Impact of climate signal on Precipitation Variability in India and Southeast Asia: An Analysis Using TRMM Data

Pooja Priya Jeyanthi Sethumathavan
201679518

Dr Stephen Griffiths

Submitted in accordance with the requirements for the
module MATH5872M: Dissertation in Data Science and Analitics
as part of the degree of

Master of Science in Data Science and Analystics

The University of Leeds, School of Mathematics

September 2024

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Remember to include in here a signed and scanned pdf copy of your academic integrity form.

Abstract

The dissertation examines the influences of large-scale climatic phenomena, such as El Niño-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD), on the precipitation pattern over India and Southeast Asia for a continuous period of 22 years from 1998 to 2019. In this context, the present study uses time series, correlation analysis, and Empirical Orthogonal Function (EOF) analysis in order to analyse the temporal and spatial variability of rainfall across these regions using TRMM data. The analysis indicates highly significant relations between the phases of ENSO and precipitation anomalies, showing strong negative correlations to imply that generally, El Niño is associated with below-normal rainfall, while La Niña is associated with above-normal precipitation. The influence of the IOD is rather regional; parts of Southeast Asia, especially Indonesia and Malaysia, are significantly responsive to its phases. EOF analysis focuses on the dominant patterns of spatial rainfall variability and how seasonal monsoons and large-scale climatic drivers shape precipitation over these regions. Accordingly, the research contributions are targeted at enhancing the knowledge of climate variability and the resulting implications for the management of water resources, agriculture, and disaster preparedness that are very sensitive to the seasonal rainfall patterns.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Rainfall is crucial in terms of contributing to the economy, agriculture, and health of the environment in general. Climate variability influences rainfall strongly, and the effects are evident in agriculture, drought, and flooding. Precipitation is one of the most vital components of Earth's water cycle, where its irregularities have impacts on ecosystems, agriculture, and water resources, thus affecting human livelihoods, Stseo (2020). However, the patterns of precipitation are not uniform, and therefore depend on several factors which include large-scale climate phenomena such as El Niño-Southern Oscillation (ENSO) and the Indian Ocean Dipole (IOD).

A periodic sea surface temperature fluctuation with simultaneous changes of atmospheric high and low pressure over the equatorial Pacific Ocean as shown in the Figure 1.1. Phases include the following:

- **El Niño:** During an El Niño, trade winds are weaker than usual. This allows warmer than usual sea surface temperatures to spread across the central and eastern Pacific toward the South American coast. This warming of the ocean leads to increased rainfall and flooding along western South America and the hotter ocean water pushes the normal seasonal weather patterns in other parts of the world out of balance. Other regions around Southeast Asia, Australia experience a comparative period of drought with low rainfall amounts and often include periods of drought. The phenomenon also tends to induce short-term increases in global temperatures through the short-term redistribution of heat from the ocean into the atmosphere, NOAA (2024).

- **La Niña:** In La Niña, a strengthening of the trade winds pushes the warm waters westward. This allows the cooler-than-average sea surface temperatures to take over the central and eastern Pacific, especially right toward the coast of South America. The typical outcome of this is above-average rainfall over Southeast Asia, Australia, and parts of South Asia that lead to a heightened flood risk, while during La Niña, the western coast of South America is usually considerably drier than normal, Stallard & E. (2024).

Another climatic phenomenon is the IOD, or Indian Ocean Dipole, referring to fluctuations in sea surface temperatures between the western and eastern parts of the Indian Ocean. Similar

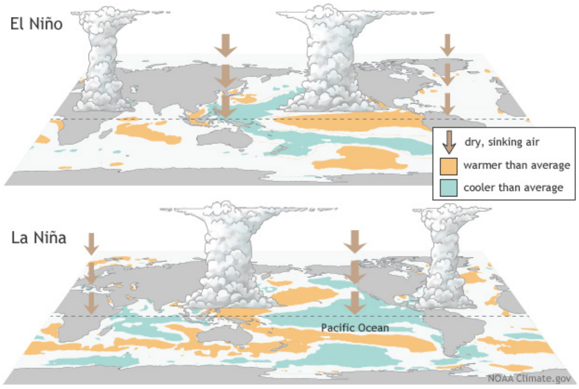


Figure 1.1: Typical pattern of ENSO

Taken from the Source: `https://www.climate.gov/enso`

to ENSO, the IOD operates on a two-phase system, NOAA Climate.gov (2020):

- **Positive IOD:** During this phase, the western Indian ocean becomes much warmer than it should be and even more on the western side. This temperature differential often results in decreased rainfall across areas like Australia and parts of Southeast Asia. Warming of waters in the Indian Ocean to its west can still be associated with increased monsoon rainfall over India, while at other times a drought would be expected for eastern Africa.
- **Negative IOD:** The negative phase of the IOD means the western central part of the Indian Ocean is cooler, and the eastern part is warmer; this normally results in above-normal rain over Australia and Southeast Asia. During some events, this may turn out to be below or near normal during much of the monsoon over India, adding to drought conditions.

ENSO and IOD are the leading global and regional drivers for climate variability, not only in terms of the patterns of rainfall and temperature but even down to the very frequency and intensity of tropical cyclones. Knowledge of global and regional precipitation patterns plays an important role in dealing with the environmental challenges that come with the contemporary issue of climate change. Such knowledge can be extremely helpful in forecasting future climatic scenarios and formulating plans to lessen the effects of climate variability on areas that are already vulnerable. In this, the satellite-based datasets play a important role. The satellite instruments have been observing the atmosphere and ocean and land surface for a long period of time. Among these qualities, they measure precipitation. One of such product is the Tropical Rainfall Measuring Mission (TRMM), offering a unique and comprehensive view of the global precipitation. The TRMM dataset is from 1998 to 2019, with high resolution, is an important complement to the ground-based observations that are usually sparse or not available in tropical regions. The word "Tropical" included in TRMM underlines the fact that the focus is on tropical regions, which naturally are much more sensitive to climate variability regarding their geographical location, NASA Global Precipitation Measurement Mission (2024).

1.1 Aims

In this paper, TRMM is utilized to investigate patterns of precipitation over a 22-year period (1998-2019) with special emphasis on the Indian subcontinent and Southeast Asia. Both regions are quite susceptible to seasonal monsoons as well as large-scale climatic phenomena; hence, their appropriateness for the purpose of studying ENSO and IOD with respect to precipitation. Objectives of this research include the following:

1. **Time Series Analysis:** The normal annual, monthly, seasonal cycle, and anomalies of precipitation in selected regions of India and Southeast Asia.

2. **Climate Signal Correlation:** To study the correlations between precipitation anomalies and climate signals such as ENSO, IOD in various regions of India and Southeast Asia, also perform a seasonal analysis in order to provide more detailed information about these relationships.

3. **EOF Analysis:** Perform EOF analysis on raw precipitation data,to identify dominant spatial patterns in rainfall from the chosen regions for an understanding of the variability and drivers of rainfall.

This will, in turn, add to the knowledge on large-scale climate phenomena that influence precipitation patterns in tropical regions often most affected by climate variability. The present work examines exactly how these patterns have changed over the past two decades, using long-term data from TRMM, something of vital importance to be expected in the prediction of future trends. Using this finding, precipitation prediction may inform water resource management and agricultural planning in these regions, that are highly dependent on such seasonal rainfall patterns, Columbia Climate School (2019).

1.2 Structure of the dissertation

In this paper, Chapter 2 gives an overview of the datasets used, especially TRMM, and how it relates to our study. In chapter 3, we talk about preliminary analysis, including global. Chapter 4 and 5 deals with regional precipitation analysis and anomalies. Chapters 6-7 present correlation of precipitation anomalies with climate signals such as ENSO and IOD, adding statistical tests of significance where necessary. Chapter 8 introduces EOF analysis and its application to the investigation of precipitation spatial and temporal patterns for the study regions.

Chapter 2

Dataset Overview

The Tropical Rainfall Measuring Mission, or TRMM, was a pioneering research satellite that operated between 1997 and 2015. It was designed to study better the distribution and variability of precipitation within the tropics. It has played an important part in studying the water cycle in the current climate system. Although the TRMM was decommissioned on April 15, 2015, its legacy continues in the form of continued data afforded by the Integrated Multi-Satellite Retrievals for GPM (IMERG) product, Climate Data Guide (2023). IMERG combines data from the GPM mission and other satellites, extending the precipitation record up to 2019. This allows seamless continuation of this dataset, building on the TRMM legacy even long after the satellite mission ended. In the end, we will also discuss about the El Nino Index dataset and IOD index dataset.

2.1 TRMM Instruments and Data Products

Some of the main instruments used to measure precipitation included:

- **Precipitation Radar (PR):** Provided three-dimensional profiles of precipitation.
- **TRMM Microwave Imager (TMI):** A multi-channel radiometer used for rain intensity measurement.
- **Visible and Infrared Scanner (VIRS):** Assisted in identifying cloud coverage and provided background data for the other instruments.

Complementing the data from the above-mentioned instruments were two other instruments, namely Clouds and Earth's Radiant Energy System-CERS and the Lightning Imaging Sensor-LIS. These instruments were crucial to the TRMM Combined Instrument (TCI) calibration dataset called TRMM 2B31 used in the TRMM Multi-satellite Precipitation Analysis (TMPA), NASA Global Precipitation Measurement Mission (2024). Three major types of datasets produced by TRMM; they are Daily TMPA 3B42 , 3-hourly TMPA 3B42 and Monthly TMPA 3B43. In this paper, we have selected the monthly averages from TRMM 3B43 to be used because they are robust and averaged. Hence, they are better suited for use in any long-term climatological study.

2.2 The Use of the TRMM 3B43 Dataset

The selection of the appropriate dataset is considered one of the important steps in climatological research, while considering a long-term precipitation pattern. Resolution, accuracy, and reliability of data may affect the outcome or conclusion of the study significantly. To support the choice of monthly dataset, we examined several key studies. The decision to use the TRMM 3B43 is further supported by relevant literature that highlights its suitability for similar research.

Koad & Jaroensutasinee (2020),the dataset used in this paper were sourced from TRMM Multi-satellite Precipitation Analysis (TMPA) Version 7, more specifically from the 3-hourly 3B42. The authors aggregated this high-frequency data into monthly precipitation estimates that could capture the annual cycles of precipitation across tropical and subtropical regions. The methodology gives an added edge in point analysis through the capturing of short-term variability and exacting precipitation events that are fundamental in regions with complicated precipitation patterns. Using the bimodal von Mises distribution, it was possible for this study to extract precipitation cycles in different geographical areas and underlined how these precipitation cycles vary according to geographical location and environmental conditions. Therefore, because Koad and Jaroensutasinee did their analysis with a lot of detail using 3-hourly data, we could question which dataset to base our research from, whether to base it on 3-hourly TMPA 3B42 that involves aggregation and complex statistical modelling or the more straightforward pre-aggregated monthly TRMM 3B43 data? This choice is all the more critical in view of the wider spatial and temporal focus of our study on India and Southeast Asia. While 3-hourly data are more granular and capture finer details in the variability of the precipitation, on the other hand, TRMM 3B43 is much easier dataset to operate and interpret, especially in long-term climatological studies that focus on understanding large-scale patterns and trends.

To this question, we will refer to the study Scheel et al. (2010), which analysed the ability of TMPA 3B42 to estimate at a daily scale with a resolution of $0.25^\circ \times 0.25^\circ$ precipitation rates for the Central Andes. The results shows,TMPA generally performs well in capturing the occurrence of strong precipitation events,as data gets aggregated into coarser temporal resolutions, namely weekly, biweekly, and monthly, the correlation against the ground improves considerably. This points to the fact that, monthly TMPA data gives a decent estimate for long-term climatological analysis, when a daily resolution in great detail is not required. Explicitly, the authors use the Pearson correlation coefficient as a measure of linear association between the TMPA estimates and the ground-based gauge measurements. This is a fundamental finding because it indicates that the use of the TRMM 3B43 dataset would not lead to the loss of accuracy in our analysis. Rather, it provides a simplification of data management and interpretation.

Singh et al. (2019),this study further supports the use of TRMM monthly data by directly comparing the TRMM 3B42 data with IMD ground measurements over Maharashtra, India, in the monsoon months of 2004-2013. The study suggest that the TRMM dataset, particularly when considered over a longer temporal scale such as weekly or monthly, is well-suited for regional precipitation analyses in India, especially in areas of complex geography such as the Western Ghats. They have, therefore, arrived at a decision that the TRMM data can be a viable alternative to ground observation. This study also applies the Pearson correlation coefficient in assessing the relationship of satellite-derived and ground-based precipitation data. This finding has direct applicability to my research, since the reliability of TRMM data in the Indian context is verified, and we should be assured that TRMM 3B43 is suitable for our analysis related to precipitation patterns in India and Southeast Asia.

These three studies collectively satisfy the rationale underlying the use of the TRMM 3B43 monthly dataset for our study. This choice allows the focusing on a much larger time pattern and relationships that are particularly central to our studies, such as the impacts of ENSO and IOD on precipitation patterns, without the complexity and extra processing required by 3-hourly data. Since this dataset has already been validated for application in similar research contexts, we are assured our results will be accurate and reliable. Hence, the TRMM 3B43 Dataset would best fit our research objectives.

2.3 Data Processing and Handling

We established the rationale for the use of the TRMM 3B43 dataset, what would follow next would be an understanding of how this data is processed and handled. The TRMM 3B43 data are a monthly version of the 3B42 dataset; Monthly precipitation values are aggregated from daily and 3-hourly data that make up dataset 3B43. This is done by summing or averaging the values over an entire month for increased stability with less noise compared to daily and 3-hourly observations.Thus, the TMPA algorithm integrates the data products of several sources: microwave sensors and infrared. This creates, however, the possibility of inconsistency or irregularity within the dataset, as there are many different sources these data are coming from. While microwave estimates are more precise,they are less frequent, hence, infrared data, though more frequent but not as accurate, needs to be used to fill in temporal gaps, Huffman et al. (2015). A balance between accuracy and coverage is important in providing confidence in the TMPA dataset. Over time, in the process of new satellites replacing old ones that have been decommissioned - such as TRMM in 2015, sources of data have changed. This has usually resulted in possible lapses in data coverage, in the case of TRMM's monthly dataset, 3B43, missing data points are usually simply filled with a value rather than being interpolated over. This is because interpolation, though useful in some contexts, may introduce artificial data that may not reflect the actual precipitation patterns, especially in scientific research and analysis. TRMM therefore fills the missing value with an extreme negative value of "-9999.9003" as a fill value to fill the missing value and mark the gaps. Specifically, this shows the most refined and precise version

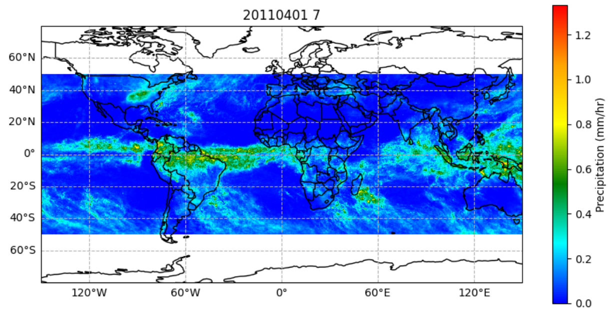


Figure 2.1: Global precipitation patterns for March 2011

of the TRMM 3B43 dataset from the final release of TRMM data products, Version 7. This version includes a number of enhancements over its predecessor and past versions and improves reliability for long-term studies of the dataset

2.4 Dataset Extraction and Information

TRMM 3B43 monthly data were downloaded from the NASA Goddard Earth Sciences Data and Information Services Centre (GES DISC) via `https://disc.gsfc.nasa.gov/datasets/TRMM_3B43_7/summary`. We use the option "Get file subsets using OPeNDAP" to access the data to choose time range and variables precisely. Then, we selected a "precipitation" variable and downloaded data for 22 years in the NetCDF format, which is one standard format for climate data, OPENDAP HyrAx (2024). In our application, we used Python scripting along with the command-line utility program **wget** for automatizing the download. The environment was set up with necessary authentication files like `.netrc`, `.urs_cookies`, and `.dodsrc`, allowing a safe and smooth download of the data, NASA GES DISC (2024*b*). This procedure enabled us to download and organize in a very effective way the 264 monthly datasets that later on would be used for analysis. The dataset was provided in NetCDF4 file format. The dimensions of the dataset are 1440×400 grid point (longitude * latitude). So, in total it has 5,76,000 grid cells. As shown in the Figure 2.1, notice that the pattern of La Niña is evident, we could see the higher precipitation around the western pacific regions and lower than average around the eastern and central pacific regions.

The TRMM 3B43 dataset provides monthly averaged precipitation across a global grid, with each grid cell representing a specific geographic location. The variables include, precipitation (mm/hr), nlon (longitude), and nlat (latitude).The geographical coverage of $50°N$ to $50°S$ and $180°E$ to $180°W$. The resolution stands at $0.25° \times 0.25°$. The Fill value is denoted as "-9999.9" NASA GES DISC (2024*a*). The resolution is coarse for localized studies; however, it is acceptable in defining the larger-scale climatic trends. The TRMM 3B43 monthly dataset may be especially useful in the study of long-term climate features, seasonal cycles, and annual trends. Unlike the daily or 3-hourly datasets, which are susceptible to short-term variability in space and time, this monthly dataset carries a more robust view in carrying the precipitation patterns. Selection of TRMM 3B43 in our study is justified because this dataset represents a suitability for the capture of long-term trends. Secondly, this dataset carries along gauge-based corrections which enhance the accuracy of the data. While IMERG dataset indeed is of higher spatial resolution with improved temporal consistency of records, TRMM provides a continuous record from 1998 into the present that is quite indispensable for the analysis of long-term precipitation trends. Benefits of using TRMM data include coverage of global tropics often under-monitored by ground-based stations. Also, the incorporation of ground-based rain gauge data enhances the accuracy of the satellite-derived precipitation estimates. The disadvantages of using TRMM might include that the $0.25^\circ \times 0.25^\circ$ resolution could be too coarse for localized studies or to understand microclimates. Also, while the monthly data is adequate for climate studies, it lacks more detailed temporal features captured by daily and 3-hourly data. Since the focus of our study is long-term climatological trends, this limitation imposed by coarse resolution is easily overcome by the benefits of TRMM 3B43, Climate Data Guide (2023).

2.5 ENSO Index and IOD Index Dataset Overview

The Multivariate ENSO Index Version 2 (MEI.v2) is widely used which defines the intensity and phase of ENSO events . MEI multivariately combines six meteorological and oceanographic variables: sea-level pressure, surface winds, SST, surface air temperature, and cloudiness of the tropical Pacific Ocean (30°N-30°S, 100°E-70°W). After the collection of variables, PCA is applied to these variables, and the first principal component is used to compute the MEI values, representing ENSO variability, Team (2024). It provides a robust measure of the ENSO's impact in terms of value. The Values of the index represent deviations in the long-term mean for a defined two-month period, such as DJ-December-January, JF-January-February. For positive values, conditions reflect El Niño, while for negative values, they indicate La Niña. The original MEI.v2 was downloaded from its official website at NOAA: `https://psl.noaa.gov/enso/mei/`. Because the MEI.v2 data record every two months and our precipitation anomalies are on a month-to-month basis, we then interpolated the MEI.v2 index for the monthly time series. Such interpolation is obtained by averaging the values between consecutive bimonthly periods. After processing, the interpolated data is saved for further processing in a CSV file. First, to check that our interpolated MEI.v2 time series was accurate, we plotted the anomaly index and compared it against the official NOAA MEI.v2 anomaly plot, refer to the Figure 2.2. The comparison indeed revealed that the interpolation was carried out without deviation from the trends and anomalies presented by the official data.

The Indian Ocean Dipole is quantified using Dipole Mode Index describing anomalies of sea surface temperature between two regions:

1. Western Indian Ocean: Usually falls between 50°E - 70°E and 10°S - 10°N.

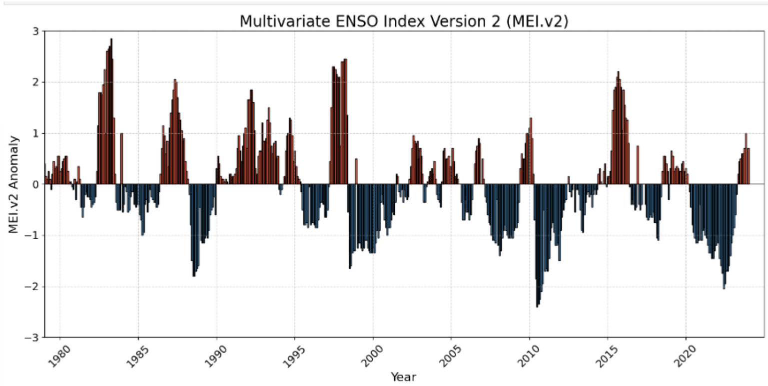


Figure 2.2: MEI.v2 index Anomaly Plot

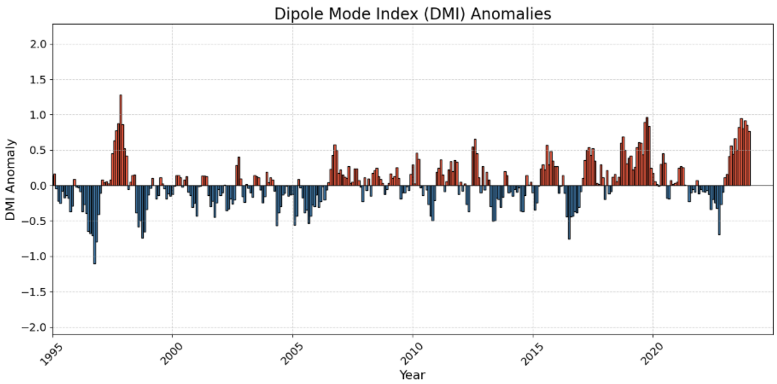


Figure 2.3: DMI index Anomaly Plot

2. Eastern Indian Ocean: Usually measured at 90°E - 100°E and 10°S - 0°N.

DMI is the anomaly difference between the west sea surface temperature and that of the east. Positive phases of the IOD are underlined by positive values of DMI, NASA Jet Propulsion Laboratory (2024). Conversely, negative phases of the IOD have their values of DMI being negative, with conditions being reversed. In this analysis, we used the DMI data obtained from the NOAA website, `https://psl.noaa.gov/gcos_wgsp/Timeseries/DMI/`,which ranges from 1870 to 2024, refer to the Figure 2.3. It contains monthly values of DMI anomalies, which are deviations from the long-term average value. It has 13 columns in total, where the first column is the year, and the rest of the twelve columns are for the month-wise DMI values from January to December. This index would be helpful for understanding the sea surface temperature spatial distribution of the Indian Ocean, and its historical analysis helps to achieve comprehensive understanding about the role of IOD in climate variability. Thus, it acts as a real-time indicator of the intensity of positive or negative phases of IOD events. For the given analysis, the filtered data from 1998 to 2019 of DMI was used, in which we can do a detailed analysis of the impact of IOD.

Chapter 3

Preliminary Analysis of Global Data

The TRMM 3B43 dataset offers a valuable insight into global precipitation patterns over an extended period. A complete grasp of the variability if precipitation across various regions and time periods is obtained by considering both spatial and temporal variability. We plot the global precipitation values as our first step in visualizing the vast data present and to have a wholesome view of what we deal with. It provides the initial and, at the same time, primary step in researching the distribution and variability of precipitation on Earth, as one gets more specific to local studies of single or seasonal analyses. Let's start with visualizing the precipitation across the entire tropical part of the world, which will show us the areas with heavy rain to the areas with low rain, we could visualize these with the use of global spatial plot with linear and logarithmic scale. As shown in the Figure 3.1, It display the range of precipitation estimated for January 2000 using a linear scale to highlight different aspects of global precipitation patterns.The left plot shows the precipitation on a linear scale to give an overall wide-ranging view of the general precipitation. The right-hand-side plot is on a logarithmic scale, with values below 0.0000001 mm/hr excluded to avoid taking the logarithm of zero or very small values; this logarithmic scale emphasizes more subtle variations that might be overlooked in low-precipitation areas.

3.1 Global Total Monthly Precipitation Analysis

Global precipitation is fundamentally the very core upon which the study of climate change, water cycle dynamics, and the effects of weather on ecosystems and human activities are hinged. Hence, it shall be calculated both on a total annual and monthly global scale to highlight various trends and patterns, whether an increasing or decreasing trend. Because original precipitation values given in the monthly TRMM 3B43 dataset already are the average precipitation rates for each grid cell over the month, taking the averages of the averaged values do not provide additional meaningful information. First, we are going to start by calculating the total monthly precipitation. The intra-seasonal variation in the precipitations can thus be detected from analysing the monthly precipitation over the globe, for example: we could we a time shift in the monsoon or increase in the dry and wet seasons. Calculations involved converting precipitation rate,

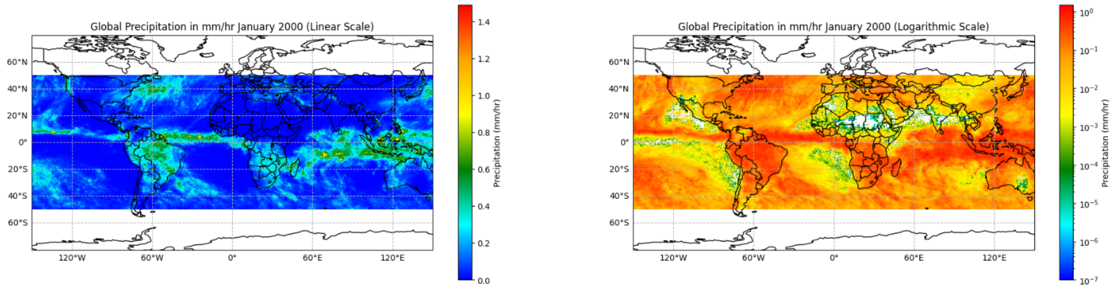


Figure 3.1: Global Tropical spatial plot with linear and logarithmic scale

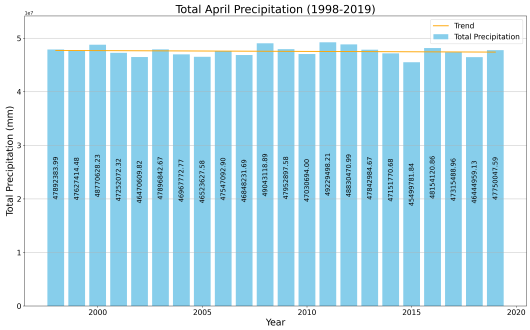
mm/hr to total precipitation in each month by multiplying the rate with the total number of hours in that month for every grid cell. Then these values are summed across all the grid cells to get the global total for every month from 1998 to 2019, refer to the Figure 3.2. The standard formula has been used to calculate total monthly precipitation as, Total Monthly Precipitation (mm) $=$ Precipitation Rate (mm/hr) $\times$ Number of Hours in the Month.

3.2 Global Annual Precipitation Analysis

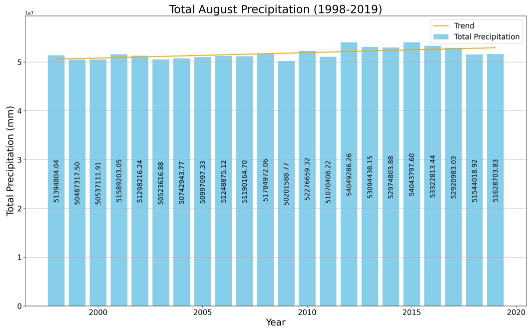
We will now calculate the global total annual precipitation.This analysis identifies the interannual variability in global precipitation patterns, which might result in a particularly wet year one year and a comparatively dry year the following. On the calculations, the total annual precipitation is obtained by adding up the monthly precipitation values throughout the year. This represents the sum of precipitation that fell for the whole year globally, refer to the Figure 3.3. Total annual precipitation was then computed from a standard formula, Total Annual Precipitation (mm) $=$ $\sum_{m=1}^{12}$ Total Monthly Precipitation (mm).

3.3 Global Spatial Plots

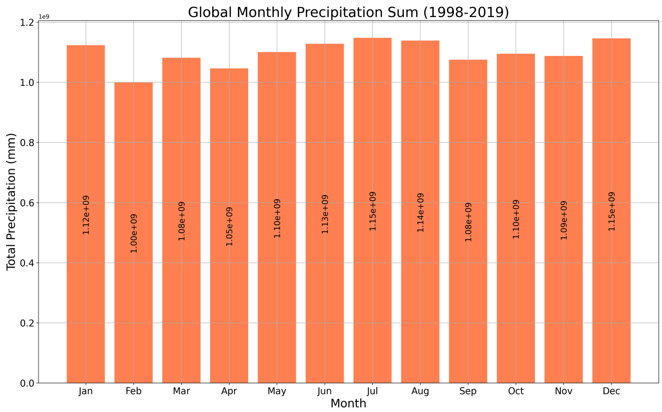
Month-to-month spatial variability may provide valuable insights into the seasonal shifts and effects of large-scale climatic events such as those associated with El Niño and La Niña. In this course of study, we have produced 264 monthly precipitation global spatial plots for the period between January 1998 and December 2019. This can enable us to observe the temporal variability in the monthly and yearly variation of the precipitation patterns in detail, not only geographically but also temporally. As shown in the Figure 3.4, it demonstrate how different the global precipitation pattern is during these two extreme climate events and the spatial variability in precipitation during these phenomena, showing large shifts in precipitation distribution across the tropics. This monthly spatial analysis complements the global total and average precipitation analyses by showing the visual and temporal context.



(a) Global Tropical April precipitation



(b) Global Tropical August Precipitation



(c) Global Tropical Monthly Precipitation

Figure 3.2: Monthly Precipitation Trends

3.4 Understanding Spatial and Temporal Variability

Spatial and temporal variability is an integral part of the analysis of precipitation data. It is meant that the pattern concerning the location of precipitation within a region, such as rainfall over different parts of the world. For example, North and South India show large variation in the distribution of rainfall due to their topographic features, which, again, is a spatial variation. Temporal distribution applies to the change over time; it may be days, months, or years with variations in the quantum of precipitation or a change in the normal pattern. In this regard, one location's quantum of rain might be very different from season to season, or the rain changes from year to year; this reflects temporal variability.

Intra-seasonal variation refers to those changes occurring within seasons. Even within the same wet season of the monsoon, there could be periods of heavy rainfall amidst a somewhat dry spell. This intra-seasonal variability is what's important in understanding the dynamics of seasonal precipitation. By "interannual variations" are meant those which take place from one year to another. In rainfall of monsoon, there might be a year that is extremely wet, while the next year will be rather dry. Such interannual variability is crucial to comprehend long-term tendencies and get ready for consequences brought about by climate variability because the following terms are commonly used in this study it is important to give their definition first.

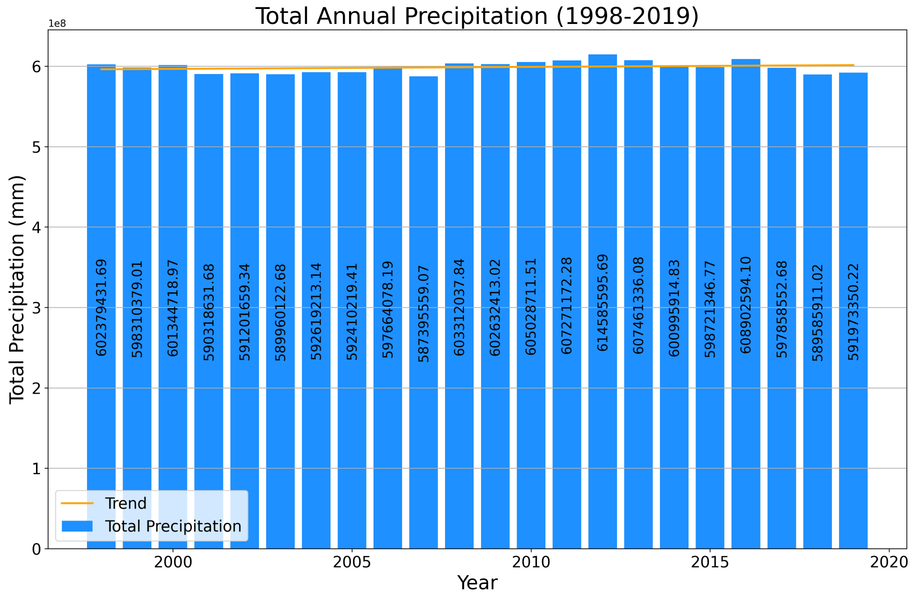
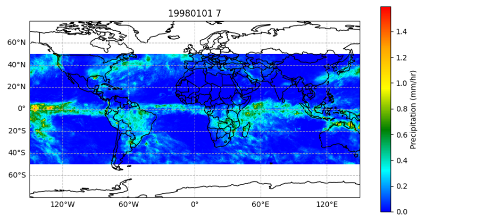
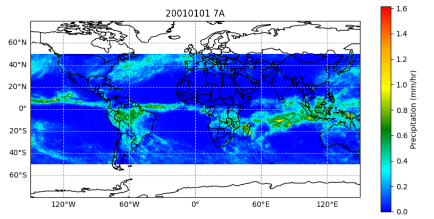


Figure 3.3: Global Total Precipitation for the Span of 22 Years



(a) Global spatial plot for January 1998 (El Niño)



(b) Global spatial plot for January 2000 (La Niña)

Figure 3.4: Comparison of global spatial plots during El Niño (1998) and La Niña (2000) events.

Spatial distribution refers to precipitation spread over different areas, whereas temporal distribution refers to changes in precipitation over time. Spatial variability describes variation across locations, temporal variability refers to variation across different time periods. Intra-seasonal variation has fluctuations within a single season, at the same time, interannual refers to one year as opposed to another. This chapter outlines in detail the large-scale features of precipitation patterns of the world in general and of regional variability of precipitation in particular, defining these terms in the early stage of the chapter, and gives a comprehensive view about the complex dynamics involved in precipitation variability.

Chapter 4

Time Series Analysis

Time Series analysis is a statistical method to analyze the sequence of data points which are ordered on a time basis, or observations taken sequentially, over regular space-time intervals. In climatology, time series analysis is especially important because it enables researchers to understand how different climate variables (e.g., temperature – precipitation and pressure) evolve over various timescales from a day up to decades, Tableau (n.d.). Time series analysis has been used in our study to determine the interactions between rainfall patterns and two of the large-scale climatic phenomena, i.e., El Niño Southern Oscillation (ENSO) and Indian Ocean Dipole (IOD). Such climate drivers have far-reaching effects on regional and global precipitation patterns, affecting everything from seasonal monsoon rains to long-lasting wet spells or multi-year droughts. With our time series analysis, we hope to explore these relationships and detect patterns, and see how they have evolved over the years. Time series analysis allows the decomposition of complex climate data into trends, seasonal cycles, and irregular variation components, making it easier to establish the contribution of each factor to overall climate variability. We will be able to quantify the strength and variability of such relationships across various regions and time periods by correlating time series of precipitation with the indices of ENSO and IOD.

In the broader scientific context of placing our research, (Athira et al. 2023) this present work tries to present a comprehensive study of the variability in the Indian Summer Monsoon Rainfall (ISMR) and its relationship with ENSO and other climate factors. The paper uses EOF analysis to identify two dominant spatial modes of ISMR which describe the dominant spatial variability patterns of ISMR. The paper focuses on how the ENSO-ISMR relationship has changed both in time and location, pointing at its temporal variability over different time periods and its spatial variability across different regions of the country. The author performed wavelet analysis for the time series of rainfall to understand the periodicity and timescales associated with these patterns; they also performed correlation analyses to find the relations of these patterns with different climate indices, such as ENSO, IOD and others. According to this study, during recent decades, the ENSO-ISMR association weakened, specifically over central India. Similar relationships are examined in our research, although we pay particular attention to the relationship among precipitation, ENSO and IOD. Our work tries to analyze the relationship between precipitation patterns (at both monthly and annual scales) and the climate signals from ENSO and IOD, while the aforementioned study examines the larger ISMR across the monsoon season. Techniques like correlation coefficients for measuring direction and strength of the relationships, EOF analysis to pinpoint predominant spatial modes of variability in precipitation. The above outcome of the research paper motivates us to investigate not just the general relationship between ENSO and precipitation, but also potential regional and seasonal variations in this relationship, and its interactions with other climatic signals such as the IOD. By developing a time series of this precipitation data, we can compare monthly, yearly, and seasonal trends and anomalies. This will allow us to see if there are any anomalies around the time that may be associated with an ENSO or IOD event.

4.1 Structuring the Analysis

Given the importance and relevance of time series analysis in climatology, the structuring of our analysis is designed by keeping the geographical focus areas of our study in mind, since understanding regional climate is important to unlock the climatic trends and patterns. The six regions, across India and Southeast Asia, differ in their climatic patterns and seasonal precipitation regimes. Such regions have been selected so that these capture variations in different climatic zones most likely to occur in response to ENSO, IOD, and other climate signals. For the robustness of our analysis, we provide the number of grid count in each region. We do this because any analysis of the data, region by region, is inconsistent due to the differing number of grid points in each location. Those locations with fewer points could introduce bias and might underestimate or overestimate precipitation. Similarly, the analysis of regions that have more than one grid point will be able to produce a much more comprehensive understanding of precipitation trends-and, quite often, a far more accurate one. The geographical coordinates of these regions, in latitude,longitude, and grid points are defined in the Table 4.1.

Region: Typical Climate	**Latitude range**	**Longitude range**	**Grid points**
North India: The region has a well-diversified climate from hot summers, cold winters, and a rainy season during the middle of the year. The region has typical characteristics of monsoon rainfall and winter precipitation, especially within the foothills of the Himalayas, World Bank Climate Change Knowledge Portal (2024*a*).	24.125°N to 33.875°N	70.125°E to 97.125°E	4360 grid points
Central India: The region has a hot climate with the majority of annual rainfall during its midyear wet season. It heavily relies on this rainfall for agriculture, World Bank Climate Change Knowledge Portal (2024*a*).	20.875°N to 24.125°N	69.875°E to 84.125°E	812 grid points

South India: The southern part of India has a warm climate as a whole with fairly well-distributed rainfall throughout the year when compared with the north. It receives substantial rainfall in two separate periods in the year i.e., June - September and October - December. The presence of two major rainy seasons in South India makes it a vital region for examining how various climate signals such as ENSO and IOD influence the rainfall variability at different times of the year, World Bank Climate Change Knowledge Portal (2024*a*).	7.875°N to 20.625°N	74.875°E to 88.125°E	2808 grid points
Thailand: The region has a tropical climate, warm throughout the year and a marked rainy season. Weather over the nation is determined by both regional and global climate patterns. Its climate is especially sensitive to any variation in global climate signals. Any research on this area should be considered in expanding the knowledge of how these signals impact precipitation in the tropics, World Bank Climate Change Knowledge Portal (2024*d*).	4.875°N to 20.875°N	96.875°E to 106.125°E	2470 grid points
Indonesia: Indonesian climate is equatorial, and thus quite uniform throughout the year. The warm temperatures and high humidity remain constant throughout the year. However, there is much seasonal variation in rainfall, which occurs in a somewhat predictable seasonal pattern. Due to Indonesia's size and location, the country is an important region for both ENSO and IOD impacts, World Bank Climate Change Knowledge Portal (2024*b*).	-11.125°S to 5.875°N	94.875°E to 140.875°E	12,765 grid points
Malaysia: The region lies in the tropical rainforest climate; there is heavy rainfall throughout the year, and the temperature is also comparatively high. The weather is characterized by little variation in temperature, accompanied by heavy rainfall, World Bank Climate Change Knowledge Portal (2024*c*).	0.875°N to 6.875°N	98.875°E to 120.125°E	2150 grid points

Table 4.1: Climate characteristics of different regions in India and Southeast Asia

Thus, latitude and longitude specification enable the exact study of regions and ensures that the analysis captures the climatic and geographical features representative of each area. These definite geographical boundaries of the regions allow temporal analysis in climate data. For example, we trace how, over time, the influence of ENSO on precipitation has been changing over a certain region. These grid count points are indicative of the geographical extent of each area and thus stand in validation that the selected boundaries capture enough data points to represent the regional precipitation correctly.

4.2 Regional Focused Precipitation Extraction

Following the analysis of the grid count, our research proceeds with the extraction of data on precipitation for the six regions. We then convert that extracted regional precipitation data into time series. For this purpose, we first import the NetCDF files of monthly and yearly TRMM 3B43 precipitation data. These were files with detailed precipitation rates with latitude and longitude coordinates. Before proceeding with the extraction, we cleaned the data of missing values. We replaced the `_FillValue`, which denotes the case of missing or invalid data, with `NaN` so that the analysis performed afterward does not bias based on these gaps. Then, we count both before and after this replacement the number of `NaNs` and zeros as a means to maintain quality in data. We used the latitude and longitude bounds of the following six regions: North India, Central India, South India, Thailand, Indonesia, and Malaysia. The above-specified bounds in each region were then used to filter the grid points falling within the region. This was done to extract the precipitation values in each; after extracting the precipitation data for each region, we then structured the extracted data into a format more usable. The structuring of data was performed by organizing it by year, month, region, latitude, longitude, and precipitation value, ensuring that each data point correctly corresponded to its latitude and longitude coordinate. This structured data is then compiled into a comprehensive `DataFrame` to serve as a basis for creating the time series. The organized regional precipitation data had been exported into a `CSV` file format for further uses.

4.3 Creating Time Series Data

We performed extensive time series analysis of precipitation with temporal and spatial averages to detect the temporal trends in precipitation. In doing this, the methodology will enable us to temporally and spatially aggregate the data, and thus, provide a coherent and clear picture of how the precipitation varies month-to-month and from year to year in different climatic zones. This was performed to perceive short-term variability and long-term trends of precipitation, and hence became of importance to the study of large-scale climatic phenomena such as ENSO and IOD in relation to the regional rainfall patterns.

Spatial Averages: The average of the precipitation for each month has been calculated for each region by averaging the precipitation values over all the grid points falling inside a particular region. This yielded one representative single value of precipitation for the whole region that included the integrated climatic condition of that region. The formula used in calculating the spatially averaged precipitation of a particular month is given by,

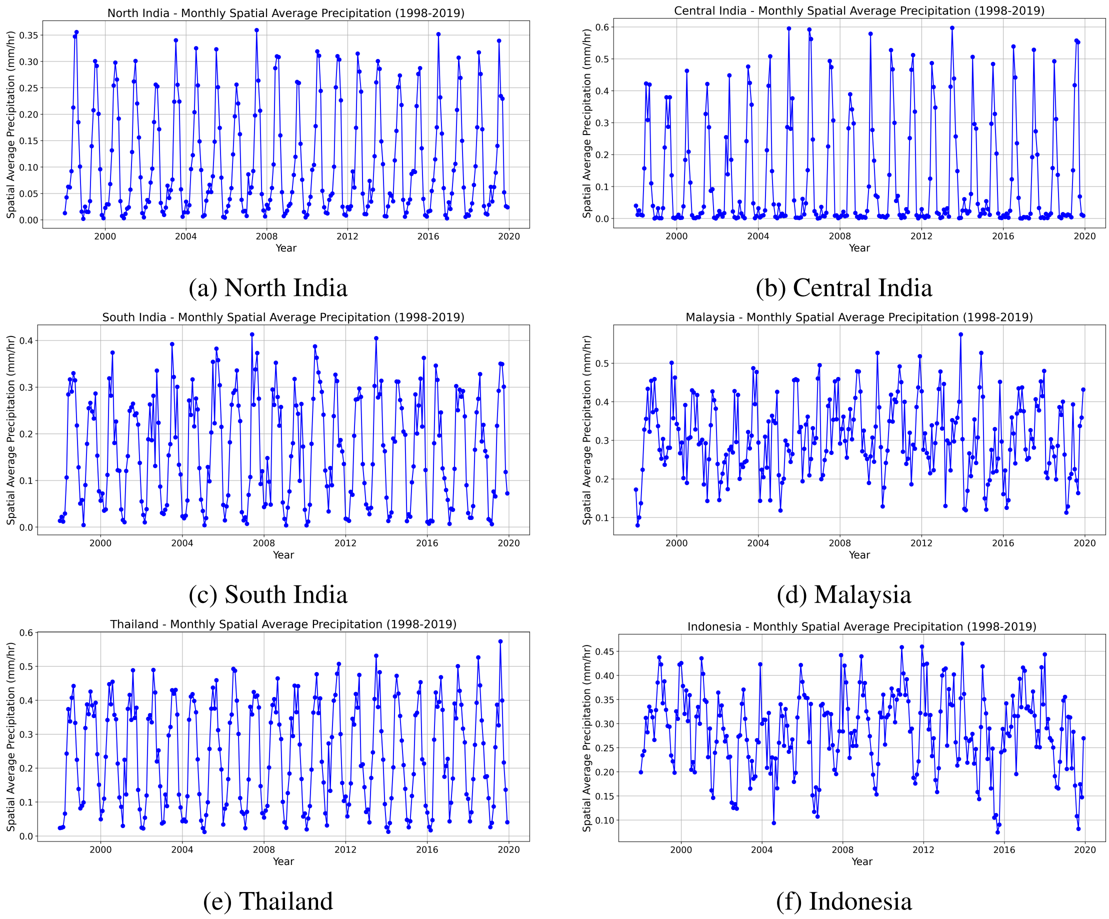
$$P_{\text{avg}} = \frac{1}{n} \sum_{i=1}^{n} P_i$$

where P_{avg} represents the spatial average of precipitation during that month. P_i is the precipitation value over a region at all respective grid points. n is the total number of grid points in the region. The above equation provides only the summation of all precipitation values of the grid points in the region, and then division by the number of grid points.

Temporal Averages: This has been computed using a time average of the precipitation values-in this case, over a month or over the season. This methodology smoothens out the short-term fluctuations and brings out the long-term precipitation trends. Focus on time averages allows the derailing from the real trends of rainfall masked by variability in short term and helps in understanding whether the rainfall in a particular region is increasing or decreasing over the years.

These spatial averages were set into a time series format. Each entry in this time series corresponds to a particular year and month, including the calculated spatial average as the representative value for that region. Besides, an extra column 'Date' was added to enable easy plotting and analysis by combining the year and month into one datetime value. First, we extracted the time series data for each region, then exported them in .csv format for reference and future analyses. We then created time series plots for each region as a means of visualization of temporal patterns in precipitation. Indeed, different time series plots reflect the monthly spatial average of the precipitation throughout the years from 1998 to 2019, hence clearly showing how precipitation has varied over time. Such plots allow us to observe trends, seasonal variations, and anomalies that may correspond to some of the most important climatic events.

For instance look at the Figure 4.1, the Malaysian plot is dominated by great month-to-month variability without a clear trend, indicating that precipitation patterns there are influenced by many factors. There is no pronounced seasonal cycle in the plot; thus, this indicates that rainfall is possible during all months of the year with strong peaks likely due to strong monsoons or anomalies. Contrasting this is the plot of North India, which reflects marked seasonal variation-thus peaking at times of the monsoon season and relatively no rainfall during dry seasons. Central India also shows a monsoonal pattern; there is very high rainfall during the monsoon season, with no obvious long-term trend. Indonesia, with its equatorial climate, shows the least seasonality, as much as any area can, with rainfall relatively well-distributed throughout the year, while South India and Thailand show regional temporal patterns related to monsoon peaks. These temporal analyses allow an understanding of the range of variations in rainfall variability at different regions and can be complemented with spatial mapping for geographical distribution.



(a) North India

(b) Central India

(c) South India

(d) Malaysia

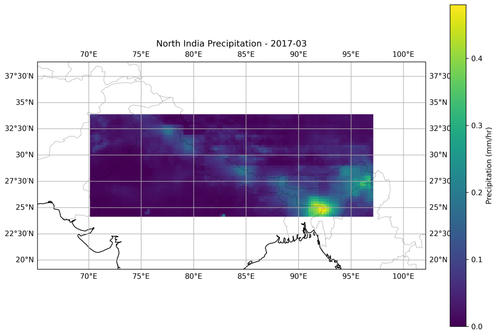
(e) Thailand

(f) Indonesia

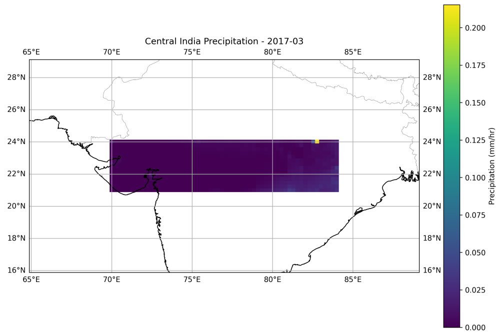
Figure 4.1: Monthly spatial average precipitation (1998-2019)

4.4 Spatial Mapping of Precipitation Data

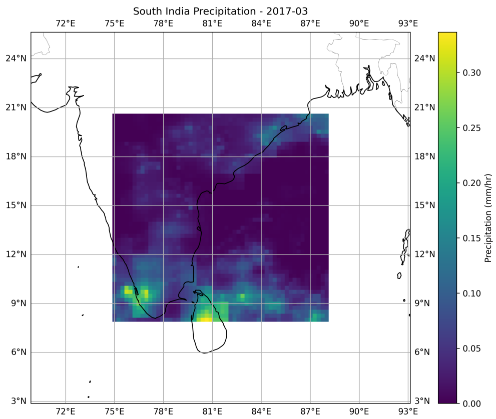
After creating the time series, we also focused on spatial mapping of precipitation data for each region. This step is an important aspect in order to identify the spatial pattern and hence know how different areas within a region experience different precipitation levels. These are 264 spatial maps for each region, showing visually how the distribution of rainfall is across each region using extracted precipitation data. The maps will pinpoint high and low-precipitation areas, hence giving a clear overview of the way rainfall is distributed across the region. These maps help us identify the spatial pattern, which might not be easy to realize from time series data. For instance, some regions within any area may have more rain than others, or the rainfall might shift from one region to another progressively. Therefore, in constructing spatial precipitation maps for regions of varied climatic characteristics, one needs to take natural variability in precipitation levels over those regions. Perhaps a standardized color scale will not be indicative or will fail to indicate the variation in precipitation of each particular region, especially when the regions are under very different ranges of precipitation.



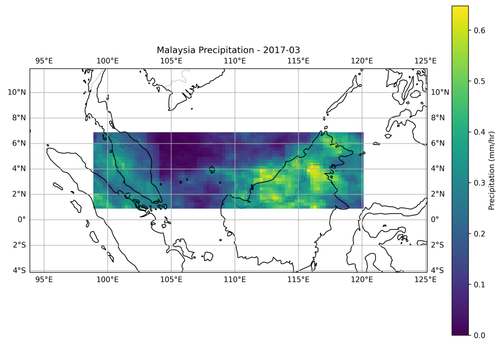
(a) North India



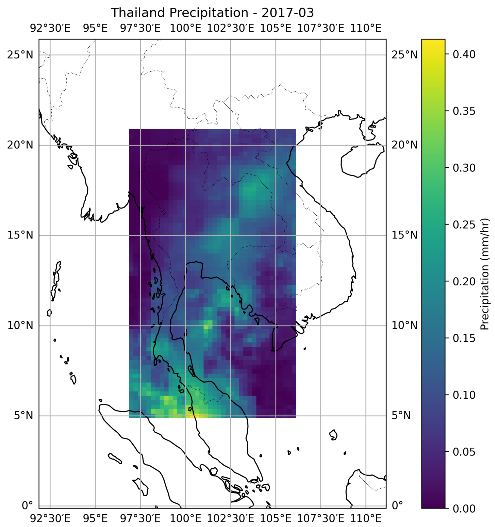
(b) Central India



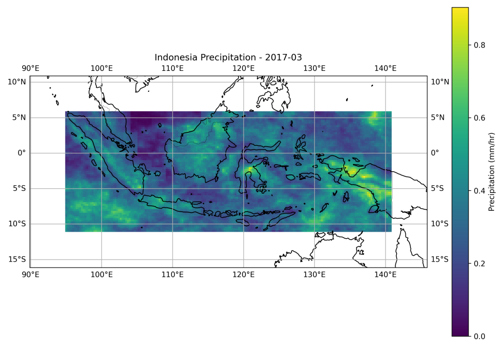
(c) South India



(d) Malaysia



(e) Thailand



(f) Indonesia

Figure 4.2: 2017 March spatial map for North India, Central India, South India, Malaysia, Thailand, and Indonesia

As a concrete example, where a standardized color scale set with an upper limit to capture higher precipitation in regions such as Indonesia has been employed, it may fail to express the subtleness of a region like Malaysia, with a typical precipitation range between 0.1 to 0.6

mm/hr. This could mask significant information in the spatial distribution of precipitation within those regions that normally receive less rainfall by using a standardized scale. For that reason, we allow the color scale to update dynamically based on the actual precipitation data coming from each region, which can make the maps informative and reflect relative differences in precipitation. We do this because such a method allows us to underline great spatial variations within each region; hence, the maps become more useful to understand the patterns of localized precipitation.

From the Figure 4.2,we have noticed from the spatial maps that the regions belonging to Malaysia, Indonesia, and Thailand always receive higher precipitation amounts compared to Central India, North India, and South India. Just from the spatial map, we would probably be able to say that total precipitation in central, north, and south India will be comparatively lower than the total precipitation in Malaysia, Indonesia, and Thailand. However, these maps give a full view of the spatial distribution but fail to elucidate the strength of time variability. Next, we'll calculate the total annual rainfall, the trend in monthly rainfall over several years and the seasonal cycle in order to quantify and proceed with further analysis of precipitation variability. These will enable us to understand how not only the total amount of precipitation varies, but also how it is distributed on various temporal scales, and how it would be influenced due to climatic factors such as ENSO and IOD.

4.5 Regional Analysis

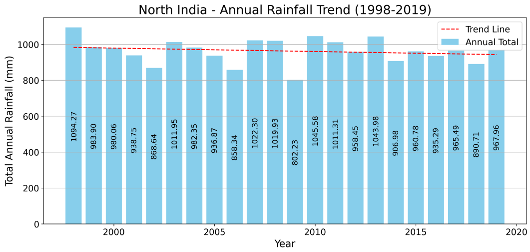
This regional analysis will try to bridge the gaps between the spatial averages and temporal trends, with the help of their seasonal cycles, in order to provide a broad understanding of the change undergone by precipitation patterns throughout the last two decades. The present analysis not only highlights the variability and trends in regional rainfall but also lays the ground for further investigation of driving climatic factors underlying these precipitations: namely, the ENSO and IOD.

4.5.1 Calculation of Total Annual Rainfall

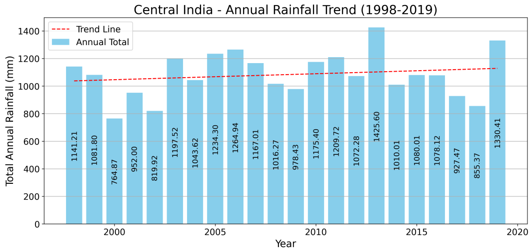
The first step in our regional analysis is to calculate the amount of total annual rainfall received by each region. This measure represents the sum of rainfall that a region receives during the course of a year and is highly important for the understanding of the overall availability of water in the area. First, in order to determine the total annual rainfall, the spatially averaged precipitation rates measured in mm/hr had to be converted into total monthly precipitation by multiplying each value both by 24 (hours per day) and by the number of days in the respective month. Thus, the formula will look as follows:

$$\text{Monthly Total (mm)} = \text{Spatial Average} \times 24 \times \text{Days in Month}$$

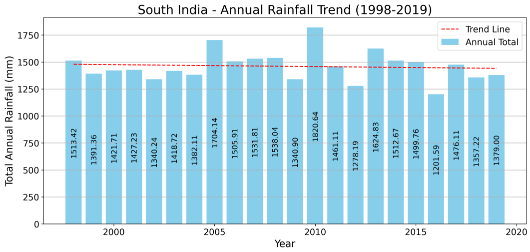
.



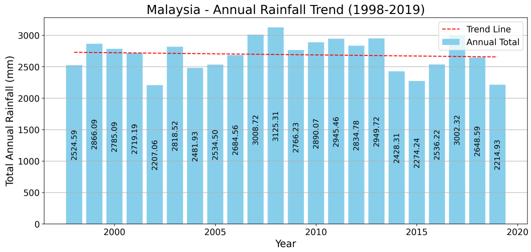
(a) North India



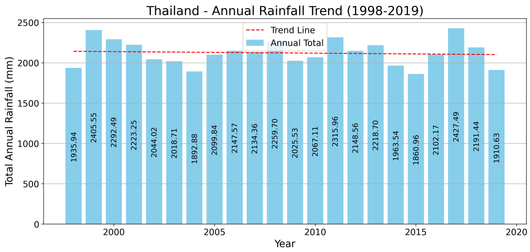
(b) Central India



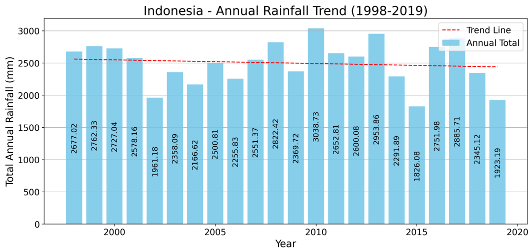
(c) South India



(d) Malaysia



(e) Thailand



(f) Indonesia

Figure 4.3: Calculation of Total Annual Rainfall

Monthly totals were summed to obtain total annual rainfall for each year. We found this method gave an exact representation of yearly precipitation in every region, showing the trend within the 22 studied years. From the total annual rainfall, we were able to observe regions which were falling under wetter or drier years, probably indicating changing climatic conditions. To examine any increasing, decreasing, or stable trend in the total annual rainfall over time, a linear regression was performed on the data. Bars showing the variation in annual rainfall with a trend line added in order to visualize long-term changes were plotted. In respect to climate change, this analysis is important for interpretation of how water resources are affected and changing in various regions. As shown in the Figure 4.3, distinct patterns are shown from the total annual rainfall across different regions. On average, there is a variability from year to year and no significantly long-term annual rainfall trend in Central India. In fact, years like 2013 and 2019 exhibit peaks, which again could be related to anomalous climatic events. Annual rainfall over Indonesia appears mainly stable with slight variations inter-annually. There are no substantial increasing or decreasing long-term trends, which underlines a fairly steady climate with no major disruptions. The trend in annual rainfall in Malaysia has shown a considerable