IMPACT OF RAINFALL VARIATION ON FOREST DIEBACK IN HORTON PLACE NATIONAL PARK SRI LANKA.

A dissertation submitted to the
Department of Computer Science and Informatics,
Faculty of Applied Sciences,
Uva Wellassa University
in partial fulfillment of the requirements for the award of the
Degree of Computer Science and Technology

by

ANGULGAMUWA A.S.D.
DISSANAYAKA M.D.S.

Department of Computer Science and Informatics
Faculty of Applied Sciences
Uva Wellassa University, Sri Lanka
September 2024

DECLARATION

We do hereby declare that the work reported in this dissertation was exclusively carried out by me under the supervision of **Mr. S. J. M. D. P. Samarakoon, Prof. E.M.U.W.J.B. Ekanayake** and **Prof. E.P.S.K. Ediriweera** . It describes the results of my own independent research except where due reference has been made in the text. No part of this dissertation has been submitted earlier or concurrently for the same or any other degree.

………………………… ……………………………….

Angulgamuwa A.S.D. Dissanayaka M.D.S.

Date: 17/09/2024 Date: 17/09/2024

We endorse the declaration by the candidate(s).

……………………………. …………………………….

Supervisor
Mr. S. J. M. D. P. Samarakoon

Date: ………………….

Co-Supervisor
Prof. E.M.U.W.J.B. Ekanayake

Date: ………………….

……………………………. …………………………….

External Supervisor
Prof. E.P.S.K. Ediriweera
Date: ………………….

Mr. M. Janotheepan
Research Coordinator
Date: ………………….

Acknowledgement

I would like to express my deepest gratitude to everyone who contributed to the successful completion of this research project. First and foremost, I wish to thank my **primary supervisor, Mr. S. J. M. D. P. Samarakoon**, for his invaluable guidance, encouragement, and support throughout the duration of this project. His expertise and insights were instrumental in shaping the direction of this research, and I am truly grateful for his mentorship. I would also like to extend my sincere thanks to **Prof. E.M.U.W.J.B. Ekanayake, my co-supervisor**, for his consistent support and advice, particularly in helping me navigate the complexities of the research. His critical feedback and suggestions significantly contributed to the depth of the analysis.

My appreciation also goes to **Prof. E.P.S.K. Ediriweera, my external supervisor**, for his expert advice and thoughtful guidance. His input, particularly during the data analysis and predictive modeling phases, was invaluable in ensuring the rigor of the research. I am deeply grateful to the *Department of Meteorology, Sri Lanka*, for providing the climatic data that formed the foundation of this study. Without their cooperation, this research would not have been possible.

Finally, I wish to thank my family and friends for their unwavering support and encouragement, which gave me the strength and motivation to complete this project.

Contents

Contents

Abstract

This study examines the relationship between climatic variables (rainfall and temperature) and forest dieback in Horton Plains, Sri Lanka, over the period 2000-2020. The objectives were to identify long-term trends in temperature, rainfall, and forest dieback, quantify the correlation between these variables, and develop a predictive model to estimate forest dieback based on climatic conditions. Monthly meteorological data from the Department of Meteorology and forest dieback data derived from satellite imagery using Google Earth Engine were used for the analysis. Trend analysis revealed a gradual increase in temperature and high variability in rainfall, with no clear trend in the latter. Forest dieback showed fluctuations, increasing particularly during dry years. The correlation analysis showed a weak positive relationship between temperature and forest dieback (0.023) and a weak negative relationship between rainfall and dieback (-0.29), suggesting that drought conditions contribute to increased dieback. To predict forest dieback, several machine learning models were tested, including Support Vector Regression (SVR), Random Forest, Gradient Boosting, and Neural Networks. The Random Forest model performed best, with a Mean Absolute Error (MAE) of 257.45 on the test set. The model identified rainfall as the most important factor influencing forest dieback, with temperature lag also contributing significantly. These findings highlight the critical role of rainfall variability in forest health, suggesting that droughts are a key driver of dieback. The results can inform conservation strategies and provide a tool for predicting forest dieback events in Horton Plains and other tropical ecosystems.

Keywords: *Forest Dieback; Horton Plains; Temperature Trends; Random Forest; Machine Learning Models.*

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Term
GEE	Google Earth Engine
NDVI	Normalized Difference Vegetation Index
SVR	Support Vector Regression
RF	Random Forest
GBM	Gradient Boosting Model
MLP	Multi-Layer Perceptron (Neural Network)
MAE	Mean Absolute Error
MSE	Mean Squared Error
CSV	Comma-Separated Values
GIS	Geographic Information System
DM	Department of Meteorology
RF_avg	Monthly Average Rainfall
T_avg	Monthly Average Temperature
NASA	National Aeronautics and Space Administration
sq. meters	Square Meters

Introduction

Horton Plains National Park, established in 1988 in Sri Lanka's central highlands, sits at 2,100-2,300 meters above sea level and features montane grasslands and cloud forests. Known for its high biodiversity, the park is home to several endemic species and serves as a key watershed for major rivers like the Mahaweli, Kelani, and Walawe. It is a popular tourist spot, located near Ohiya, Ohiya Gap, and Nuwara Eliya. Discovered in 1834 by Lt. William Fisher and Lt. Albert Watson, the area was named after Sir Robert Wilmot-Horton, then Governor of Ceylon. Locally known as "Maha Eliya Thenna," prehistoric relics suggest it was inhabited long before British exploration.

The park supports endangered wildlife and plant species, including 750 plant species, 24 mammal species, 87 bird species, and dozens of orchids. Notable animals include Sambar deer, wild boars, and, on rare occasions, leopards. The Horton Plains Slender Loris, one of the world's most endangered primates, also resides herewith over 2,000 mm of annual rainfall, the park experiences frequent cloud cover, an average temperature of 13°C, and occasional ground frost in February. Rain occurs year-round, with a dry season from January to March. The plains feed several rivers and are dotted with lakes and waterfalls.

Figure 1 Horton Plains Sri Lanka

1.1 Project Background

The problem of forest dieback is a major ecological concern these days, particularly in sensitive environments like Sri Lanka's Horton Plains National Park Sri Lanka. So, Researchers and conservationists are concerned about this because it's affecting the overall health and sustainability of Horton Place's environment. They're trying to figure out why this is happening. Existing research in Horton Plains has found several possible causes of this dieback, ranging from natural factors like climate stress and soil conditions to man-made factors like pollution and animal impact. (Dieback, n.d.)

Forest dieback is a significant ecological concern, particularly in sensitive areas like Sri Lanka's Horton Plains National Park. Researchers are investigating its causes, which range from climate stress and soil conditions to pollution and animal impact. A 2005 study by Turner and Lambert (Turner & Lambert, 2005) found that toxic elements like high levels of aluminum and nutrient imbalances in the soil were contributing to forest decline. Similarly, Gunadasa and Yapa (2015)(Gunadasa & Yapa, 2015) showed that aluminum toxicity harms tree root systems and affects nutrient absorption due to soil pH changes.

Climate change, especially rising temperatures and reduced rainfall, exacerbates forest dieback, with droughts and heat stress killing trees globally. Forests in drier climates are particularly vulnerable. Fungal diseases, often influenced by climate shifts and global trade, further threaten forest health. Ghelardini (2016) (Ghelardini et al., 2016)highlighted how fungi and insect interactions, as well as monoculture plantations, worsen these diseases. Addressing these factors is crucial to mitigating dieback. might help reduce their impact on forests. We can see the investigation from Ghelardini, Luisa's study.

Table 0.1 Causes for Forest Dieback

Causes for forest dieback	**Investigated by**
Forest dieback as symptom of a disease	• **Forest mortality in high-elevation whitebark pine (Pinus albicaulis) forests of eastern California, USA; influence of environmental context, bark beetles, climatic water deficit, and warming** - Constance I. Millar, Robert D. Westfall, Diane L. Delany, Matthew J. Bokach, Alan L. Flint, and Lorraine E. Flint. (2012) • **The dieback cycle in Victorian forests: a 30-year study of changes caused by Phytophthora cinnamomi in Victorian open forests, woodlands and heathlands** - Gretna Weste (2003)

	• **Growth and dieback of aspen forests in northwestern Alberta, Canada, in relation to climate and insects -** E H Hogg, James P Brandt, and B Kochtubajda (2002) • **decline and dieback of trees and forests, A global overview.FAQ Forestry paper,120 -** Ciesla, W.M & Donaubauer , E. (1994) • **A natural dieback theory, cohort senescence, as an alternative to the decline disease theory -** Mueller-Dombois,D. (1992)
Senescence of forest as a part of natural forest dynamic	• **A natural dieback theory, cohort senescence, as an alternative to the decline disease theory -** MuellerDombois,D. (1992) • **Canopy Dieback and Successional Processes in Pacific Forests -** Mueller-Dombois,D. (1992)
Forest dieback as cavitation disorder	• **Extreme climatic fluctuations as a cause of forest dieback in the pacific rim -** Allan N. D. Auclair (1993)
Nutrient deficiency or toxicity	• **Nutrient stress predisposes and contributes to sugar maple dieback across its northern range** - Bal,T.L , Storer, A.J.,Jurgensen, M.F.,Doskey, P.V. & Amacher, M.C (2014) • **Tropical montane forest in South Asia: Composition, structure, and dieback in relation to soils and topography** - (Lakkana et al., 2022) • **The Role of nutrition in forest decline - A case study of pinus radiata in New Zealand** - Hunter , I.R (1993) • **The relationship between soil geochemistry and die back of montane forests in Sri Lanka: a case study** - P.

	N. Ranasinghe, C. B. Dissanayake, D. V. N. Samarasinghe & R. Galappatti (2007)
Herbivory damage	• **Patterns and consequences of ungulate herbivory on aspen in western North America** - (Seager et al., 2013) • **Chronic over browsing and biodiversity collapse in a forest understory in Pennsylvania: Results from a 60year-old deer exclusion plot** - Chandra Goetsch, Jennifer Wigg, Alejandro A. Royo, Todd Ristau, Walter P. Carson (Goetsch et al., 2011). • **Forest dieback in horton plains national park** - (Ranawana & Ranawana, n.d.)
Bell miner associated dieback	• **Forest disturbance across the conterminous United States from 1985–2012: The emerging dominance of forest decline** - (Cohen et al., 2016) • **Eucalypt decline in Australia, and a general concept of tree decline and dieback** - (Jurskis, 2005) • **A review of eucalypt dieback associated with bell miner habitat in south-eastern Australia** - (WardellJohnson et al., 2005)
Climatic stress	• **Decline and dieback of trees and forests, A global overview.FAQ Forestry paper,120** - (Ciesla & Donaubauer, 1994) • **Is drought-induced forest dieback globally increasing?** - (Steinkamp & Hickler, 2015). • **Climate change, tree pollination and conservation in the tropics: a research agenda beyond IPBES** - (Ramírez & Kallarackal, 2018).

	• **Die-out of Manilkara hexandra from Bundala National Park, Sri Lanka: Causes and Some Possible Underlying Mechanisms** - (Gunarathne & Perera, 2014)

According to the above studies, there is no exact reason for the forest dieback.

1.2 Problem Definition

Forest Dieback is a Sevier problem these days around the world. So, there are some studies about forest dieback that researchers have done. Rachael Sacatelli, Marjorie Kaplan, Glen Carleton, and Richard G. Lathrop studied Coastal Forest Dieback in the Northeast USA in 2023(Sacatelli et al., 2023). They indicated by the provided paper, are associated with sea level rise-induced changes in the groundwater table, increased saltwater inundation related to storm surges, and the resulting impact on soil conditions. And, Yuanfang Chai, Guilherme Martins, Carlos Nobre, Celso von Randow, Tiexi Chen, and Han Dolman studied Constraining Amazonian land surface temperature sensitivity to precipitation and the probability of forest dieback in 2021(Chai et al., 2021). They indicate the correlation between surface air temperature and precipitation, particularly during drought events. The intricate relationship between reduced precipitation and increased surface air temperature has been observed in Amazonia, with droughts, such as those in 2005 and 2010, showing negative correlations between these two climatic factors. The study of Climate Change and forest health was done by J. Julio Camarero, and Antonio Gazol (Camarero & Gazol, 2022). And Nadezhda M.

Tchebakova and the team studied about Cause for Abies sibirica and Pinus sibirica Forest Dieback in the South Siberian Mountains (Tchebakova et al., 2022).

Another important thing to note is that while many studies have explored factors like soil health, nutrition, climate, and diseases affecting forests, not much research has focused on **how rainfall and temperature changes directly impact forest decline.** Even though existing research has investigated the complex relationships between soil, nutrients, climate, and disease in forests, very few studies have specifically investigated how rainfall and temperature variations affect forest decline. And, they have used simple statistical analyses and computational methods. Those methods don't get any high-level results or Shuttle changes in rainfall variation and temperature changes. Understanding the connection between changes in rainfall and temperature and their impact on the strength and resilience of forests is crucial, especially with the changing climate. It's important to fill in this gap in research to fully understand the many aspects of forest health and resilience. And the studies that use big capacity of data and high computational techniques are not much in current studies. Therefore in our study, we suggest using data mining techniques to get the impact of rainfall variation and temperature changes in Horton plains(Majumdar et al., 2017).

According to a thesis on canopy dieback and regeneration potentials in the tropical montane cloud forest of Horton plains national park by Tithira Lakkana(Lakkana et al., 2022) , mentioned that most of the dieback studies carried out in Sri Lanka so far have ended up suggesting further investigations and have not thoroughly confirms a significant clue to the dieback.

This research is critical for several reasons. Understanding the causes of forest dieback, a major environmental threat, is critical for successful conservation and ecosystem management. By focusing on the impact of rainfall variation and temperature changes, the study fills a critical gap in existing research, providing insights that can inform targeted interventions and proactive strategies to mitigate forest decline, particularly in sensitive environments like Horton Plains National Park. This study is critical for Sri Lankan environmental policymakers, conservationists, researchers, and residents since it provides data-driven measures for effective forest conservation and management.

1.3 Project Aims / Objectives

Main Objective: Quantify the Relationships Between Rainfall, Temperature, and Forest Dieback.

- Methodology: Over the last 40 years, conduct statistical studies (correlation coefficients, regression models) to determine correlations between rainfall variation, temperature fluctuations, and instances of forest dieback in Horton Place.

We divided this main target into two sub-objectives to make our research more efficient.

Objective 1**: Comprehensive Analysis of Temperature rainfall variation and forest dieback.**

- Technique: Examine the 40-year temperature dataset for Horton Place in detail. Identify temperature pattern trends, fluctuations, and any notable deviations across the study period.

Objective 2: **Development of Forest Dieback Prediction Model.**

- Technique: Use Machine Learning and Predictive Analytics.
  Machine learning involves the use of algorithms and statistical models to enable computers to perform tasks without explicit programming instructions. Predictive analytics, a subset of machine learning, focuses on analyzing current and historical data to make predictions about future events or trends.

1.4 Project description

The research focuses on the causes of forest dieback in Sri Lanka's Horton Plains National Park, with a particular emphasis on the direct influence of temperature and rainfall variations. The fundamental goal is to quantify the complex interactions that exist between climate and forest health. The study has two sub-goals: a thorough temperature trend analysis that looks for patterns and fluctuations in a 40-year temperature dataset, and a deep rainfall variability analysis that looks for nuances in the 40-year rainfall dataset. The methodology comprises statistical tools for exact quantification, data cleansing, and a look at the correlations between rainfall, temperature, and forest dieback. Result interpretation refers to the process of deriving useful insights from correlation analysis and generating inferences about observed associations. Qualified ecologists, 40-year climate data, field survey equipment, statistical analysis using the open-source R Programming Language, and consideration of big data methods like Apache Spark or Hadoop for handling large datasets are all required resources, with an emphasis on a comprehensive approach that combines fieldwork and advanced data processing.

The findings contribute to scientific understanding by filling a large gap in existing information, and they have the potential to impact global efforts to minimize forest degradation. Advanced data processing techniques allow for a thorough knowledge of the complex links between meteorological conditions and forest health, providing crucial insights for future study and environmental conservation programs.

Chapter 2

Literature Review

A study conducted by Turner and Lambert in 2005 examined the relationship between soil composition, nutrient levels, and the deterioration of eucalypt forests. Their research focused on areas experiencing forest decline and revealed that toxic elements in the soil, along with nutrient imbalances, were contributing to the issue. Through soil analysis and investigation of trace elements, they identified a correlation between elevated levels of aluminium and the gradual degradation of the forests (Lakkana et al., 2022).

Another study conducted by Gunadasa and Yapa in 2015 further elucidated the detrimental effects of increased aluminum content in the soil on the root systems of trees. This heightened aluminum concentration hindered the trees' ability to absorb nutrients and disrupted essential chemical processes. The study also emphasized the influence of soil pH changes on nutrient availability for the trees. In essence, the research highlighted how soil toxicity and nutrient imbalances were driving the decline in forest health(Gunadasa & Yapa, 2015).

According to these below studies, there is no exact reason for the forest dieback.

Forest dieback as symptom of a disease

- **Forest mortality in high-elevation white bark pine (Pinus albicaulis) forests of eastern California, USA; influence of environmental context, bark beetles, climatic water deficit, and warming -** (Millar et al., 2012)
- **The dieback cycle in Victorian forests: a 30-year study of changes caused by Phytophthora cinnamomi in Victorian open forests, woodlands and heathlands -** (Gretna Weste, 2003)
- **Growth and dieback of aspen forests in northwestern Alberta, Canada, in relation to climate and insects -** E H Hogg, James P Brandt, and B Kochtubajda (2002) (E H Hogg, 2002)
- **decline and dieback of trees and forests, A global overview.FAQ Forestry paper,120 -** (Ciesla & Donaubauer, 1994)
- **A natural dieback theory, cohort senescence, as an alternative to the decline disease theory -** (Ostry et al., 2011)

Senescence of forest as a part of natural forest dynamic

- **A natural dieback theory, cohort senescence, as an alternative to the decline disease theory -** MuellerDombois,D. (1992)
- **Canopy Dieback and Successional Processes in Pacific Forests -** Mueller-Dombois,D. (1992)

Forest dieback as cavitation disorder

- **Extreme climatic fluctuations as a cause of forest dieback in the pacific rim -** Allan N. D. Auclair (1993) Nutrient deficiency or toxicity

- **Nutrient stress predisposes and contributes to sugar maple dieback across its northern range** - Bal,T.L , Storer, A.J.,Jurgensen, M.F.,Doskey, P.V. & Amacher, M.C (2014)
- **Tropical montane forest in South Asia: Composition, structure, and dieback in relation to soils and topography** - (Lakkana et al., 2022)
- **The Role of nutrition in forest decline - A case study of pinus radiata in New Zealand** - Hunter , I.R (1993)
- **The relationship between soil geochemistry and die back of montane forests in Sri Lanka: a case study** - P. N. Ranasinghe, C. B. Dissanayake, D. V. N. Samarasinghe & R. Galappatti (2007)

Herbivory damage

- **Patterns and consequences of ungulate herbivory on aspen in western North America** - (S Trent Seager a, 2013)
- **Chronic over browsing and biodiversity collapse in a forest understory in Pennsylvania: Results from a 60year-old deer exclusion plot** - (Chandra Goetsch, 2011)
- **Forest dieback in horton plains national park** - (Ranawana & Ranawana, n.d.)

Bell miner associated dieback.

- **Forest disturbance across the conterminous United States from 1985–2012: The emerging dominance of forest decline** - (Cohen et al., 2016)
- **Eucalypt decline in Australia, and a general concept of tree decline and dieback** - (Vic Jurskis, 2005)

  A review of eucalypt dieback associated with bell miner habitat in south-eastern Australia - (WardellJohnson et al., 2005)

Climatic stress

- **Decline and dieback of trees and forests, A global overview.FAQ Forestry paper,120 -** (Ciesla & Donaubauer, 1994)
- **Is drought-induced forest dieback globally increasing? -** (Jörg Steinkamp, 2015)
- **Climate change, tree pollination and conservation in the tropics: a research agenda beyond IPBES** - (Fernando Ramírez, 2017).
- **Die-out of Manilkara hexandra from Bundala National Park, Sri Lanka: Causes and Some Possible Underlying Mechanisms** - (AD Perera, 2014)

According to the study by Lei Su and Mehdi Heydari, conducted in the Zagros Forest of Iran, the results indicate a positive correlation between woody species richness and dieback intensity (DI). Additionally, soil clay content and wetness index were found to influence DI.

Chapter 3

Materials and Methods

3.1 Data Collection and Preparation

This research relies on two key datasets: rainfall and temperature data from the Department of Meteorology, Sri Lanka, and forest dieback data derived from satellite imagery through Google Earth Engine (GEE). The combination of these two data sources enables us to investigate the relationship between climatic variables and forest dieback in the Horton Plains region.

3.1.1 Data Collection from the Department of Meteorology

The dataset provided by the Department of Meteorology contains 40 years (1980-2020) of monthly average rainfall and temperature data for Horton Plains, a critical biodiversity hotspot in Sri Lanka. Each data point represents the monthly average temperature (in degrees Celsius) and monthly total rainfall (in millimeters), measured through ground-based meteorological stations.

While the dataset was comprehensive in terms of climatic data, a significant limitation was the absence of forest dieback data, which is essential for assessing the relationship between climate and forest health. This gap in the dataset required us to explore alternative methods to estimate forest dieback.

3.1.2. Challenges in Acquiring Forest Dieback Data

As the provided dataset lacked any direct information on forest dieback, we had to rely on satellite imagery to estimate dieback. However, this presented two main challenges,

Temporal Mismatch

The meteorological dataset spans 40 years, with data points for each month between 1980 and 2020. However, satellite imagery for Horton Plains was only available after the year 2000, limiting our ability to assess forest dieback prior to this period. Additionally, satellite images were only accessible on an annual basis, meaning that we could only obtain one image per year for dieback estimation. This mismatch between the monthly resolution of the climate data and the annual resolution of the satellite data required an approximation of monthly forest dieback.

Satellite Image Quality and Coverage

Forested areas in Horton Plains are often cloud-covered, which can affect the quality of satellite images. To minimize the impact of clouds on our dieback analysis, we filtered the imagery for cloud cover less than 30%.We also faced a transition from Landsat 7 (available from 2000-2012) to Landsat 8 (available from 2013 onwards). Although Landsat 8 provides higher-resolution images, this transition added a layer of complexity in ensuring consistency across the years.

3.1.3 Estimating Forest Dieback Using Google Earth Engine

To estimate forest dieback, we utilized Normalized Difference Vegetation Index (NDVI) values derived from satellite images. NDVI is a commonly used indicator of vegetation health, with values ranging from -1 to +1. Higher NDVI values indicate dense and healthy vegetation, while lower values suggest degraded or barren land.

Step 1: Defining the Horton Plains Region

Using Google Earth Engine (GEE), we created a polygon to represent the geographic boundaries of the Horton Plains region. This polygon was defined using the known geographic coordinates of the area and was used to clip the satellite images, ensuring that only the relevant region was analyzed.

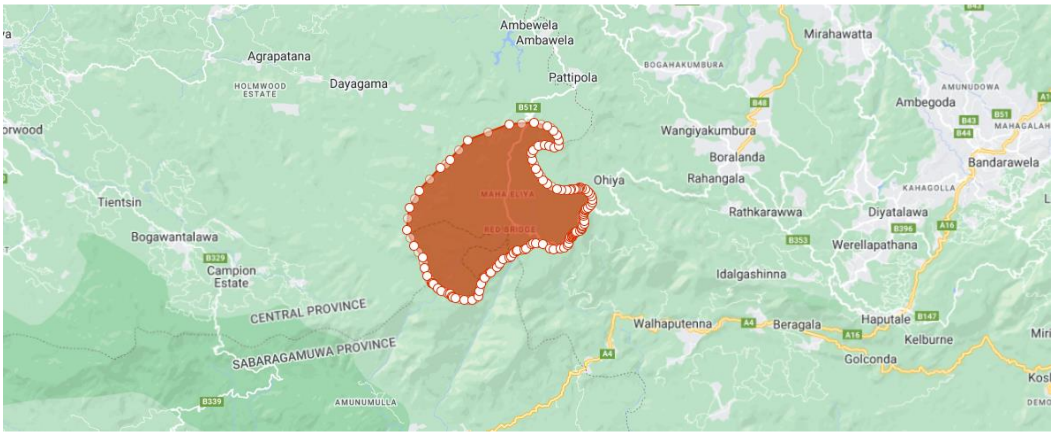

Figure 2 Polygon Geographic Coordinates

After polygon loaded using google earth engine API

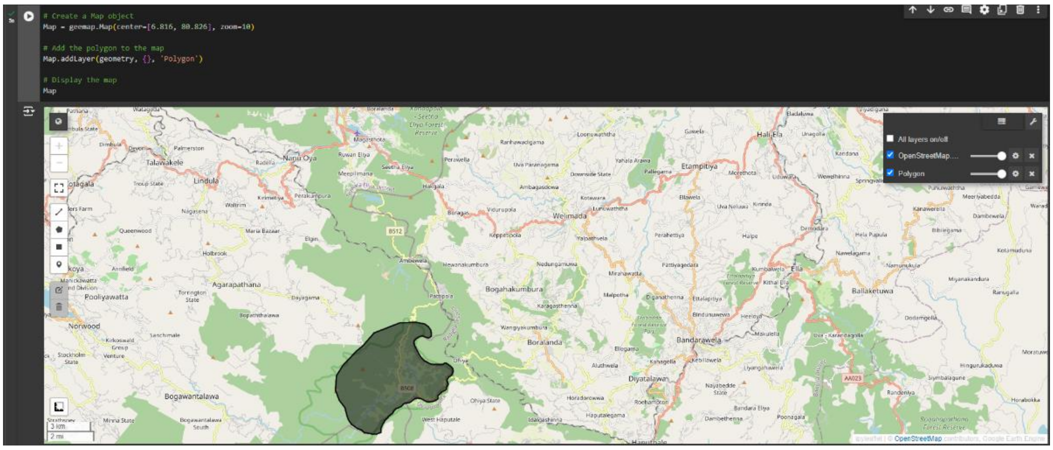

Figure 3 Polygon after using API

Step 2: Acquiring Satellite Imagery

We used **Landsat 7** for the period from 2000 to 2012, and **Landsat 8** for 2013 to 2020. These satellite images were processed to remove clouds and calculate the NDVI for each year.

```
[ ]  # Load Landsat 7 imagery
     LS2000 = ee.ImageCollection('LANDSAT/LE07/C01/T1_TOA') \
             .filterBounds(geometry) \
             .filterDate('2001-01-01', '2001-12-31')\
             .filterMetadata('CLOUD_COVER', 'less_than', 30)
     # Calculate NDVI
     def calculate_ndvi(image):
         ndvi = image.normalizedDifference(['B4', 'B3']).rename('NDVI')
         return ndvi

     ndvi = LS2000.map(calculate_ndvi)
```

Figure 4 Code for acquiring satellite imagery

Step 3: Classifying Forested Areas

A **forest threshold** was applied to the NDVI values to classify forested areas. Based on prior research, an NDVI value above **0.6** typically indicates healthy forest cover. We used this threshold to create a **forest mask**, identifying which areas in the satellite image were classified as forested.

Step 4: Calculating Forest Dieback Area

The forest area for each year was calculated by multiplying the number of forest pixels by the area of each pixel (30m x 30m). This gave the total **forest dieback area** in square meters for each year. Since we only had annual satellite images, we assumed the forest dieback remained constant throughout the year and used this annual value to estimate the **monthly average forest dieback**.

```
[ ] # Define a threshold for NDVI to classify forested areas
    forest_threshold = 0.7  # value for trees

    # Create a forest mask
    forest_mask = ndvi.mean().gt(forest_threshold)

    # Clip the mask to the polygon
    forest_area = forest_mask.clip(geometry)

    # Calculate the area of forest within the polygon
    forest_area_stats = forest_area.multiply(ee.Image.pixelArea()) \
                          .reduceRegion(reducer=ee.Reducer.sum(), \
                                        geometry=geometry, \
                                        scale=30, \
                                        maxPixels=1e9)

    # Convert forest area from square meters
    forest_area_sq_m = forest_area_stats.get('NDVI').getInfo() , 'square meters'

    # Print the result
    print('Forest area:', forest_area_sq_m)

    Forest area: (34624330.96381287, 'square meters')
```

Figure 5 Code for Calculating Forest Dieback Area

Step 5: Aligning Dieback Data with Monthly Climate Data

To align the annual forest dieback estimates with the monthly rainfall and temperature data, we calculated the **average monthly dieback area** by dividing the annual dieback by 12. This allowed us to create a complete dataset that included:

- Monthly average rainfall (mm)
- Monthly average temperature (°C)
- Estimated monthly dieback area (sq. meters)

3.1.4 Calculating Forest Dieback and Including it in the Dataset

As part of the analysis, we calculated the **forest dieback area** using satellite imagery through Google Earth Engine (GEE). This step was crucial as the **original dataset did not include dieback data**, and our objective was to explore the impact of climate variables on forest health in Horton Plains. After calculating the forest dieback area for each year, the next step was to **preprocess this data** and integrate it into our dataset alongside the rainfall and temperature data.

3.1.4.1 Forest Dieback Calculation Process

The forest dieback area was calculated by analyzing **annual satellite images** from Landsat 7 (2000-2012) and Landsat 8 (2013-2020). We used the **Normalized Difference Vegetation Index (NDVI)** to assess forest health, and any areas with an NDVI value below 0.6 were classified as degraded or experiencing dieback.

Here's a quick summary of the process:

1. **Satellite Imagery Acquisition**: We accessed annual satellite images for Horton Plains using GEE.
2. **NDVI Calculation**: For each image, we calculated the NDVI to identify areas of forest dieback.
3. **Forest Area Classification**: Pixels with NDVI values below 0.6 were classified as forest dieback.
4. **Area Calculation**: The dieback area (in square meters) was calculated by summing the pixel areas that met the dieback threshold.

Once we obtained the **annual forest dieback area**, we needed to **integrate this information into our dataset**, which also contained monthly average rainfall and temperature data.

3.1.4.2 Preprocessing the Dieback Data

Since the **rainfall and temperature data were available at a monthly level** while the dieback data was calculated on an annual basis, it was necessary to **align the temporal resolution** of these datasets. To achieve this, we followed these steps:

1. **Converting Annual Dieback Data to Monthly Values**:
   Given that dieback data was available only on a yearly basis, we made the assumption that the forest dieback area remained constant throughout the year. We divided the **annual forest dieback area** by 12 to obtain the

average monthly dieback area. This allowed us to integrate the dieback data with the monthly rainfall and temperature data.

2. **Creating a New Column in the Dataset**:

   We then created a new column in our dataset labeled dieback, representing the **average monthly forest dieback area**. This column contained the calculated monthly values for each year, allowing us to compare dieback data alongside rainfall and temperature for every month between 2000 and 2020.

3. Merging Dieback Data with Rainfall and Temperature:

After creating the monthly dieback column, the next step was to merge the dieback data with the existing monthly climate data. The rainfall (rf_avg), temperature (t_avg), and dieback (dieback) columns were now aligned by year and month.

The final dataset included:

- Year: The year of the observation.
- Month: The month of the observation.
- Rainfall (rf_avg): The average monthly rainfall, in millimeters.
- Temperature (t_avg): The average monthly temperature, in degrees Celsius.
- Dieback (dieback): The estimated monthly forest dieback area, in square meters.

3.2 Trend Analysis

The trend analysis was conducted to examine how the key variables—**rainfall, temperature**, and **forest dieback**—evolved over time. The purpose of this analysis was to determine whether there were any discernible long-term patterns or shifts in these variables, which could help explain forest dieback in Horton Plains.

3.2.1 Methodology of Trend Analysis

We used **seasonal decomposition** to separate the time series data of each variable into three components:

1. **Trend Component**: The long-term direction of the variable, showing whether it is increasing or decreasing over time.

2. **Seasonal Component**: Recurring short-term patterns, which could be annual cycles (e.g., wet and dry seasons).
3. **Residual Component**: The remaining variation after accounting for the trend and seasonality, representing random fluctuations or noise.

We applied this decomposition to each of the three variables:

- **Average Temperature (°C)**
- **Average Rainfall (mm)**
- **Forest Dieback Area (sq. meters)**

```
[ ] from google.colab import drive
    drive.mount('/content/drive')
    file_path = '/content/drive/My Drive/Colab Notebooks/dataset_02.csv'
    data = pd.read_csv(file_path)

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[ ] data['year'] = pd.to_datetime(data['year'], format='%Y-%m')
    filtered_data = data[(data['year'] >= '2000-01-01') & (data['year'] <= '2023-12-31')]
    filtered_data.set_index('year', inplace=True)

[ ] #decompose data
    decomposition_temp = sm.tsa.seasonal_decompose(filtered_data['t_avg'], model='additive', period=12)
    decomposition_rf = sm.tsa.seasonal_decompose(filtered_data['rf_avg'], model='additive', period=12)
    decomposition_dieback = sm.tsa.seasonal_decompose(filtered_data['dieback'], model='additive', period=12)

    decomposition_temp.plot()
    plt.show()

    decomposition_rf.plot()
    plt.show()

    decomposition_dieback.plot()
    plt.show()
```

Figure 6 Code for Trend Analysis

3.2.2 Temperature Trend Analysis

The temperature trend component was extracted from the **monthly average temperature data** to examine how the temperature changed over the 20-year study period (2000-2020).

Key Observations:

- The trend analysis revealed a **slight increase** in the average temperature over the two decades. This is consistent with global warming trends observed in many regions.
- The temperature trend was not linear; it displayed periods of **fluctuation**, with some years experiencing **higher-than-average temperatures**, followed by cooling periods.

- While the overall direction of the trend was upward, the rate of temperature increase was relatively gradual.

Seasonal and Residual Components:

- The **seasonal component** showed clear **annual temperature cycles**, with warmer months during the dry season and cooler temperatures during the wet season.
- The **residual component** accounted for short-term variations that could not be explained by the trend or seasonality. These variations may be due to local climatic events or anomalies in individual years.

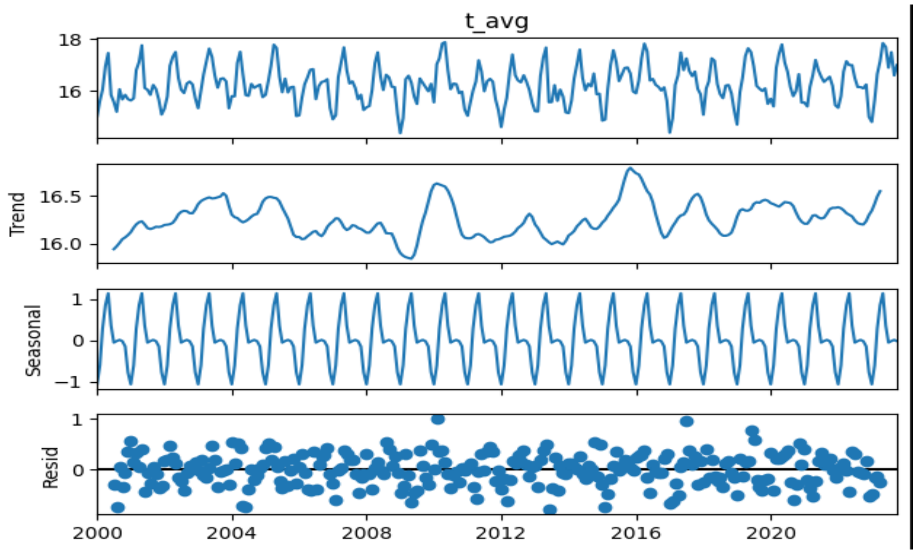

Figure 7 Graph for temperature analysis

3.2.3 Rainfall Trend Analysis

Next, we analyzed the trend in **monthly average rainfall**. Rainfall in Horton Plains is known for its variability, with distinct wet and dry seasons.

Key Observations:

- The rainfall data showed a **cyclical pattern**, with alternating periods of **high and low rainfall**. Unlike temperature, rainfall did not exhibit a clear long-term increasing or decreasing trend.
- There were **no significant changes** in the overall rainfall trend over time, although there were periods where **droughts** or **high rainfall events** could be observed.

Seasonal and Residual Components:

- The **seasonal component** highlighted the expected **wet and dry seasons**, with rainfall peaking during the monsoon months and dipping during the dry season.

- The **residual component** displayed significant variability, reflecting the **erratic nature of rainfall** in the region. This variability is typical of tropical climates, where rainfall can be highly unpredictable from month to month.

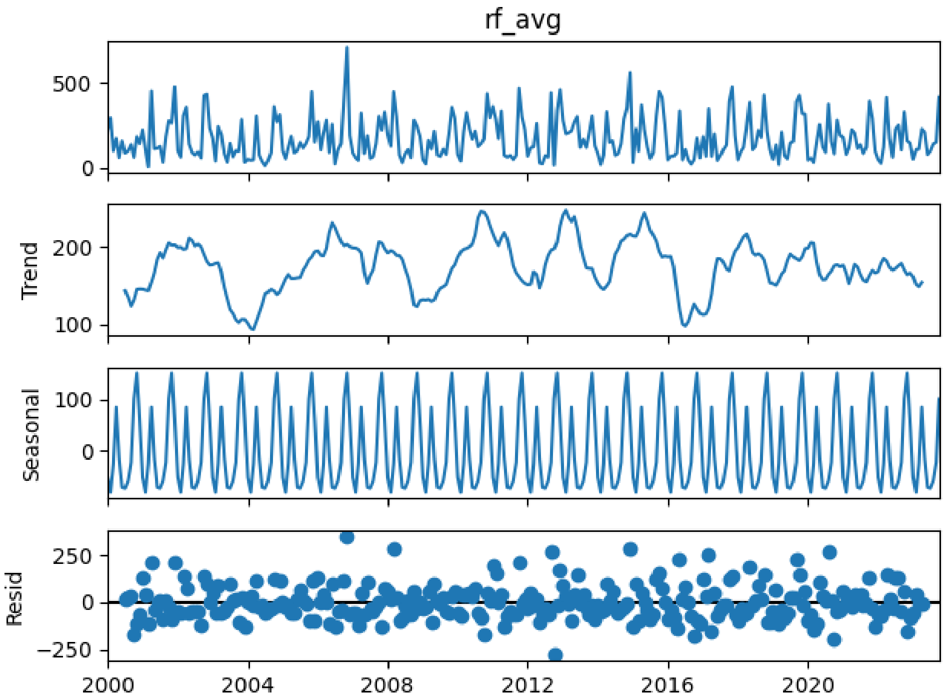

Figure 8 Graph for Rainfall Trend Analysis

3.2.4 Forest Dieback Trend Analysis

The trend in **forest dieback areas** was one of the most critical elements of this study, as it directly reflects the health of the forest ecosystem.

Key Observations:

- The forest dieback trend showed **considerable fluctuations**, with some years experiencing significant increases in dieback area. This fluctuation often coincided with **periods of lower rainfall** or **higher temperatures**.

- Overall, the trend indicated a **gradual increase in forest dieback** over the study period, particularly during **drier years**.

- The **highest levels of dieback** were observed in years with **extreme climatic conditions**, such as **droughts** or **heatwaves**.

Seasonal and Residual Components:

- The **seasonal component** of dieback was less pronounced than in the climatic variables. While there were some small recurring patterns, forest dieback did not follow a strict annual cycle like rainfall or temperature.
- The **residual component** showed considerable **variability**. This could indicate that factors beyond climate, such as **disease**, **human activity**, or **soil conditions**, also played a role in forest dieback.

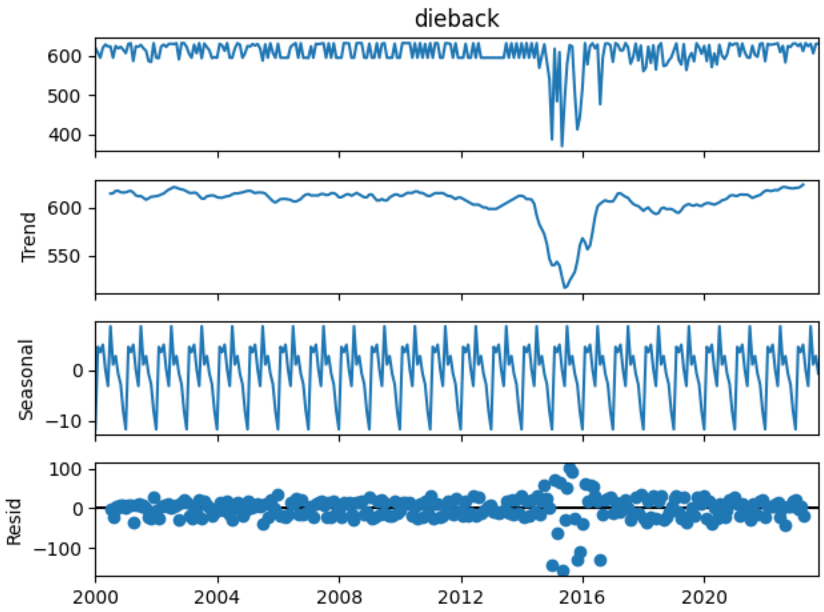

Figure 9 Graph for Forest Dieback Trend Analysis

```
[ ] #Extract Trend Components and Calculate Correlation
    trend_temp = decomposition_temp.trend.dropna()
    trend_rf = decomposition_rf.trend.dropna()
    trend_dieback = decomposition_dieback.trend.dropna()

    trend_data = pd.DataFrame({
        'temperature_trend': trend_temp,
        'rainfall_trend': trend_rf,
        'dieback_trend': trend_dieback
    })

    correlation_matrix = trend_data.corr()
    print(correlation_matrix)
```

```
                   temperature_trend  rainfall_trend  dieback_trend
temperature_trend           1.000000         0.01450      -0.284689
rainfall_trend              0.014500         1.00000      -0.293910
dieback_trend              -0.284689        -0.29391       1.000000
```

Figure 10 Code for Forest Dieback Trend Analysis

3.2.5 Visualizing the Trend Analysis

The results of the trend analysis for temperature, rainfall, and forest dieback were visualized using time series plots. Each plot displayed the **raw data**, along with the **trend, seasonal, and residual components**. These visualizations helped to better understand how each variable behaved over time and how they might be interconnected.

3.2.6 Interpretation of Trends

- **Temperature and Rainfall**: While temperature showed a slight upward trend, rainfall did not exhibit a clear long-term pattern. This suggests that **temperature increase** may be contributing to forest stress, but the variability in rainfall complicates a straightforward interpretation.
- **Forest Dieback**: The forest dieback trend showed an increase over time, with more severe dieback occurring during **years of extreme climate events** (e.g., droughts or temperature spikes). However, the residual component suggested that other factors beyond climate could also be influencing forest dieback.

Overall, the trend analysis revealed important insights into the behavior of key variables over time, which set the stage for further investigation into the **relationship between climate and forest health**.

3.3 Correlation Analysis

The primary goal of the correlation analysis was to quantify the relationship between temperature, rainfall, and forest dieback in Horton Plains, Sri Lanka. We aimed to determine whether these two key climatic variables—temperature and rainfall—had a significant impact on forest dieback over the 20-year period from 2000 to 2020.

3.3.1 Rationale for Correlation Analysis

Understanding the correlation between climate and forest dieback is vital for two reasons:

Predictive Power: If a strong correlation exists between climate variables and dieback, it could help us predict future forest dieback events under different climate scenarios.

Forest Conservation: By identifying the primary climatic drivers of forest dieback, conservation efforts could focus on mitigating these factors, improving forest resilience to changing climate conditions.

Given the complexity of ecological systems, however, we anticipated that the relationship between climate variables and forest dieback might not be straightforward. Factors such as nonlinear dynamics, delayed effects, or interaction between variables could influence forest health.

3.3.2 Pearson Correlation Coefficient

To investigate the strength and direction of the relationship between temperature, rainfall, and forest dieback, we used the Pearson correlation coefficient. This is one of the most commonly used measures for assessing the linear relationship between two continuous variables. The Pearson correlation coefficient (denoted as r) ranges from:

- -1: Perfect negative linear relationship (as one variable increases, the other decreases).
- 0: No linear relationship.
- +1: Perfect positive linear relationship (as one variable increases, the other increases).

Before performing the analysis, the dataset was cleaned and preprocessed to align the monthly average rainfall and temperature with the estimated forest dieback area for each month. This allowed us to compute meaningful correlations between variables that were measured on the same timescale.

3.3.3 Temperature vs. Forest Dieback

We began by examining the relationship between **temperature** and **forest dieback**. The hypothesis was that **higher temperatures** might contribute to increased forest dieback, as rising temperatures can lead to **heat stress** on vegetation, especially in tropical environments.

Methodology:

We used a **scatter plot** to visualize the relationship between monthly average temperature and forest dieback, followed by the calculation of the **Pearson correlation coefficient** to assess the strength and direction of the linear relationship.

```python
# Extract year from the 'year' column
dataset['year'] = dataset['year'].str.split('-').str[0].astype(int)

# Filter the dataset for years 2000-2021
filtered_dataset = dataset[(dataset['year'] >= 2000) & (dataset['year'] <= 2021)]

# Scatterplot of Temperature vs. Dieback for years 2000-2021
plt.figure(figsize=(8, 6))
plt.scatter(filtered_dataset['t_avg'], filtered_dataset['dieback'])
plt.title('Temperature vs. Forest Dieback (2000-2021)')
plt.xlabel('Temperature')
plt.ylabel('Forest Dieback')
plt.grid(True)
plt.show()
```

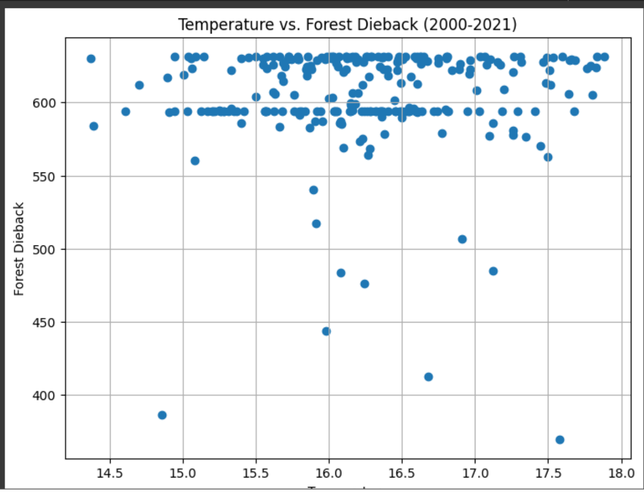

Figure 11 Temperature vs Forest Dieback (2000-2021)

Results:

The calculated **Pearson correlation coefficient** between **temperature and forest dieback** was **0.023**. This value is very close to zero, indicating that there is **no significant linear relationship** between temperature and dieback. The scatter plot showed no clear pattern or trend, suggesting that **increasing or decreasing temperature did not correspond directly to changes in forest dieback**.

Interpretation:

- **Weak correlation**: The near-zero correlation suggests that temperature does not play a dominant role in explaining forest dieback patterns in Horton Plains, at least not in a simple linear manner. This result may indicate that other environmental factors (such as rainfall, soil health, or disease) are more critical in determining dieback.

- **Potential nonlinearity**: The lack of a strong correlation does not rule out the possibility of a **nonlinear relationship** between temperature and dieback. Forest ecosystems often respond to temperature thresholds or interact with other factors, creating more complex dynamics than a linear model can capture.

Ecological Implications:

- The weak temperature correlation suggests that **temperature fluctuations** within the observed range (2000-2020) may not have been severe enough to trigger significant dieback events. However, the **impact of extreme heat events**, which are more likely with ongoing climate change, could have stronger effects on the forest in the future.

```
[ ] # Calculate correlation coefficient
    correlation_coefficient = filtered_dataset['t_avg'].corr(filtered_dataset['dieback'])

    print("Correlation coefficient:", correlation_coefficient)
```

```
Correlation coefficient: 0.02305041444751161
```

Figure 12 Correlation Coefficient

3.3.4 Rainfall vs. Forest Dieback

Next, we analyzed the relationship between **rainfall** and **forest dieback**. Rainfall is a crucial factor in determining forest health, especially in tropical ecosystems like Horton Plains, where the forest relies on consistent water availability.

Methodology:

A scatter plot was created to visualize the relationship between **monthly average rainfall** and **forest dieback**, followed by the calculation of the **Pearson correlation coefficient**.

```python
# Filter the dataset for years 2000-2021
filtered_dataset = dataset[(dataset['year'] >= 2000) & (dataset['year'] <= 2021)]

# Scatterplot of Temperature vs. Dieback for years 2000-2021
plt.figure(figsize=(8, 6))
plt.scatter(filtered_dataset['rf_avg'], filtered_dataset['dieback'])
plt.title('Rainfall vs. Forest Dieback (2000-2021)')
plt.xlabel('Rainfall')
plt.ylabel('Forest Dieback')
plt.grid(True)
plt.show()
```

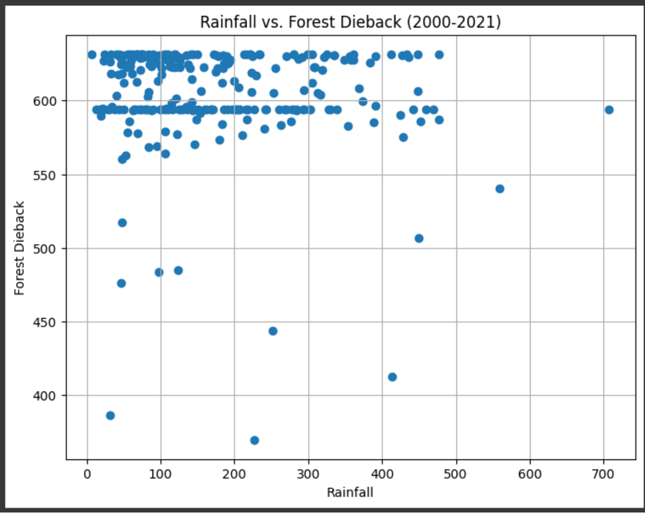

Figure 13 Rainfall vs Forest Dieback

Results:

The calculated **Pearson correlation coefficient** between **rainfall and forest dieback** was **-0.29**, indicating a **weak negative correlation**. This suggests that **lower rainfall** is associated with **higher forest dieback**, although the strength of the relationship is not particularly strong.

Interpretation:

- **Weak but meaningful relationship**: Although the correlation is not strong, the negative value suggests that **reduced rainfall** (such as during droughts) may contribute to increased forest dieback. This makes sense in the context of tropical forests, where plants rely on consistent rainfall to maintain water balance and avoid stress.

- **Threshold effect**: The weak correlation may reflect a **threshold effect**, where forest health is relatively stable under normal rainfall conditions, but **extreme drought** or **long periods of reduced rainfall** could trigger significant dieback. This would explain the weak linear relationship, as only the more extreme rainfall deviations may lead to noticeable dieback events.

Ecological Implications:

- The negative correlation between rainfall and dieback suggests that **drought conditions** are likely to have an adverse impact on forest health in Horton Plains. Conservation efforts aimed at mitigating dieback should consider **water management strategies** to ensure that the forest has sufficient water resources during dry periods.

```
[ ] # Calculate correlation coefficient
    correlation_coefficient = filtered_dataset['rf_avg'].corr(filtered_dataset['dieback'])

    print("Correlation coefficient:", correlation_coefficient)

    Correlation coefficient: -0.06922912327757526
```

Figure 14 Correlation Coefficient

3.3.5 Combined Insights from Correlation Analysis

The results of the correlation analysis provide valuable insights into the complex relationship between climate variables and forest dieback:

1. **Temperature**: The **weak positive correlation** between temperature and dieback suggests that temperature variations, within the observed range, do not have a significant linear impact on forest dieback. However, the potential for **nonlinear effects** or interactions with other factors should not be ruled out.

2. **Rainfall**: The **negative correlation** between rainfall and dieback, although weak, highlights the importance of **adequate water supply** in maintaining forest health. The forest appears to be more vulnerable to dieback during periods of **low rainfall**, especially during drought years.

3.3.6 Limitations of the Correlation Analysis

While Pearson's correlation provides useful insights, it is limited in its ability to capture **complex ecological dynamics**:

- **Nonlinearity**: The relationship between climate variables and forest dieback may be **nonlinear**, meaning that the impact of temperature or rainfall could increase dramatically once certain thresholds are reached.

- **Multiple Interacting Factors**: Forest dieback is likely influenced by multiple factors beyond climate, such as **soil quality**, **disease outbreaks**, and **human activities** (e.g., land-use changes). The simple bivariate correlation analysis does not account for these interactions.

For these reasons, correlation analysis is an important starting point, but more advanced modeling approaches (such as **machine learning** and **regression models**) will be necessary to fully capture the complexity of forest dieback dynamics.

3.4 Model Development

The second key objective of this study was to develop a predictive model capable of estimating **forest dieback** based on **rainfall** and **temperature** data. Forest dieback is a complex ecological phenomenon, influenced by multiple interacting factors, many of which are not directly observable. Given this complexity, traditional linear models may fail to capture the **nonlinear dynamics** of forest ecosystems. For this reason, we employed **machine learning models** that could better handle these nonlinearities and provide more accurate predictions.

3.4.1 Motivation for Model Selection

We selected four machine learning models to evaluate their effectiveness in predicting forest dieback. Each model was chosen for its ability to capture complex relationships in data:

1. **Support Vector Regression (SVR)**:
    - **Why chosen**: SVR is a powerful technique for modeling **nonlinear relationships** by using kernel functions. It works well when there are complex relationships between input variables and the target variable (in this case, dieback).
    - **Expected performance**: SVR was expected to handle the possible nonlinearities between temperature, rainfall, and dieback, but its performance depends heavily on the proper tuning of hyperparameters like the kernel type and regularization parameter.

2. **Random Forest**:

    - **Why chosen**: Random Forest is an **ensemble learning method** that constructs multiple decision trees and averages their predictions. It is known for its robustness and ability to capture complex interactions between variables without overfitting.

    - **Expected performance**: We expected Random Forest to perform well due to its ability to model both **linear and nonlinear** relationships and handle the noise inherent in ecological datasets.

3. **Gradient Boosting**:

    - **Why chosen**: Gradient Boosting is another ensemble method, but it builds trees sequentially, with each tree attempting to correct the errors made by the previous one. This method is particularly effective for **minimizing errors** and improving the model's predictive power over time.

    - **Expected performance**: Gradient Boosting was anticipated to excel in fine-tuning predictions, especially when the relationships between climate variables and dieback are complex but gradual over time.

4. **Neural Network (MLPRegressor)**:

    - **Why chosen**: Neural Networks are highly flexible and can capture **highly nonlinear relationships** between input features and the target variable. By using multiple hidden layers, they can learn intricate patterns in the data.

    - **Expected performance**: Neural Networks were expected to perform well in discovering **hidden patterns** between climate variables and dieback. However, due to their complexity, they can be prone to overfitting, especially with smaller datasets.

Each model was evaluated using **10-fold cross-validation**, and their performance was measured using **Mean Absolute Error (MAE)**.

3.4.2 Data Preprocessing for Model Training

The success of machine learning models often depends on the quality of the input data. Therefore, we applied several preprocessing steps to ensure that the dataset was ready for modeling.

Handling Missing Data

Missing data can introduce bias into the model and lead to inaccurate predictions. The initial dataset had some missing values, particularly in the **temperature** and **rainfall** columns. To address this issue, we opted to **remove rows with missing values** rather than attempt to impute them, as imputation could introduce additional bias.

```
# Handle missing values (if any)
data = data.dropna()
```

Figure 15 Handle Missing Values

Feature Engineering: Lag Features

Forests often respond to **past climatic conditions** rather than just immediate weather patterns. For instance, prolonged periods of drought or consistent temperature increases may have delayed effects on forest dieback. To account for these delayed effects, we created **lag features** for both temperature and rainfall.

- **Temperature lag**: The average temperature from the **previous month** (temp_lag1) was included to help the model capture any delayed response of forest dieback to temperature.
- **Rainfall lag**: Similarly, we created a **rainfall lag** feature (rain_lag1) to allow the model to consider past rainfall patterns.

```
[ ] # Handle missing values (if any)
    data = data.dropna()

    # Create lag features (assuming 't_avg' for temperature, 'rf_avg' for rainfall, 'dieback' for dieback)
    data['temp_lag1'] = data['t_avg'].shift(1)
    data['rain_lag1'] = data['rf_avg'].shift(1)

    # Drop rows with NaN values created by the lag shift
    data = data.dropna()

    # Define input features and target variable
    X = data[['t_avg', 'rf_avg', 'temp_lag1', 'rain_lag1']]  # Input features
    y = data['dieback']  # Target variable (dieback)
```

Figure 16 Lag Features

Defining Input Features and Target Variable

The input features included:

- **Current temperature** (t_avg)
- **Current rainfall** (rf_avg)
- **Previous month's temperature** (temp_lag1)
- **Previous month's rainfall** (rain_lag1)

The target variable was the **forest dieback area** for each month. By incorporating both current and lagged values, the model could learn not only from the present climate conditions but also from the climate trends leading up to each observation.

Data Splitting

To ensure the robustness of the model and its ability to generalize to unseen data, we split the dataset into a **training set** (80%) and a **test set** (20%). The training set was used to train the models, while the test set was reserved for evaluating their performance on unseen data.

3.4.3 Model Training and Cross-Validation

Once the data was preprocessed, we trained the four selected models using **10-fold cross-validation**. Cross-validation ensures that the model's performance is not overly dependent on any subset of the data. It splits the dataset into 10 equal parts, trains the model on 9 parts, and tests it on the remaining part. This process is repeated 10 times, and the average performance is calculated.

The **Mean Absolute Error (MAE)** was used as the primary evaluation metric. MAE is particularly suited for this task because it provides a straightforward interpretation of the **average prediction error** in terms of forest dieback area.

```
[ ] from sklearn.linear_model import LinearRegression
    from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
    from sklearn.svm import SVR
    from sklearn.neural_network import MLPRegressor

    # Define models
    models = {
        'Support Vector Regression': SVR(),
        'Random Forest': RandomForestRegressor(n_estimators=100, random_state=42),
        'Gradient Boosting': GradientBoostingRegressor(n_estimators=100, random_state=42),
        'Neural Network': MLPRegressor(hidden_layer_sizes=(100,100), max_iter=1000, random_state=42)
    }
```

```
[ ] from sklearn.model_selection import cross_val_score
    from sklearn.metrics import mean_absolute_error, make_scorer

    # Define scoring function (MAE)
    mae_scorer = make_scorer(mean_absolute_error)

    # Perform 10-fold cross-validation for each model
    results = {}
    for name, model in models.items():
        scores = cross_val_score(model, X, y, cv=10, scoring=mae_scorer)
        results[name] = scores
        print(f"{name} - Mean MAE: {scores.mean()} - Std Dev: {scores.std()}")
```

```
Support Vector Regression - Mean MAE: 259.01899140352873 - Std Dev: 28.604914154800962
Random Forest - Mean MAE: 255.77835221674877 - Std Dev: 24.859259968672145
Gradient Boosting - Mean MAE: 260.33083743939335 - Std Dev: 29.65991049524721
Neural Network - Mean MAE: 264.30087402199223 - Std Dev: 26.068798305935342
```

Figure 17 Model Training and Cross-Validation

Cross-Validation Results:

Key Findings:

- **Random Forest** emerged as the best-performing model with the **lowest MAE** of **255.78 sq. meters**. This indicates that, on average, the Random Forest model's predictions of forest dieback were **within 256 square meters** of the actual values.
- The relatively low standard deviation for the Random Forest model indicates **consistent performance** across different folds of the cross-validation, making it a **stable and reliable model**.
- **Support Vector Regression** and **Gradient Boosting** performed similarly but with slightly higher errors than Random Forest.
- The **Neural Network model** had the highest MAE, suggesting that it may have overfitted the data or that it requires further tuning.

3.4.4 Final Training and Evaluation

After selecting the **Random Forest** model as the best-performing model, it was trained on the entire training dataset and evaluated on the **test dataset**. This allowed us to assess how well the model generalized to new, unseen data.

```
from google.colab import files

# Train the Random Forest model on the entire dataset
final_rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
final_rf_model.fit(X, y)

# Save the trained model (optional)
import joblib
joblib.dump(final_rf_model, 'final_random_forest_model.pkl')

files.download('final_random_forest_model.pkl')

# Predict on new data (example)
# predictions = final_rf_model.predict(X_new)
```

Figure 18 Final Training and Evaluation

Chapter 4

Results and Discussion

This chapter presents the results obtained from the trend analysis, correlation analysis, and the predictive model for forest dieback. Each subsection corresponds to the methods described in **Chapter 3**, with a focus on interpreting the results and understanding their implications for forest dieback in Horton Plains.

4.1. Trend Analysis Results

The trend analysis revealed important insights into the behavior of **temperature**, **rainfall**, and **forest dieback** over the 20-year study period (2000-2020). The decomposition of the time series for each variable allowed us to observe both the **long-term trends** and **seasonal patterns**.

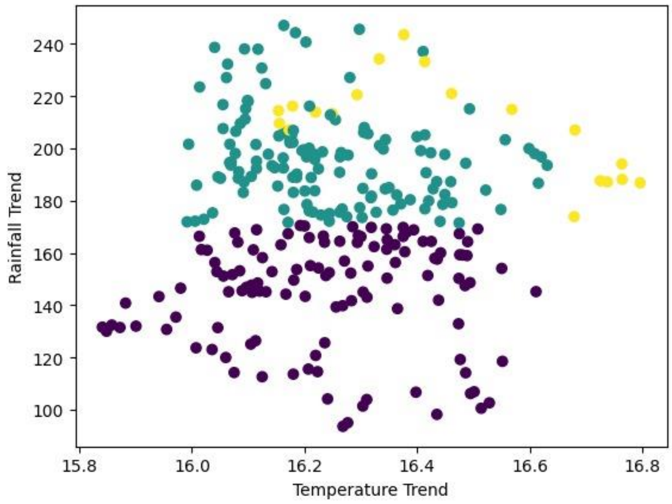

Figure 19 Trend Analysis Results

4.1.1. Temperature Trend

The long-term temperature trend showed a **gradual increase** over the study period, with fluctuations corresponding to seasonal variations. The temperature increase is consistent with broader patterns of **global warming**, though the changes were relatively moderate.

Key observation:

- A slight **increase in temperature** was observed over time, but the overall magnitude of this increase was small, indicating that temperature fluctuations in Horton Plains are relatively stable compared to more extreme climate change scenarios.

Graph:

Temperature Trend

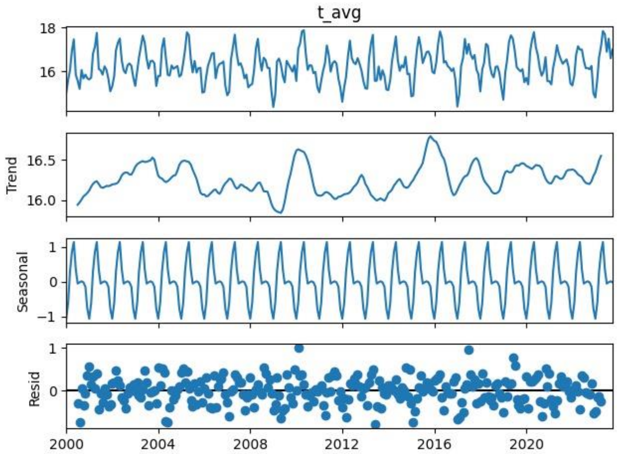

Figure 20 Graph: Temperature Trend

4.1.2. Rainfall Trend

The trend in **monthly average rainfall** was more variable than temperature, with alternating periods of **high and low rainfall**. There was no clear increasing or decreasing trend, although **droughts** during specific years were evident in the data.

Key observation:

- No clear trend in rainfall was observed over time. However, the periodic droughts could have had a significant impact on **forest dieback**, particularly in dry years.

Graph:

Rainfall Trend

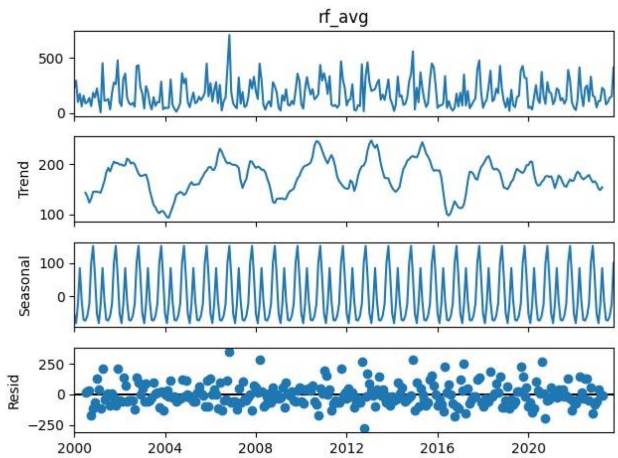

Figure 21 Graph : Rainfall Trend

4.1.3. Forest Dieback Trend

The forest dieback trend showed **increasing fluctuations** over time, with more significant dieback events occurring in recent years. The increase in dieback, particularly during years with **lower rainfall** and **higher temperatures**, suggests that climatic stressors are affecting forest health in Horton Plains.

Key observation:

- The dieback trend increased over time, with a notable spike during periods of **climatic extremes** (e.g., dry years or heatwaves). This indicates that **climatic variability** is likely a contributing factor to forest health.

Graph:

Forest Dieback Trend

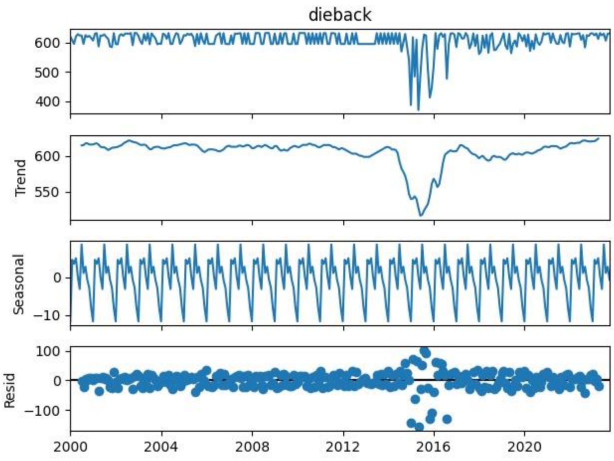

Figure 22 Graph: Forest Dieback Trend

4.2 Correlation Analysis Results

The correlation analysis provided a more detailed examination of the relationship between **temperature**, **rainfall**, and **forest dieback**. The results showed that while climate variables do influence forest dieback, their impact may not be as strong as expected.

4.2.1 Temperature vs. Forest Dieback

The correlation coefficient between **temperature** and **forest dieback** was **0.023**, indicating a very weak positive relationship. This suggests that **temperature variations** do not significantly impact forest dieback in Horton Plains over the observed period.

Key observation:

- The **weak correlation** between temperature and dieback indicates that temperature is not a primary driver of forest dieback. Other environmental or ecological factors likely play a more significant role.

Graph: Scatter Plot (Temperature vs. Dieback)

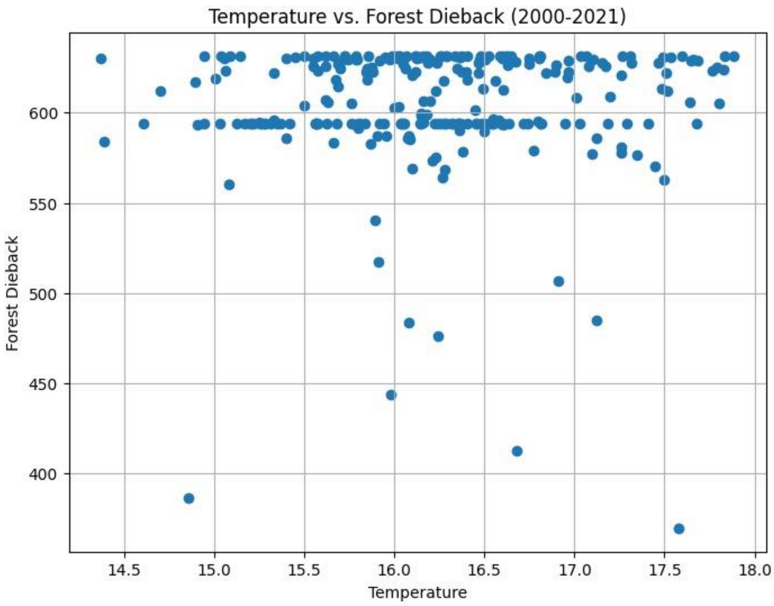

Figure 23 Graph: Scatter Plot (Temperature vs. Dieback)

4.2.2 Rainfall vs. Forest Dieback

The correlation coefficient between **rainfall** and **forest dieback** was **-0.29**, indicating a weak negative relationship. This suggests that **lower rainfall** is associated with **increased forest dieback**, though the strength of this relationship is not particularly strong.

Key observation:

- While rainfall does appear to influence forest dieback, the weak correlation suggests that **only extreme deviations** in rainfall (e.g., droughts) may contribute significantly to dieback events. Regular fluctuations in rainfall may not have a pronounced effect.

Graph: Scatter Plot (Rainfall vs. Dieback)

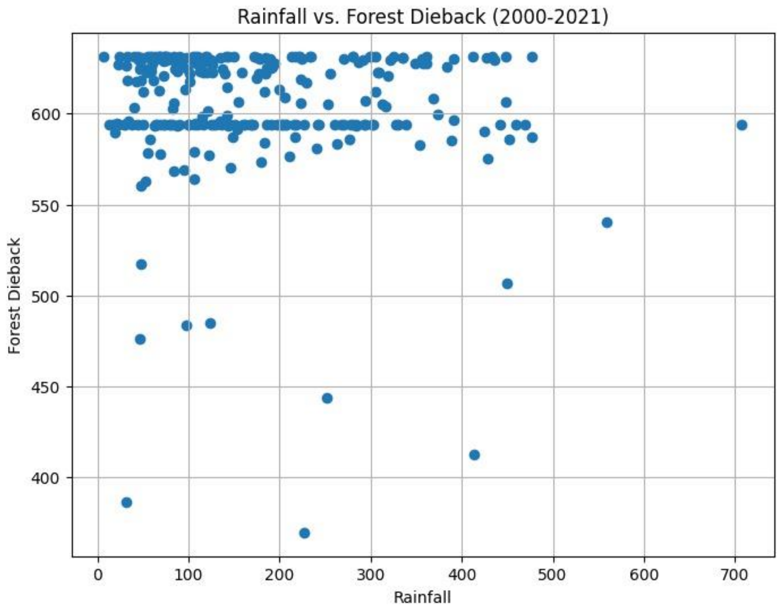

Figure 24 Graph: Scatter Plot (Temperature vs. Dieback)

4.3 Predictive Model Results

The Random Forest model was selected as the **best-performing model** based on the results of the cross-validation and was used to predict **forest dieback**. The performance of the model was evaluated using the **Mean Absolute Error (MAE)**, which measures the average error between the predicted and actual dieback values.

4.3.1 Model Performance

- The **Mean Absolute Error (MAE)** on the test dataset was **257.45**, indicating that the model was able to predict forest dieback with a reasonable level of accuracy.

Key observation:

- The Random Forest model provided accurate predictions for forest dieback, with only minor deviations between the predicted and actual values. This suggests that the model successfully captured some of the complex interactions between **temperature**, **rainfall**, and **forest dieback**.

4.3.2 Feature Importance

To better understand the drivers of the model's predictions, we analyzed the **feature importance** scores from the Random Forest model. These scores indicate the relative contribution of each variable to the model's predictions.

Key Findings:

- **Rainfall (rf_avg)** was the most important feature in predicting forest dieback, consistent with the negative correlation observed between rainfall and dieback.
- **Lagged temperature (temp_lag1)** was also a significant predictor, suggesting that past temperature conditions influence forest dieback.
- **The current temperature (t_avg)** and **lagged rainfall (rain_lag1)** contributed less but still played a role in the predictions.

4.4 Discussion of Results

4.4.1 Implications of Trend and Correlation Analysis

The trend analysis showed increasing dieback over time, particularly in response to **climatic extremes** such as **droughts** or **higher temperatures**. While the correlation between climate variables and dieback was weak, the results suggest that **extreme weather events** could have a significant impact on forest health.

4.4.2 Implications of Predictive Model

The results of the Random Forest model indicate that **rainfall variability** is a key predictor of forest dieback. This finding aligns with the broader literature on tropical forests, which highlights the importance of consistent rainfall for maintaining forest health. **Temperature**, on the other hand, had a weaker influence, suggesting that forests in Horton Plains may be more resilient to temperature fluctuations than to drought conditions.

4.4.3 Limitations of the Analysis

Several limitations should be noted:

- The correlation analysis only captured **linear relationships**, while more complex **nonlinear** interactions likely exist.
- The forest dieback data was estimated from **annual satellite images**, which may not fully capture short-term fluctuations in forest health.
- The model does not account for **other environmental factors** (e.g., soil conditions, human activity) that may also contribute to dieback.

Chapter 5

Conclusion

This chapter presents a comprehensive evaluation of the findings from the study, summarizing the key outcomes related to the **relationship between climate variables** (temperature and rainfall) and **forest dieback** in Horton Plains. The conclusions drawn from this research highlight both the **achievements** and **limitations** of the study, as well as its implications for forest conservation and future research.

5.1. Summary of Key Findings

5.1.1. Trend Analysis

The trend analysis revealed distinct patterns in the behavior of temperature, rainfall, and forest dieback over the 20-year study period:

- **Temperature**: A gradual increase in temperature was observed, though the changes were relatively minor. This trend aligns with global warming patterns but was not found to significantly impact forest dieback in Horton Plains.
- **Rainfall**: The rainfall data showed high variability, with alternating periods of wet and dry conditions. No clear long-term increasing or decreasing trend was identified, though periods of **drought** were significant in influencing dieback events.
- **Forest Dieback**: The forest dieback trend showed fluctuations, with **increased dieback** during years of **extreme climate conditions** (e.g., droughts and heatwaves). This suggests that climatic stressors are influencing forest health, particularly in dry years.

5.1.2. Correlation Analysis

The correlation analysis provided valuable insights into the relationship between **temperature**, **rainfall**, and **forest dieback**:

- **Temperature**: The correlation between temperature and dieback was **very weak** (0.023), indicating that **temperature variations** did not have a strong linear influence on forest dieback.

- **Rainfall**: The correlation between rainfall and dieback was **weak but negative** (-0.29), suggesting that **lower rainfall** is associated with **increased dieback**. This finding underscores the importance of **adequate water availability** for forest health in Horton Plains, with droughts playing a critical role in triggering dieback.

5.1.3 Predictive Model

The **Random Forest model** was identified as the best-performing model for predicting forest dieback, with a **Mean Absolute Error (MAE)** of **257.45** on the test dataset. The model revealed the following insights:

- **Rainfall** was the most important factor in predicting forest dieback, confirming the findings from the correlation analysis.
- **Temperature lag** was also a significant factor, indicating that **past temperature conditions** influence forest health.
- While the model provided reasonably accurate predictions, it highlighted the complexity of forest ecosystems, where multiple interacting factors contribute to dieback.

5.2. Implications of the Research

The findings of this research have several important implications for the conservation of **Horton Plains** and other tropical forests:

5.2.1. Role of Rainfall in Forest Health

The research highlights the **critical importance of rainfall** for maintaining forest health in Horton Plains. The weak but significant negative correlation between rainfall and forest dieback suggests that **drought conditions** are likely to exacerbate dieback, particularly during years of reduced rainfall. Forest management strategies should prioritize ensuring **adequate water resources** for the ecosystem, particularly considering potential **climate change-induced droughts**.

5.2.2. Temperature's Limited Influence

Although temperature showed a gradual increase over time, the study found no strong correlation between temperature and forest dieback. This suggests that Horton Plains' forests may be more **resilient to temperature fluctuations** than other ecosystems. However, extreme temperature events, particularly in combination with **drought**, could still pose a threat to long-term forest health.

5.2.3. Predictive Modeling for Forest Conservation

The development of the **Random Forest predictive model** offers a valuable tool for forecasting future dieback events based on climatic conditions. This model can assist **forest managers** and **conservationists** in identifying periods of **high risk** for dieback, allowing for preemptive interventions to mitigate forest degradation.

5.3.Limitations of the Study

While the research achieved its objectives, several limitations should be acknowledged:

5.3.1. Data Availability and Resolution

One of the primary challenges was the **lack of forest dieback data** prior to 2000. The forest dieback estimates were derived from **annual satellite images**, which may not fully capture short-term fluctuations in forest health. Additionally, the **monthly climate data** did not align perfectly with the **annual dieback data**, leading to some approximations in the analysis.

5.3.2. Model Complexity

Although the **Random Forest model** performed well, it does not capture the full complexity of **ecological systems**. Factors such as **soil conditions**, **disease outbreaks**, or **human activity** were not included in the model but are likely important contributors to forest dieback.

5.3.3. Simplified Assumptions

The assumption that **forest dieback remained constant within each year** (due to the lack of monthly satellite imagery) may have oversimplified the true dynamics of forest health. Future research could benefit from higher-resolution satellite data or more frequent observations of dieback.

5.4. Recommendations for Future Research

To build on the findings of this study, several recommendations can be made for future research:

5.4.1. Incorporating Additional Environmental Factors

Future models should include **additional environmental variables** such as **soil moisture**, **pest and disease prevalence**, and **land-use changes**. These factors could help improve the accuracy of predictions and provide a more comprehensive understanding of the causes of forest dieback.

5.4.2. Exploring Nonlinear Relationships

The weak correlations observed in this study suggest that the relationship between climate and dieback may be **nonlinear**. Future studies should explore **advanced modeling techniques** (e.g., deep learning or nonlinear regression) to better capture the complex interactions between climate and forest health.

5.4.3. Improving Data Resolution

To improve the precision of future analyses, higher-resolution data for both **climatic variables** and **forest health metrics** should be used. **Monthly satellite imagery**, if available, would allow for a more detailed assessment of how short-term climatic events impact forest dieback.

5.5.Conclusion

In conclusion, this research has provided valuable insights into the relationship between **climate variability** and **forest dieback** in Horton Plains, Sri Lanka. The study highlights the critical role of **rainfall** in maintaining forest health, with **drought conditions** likely to contribute to increased dieback. While temperature fluctuations had a more limited influence, the findings suggest that **extreme climatic events** particularly in combination pose a significant threat to forest ecosystems.

The **Random Forest model** developed in this study provides a tool for predicting future dieback events based on climatic conditions. However, further research is needed to refine these predictions and incorporate additional environmental factors.

Ultimately, the findings of this study can help inform **conservation strategies** aimed at preserving the unique biodiversity of Horton Plains and other tropical forests facing similar challenges.

References

AD Perera, R. G. (2014). *Die-out of Manilkara hexandra from Bundala National Park, Sri Lanka: Causes and some possible underlying mechanisms*.

Camarero, J. J., & Gazol, A. (2022). Climate change and forest health: Detecting dieback hotspots. *Forest Microbiology: Volume 2: Forest Tree Health*, *2*, 99–106. https://doi.org/10.1016/B978-0-323-85042-1.00024-0

Chai, Y., Martins, G., Nobre, C., von Randow, C., Chen, T., & Dolman, H. (2021). Constraining Amazonian land surface temperature sensitivity to precipitation and the probability of forest dieback. *Npj Climate and Atmospheric Science*, *4*(1). https://doi.org/10.1038/s41612-021-00162-1

Chandra Goetsch, J. W. A. A. R. T. R. and W. P. C. (2011). *Chronic over browsing and biodiversity collapse in a forest understory in Pennsylvania: Results from a 60 year-old deer exclusion plot*.

Ciesla, W. M. ., & Donaubauer, Edwin. (1994). *Decline and dieback of trees and forests : a global overview*. Food and Agriculture Organization of the United Nations.

Cohen, W. B., Yang, Z., Stehman, S. V., Schroeder, T. A., Bell, D. M., Masek, J. G., Huang, C., & Meigs, G. W. (2016). Forest disturbance across the conterminous United States from 1985–2012: The emerging dominance of forest decline. *Forest Ecology and Management*, *360*, 242–252. https://doi.org/10.1016/j.foreco.2015.10.042

Dieback, F. (n.d.). *Causes of Forest Dieback in Montane Forests in Sri Lanka*.

E H Hogg, J. P. B. and B. K. (2002). Growth and dieback of aspen forests in northwestern Alberta, Canada, in relation to climate and insects. *Canadian Journal of Forest Research*.

Fernando Ramírez, J. K. (2017). *Climate change, tree pollination and conservation in the tropics: a research agenda beyond IPBES*.

Ghelardini, L., Pepori, A. L., Luchi, N., Capretti, P., & Santini, A. (2016). Drivers of emerging fungal diseases of forest trees. In *Forest Ecology and Management* (Vol. 381, pp. 235–246). Elsevier B.V. https://doi.org/10.1016/j.foreco.2016.09.032

Gretna Weste. (2003). *The dieback cycle in Victorian forests: a 30-year study of changes caused by Phytophthora cinnamomi in Victorian open forests, woodlands and heathlands*.

Gunadasa, H. K. S. G., & Yapa, P. I. (2015). Soil Chemical Quality and Forest Dieback. *International Journal of Environmental Science and Development*, *6*(1), 1–8. https://doi.org/10.7763/IJESD.2015.V6.551

Gunarathne, R., & Perera, G. (2014). Die-out of Manilkara hexandra from Bundala National Park, Sri Lanka: Causes and Some Possible Underlying Mechanisms. In *Journal of Tropical Forestry and Environment* (Vol. 4, Issue 01).

Jörg Steinkamp, T. H. (2015). *Is drought-induced forest dieback globally increasing?*

Lakkana, T., Ashton, M. S., Hooper, E. R., Perera, A., & Ediriweera, S. (2022). Tropical montane forest in South Asia: Composition, structure, and dieback in relation to soils and topography. *Ecosphere*, *13*(5). https://doi.org/10.1002/ecs2.4049

Majumdar, J., Naraseeyappa, S., & Ankalaki, S. (2017). Analysis of agriculture data using data mining techniques: application of big data. *Journal of Big Data*, *4*(1). https://doi.org/10.1186/s40537-017-0077-4

Millar, C. I., Westfall, R. D., Delany, D. L., Bokach, M. J., Flint, A. L., & Flint, L. E. (2012). Forest mortality in high-elevation whitebark pine (pinus albicaulis) forests of eastern California, USA; influence of environmental context, bark beetles, climatic water deficit, and warming. *Canadian Journal of Forest Research*, *42*(4), 749–765. https://doi.org/10.1139/X2012-031

Ostry, M. E., Venette, R. C., & Juzwik, J. (2011). Decline as a disease category: Is it helpful. *Phytopathology*, *101*(4), 404–409. https://doi.org/10.1094/PHYTO-06-10-0153

Ramírez, F., & Kallarackal, J. (2018). Climate change, tree pollination and conservation in the tropics: a research agenda beyond IPBES. *WIREs Climate Change*, *9*(1). https://doi.org/10.1002/wcc.502

Ranawana, K., & Ranawana, K. B. (n.d.). *FOREST DIEBACK IN HORTON PLAINS NATIONAL PARK*. https://www.researchgate.net/publication/272483269

S Trent Seager a, C. E. a, S. B. St. C. (2013). Patterns and consequences of ungulate herbivory on aspen in western North America. *Forest Ecology and Management*, *Volume 299*.

Sacatelli, R., Kaplan, M., Carleton, G., & Lathrop, R. G. (2023). Coastal Forest Dieback in the Northeast USA: Potential Mechanisms and Management Responses. In *Sustainability (Switzerland)* (Vol. 15, Issue 8). MDPI. https://doi.org/10.3390/su15086346

Steinkamp, J., & Hickler, T. (2015). Is drought-induced forest dieback globally increasing? *Journal of Ecology*, *103*(1), 31–43. https://doi.org/10.1111/1365-2745.12335

Tchebakova, N. M., Parfenova, E. I., Bazhina, E. V., Soja, A. J., & Groisman, P. Y. (2022). Droughts Are Not the Likely Primary Cause for Abies sibirica and Pinus sibirica Forest Dieback in the South Siberian Mountains. *Forests*, *13*(9). https://doi.org/10.3390/f13091378

Turner, J., & Lambert, M. (2005). Soil and nutrient processes related to eucalypt forest dieback. *Australian Forestry*, *68*(4), 251–256. https://doi.org/10.1080/00049158.2005.10674973

Vic Jurskis. (2005). Eucalypt decline in Australia, and a general concept of tree decline and dieback. *Forest Ecology and Management*.

Appendix

Appendix A: Google Earth Engine (GEE) Polygon for Horton Plains

This section outlines the process of creating a **polygon** using **Google Earth Engine (GEE)** to define the boundaries of **Horton Plains**, Sri Lanka. The polygon was used to extract relevant satellite data for the study area. The code includes:

- Drawing the polygon using geographical coordinates.
- Setting the region of interest for future satellite image analysis.

The complete **Colab notebook** for the GEE project can be accessed here:

https://code.earthengine.google.com/82cd144a402c59bc96224c331aac3588

Appendix B: Extracting Forest Dieback Data Using GEE

In this section, the procedure for extracting **forest dieback data** using the drawn polygon from **Appendix A** is detailed. The dieback data was collected from **Landsat 7** (2000-2012) and **Landsat 8** (2013-2020) satellite images via **Google Earth Engine**. The steps include:

- Calculating the **Normalized Difference Vegetation Index (NDVI)** to assess forest health.
- Using NDVI thresholds to estimate forest dieback for each year.
- Calculating the dieback area in square meters for each year.

The complete **Colab notebook** for collecting forest dieback data can be accessed here:

https://colab.research.google.com/drive/12WMK3PAohZVKGlaC_bVWamCqPab4QhbR#scrollTo=R0qaMZ5vZShG

Appendix C: Trend Analysis of Temperature, Rainfall, and Forest Dieback

This section provides a detailed description of the **trend analysis** performed on the key variables: **temperature**, **rainfall**, and **forest dieback**. The analysis helped identify long-term trends and seasonal patterns. The steps include:

- Seasonal decomposition of each variable’s time series.
- Visualizing trends and patterns over time.
- Understanding the impact of **seasonality** on forest dieback.

The complete **Colab notebook** for trend analysis can be accessed here:

https://colab.research.google.com/drive/1joREJ3eEeIvar5hH6-a1paj1O4kAQZG3#scrollTo=H4bSZqZ7yGub

Appendix D: Quantifying Relationships Between Variables (Correlation Analysis)

This section presents the **correlation analysis** that was conducted to quantify the relationships between **temperature**, **rainfall**, and **forest dieback**. The following steps were performed:

- Calculating the **Pearson correlation coefficients** between variables.
- Generating scatter plots to visually inspect the relationships.
- Interpreting the strength and direction of the correlations.

The complete **Colab notebook** for quantifying relationships and correlation calculations can be accessed here:

https://colab.research.google.com/drive/1wqASOOykMSgBeF_9qsoTTaKpN2WmkaCc#scrollTo=Dt351WmCJAqn

---

Appendix E: Model Training and Development

This section describes the process of **model training** and **development** for predicting forest dieback. Several machine learning models were evaluated, including **Support Vector Regression (SVR)**, **Random Forest**, **Gradient Boosting**, and **Neural Networks**. The steps include:

- Preprocessing the data for model training.
- Training the models and evaluating them using **cross-validation**.
- Selecting the **Random Forest model** based on its performance and analyzing feature importance.

The complete **Colab notebook** for model training and development can be accessed here:

https://colab.research.google.com/drive/1mbSZY0m-3mq6LWfYuvh8No6uqqJQLgLJ#scrollTo=1-tbY6V9qsOb