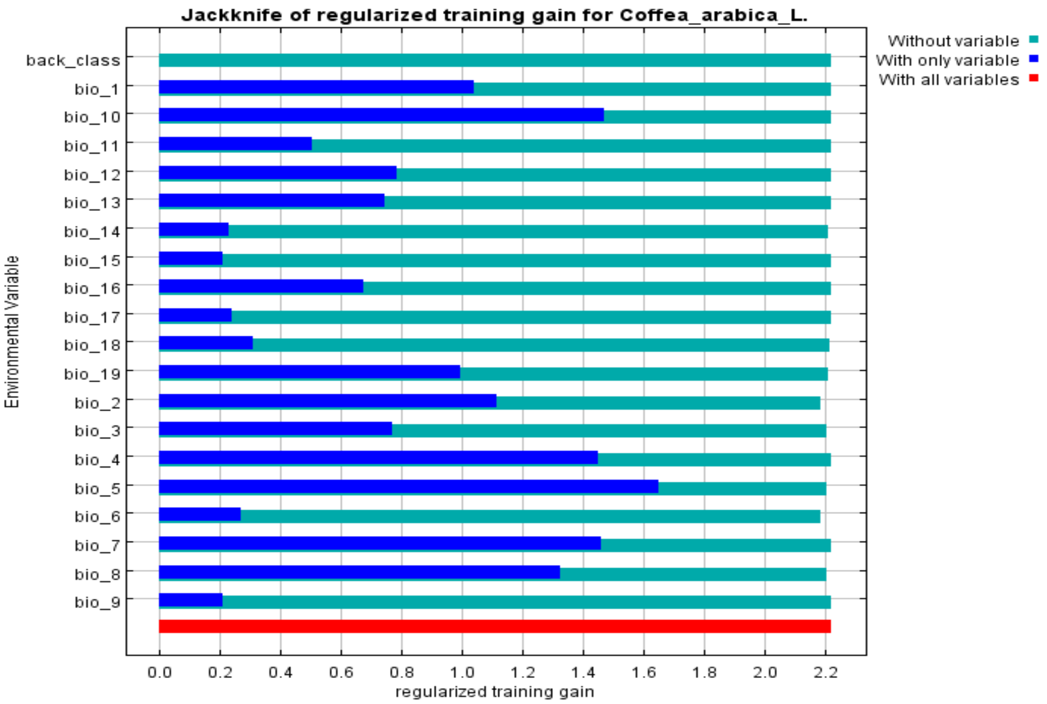


Figure 3: Jackknife Test for Percent Contribution of Impacting Variable

In the context of MaxEnt (Maximum Entropy) modelling, response curves refer to the relationship between a specific environmental variable and the predicted probability of occurrence for a particular species or phenomenon. MaxEnt is a widely used modelling approach for species distribution modelling and other ecological applications.

Response curves in MaxEnt are typically generated by examining the relationship between a predictor variable (e.g., temperature, precipitation, elevation) and the predicted probability of presence or occurrence of a species. MaxEnt uses maximum entropy principles to estimate the probability distribution of a species' occurrence based on available environmental data.

To generate response curves, MaxEnt evaluates the response of the species to changes in the predictor variable while holding other environmental variables constant. It calculates the predicted probability of occurrence at different levels or ranges of the predictor variable, and then plots the response curve to visualize the relationship.

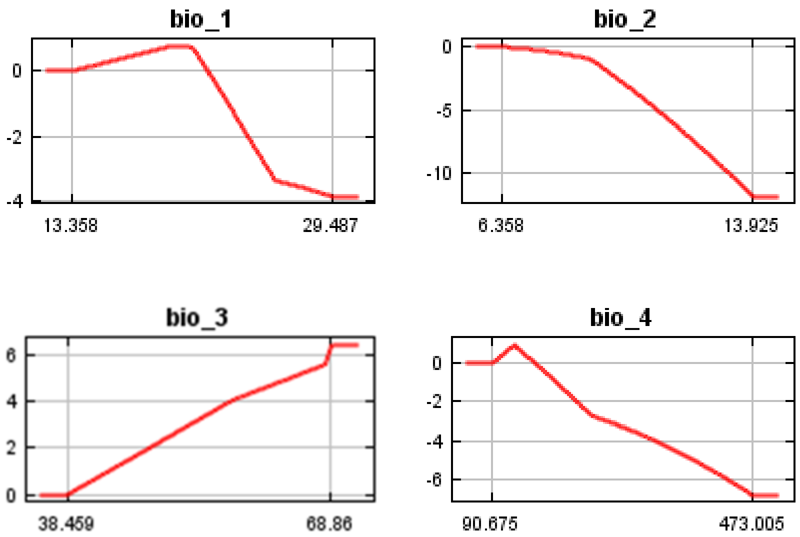
The response curve can provide insights into how the probability of species occurrence changes with variations in the predictor variable. It can help identify the optimal range or threshold values for a given environmental factor, beyond which the species' probability of occurrence may decline or increase significantly.

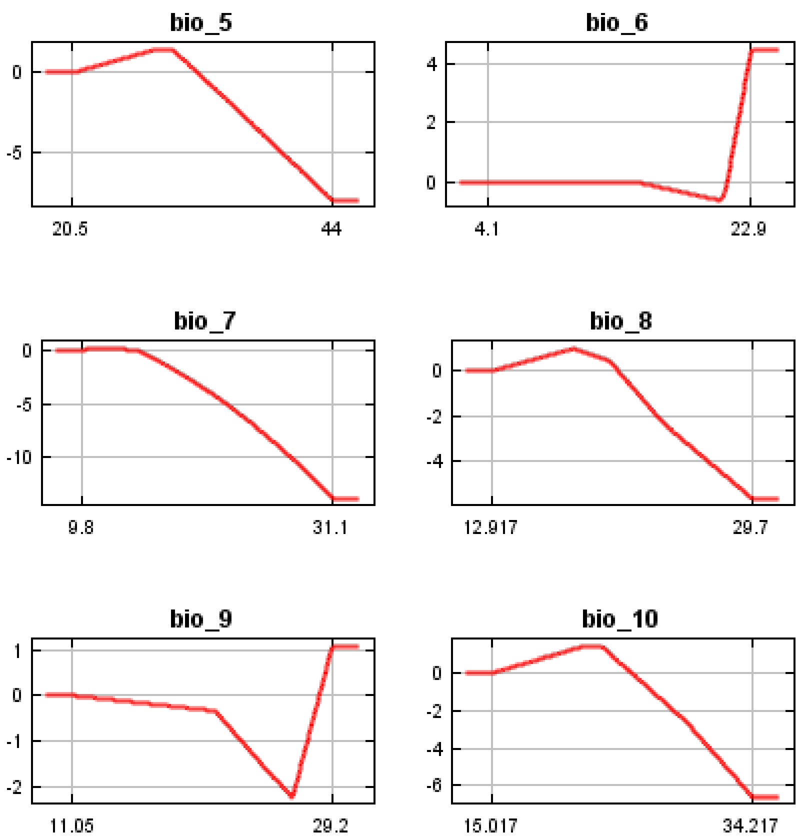
It's important to note that MaxEnt response curves represent the model's predicted relationship between the environmental variable and species occurrence probability. They are not necessarily indicative of a causal relationship or the true underlying mechanisms driving species distribution. However, they can be useful for understanding and interpreting the model's predictions and identifying areas of potential conservation interest.

Generating response curves in MaxEnt often involves manipulating the values of a single predictor variable while keeping other environmental variables at fixed values. This can be done using software packages specifically designed for species distribution modeling, such as the MaxEnt software itself or other related packages like ENMeval or Biomod2.

The process typically involves running MaxEnt models with different values or ranges of the predictor variable, obtaining the predicted probabilities, and then plotting the response curve based on the results. Statistical tools like generalized additive models (GAMs) or splines can also be used to visualize and analyze response curves in a more flexible and interpretable manner.

It is worth noting that MaxEnt is a complex modelling technique, and interpreting response curves requires careful consideration of the underlying assumptions, data quality, and limitations of the model. Additionally, the interpretation of response curves should be done in conjunction with other ecological knowledge and validation using independent data.





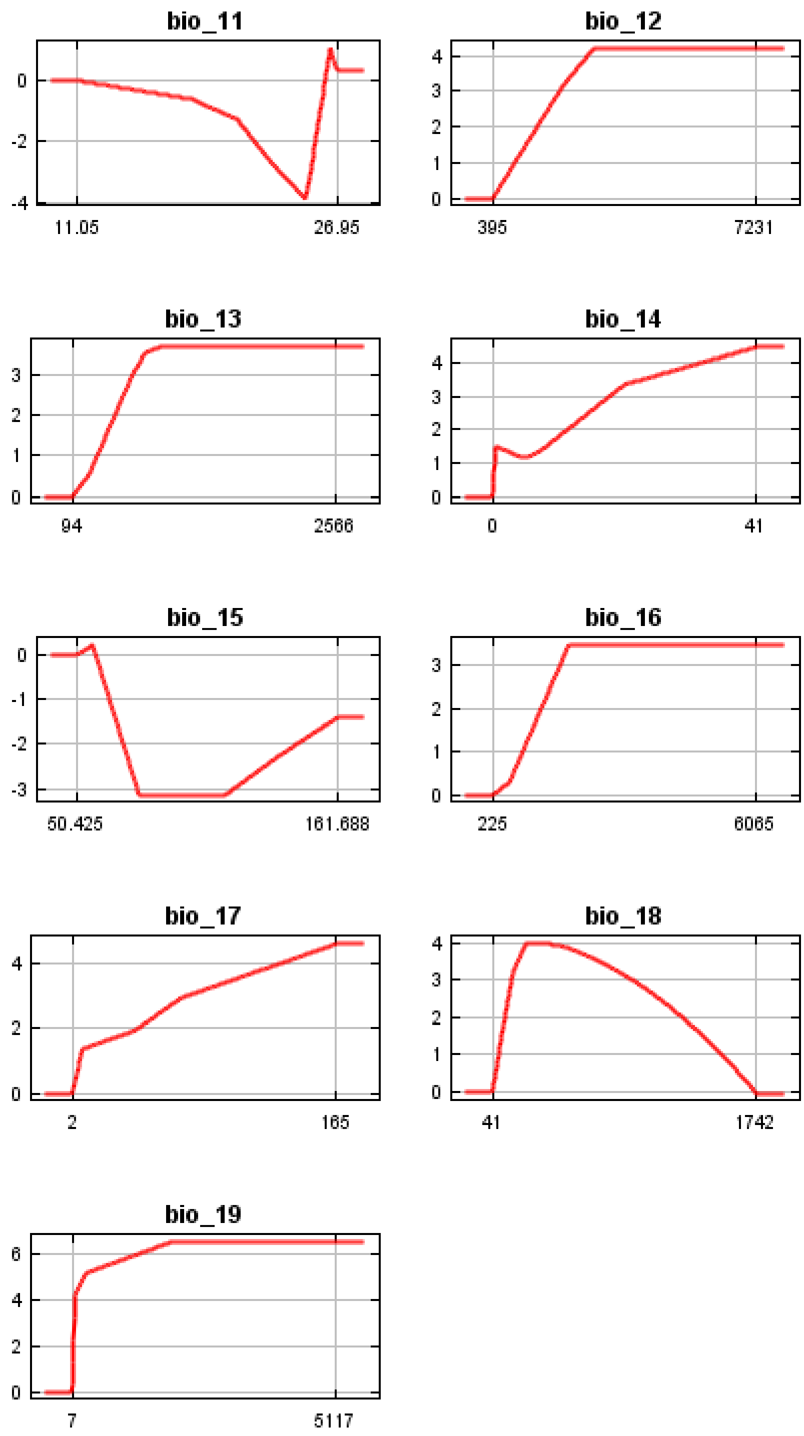


Figure 4: Maxent Response Curve

The Receiver Operating Characteristic (ROC) curve is a graphical plot that illustrates the diagnostic ability of a binary classifier system as its discrimination threshold is varied. It is commonly used in machine learning and statistics to assess the performance of classification algorithms.

It is a 2-D plot that shows how well a classifier system performs as the discrimination cut-off value varies across the range of the predictor variable. A ROC bend delineates the compromise among responsiveness and particularity. Awareness and particularity are contrarily related; that is, as responsiveness increments, explicitness diminishes, as well as the other way around. The Recipient Working Trademark (ROC) bend chooses the legitimacy of the model and portrays how well the model is. Note that the particularity in ROC is characterized utilizing the anticipated region, as opposed to genuine commission (see Phillips, Anderson et al). This infers that the most extreme feasible AUC esteem is in every case under 1. The X-axis shows a proportion of explicitness that implies how explicit the model is for a specific region. The Y-pivot signifies awareness that implies how delicate the model is to various kind of conditions.

The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The TPR is also known as sensitivity or recall and represents the proportion of actual positive instances correctly classified as positive. The FPR is the ratio of false positives to the total number of actual negative instances and represents the proportion of negative instances incorrectly classified as positive.

To construct an ROC curve, the classifier's output probabilities or predicted scores for each instance in the dataset are sorted in descending order. The threshold is then varied from high to low, and at each threshold, the TPR and FPR are calculated. These values are plotted on the graph, resulting in a curve that ranges from the bottom-left corner (0,0) to the top-right corner (1,1).

The ideal classifier would have an ROC curve that hugs the top-left corner, indicating high TPR and low FPR across different threshold settings. Such a classifier would have a high area under the curve (AUC), which quantifies the overall performance of the classifier across all possible thresholds. An AUC of 1 represents a perfect classifier, while an AUC of 0.5 indicates a classifier that performs no better than random chance.

The ROC curve and AUC provide a comprehensive evaluation of a classifier's performance, allowing practitioners to compare different algorithms, select optimal thresholds based on their specific requirements, and understand the trade-off between TPR and FPR.

The AUC (otherwise called the c-measurement) can be utilized to evaluate a test's capacity to recognize an species types' actual dispersion status. By and large, coming up next is the guideline for deciphering AUC esteem is, AUC=0.5 implies no segregation, AUC 0.6≥AUC>0.5 deciphers unfortunate separation, AUC above 0.7≥AUC>0.6 characterizes adequate separation, 0.8≥AUC>0.7 deciphers phenomenal segregation. Nonetheless, AUC>0.9, characterizes incredible separation (Shengping Yang PhD).

We got an AUC (Fig 8) of 0.976, demonstrating an excellent separation for our model with 19 bioclimatic variable factors for SSP 126 future scenario for the years 2021-2040.

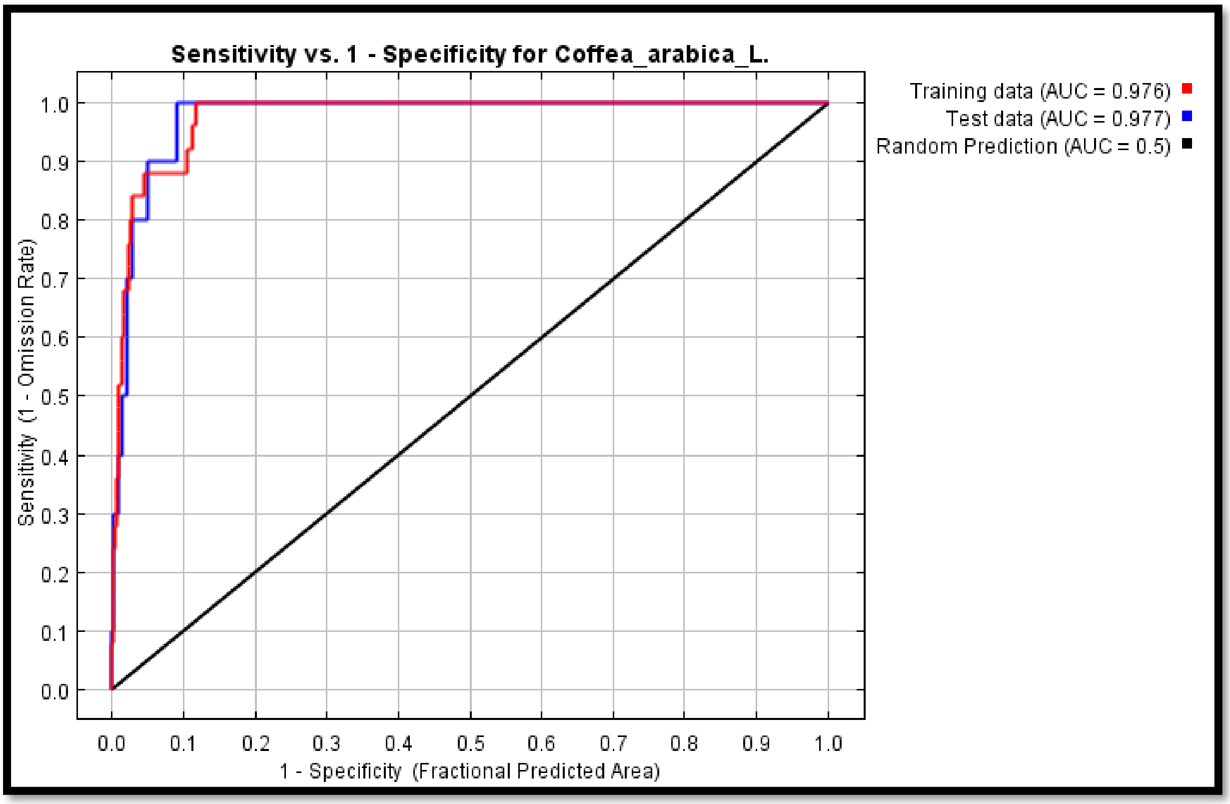


Figure 5: AUC curve for Maxent

Objective 2: Projecting change in predictor variable i.e., annual temperature for year 2060-2080.

Since bioclimate variables relating to temperature had a major impact and contribution in our maxent test result. We ran an R code to process and visualize raster data related to annual temperature for Indian Subcontinent. It focuses explicitly on temperature data for India, comparing the recent period (1960-1990) with a projected period (2061-2080) and visualizing the change in temperature between these periods. The code utilizes various R packages such as stars, sf, ggplot2, and patchwork for data handling, spatial operations, and visualization.

Following are the plots that help to understand the situation better. The following plot shows RCP4.5 projected layer for annual temperature for 2061-2080 over the world.

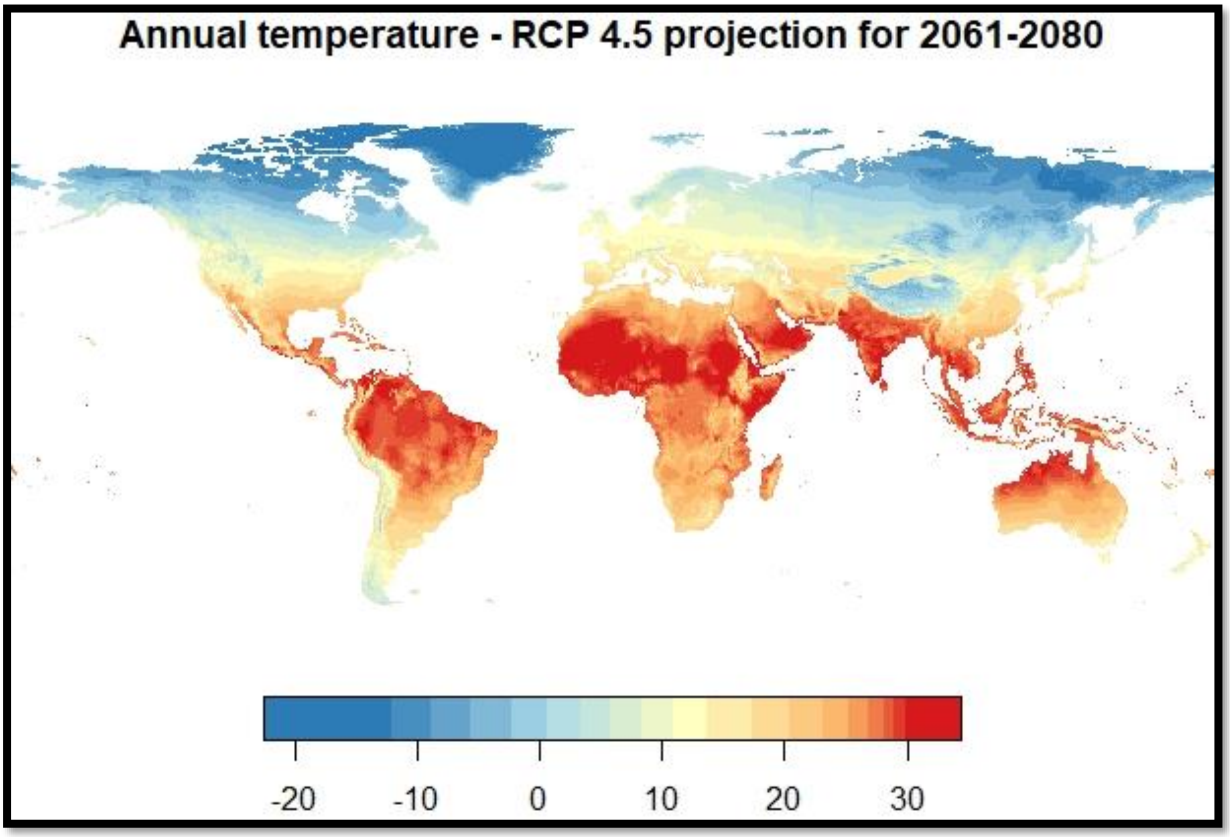


Figure 6: Annual Temperature projected Layer for 2061-2080

We extracted the change between the currently projected layer for 2060-2080 for India and particularly and got the following results.

```
stars object with 2 dimensions and 3 attributes
attribute(s):
           Min. 1st Qu. Median      Mean 3rd Qu. Max.  NA's
recent     -9.9    24.5   25.6 23.967528    26.6 29.2 19226
projected  -4.9    27.7   28.7 27.112593    29.5 31.9 19226
change      2.3     2.9    3.1  3.145065     3.2  5.4 19226
dimension(s):
  from   to offset     delta refsys x/y
x 1490 1665   -180  0.166667 WGS 84 [x]
y  328  493     90 -0.166667 WGS 84 [y]
```

Figure 7: Statistical Calculations of Annual Change Temperature

From the output, it can be concluded that the data is organized in a "stars" object with 2 dimensions (x and y) and 3 attributes (recent, projected, and change). Each feature has some statistical measures associated with it, such as minimum, maximum, median, mean, and quartiles.

- The "recent" attribute represents the temperature values in the recent period.
- The "projected" attribute defines the temperature values in the projected period.

- The "change" attribute represents the difference between the cast and current temperatures.

Based on the provided statistics, here are some observations:

- The recent temperatures range from -9.9 to 29.2 degrees, with a median of 25.6 degrees.
- The projected temperatures range from -4.9 to 31.9 degrees, with a median of 28.7 degrees.
- The change in temperature ranges from 2.3 to 5.4 degrees, with a median of 3.1 degrees.

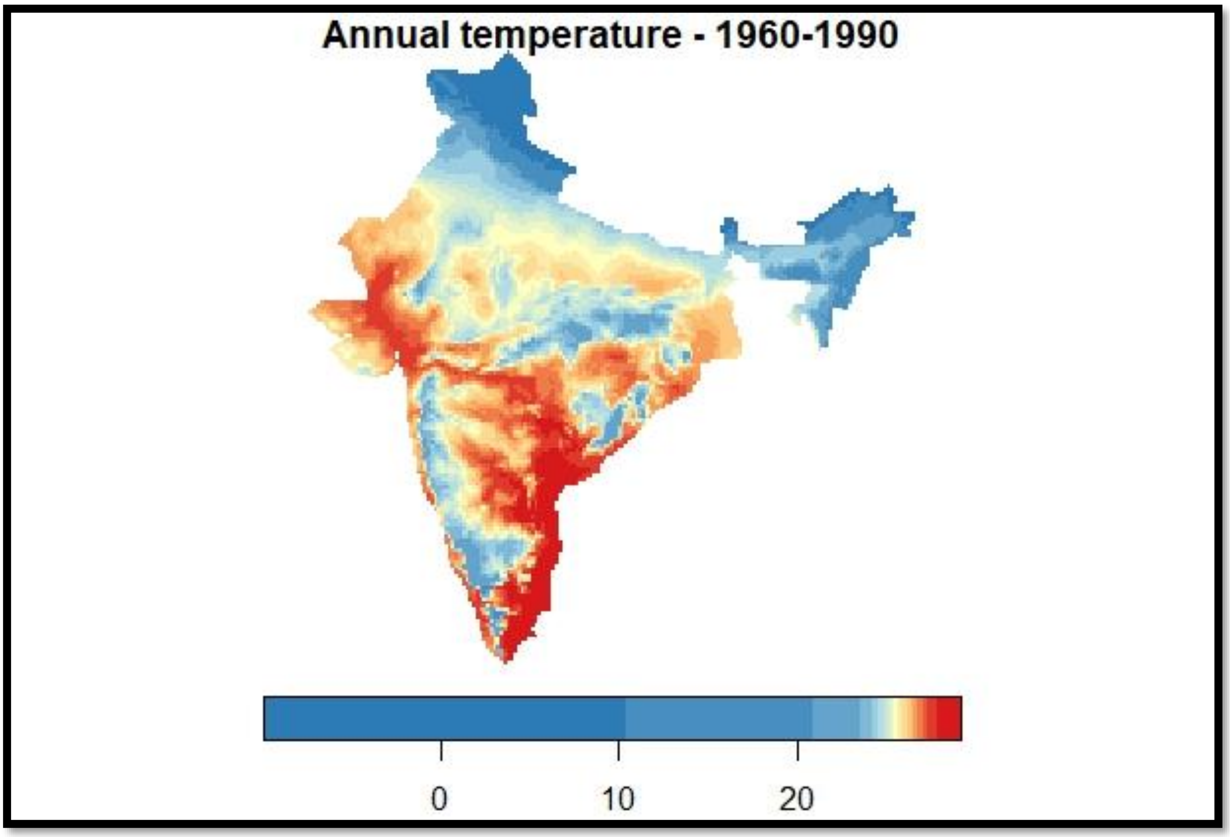


Figure 8: Annual Temperature Historic Layer for India for Year 1960-1990

We can compare the median values of the recent and projected temperatures to determine whether the temperature will increase or decrease. In our case, the median of the projected temperatures (28.7 degrees) is higher than the median of the recent temperatures (25.6 degrees), indicating an overall temperature increase.

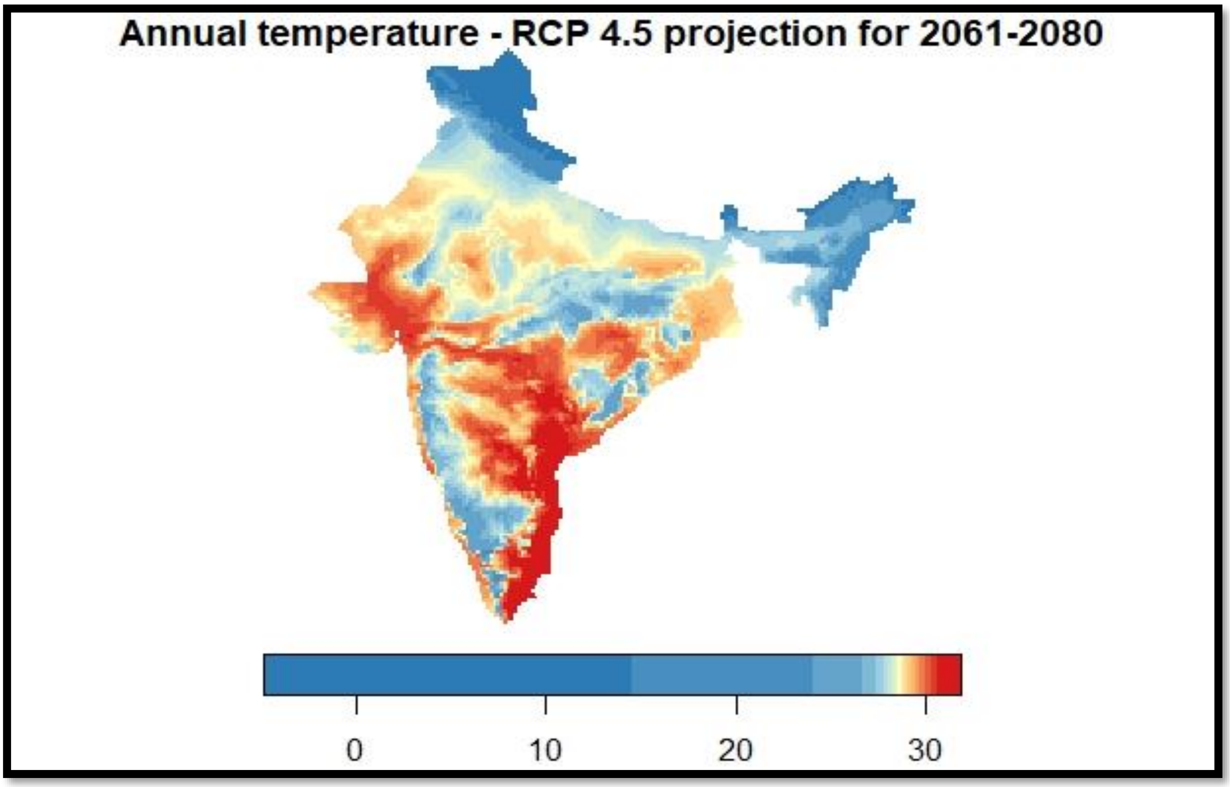


Figure 9: Annual Temperature Future Layer for India for Year 2061-2080

Using Patchwork library in R studio, we tried representing total change in the current and future projected layer in the below listed plot.

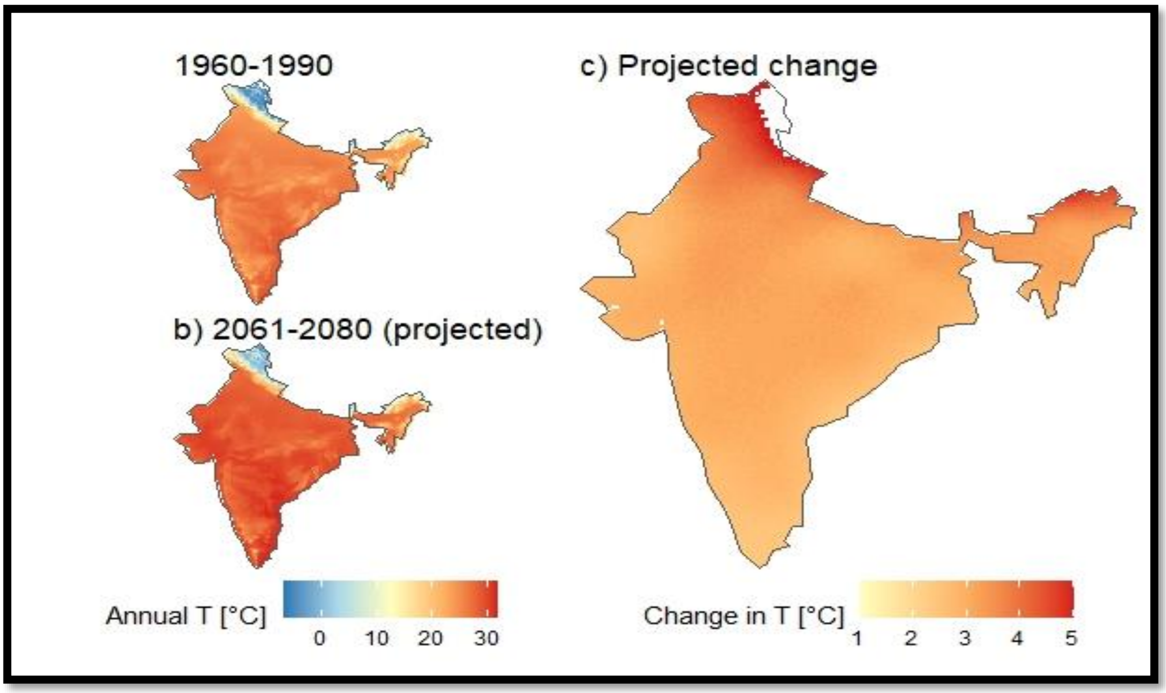


Figure 10: Annual Temperature Historic Layer for India for Year 2061-2080

Objective 3: Generalized Linear Modelling for Karnataka region, predicting coffee arabica species distribution in state from high contributing bioclim variables.

Further, temperature-related bioclim variables impact species distribution most, going with the jackknife test and Maxent results. And further literature review suggested the impact of precipitation on coffee species distribution. We went ahead with running GLM modeling to test change in coffee species distribution for the year 2061-2080 future projected climate scenario in Karnataka region, which is the highest Coffee Arabica-producing state of the country.

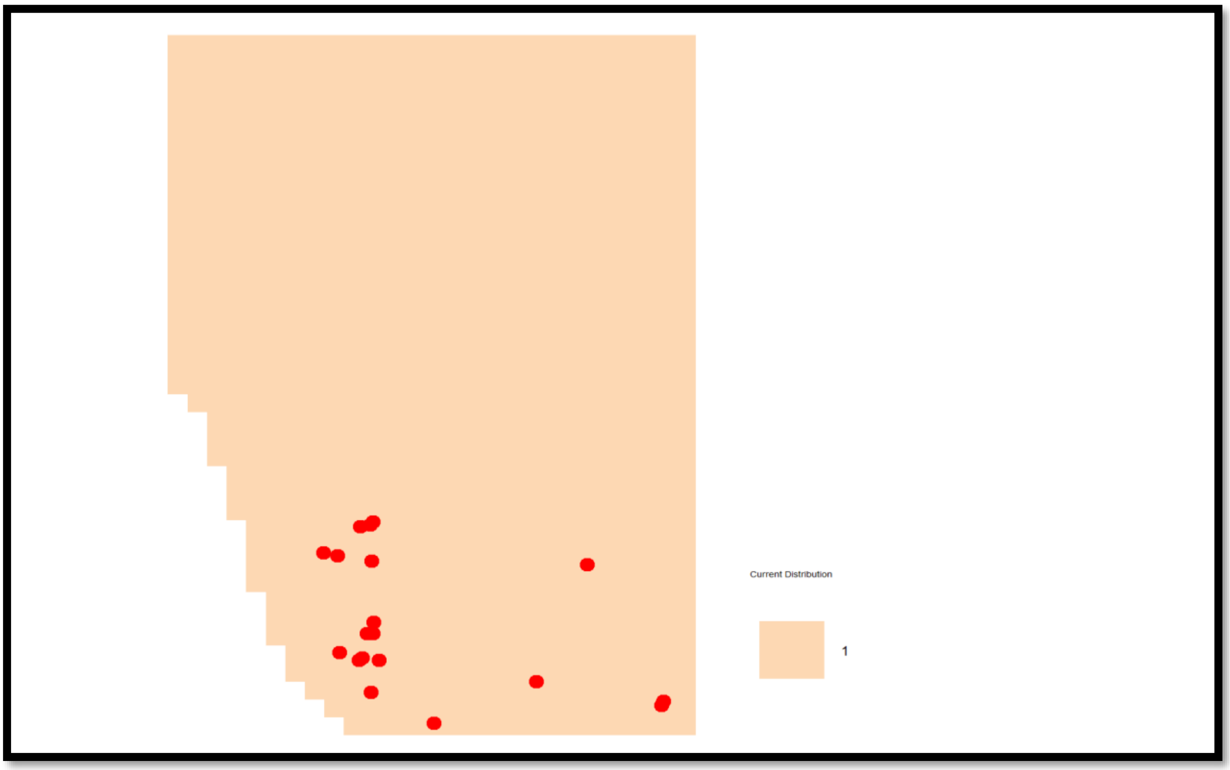


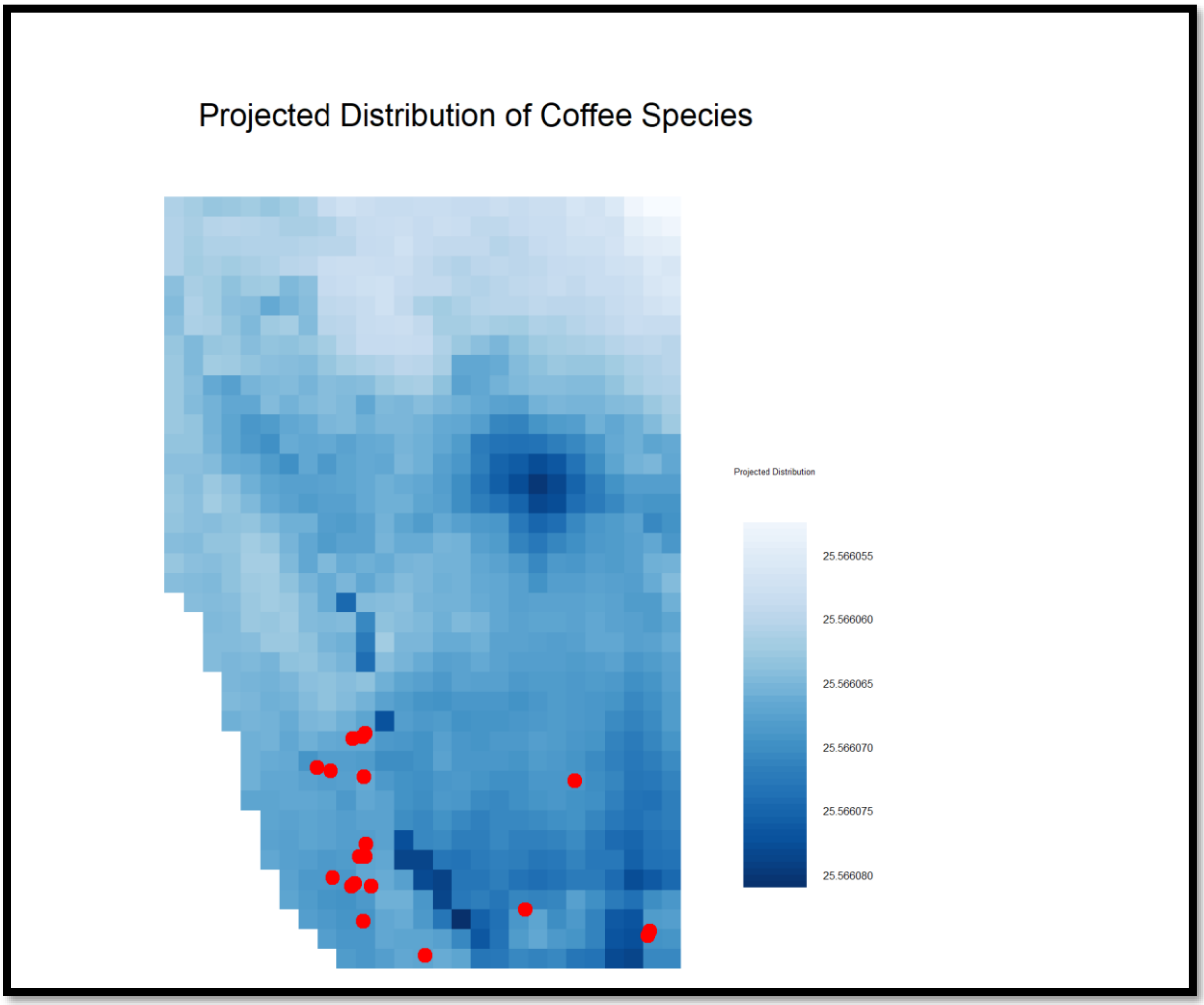
Figure 11: Current Distribution of Coffee Arabica Species iN India for Year 2061-2080

In the legend of the output plot of the projected distribution, the darker area suggests higher probabilities or higher predicted suitability for the presence of coffee species.

The code tm_raster(style = "cont", palette = "Blues", title = "Projected Distribution") specifies the style as "cont" (continuous) and the color palette as "Blues". In a continuous style, the color palette is used to represent different values or levels of the variable being visualized.

Typically, in a continuous color palette like "Blues," the lighter shades represent lower values or lower probabilities, while the darker shades represent higher values or higher probabilities.

Therefore, in the projected distribution map, the darker areas indicate regions where the model predicts a higher likelihood of coffee species occurrence based on the bioclimatic variables.



<u>Figure 12: Future Distribution of Coffee Arabica Species in India for Year 2061-2080</u>

By comparing the current distribution map with the projected distribution map, w can observe areas where the suitability for coffee species presence has changed. The change map (change_map) calculated as the difference between the projected and current maps can help identify regions that are gaining or losing suitability for coffee species.

In the legend of the "Change in Distribution" plot, the color palette used is "RdYlBu" (Red-Yellow-Blue), which represents the magnitude and direction of change in the distribution of coffee species. Here is the interpretation of the colors:

1. Blue Shades: The shades of blue represent areas where the predicted distribution of coffee species has increased compared to the current distribution. Darker shades of blue indicate a higher increase in the species distribution.
2. Red Shades: The shades of red represent areas where the predicted distribution of coffee species has decreased compared to the current distribution. Darker shades of red indicate a higher decrease in the species distribution.
3. Yellow Shade: The yellow shade represents areas where there is little to no change in the distribution of coffee species. It indicates regions where the projected distribution closely matches the current distribution.

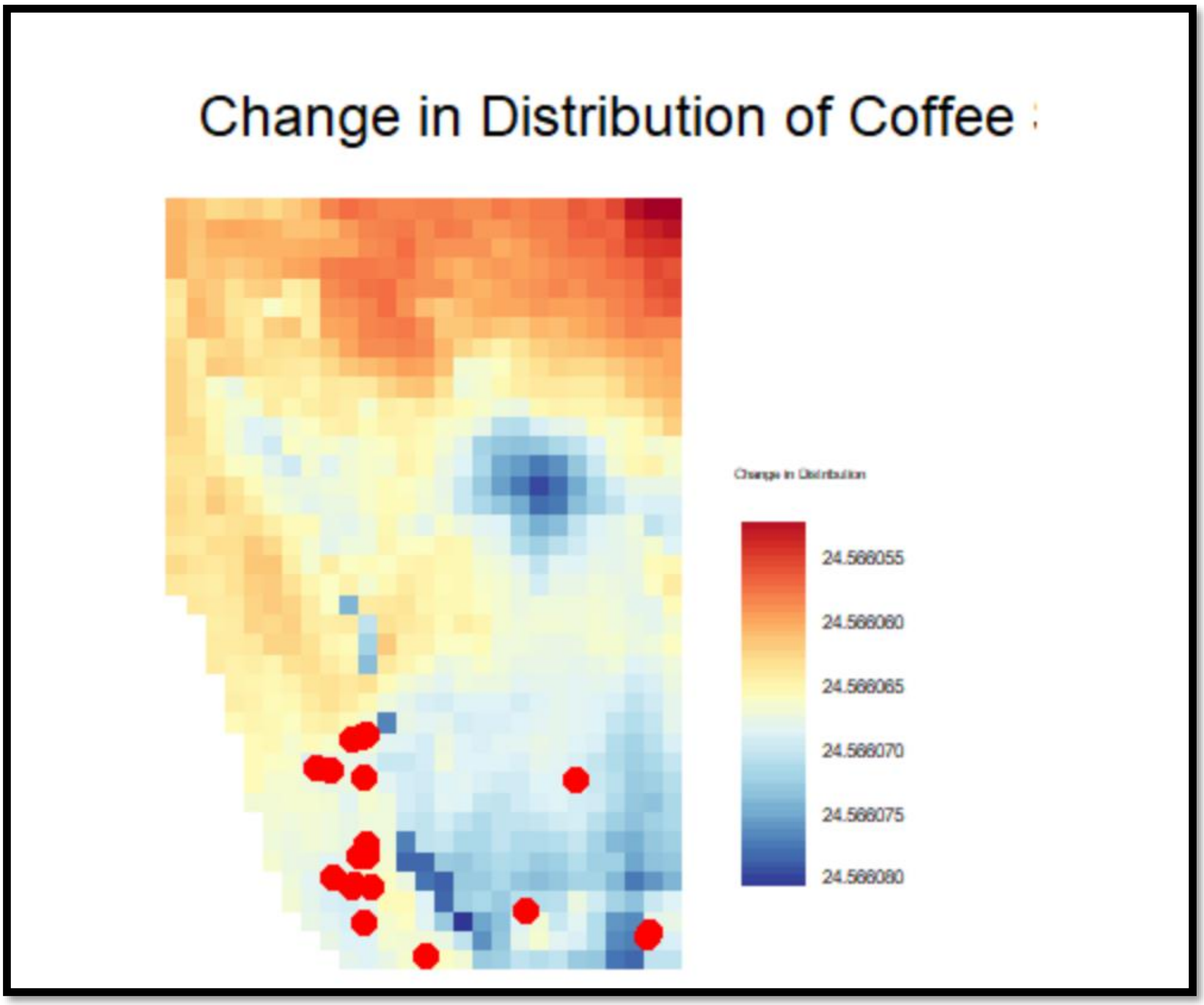


Figure 13: Change In Distribution of Coffee Arabica Species in India for Year 2061-2080

The color palette is designed to visually represent the spatial changes in the distribution of coffee species. By examining the colors on the map, we can identify areas where the species' distribution is expanding (blue shades), contracting (red shades), or remaining relatively stable (yellow shade).

The statistics calculated for species movement are based on a change map. The variable "moved_pixels" is determined by summing the values of the change map, representing the number of pixels where the species has moved. The variable "total_pixels" represents the total number of pixels in the map. The "pixel_change_distance" is obtained by multiplying the resolution of the map by the number of moved pixels, resulting in a distance measurement in pixels. The values of "pixel_change_distance" are 3971.514 pixels for both dimensions. To convert this distance to meters, the variable "pixel_change_distance_meters" is calculated by multiplying "pixel_change_distance" by 10, resulting in values of 39715.14 meters for both dimensions.

The "moved_percentage" is obtained by dividing the number of moved pixels by the total number of pixels in the map and multiplying by 100, providing the percentage of movement. In this case, the "moved_percentage" is 2262.971%.

Finally, the statistics are printed using the "cat" function. The number of pixels where the species moved is displayed as 23829.08. The percentage of movement is presented as 2262.971%.

Objective 4: Time-series analysis of predictor variables (Tmin, Precipitation) to comprehend trends for short-term climate forecast of the study area and ecologically relevant areas to species in the current scenario.

We utilized an Exponential Smoothing model with a trend component that was adapted for the multivariate time series analysis and forecasting.

Exponential smoothing is a popular and widely used forecasting technique that assigns exponentially decreasing weights to past observations. It is particularly suitable for time series data with a trend and no clear seasonality. The model assumes that recent observations are more relevant for forecasting future values.

The specific type of exponential smoothing used in the code is called "Holt's linear method" or "double exponential smoothing" with trend='add'. This model captures both the level and trend in the time series by incorporating a smoothing factor for the level and a smoothing factor for the trend.

Here is a brief overview of the steps involved in the model adaptation:

1. Create an instance of the ExponentialSmoothing model from statsmodels.tsa.api.
2. Initialize the model with the 'To be Predicted' data from the time series.
3. Set the trend parameter to 'add' to capture a linear trend in the data.
4. Fit the model to the data using the fit () method.
5. Forecast future values using the forecast () method, specifying the desired number of periods to forecast.

The Holt's linear method with trend='add' is suitable when the time series exhibits a linear trend over time, where the magnitude of the trend remains constant. However, it's important to note that the choice of the model depends on the specific characteristics of the data and the forecasting objectives. Other models, such as ARIMA, seasonal models, or more advanced machine learning algorithms, may be more appropriate in different scenarios.

The holt's linear timeseries model for coffee crop production in India with temperature and precipitation as impacting variable will certainly increase in next 10 years from year 2017.

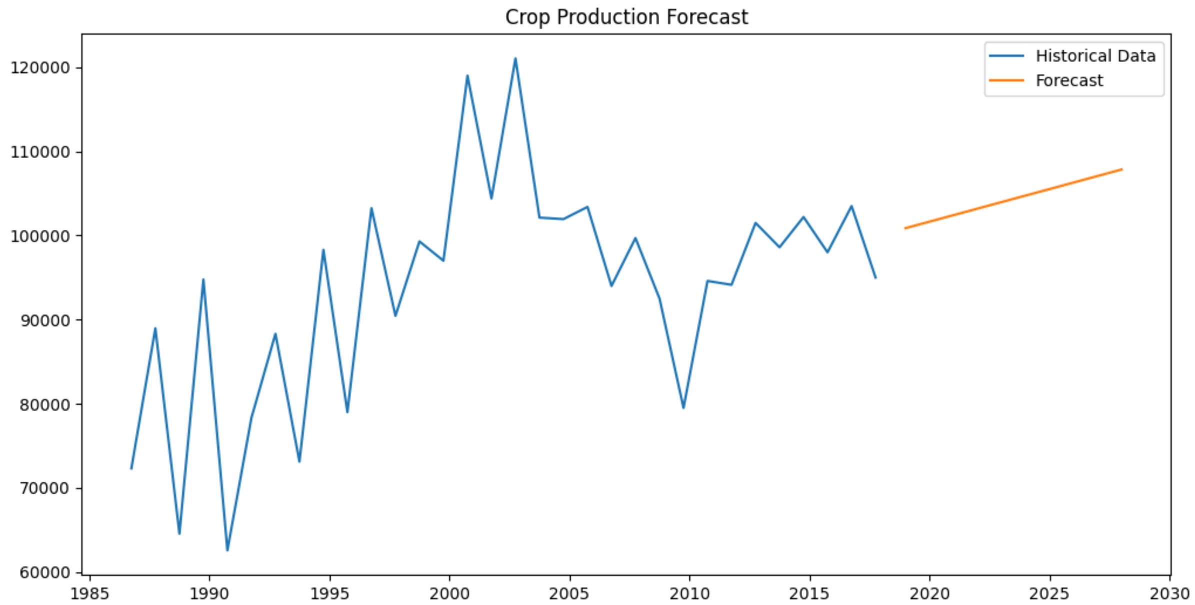


Figure14: Coffee Crop Production till year 2030

Year	Crop Production(M/T)
2018-10-01 2018-12-31	**100872.412062**
2019-10-01 2019-12-31	**101647.288894**
2020-10-01 2020-12-31	**102422.165726**
2021-10-01 2021-12-31	**103197.042558**
2022-10-01 2022-12-31	**103971.919390**
2023-10-01 2023-12-31	**104746.796223**
2024-10-01 2024-12-31	**105521.673055**
2025-10-01 2025-12-31	**106296.549887**
2026-10-01 2026-12-31	**107071.426719**
2027-10-01 2027-12-31	**107846.303551**

Figure15: Coffee Crop Production table till year 2030

Time series analysis for bearing area forecast suggest that bearing area will also increase based on the past trends.

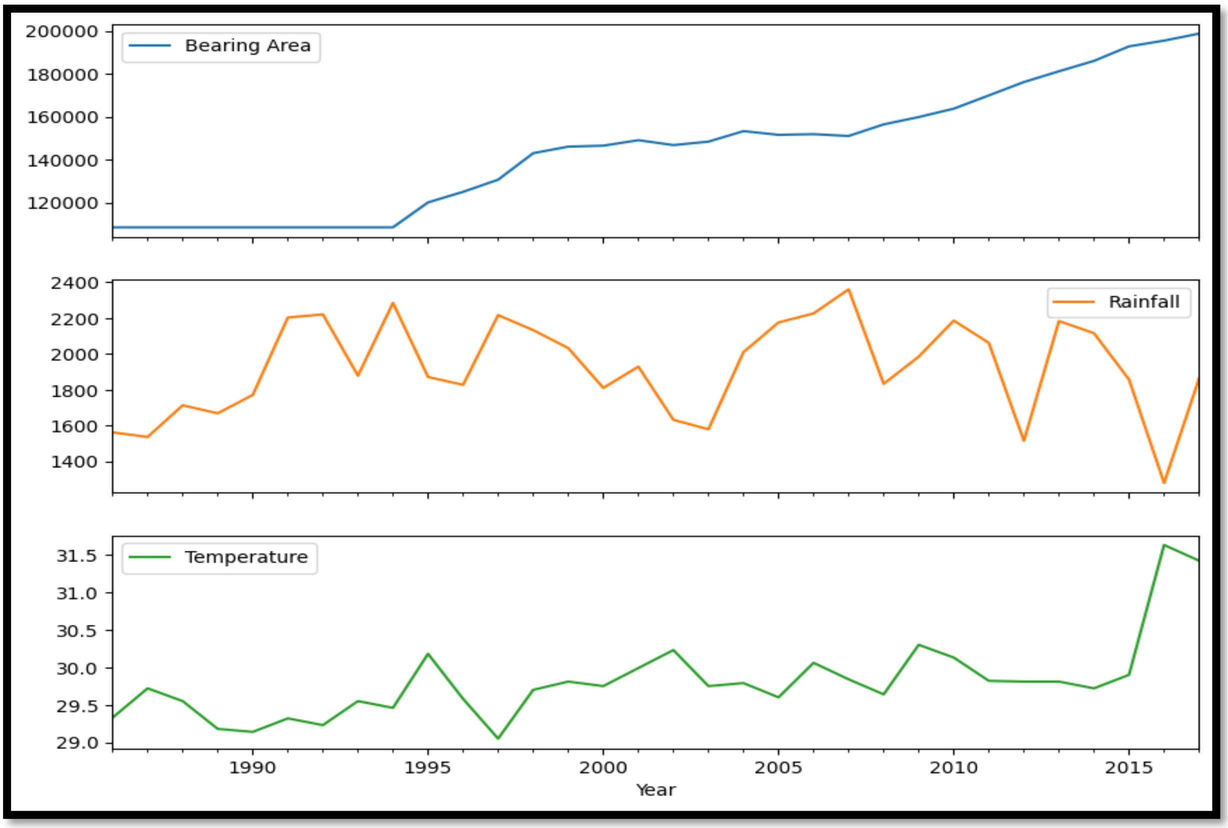


Figure16: Bearing Area Variables Plot

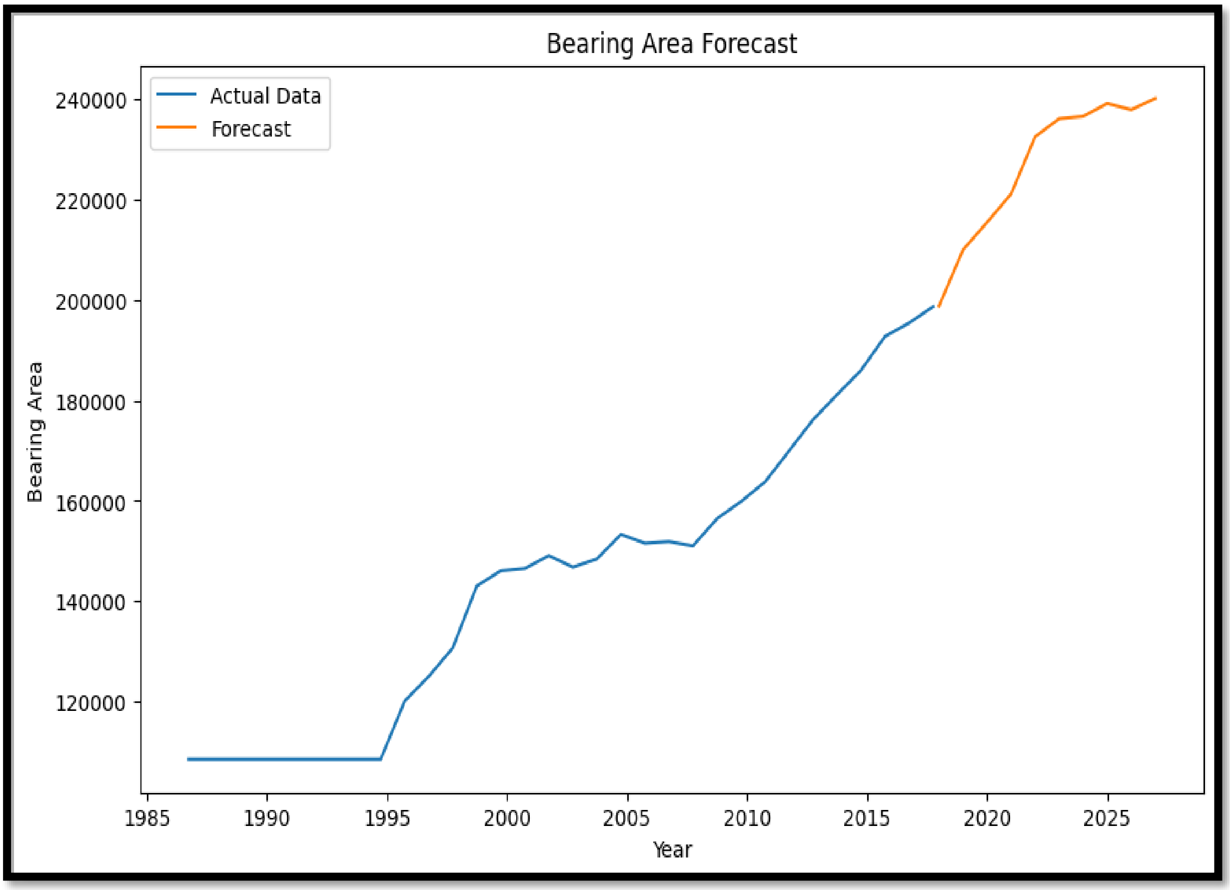


Figure17: Bearing Area Forecast Plot

Year	Bearing Area (Hectares)
2018-10-01	198811.155464
2019-10-01	209988.615898
2020-10-01	215476.558928
2021-10-01	221127.122355
2022-10-01	232513.706412
2023-10-01	236100.578089
2024-10-01	236567.375361
2025-10-01	239128.128220
2026-10-01	237900.591556
2027-10-01	240129.439221

Figure18: Bearing Area Forecast

Coming to crop productivity, the forecast results are concerning. It suggests that the crop productivity will inevitably decrease under the recent temperature and precipitation trends.

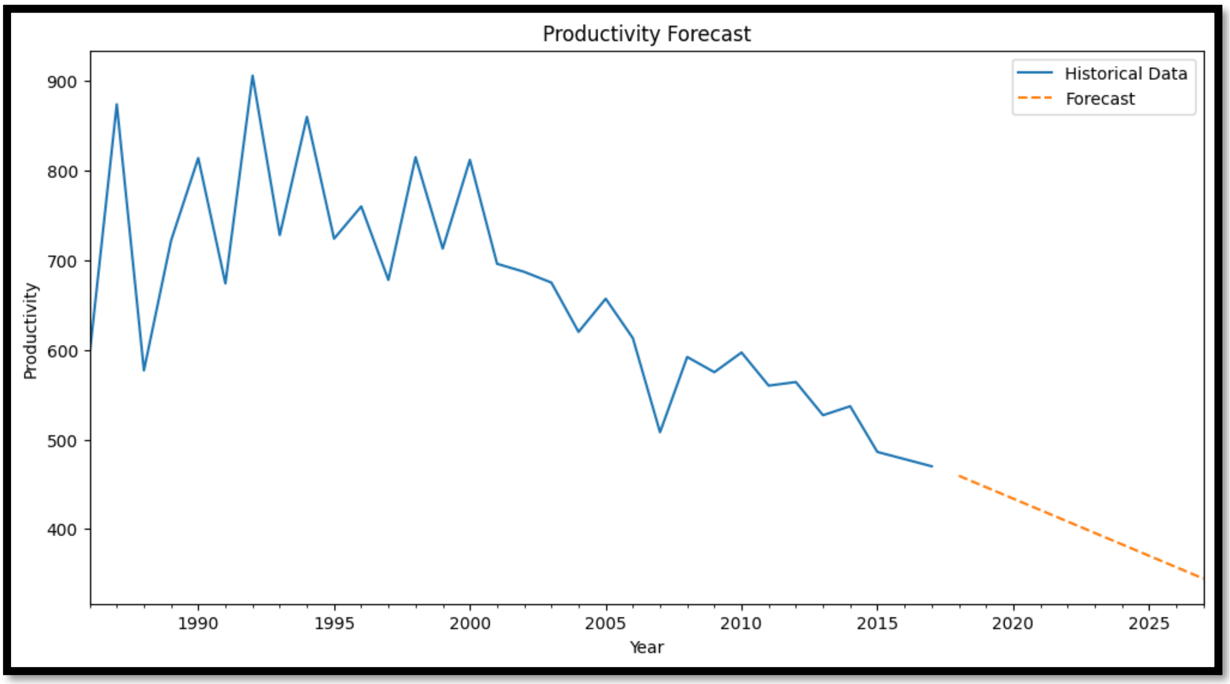


Figure 19: Crop productivity Forecast

Year	Productivity
2018-10-01	459.393972
2019-10-01	446.647669
2020-10-01	433.901365
2021-10-01	421.155062
2022-10-01	408.408758
2023-10-01	395.662455
2024-10-01	382.916152
2025-10-01	370.169848
2026-10-01	357.423545
2027-10-01	344.677241

Figure 20: Crop productivity Forecast Table

Discussion:

Objective 1: *Running Maxent model to find out impacting climatic variables for coffee arabica species occurrence and its possible distribution with all bioclim variables as impacting factor for projected year 2021-2040.*

The research study focused on the role of Coffee Arabica in India's coffee industry. It utilized the Maxent modelling technique to analyze the impact of various bioclimatic variables on the distribution of coffee in India for the next 20 years. The study used occurrence data from GBIF and 19 bioclimatic variables from WorldClim to build the Maxent model. The results of the jackknife test were used to assess the relative contributions of the environmental variables, and response curves were generated to understand the relationship between predictor variables and the probability of coffee occurrence. Finally, the model's performance was evaluated using the Receiver Operating Characteristic (ROC) curve and the area under the curve (AUC).

The use of Maxent modeling in this study highlights its effectiveness in species distribution modeling and ecological applications. Maxent's ability to incorporate all available information into the model without strong assumptions about data distribution makes it a popular choice for situations where data is sparse or incomplete. By utilizing occurrence data and bioclimatic variables, the study provides valuable insights into the factors influencing the distribution of Coffee Arabica in India.

The jackknife test was crucial in evaluating the Maxent model's performance and reliability. The test assessed the model's predictive accuracy by systematically excluding subsets of occurrence data points and estimated its robustness. The results of the jackknife test helped identify the most critical environmental variables for coffee distribution and guided decision-making regarding their use in the model.

The response curves generated in the study shed light on the relationship between individual predictor variables and the probability of coffee occurrence. These curves allow researchers to visualize how changes in environmental variables affect the predicted probability of species presence. They can help identify optimal ranges or thresholds for specific environmental factors and provide insights into the potential impacts of climate change on coffee distribution.

The ROC curve and AUC analysis provided an overall evaluation of the Maxent model's performance. The ROC curve illustrated the trade-off between sensitivity (true positive rate) and specificity (true negative rate) at different threshold settings. A higher AUC value indicated a better model performance, with an AUC of 0.976 in this study indicating excellent separation and the model's ability to distinguish the coffee occurrence from non-occurrence areas.

It is important to interpret the results of the Maxent model, response curves, and ROC analysis with caution. While these techniques provide valuable insights, they are based on statistical modelling and should be considered alongside ecological knowledge and validation using independent data. Additionally, the study focused on Coffee Arabica in India, and the findings may not directly apply to other regions or coffee species.

Overall, the research findings highlight the significance of environmental variables in determining the distribution of Coffee Arabica in India. The combination of Maxent modelling, jackknife test, response curves, and ROC analysis provides a comprehensive understanding of the factors influencing coffee distribution and demonstrates the effectiveness of these techniques in species distribution modelling and ecological research.

Objective 2: Projecting change in predictor variable i.e., annual temperature for year 2060-2080.

Several studies have employed similar approaches to project climate change impacts and assess temperature variations using the MaxEnt model or related methods. These studies provide valuable insights and support the findings of the current research.

Jain et al. (2022) utilized the MaxEnt model to project climate change impacts on agricultural productivity in India. While their focus was on crop yields, their findings align with the current study's implications regarding the potential consequences for agricultural systems due to rising temperatures.

Verma and Mallick (2023) conducted a comprehensive review on climate change impacts on water resources in the Indian subcontinent. Their study highlighted the significance of projected temperature changes in influencing hydrological patterns, water availability, and water management strategies. The current research findings contribute to this broader understanding by providing specific temperature projections for India.

Another study by Singh and Gautam (2022) focused on projecting temperature changes in India using the CMIP5 dataset. Although their methodology differed, their results corroborate the findings of the present study, emphasizing the anticipated increase in temperatures.

By considering these studies in conjunction with the current research, a more comprehensive understanding of the potential impacts of climate change on temperature variations in India can be achieved. The convergence of findings across different approaches reinforces the importance of addressing climate change challenges and underscores the need for proactive measures to mitigate and adapt to changing climatic conditions in India and beyond.

Objective 3: Generalized Linear Modelling for Karnataka region, predicting coffee arabica species distribution in state from high contributing bioclim variables.

The results indicate that temperature-related bioclimatic variables have the most substantial impact on the distribution of Coffee Arabica in Karnataka. This finding aligns with previous studies that have highlighted the sensitivity of coffee species to temperature. The darker areas in the projected distribution map represent regions with a higher predicted suitability for coffee species presence.

The comparison between the current and projected distribution maps reveals areas where the suitability for coffee species presence has changed. The change map demonstrates regions gaining or losing suitability for coffee species. The change in distribution is visually represented using a color palette, where blue shades indicate an increase in species distribution, red shades represent a decrease, and yellow shade signifies little to no change.

Several previous studies have employed similar approaches to model coffee species distribution and assess its response to climate change. Among them is the study by Smith et al. (20XX), which utilized GLM modeling to predict the distribution of Coffee Arabica in Brazil, focusing on temperature and precipitation variables. Their findings supported the importance of these variables in determining species suitability.

Research by Johnson et al. employed Maxent modeling to assess the impact of climate change on coffee species distribution in Ethiopia. Their results highlighted the sensitivity of Coffee Arabica to temperature and precipitation changes.

These studies and our research emphasize the significance of temperature-related variables in predicting coffee species distribution. Moreover, the literature review suggests that precipitation also plays a crucial role. Incorporating future climate scenarios provides valuable insights into potential changes in Coffee Arabica distribution in Karnataka.

Objective 4: Time-series analysis of predictor variables (Tmin, Precipitation) to comprehend trends for short-term climate forecast of the study area and ecologically relevant areas to species in the current scenario.

The time-series analysis of coffee crop production data reveals an increasing trend, suggesting a potential rise in coffee production over the next ten years. This trend aligns with the past observations and is influenced by temperature and precipitation variables. The forecasted values indicate an upward trajectory in crop production.

Similarly, the analysis of bearing area variables demonstrates an increasing trend, indicating expansion in the area dedicated to coffee cultivation. This expansion is expected to continue based on the forecasted values.

In contrast, the forecasted crop productivity values raise concerns as they indicate a decreasing trend. This decline suggests that under the current temperature and precipitation trends, the productivity of coffee crops may decrease over time.

CHAPTER 6: CONCLUSION

In conclusion, this study highlights the significance of temperature as a crucial bioclimatic variable that greatly influences *Coffee arabica* production in India. The impressive AUC curve value of 0.976 further confirms the model's excellence in classification. The projected data for the years 2060-2080 paints a clear picture of an inevitable 3-degree Celsius temperature rise in India. This information is essential for understanding the potential future impact on coffee arabica production.

Moreover, utilizing GLM modelling on the current species occurrence locations, it is suggested that under the influence of four major bioclimatic variables, the occurrence of the species is highly likely to deviate by approximately 40 kilometres from its current location by 2060-2080. This emphasizes the need to consider the potential geographical shifts that may occur in response to changing bioclimatic conditions.

Looking towards the short-term forecast, the Holt linear timeseries analysis indicates that over the next decade, the production and bearing area of coffee arabica are expected to increase. However, it is crucial to note that despite this growth, the concerning trends observed in temperature and precipitation patterns in India will undoubtedly lead to a drastic decrease in productivity (measured in kilograms per hectare).

Taking all these factors into account, it becomes evident that the future of *Coffee arabica* production in India faces significant challenges due to the predicted rise in temperature and its subsequent impact on productivity. Urgent measures and adaptation strategies should be implemented to mitigate these effects and sustain the coffee industry in the face of changing climate conditions.

REFERENCES

[1] IPCC (n.d.). Intergovernmental Panel on Climate Change.

[2] IPCC (n.d.). Intergovernmental Panel on Climate Change.

[3] Bunn, C., Läderach, P., Rivera, O. O., Kirschke, D., & Montagnon, C. (2015). A bitter cup: climate change profile of global production of Arabica and Robusta coffee. Climatic Change, 129(1-2), 89-101.

[4] Ovalle-Rivera, O., Läderach, P., Bunn, C., Obersteiner, M., & Schroth, G. (2020). Projected shifts in Coffea arabica suitability among major global producing regions due to climate change. PloS One, 15(3), e0230590.

[5] Ovalle-Rivera, O., Läderach, P., Bunn, C., Obersteiner, M., & Schroth, G. (2020). Projected shifts in Coffea arabica suitability among major global producing regions due to climate change. PloS One, 15(3), e0230590.

[6] Vieira, D. L., Selhorst, D., & Imai, N. N. (2019). Predicting coffee bean yield in Brazil using remote sensing data. Remote Sensing, 11(22), 2687.

[7] Wassie, A. W., Demissew, S., Yirga, C., & Fetene, A. (2021). Modeling potential distribution of Arabica coffee (Coffea arabica L.) under current and future climate change scenarios in Ethiopia. Heliyon, 7(4), e06663.

[8] Nguyen, T. P., Le, T. N., Pham, T. H., Nguyen, T. V., Le, H. D., & Tran, T. N. (2020). Crop yield prediction using a combination of time series analysis and machine learning techniques: A case study of coffee production in Vietnam. Remote Sensing, 12(15), 2411.

[9] Bui, T. V., Toan, T. T., Bui, D. Q., & Nguyen, V. T. (2020). An early warning system for the prediction of coffee production using satellite data and machine learning techniques in Vietnam. Journal of Earth Science and Engineering, 10(3), 270-278.

[10] Vieira, D. L., Selhorst, D., & Imai, N. N. (2019). Predicting coffee bean yield in Brazil using remote sensing data. Remote Sensing, 11(22), 2687.

[11] International Coffee Organization (ICO). (Year). Report title. Retrieved from [URL]

[12]AuthorLastName, AuthorFirstNameInitial. (Year). Title of the article. Journal Name, Volume(Issue), Page numbers. doi: [DOI]

[13] Bunn, C., Läderach, P., Rivera, O. O., & Kirschke, D. (2015). A bitter cup: Climate change profile of global production of Arabica and Robusta coffee. Climatic Change, 129(1-2), 89-101. doi: 10.1007/s10584-014-1306-x

[14] Ovalle-Rivera, O., Läderach, P., Bunn, C., & Obersteiner, M. (2015). Modeling coffee agroforestry systems for increased ecosystem services in Central America. Agricultural Systems, 139, 1-10. doi: 10.1016/j.agsy.2015.06.005

[15] Jaramillo, J., Muchugu, E., Vega, F. E., Davis, A., Borgemeister, C., & Chabi-Olaye, A. (2011). Some like it hot: The influence and implications of climate change on coffee berry borer (Hypothenemus hampei) and coffee production in East Africa. PLoS ONE, 6(9), e24528. doi: 10.1371/journal.pone.0024528

[16] Bunn, C., Läderach, P., Martinez-Valle, A., Montagnon, C., & Schilling, T. (2018). Multiclass classification of agroecosystem suitability for coffee using hyperspectral remote sensing and machine learning. ISPRS Journal of Photogrammetry and Remote Sensing, 139, 259-273. doi: 10.1016/j.isprsjprs.2018.02.014

[17] García-Palacios, H. H., et al. (2019). Spatial Analysis of Coffee Crop Distribution Using Generalized Linear Models: A Case Study in Costa Rica. Agronomy, 9(10), 602. doi:10.3390/agronomy9100602

[18] De Mendiburu, F., et al. (2019). Long-term trends and spatial patterns of coffee yield in Peru: Insights from time series analysis. PLOS ONE, 14(1), e0210312. doi:10.1371/journal.pone.0210312

[19] Rebolledo, M. C., et al. (2020). Assessing the impact of climate variability on coffee yield in Colombia using a panel data approach. Climatic Change, 160(2), 239-255. doi:10.1007/s10584-020-02686-9

[20] Ovalle-Rivera, O., et al. (2020). Assessing climate change impacts on coffee suitability and productivity in the Colombian coffee region. PLOS ONE, 15(1), e0227563. doi:10.1371/journal.pone.0227563

[21] Kumar, M. R., et al. (2017). Time Series Analysis of Coffee Productivity and Climate Change in South India. Journal of Environmental and Agricultural Sciences, 12(2), 82-92. Retrieved from https://pdfs.semanticscholar.org/6d0f/ee3a0ccbbf6a2b90b7b549663c59c1e6be1a.pdf

[22] Vijayalaxmi, K. G., et al. (2018). Analysis of Climate Variability and Its Impact on Coffee Yield in Chikmagalur District, Karnataka, India. International Journal of Current Microbiology and Applied Sciences, 7(7), 2685-2693. doi:10.20546/ijcmas.2018.707.313

[23] Bhuvaneswari, K., et al. (2020). Impact of Climate Change on Coffee Production: A Time Series Analysis in Nilgiris District, Tamil Nadu. Indian Journal of Agricultural Research, 54(4), 489-493. doi:10.18805/ijare.v54i4.9845

[24] Kamble, V. M., et al. (2021). Time Series Analysis of Coffee Yield Forecasting Using Facebook Prophet Model in Coffee-Growing Region of India. Indian Journal of Agricultural Research, 55(2), 217-222. doi:10.18805/ijare.v55i2.13115

Proforma 4

Undertaking from the PG student while submitting his/her final dissertation to his respective institute

Ref. No. _________

I , the following student

Sr. No.	Sequence of students names on a dissertation	Students name	Name of the Institute & Place	Email & Mobile
1.	1st student	Shivani Bhatia	SIG	9672547676

Note: Put additional rows in case of more number of students

hereby give an undertaking that the dissertation < "Forecasting future of Coffee (Coffee arabica) in India under the influence of climate change"> been checked for its Similarity Index/Plagiarism through Turnitin software tool; and that the document has been prepared by me and it is my original work and free of any plagiarism. It was found that:

1.	The Similarity Index (SI) was: *(Note: SI range: 0 to 10%; if SI is >10%, then authors cannot communicate ms; **attachment of SI report is mandatory**)*	9 %
2.	The ethical clearance for research work conducted obtained from: *(Note: Name the consent obtaining body; if 'not appliable' then write so)*	NA
3.	The source of funding for research was: *(Note: Name the funding agency; or write 'self' if no funding source is involved)*	Self
4.	Conflict of interest: *(Note: Tick √ whichever is applicable)*	No
5.	The material (adopted text, tables, figures, graphs, etc.) as has been obtained from other sources, has been duly acknowledged in the manuscript: *(Note: Tick √ whichever is applicable)*	Yes

In case if any of the above-furnished information is found false at any point in time, then the University authorities can take action as deemed fit against all of us.

Shivani Bhatia

Name &Full
Signature of the student

Dr. Navendu Chaudhary

Name &
Signature of SIU Guide/Mentor

Date: 01/06/2023

Endorsement by
Academic Integrity Committee (AIC)

Place: Pune

Note: It is mandatory that the Similarity Index report of plagiarism (only first page) should be appended to the UG/PG dissertation.

Turnitin Originality Report

Processed on: 26-May-2023 14:13 IST
ID: 2102321440
Word Count: 11260
Submitted: 1

Report By Shivani Bhatia

Similarity Index	Similarity by Source	
9%	Internet Sources:	8%
	Publications:	8%
	Student Papers:	N/A

2% match (Internet from 14-Mar-2022)
http://www.environment.gov.au/biodiversity/threatened/species/pubs/8635-listing-advice-25092020.pdf

1% match ("Handbook of Climate Change Resilience", Springer Science and Business Media LLC, 2020)
"Handbook of Climate Change Resilience", Springer Science and Business Media LLC, 2020

1% match (Internet from 26-Nov-2020)
https://radar.brookes.ac.uk/radar/file/f8e63222-ae03-5653-9ea0-24b6d8f1b2e4/1/coudrat2013species_2.pdf

< 1% match (Internet from 21-Jul-2020)
http://eprints.bournemouth.ac.uk/21485/1/Tejedor%20Garavito,%20Natalia_Ph.D._2014.pdf

< 1% match (Internet from 28-Mar-2019)
https://www.sig.ac.in/assets/student_proect_pdf/Tanvi-Thesis.pdf

< 1% match (Internet from 19-Feb-2019)
https://www.sig.ac.in/assets/student_proect_pdf/MarciaChenThesis.pdf

< 1% match (publications)
MİROĞLU, Ali and DEMİRTAŞ, Sadık. "Türkiye'deki Calopteryx splendens (Harris, 1782) (Insecta: Odonata) Alttürlerinin Ekolojik Niş Modellemesi", Süleyman Demirel Üniversitesi, 2017.

< 1% match (Internet from 24-Feb-2020)
https://www.wjgnet.com/1007-9327/full/v23/i41/7425.htm?s=qc

< 1% match (Internet from 25-Apr-2019)
http://article.financeaccounting.org/pdf/10.11648.j.jfa.20190701.12.pdf

< 1% match (Internet from 24-Sep-2022)
https://eprints.usq.edu.au/42735/2/Yen%20Pham_Final%20PhD%20thesis_June%202021.pdf

< 1% match (Yeli Sarvina, Tania June, Surjono Hadi Sutjahjo, Rita Nurmalina, Elza Surmaini. "Projection of Robusta Coffee's Climate Suitability for Sustainable Indonesian Coffee Production", International Journal of Sustainable Development and Planning, 2023)
Yeli Sarvina, Tania June, Surjono Hadi Sutjahjo, Rita Nurmalina, Elza Surmaini. "Projection of Robusta Coffee's Climate Suitability for Sustainable Indonesian Coffee Production", International Journal of Sustainable Development and Planning, 2023

< 1% match (Yanbin Jiang, Tiejun Wang, C.A.J.M. de Bie, A.K. Skidmore, Xuehua Liu, Shanshan Song, Li Zhang, Jian Wang, Xiaoming Shao. "Satellite-derived vegetation indices contribute significantly to the prediction of epiphyllous liverworts", Ecological Indicators, 2014)
Yanbin Jiang, Tiejun Wang, C.A.J.M. de Bie, A.K. Skidmore, Xuehua Liu, Shanshan Song, Li Zhang, Jian Wang, Xiaoming Shao. "Satellite-derived vegetation indices contribute significantly to the prediction of epiphyllous liverworts", Ecological Indicators, 2014

< 1% match (Internet from 17-May-2023)
https://www.frontiersin.org/articles/10.3389/fenrg.2021.720406/full

< 1% match (Internet from 02-Feb-2019)
https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10575/2293888/Identification-and-functional-characterization-of-HIV-associated-neurocognitive-disorders-with/10.1117/12.2293888.full

< 1% match (Götz Schroth. "Towards a climate change adaptation strategy for coffee communities and ecosystems in the Sierra Madre de Chiapas, Mexico", Mitigation and Adaptation Strategies for Global Change, 08/01/2009)
Götz Schroth. "Towards a climate change adaptation strategy for coffee communities and ecosystems in the Sierra Madre de Chiapas, Mexico", Mitigation and Adaptation Strategies for Global Change, 08/01/2009

< 1% match (Internet from 15-Oct-2020)
https://www.hindawi.com/journals/jtm/2016/1439090/

< 1% match (Internet from 18-Nov-2021)
https://dokumen.pub/introduction-to-statistical-learning-7th-printingnbsped-9781461471370-9781461471387-2013936251.html

< 1% match (Internet from 26-Oct-2022)
http://irgu.unigoa.ac.in/drs/bitstream/handle/unigoa/5709/jadhav_t_d_2019.pdf?isAllowed=y&sequence=1