

# Rainfall Analysis Using Satellite Data

Naveen Sabarinath Babu  
201783589

Supervised by Dr. Stephen Griffiths

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

Master of Science in Data Science and Analytics

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.

**School of Mathematics**  
FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

---

## Academic integrity statement

I am aware that the University defines plagiarism as presenting someone else's work, in whole or in part, as your own. Work means any intellectual output, and typically includes text, data, images, sound or performance.

I promise that in the attached submission I have not presented anyone else's work, in whole or in part, as my own and I have not colluded with others in the preparation of this work. Where I have taken advantage of the work of others, I have given full acknowledgement. I have not resubmitted my own work or part thereof without specific written permission to do so from the University staff concerned when any of this work has been or is being submitted for marks or credits even if in a different module or for a different qualification or completed prior to entry to the University. I have read and understood the University's published rules on plagiarism and also any more detailed rules specified at School or module level. I know that if I commit plagiarism I can be expelled from the University and that it is my responsibility to be aware of the University's regulations on plagiarism and their importance.

I re-confirm my consent to the University copying and distributing any or all of my work in any form and using third parties (who may be based outside the EU/EEA) to monitor breaches of regulations, to verify whether my work contains plagiarised material, and for quality assurance purposes. I confirm that I have declared all mitigating circumstances that may be relevant to the assessment of this piece of work and that I wish to have taken into account. I am aware of the University's policy on mitigation and the School's procedures for the submission of statements and evidence of mitigation. I am aware of the penalties imposed for the late submission of coursework.

Name Naveen Sabarinath Babu

Student ID 201783589

# **Abstract**

In this report, we have investigated the influence of climate indices like ENSO, PDO and DMI on rainfall precipitation patterns in India and Indonesia, which rely heavily on seasonal rainfall for agriculture and water management. The study takes 20 years of data from 2003 to 2023 where various smoothing techniques are used to reduce noise and show the trends of both the regional precipitation and find the correlation with climate indices. Key findings reveal a consistent and significant impact of ENSO on rainfall in both regions with a strong relation to Indonesia. Whereas, other indices like PDO and DMI have lesser influence compared to ENSO. India has a decent relationship with ENSO and a weaker relationship with PDO and DMI . Further, seasonal correlations were taken into account with the best smoothing filter of both the regions and related with ENSO, which gives SON (September, October and November ) to be related for both regions. In the predictive model, we have considered the SARIMAX model, Climatology (20-year average data) and previous year's data where the SARIMAX models, which incorporate external climatic indices, demonstrated superior accuracy over simpler models for Indonesia but for India the simpler model seems to work best. This study enhances the understanding of climatic impacts on regional rainfall and improves the reliability of rainfall predictions, which is essential for effective resource management and disaster preparedness.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Background . . . . .	1
1.2	Research Objective . . . . .	1
1.3	Overview of the chapters . . . . .	2
<b>2</b>	<b>Data Set</b>	<b>4</b>
2.1	Data Description . . . . .	4
2.2	GPM (IMERG) NASA Data Set . . . . .	4
2.3	ENSO 3.4 Dataset . . . . .	5
2.4	Dipole Mode Index (DMI) . . . . .	6
2.5	Pacific Decadal Oscillation (PDO) . . . . .	6
<b>3</b>	<b>Exploratory Data Analysis</b>	<b>7</b>
3.1	Data Overview and Processing . . . . .	7
3.2	Visualizing Data . . . . .	7
3.3	Statistical Analysis . . . . .	8
3.4	Average Monthly Precipitation Trends . . . . .	9
3.5	Regional Analysis within India . . . . .	10
<b>4</b>	<b>Analysis of Rainfall Patterns and Climate Indices</b>	<b>12</b>
4.1	Calculating Monthly Average Precipitation . . . . .	13
4.1.1	India Raw Anomaly . . . . .	13
4.1.2	Indonesia Raw Anomaly . . . . .	13
4.2	Smoothing Techniques in Time Series Analysis . . . . .	14
4.2.1	Kolmogorov-Zurbenko (KZ) Filter . . . . .	14
4.2.2	Triangle Smoothing . . . . .	17
4.2.3	Gaussian Smoothing . . . . .	20
4.2.4	Moving Average . . . . .	23
4.3	Impact of Window Size on Smoothing Precipitation Anomalies . . . . .	25
4.4	Pattern Resemblance Between Smoothed Precipitation and Climate Indices . . . . .	26
4.5	Correlation Analysis . . . . .	29
4.5.1	India . . . . .	29
4.5.2	Indonesia . . . . .	30
4.5.3	Analysis of Lagged Relationship Between ENSO and Precipitation . . . . .	32
4.6	Seasonal Relationship Between Precipitation and ENSO . . . . .	33
4.7	Conclusion . . . . .	34

<b>5 Annual rainfall forecasting with SARIMAX</b>	<b>35</b>
5.1 Introduction . . . . .	35
5.2 Modeling . . . . .	36
5.2.1 India Model . . . . .	36
5.2.2 Indonesia Model . . . . .	38
5.3 Comparison of Model Performance . . . . .	39
5.4 Overall Model Performance . . . . .	44
<b>6 Conclusion and Discussion</b>	<b>45</b>
6.1 Conclusion . . . . .	45
6.2 Discussion . . . . .	46

# List of Figures

2.1	Monthly precipitation data for September over the Indian. . . . .	5
2.2	Anomalies of ENSO 3.4 data. . . . .	5
2.3	DMI(Dipole Mode Index) anomalies. . . . .	6
2.4	Anomaly of PDO (Pacific Decade Oscillation) . . . . .	6
3.1	Monthly precipitation time series data for India over a 20 year. . . . .	8
3.2	Monthly precipitation time series data for Indonesia over a 20 year. . . . .	8
3.3	Average precipitation of Indian (2003 - 2023) . . . . .	9
3.4	Average precipitation of Indonesia (2003 - 2023) . . . . .	10
3.5	Average mothly precipitation in different region of India. . . . .	10
4.1	Monthly Precipitation Anomalies for India (2003-2023). . . . .	13
4.2	Monthly Precipitation Anomalies for Indonesia (2003-2023). . . . .	14
4.3	Smoothed Anomalies of India using a KZ Filter with a Window Size of (3,5,7).	16
4.4	Smoothed Anomalies of Indonesia using a KZ Filter with a Window Size of (3,5,7). . . . .	17
4.5	Triangle Smoothed Anomalies of India using a kernel Size of (3,5,7). . . . .	19
4.6	Triangle Smoothed Anomalies of Indonesia using a kernel Size of (3,5,7). . . . .	20
4.7	Gaussian Smoothed Anomalies of India using a Sigma Size of (3,5,7). . . . .	21
4.8	Gaussian Smoothed Anomalies of Indonesia using a Sigma Size of (3,5,7). . . . .	22
4.9	(3,5,7)-Month Moving Average of India Anomaly. . . . .	24
4.10	(3,5,7)-Month Moving Average of Indonesia Anomaly. . . . .	25
4.11	Dipole Mode Index (DMI). . . . .	26
4.12	Pacific Decadal Oscillation (PDO) Index. . . . .	26
4.13	Side-by-side comparison of DMI and PDO indices. . . . .	26
4.14	Comparing the ENSO indies with smoothing anomalies of India. . . . .	27
4.15	Comparing the ENSO indies with smoothing anomalies of Indonesia. . . . .	28
4.16	Lagged Correlation Trends between ENSO and India Precipitation Over Time .	32
4.17	Lagged Correlation Trends between ENSO and Indonesia Precipitation Over Time	33
5.1	Partial Autocorrelation and Autocorrelation Plots for India. . . . .	37
5.2	Autocorrelation and Partial Autocorrelation Plots for Indonesia. . . . .	38
5.3	Comparison of Prediction Evaluation for Indonesia and India for 2021. . . . .	40
5.4	Comparison of Prediction Evaluations for Indonesia and India for 2022. . . . .	42
5.5	Comparison of SARIMAX model predictions for Indonesia and India in 2023. .	43

# List of Tables

4.1	Correlation Coefficients of Different Smoothing Methods on Climate Indices for India. . . . .	30
4.2	Correlation Coefficients of Different Smoothing Methods on Climate Indices for Indonesia. . . . .	31
4.3	Seasonal Correlation between ENSO (ONI) and Precipitation in India and Indonesia Using the 7-Month Moving Average . . . . .	34
5.1	Performance of SARIMAX models based on AIC and BIC for different training datasets for India. . . . .	38
5.2	Model Comparison for Indonesia (Best Model in Bold, with Exogenous Variable). . . . .	39
5.3	Comparison of Model Performance Metrics for Indonesia and India for 2021. . . . .	41
5.4	Comparison of Model Performance Metrics for Indonesia and India for 2022. . . . .	41
5.5	Comparison of Model Performance Metrics for Indonesia and India for (2023). . . . .	44

# Nomenclature

<b>ENSO</b>	El Niño Southern Oscillation
<b>PDO</b>	Pacific Decadal Oscillation
<b>DMI</b>	Dipole Mode Index
<b>SARIMAX</b>	Seasonal Autoregressive Integrated Moving Average with eXogenous variables
<b>RMSE</b>	Root Mean Square Error
<b>MAPE</b>	Mean Absolute Percentage Error
<b>GPM (IMERG)</b>	Global Precipitation Measurement (Integrated Multi-satellite Retrievals for GPM)
<b>DJF</b>	December, January, February
<b>MAM</b>	March, April, May
<b>JJA</b>	June, July, August
<b>SON</b>	September, October, November
<b>SMA</b>	Simple Moving Average
<b>ACF</b>	Autocorrelation Function
<b>PACF</b>	Partial Autocorrelation Function
<b>KZ</b>	Kolmogorov-Zurbenko Filter

# **Chapter 1**

## **Introduction**

### **1.1 Background**

Rainfall is a main component of the Earth's hydrological cycle, playing a crucial role in agriculture and water resources, especially in countries like India and Indonesia. These regions have unique seasonal rainfall patterns where India largely depends on the South-West Monsoon, while Indonesia's weather is shaped by complex tropical monsoon systems.

Predicting rainfall accurately is important for planning agricultural productivity, managing water resources, and preparing for natural disasters, mainly in extreme weather like droughts and floods.

Understanding the rainfall patterns and predicting them is a long term goal in meteorology. This task is complicated by interplay between the local weather and global climatic forces, such as the El Niño Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), and the Indian Ocean Dipole (IOD) (DMI). ENSO, which operates on a timescale of 2 to 7 years (Hanley et al. (2003)), alternates between El Niño and La Niña phases and is known for its strong influence on weather patterns worldwide, particularly by altering rainfall distribution in regions like Southeast Asia and India. PDO is a longer-term oscillation, cycling over 20 to 30 years, and affects the climate primarily in the Pacific basin, including the influence on multi-decadal trends in rainfall and temperature in surrounding regions like Indonesia, where Mantua and Hare (2002) explain the salmon production affected by PDO and described about the PDO. The DMI index measures the East-West temperature gradient across the tropical Indian Ocean, which is associated with the Indian Ocean Dipole. These large-scale climate phenomena can significantly influence regional rainfall patterns (Jury (2022)), making it crucial to include them in forecasting models to improve accuracy and reliability.

### **1.2 Research Objective**

This dissertation seeks to explore and understand how global climatic indices impact rainfall patterns in India and Indonesia. These two regions are of interest for this study due to their

unique and complex rainfall patterns. Both regions experience significant seasonal rainfall influenced by global climate. India relies heavily on the South-West Monsoon, which is important as resources can be maintained, while Indonesia's tropical climate is marked by highly variable rainfall patterns. Importantly, both regions are located near the centers of action for ENSO, PDO, and DMI, making them the main region of study that affects these global climate drivers. The summarised objectives include:

**1. Investigate the Influence of ENSO, PDO, and DMI on Rainfall Variability:**

Examine how these major climatic indices correlate with and affect rainfall across various regions within India and Indonesia. Evaluate the correlation of these relationships over a 20-year period to better understand long-term trends and assess how smoothing techniques can enhance the clarity of these correlations by reducing noise in the data and evaluate the consistency of these long-term trends and the impact of global climatic indices on regional rainfall.

**2. Evaluate the Effectiveness of Different Statistical and Smoothing Techniques:**

Various smoothing techniques will be applied to analyze rainfall data and reduce noise, allowing us to uncover clearer trends. To make this analysis more engaging, we will compare different smoothing techniques, each of which highlights specific aspects of the data and reveal distinct trends. This comparison will help us understand how different methods influence the interpretation of rainfall patterns.

**3. Analyze Seasonal Variations in the Influencing of Climate Indices:**

Here, we will explore the smoothing technique with the strongest relationship with climate indices where the seasonal split are taken into account as DJF (December to February), MAM (March to May), JJA (June to August) and SON (September to November) for both precipitation data (best smoothed technique) and the most correlating climate indices. We will utilise to check for the seasonal correlation and produce the result.

**4. Develop Models for Rainfall Prediction in India and Indonesia**

We plan to develop models that can predict rainfall for the entire year, using the previous 20 years of data as a basis. Instead of focusing on short-term forecasts, like daily or even weekly predictions which are often too variable and hard to get right we'll focus on predicting monthly rainfall totals. Monthly predictions are more manageable and generally more reliable because they average out the daily ups and downs, making the patterns easier to predict.

### 1.3 Overview of the chapters

In chapter 2, we focus on the specifics of the data we are using to describe the sources of data characteristics of the precipitation records and climate indices relevant to our study. This chap-

ter sets the foundation by ensuring that readers are familiar with the types and structures of data that inform our analysis.

Chapter 3 focuses on Exploratory Data Analysis (EDA) where we carefully examine the precipitation data to uncover underlying patterns. This process helps us identify key trends and prepare the data for more detailed analysis in subsequent chapters.

In Chapter 4, we introduce anomalies and various smoothing techniques to refine our data further and analyse correlations between climate indices and precipitation. This chapter explores how different seasonal adjustments enhance our understanding of the relationships and which indices show the strongest correlation with rainfall patterns in our regions of interest.

Chapter 5 focuses on modelling, where we apply the insights and methodologies developed in earlier chapters to predict monthly rainfall for a year. This involves constructing and validating models that can forecast precipitation trends and provide valuable tools for planning and decision making in response to ENSO related climate variations.

Chapter 6 concludes the report with an explanation of the research done over the study.

# Chapter 2

## Data Set

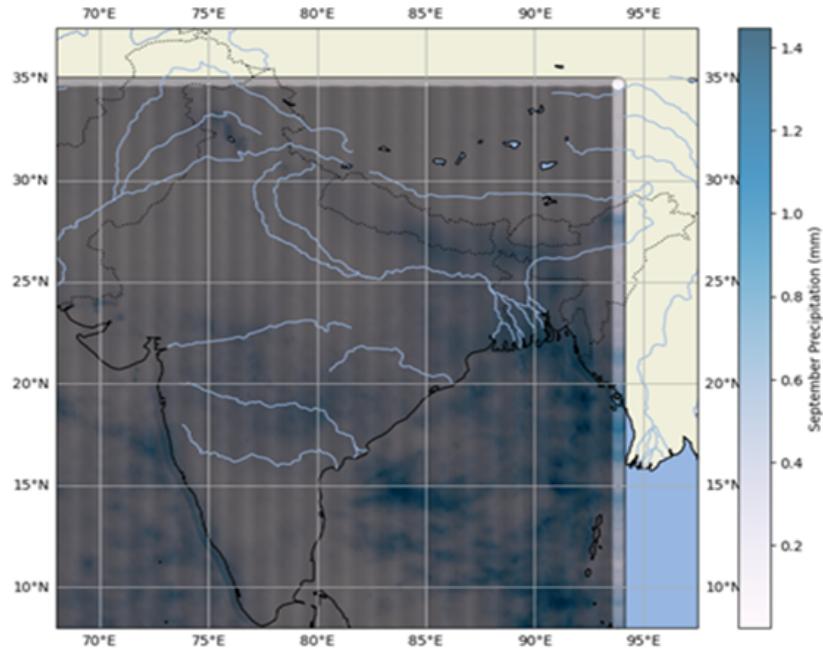
### 2.1 Data Description

Our primary dataset in this study is the Integrated Multi-Satellite Retrievals for Global Precipitation Measurement-GPM (IMERG), which captures detailed precipitation measurements. We will then explore climate indices particularly focusing on the El Niño Southern Oscillation (ENSO) to observe variations in El Niño and La Niña phases. Following this, we will examine the Dipole Mode Index (DMI) which is particularly relevant to the Indian Ocean. Lastly, we will analyse the Pacific Decadal Oscillation (PDO) which primarily affects the Pacific Ocean's climatic conditions.

### 2.2 GPM (IMERG) NASA Data Set

The (IMERG) Integrated Multi-satellite Retrievals for GPM dataset offers a sophisticated method for measuring precipitation from space. Using satellite data provides uniform spatial coverage across the globe (Huffman et al. (2015)), which is particularly beneficial over oceans and remote areas where traditional rain gauges are not feasible due to the need for physical presence and higher resource consumption. This dataset is accessible from NASA's GES (Goddard Earth Sciences) DISC (Data and Information Services Center) in the NetCDF file format (Huffman et al. (2023)), which is well-suited for managing complex multi-dimensional scientific data.

In this study, we have used monthly precipitation data in the units mm/hr (millimetre per hour) rather than daily data (see 2.1). The monthly data offers improved precision, allowing us to observe precipitation trends more effectively compared to daily data, as it provides a comprehensive view of precipitation over a one-month period.

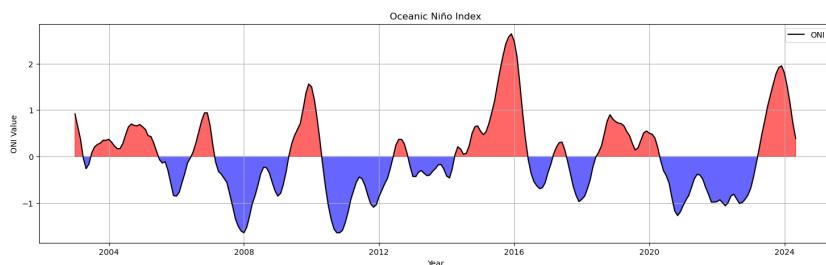


*Figure 2.1:* Monthly precipitation data for September over the Indian.

### 2.3 ENSO 3.4 Dataset

ENSO (El Niño Southern Oscillation) is the data gathered from the website National Weather Service (National Oceanic and Atmospheric Administration (NOAA) (2024)). ENSO is a critical climate phenomenon affecting global atmospheric circulation and consequently, impacting global temperature and precipitation patterns.

The dataset comprises a 3-month running mean of sea surface temperature (SST) anomalies in the Niño 3.4 (see 2.2) where there are different types of ENSO data, with historical SST data represented in blue (below normal) and red (above normal) based on five consecutive overlapping seasons. ENSO has different types of data namely, Niño-4, Niño-3 index and Niño-2 index data, where these represent weaker signals compared to Niño 3.4 (combination of Niño-4, Niño-3), which is described in the paper Hanley et al. (2003).



*Figure 2.2:* Anomalies of ENSO 3.4 data.

## 2.4 Dipole Mode Index (DMI)

The Dipole Mode Index (DMI) was selected due to its proximity to India which is collected from the online resource "Laboratory". This index measures the intensity of the Indian Ocean Dipole (IOD) through SST anomalies between the western and southeastern equatorial Indian Ocean (Sardana et al. (2022)). A positive DMI indicates a positive IOD phase, represented by the red color, and a negative DMI indicates a negative IOD phase, (in figure 2.3).

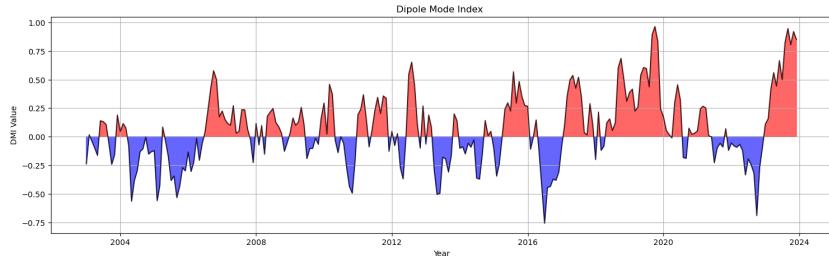


Figure 2.3: DMI(Dipole Mode Index) anomalies.

## 2.5 Pacific Decadal Oscillation (PDO)

The Pacific Decadal Oscillation (PDO) is a long-term ocean fluctuation of the Pacific Ocean which is collected from (for Environmental Information). Its inclusion in the study is strategic for understanding decadal-scale climatic variations, which can overlay more immediate phenomena like ENSO and DMI.

PDO is characterized by two phases, as seen in the figure 4.12. The warm (positive) phase, represented by red color, where SST in the central and eastern Pacific Ocean are seen warmer than usual, and the cool (negative) phase, indicated by blue color, where SSTs are cooler than usual in these regions. These phases resemble the El Niño and La Niña phenomena of ENSO, but the PDO operates on much longer timescales, with impacts extending over decades (Mantua and Hare (2002)).

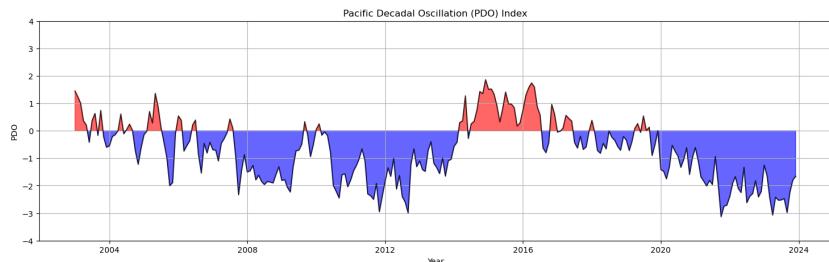


Figure 2.4: Anomaly of PDO (Pacific Decade Oscillation)

## Chapter 3

# Exploratory Data Analysis

In this chapter, the focus is on the Exploratory Data Analysis (EDA) of the precipitation data collected from NASA's satellite systems, specifically using datasets like the Global Precipitation Measurement (GPM) and IMERG products. This data is analyzed with an emphasis on regional studies for India and Indonesia, two regions that experience vastly different climatic patterns yet are both highly dependent on rainfall for agriculture, water resources, and ecosystem sustainability.

**India:** Latitude Range: 6°45'N to 37°6'N, Longitude Range: 68°7'E to 97°25'E

**Indonesia:** Latitude Range: Approximately 6°N to 11°S, Longitude Range: 95°E to 141°E

The goal of the EDA is to understand the underlying trends, seasonal variations, and statistical properties of the precipitation data over the past two decades. By examining the data through various statistical measures and visualization techniques, we can uncover key patterns about the regions.

### 3.1 Data Overview and Processing

The dataset contains monthly precipitation data specifically focused on India and Indonesia. Data points per month are substantial, amounting to 77,490 (latitude :270 , longitude:287) for India and 73,704 (latitude: 444, longitude: 160) for Indonesia, indicating high-resolution coverage.

**Data Aggregation:** Where the Monthly data are averaged to produce annual insights, with a comprehensive dataset spanning 20 years.

### 3.2 Visualizing Data

Having aggregated data, we will now proceed to visualize and analyze the monthly averaged data extended over 20 years. This will allow us to observe trends and patterns in both regions and assess if these insights offer a deeper understanding of the climatic behaviors.

**India:** The data displays a clear cyclical pattern (figure 3.1) corresponding to the monsoonal impacts, with sharp peaks during the wet season and troughs during the dry season. This pattern persists across the years, suggesting consistent climatic influences despite annual fluctuations.

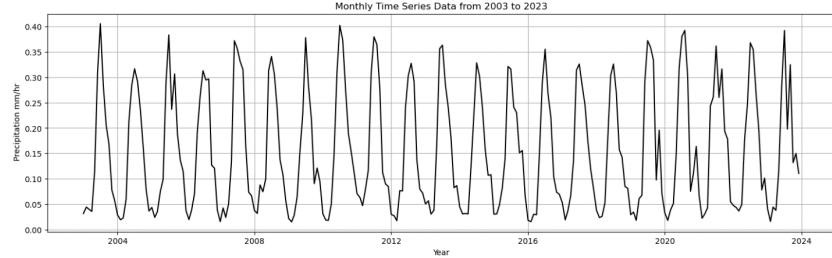


Figure 3.1: Monthly precipitation time series data for India over a 20 year.

**Indonesia:** Similar to India, Indonesia shows distinct seasonal trends but with considerable year-to-year variability (seen in figure 3.2), indicating that the influence of external climatic factors on Indonesian weather patterns has a random fluctuation over the period where we can't depict anything clearly.

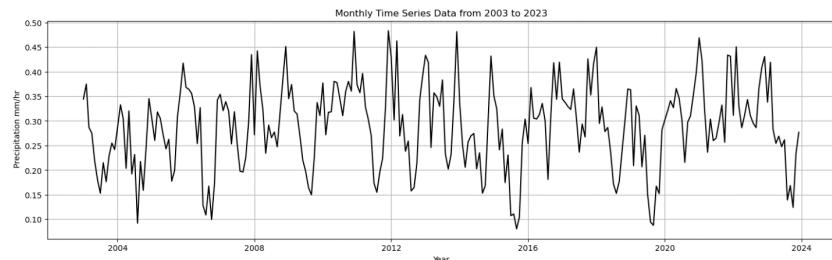


Figure 3.2: Monthly precipitation time series data for Indonesia over a 20 year.

### 3.3 Statistical Analysis

#### India:

**Mean Precipitation:** The precipitation data exhibits slight yearly variations, typically ranging between 0.14 to 0.17 mm/hr. This variance indicates the climatic fluctuations affecting rainfall annually.

**Standard Deviation:** The variability in monthly precipitation, as suggested by the standard deviation (ranging from 0.11 to 0.14), highlights the fluctuating nature of rainfall, with some years experiencing more pronounced variations.

**Extreme Values:** Rainfall values highlight significant monsoon impacts, with a maximum of 0.405 mm/hr in July 2003 and the end of winter with a minimum precipitation of 0.015 mm/hr in February 2009.

### **Indonesia:**

**Mean Precipitation:** Varies annually between about 0.21 mm/hr and 0.35 mm/hr, indicating less fluctuation compared to India.

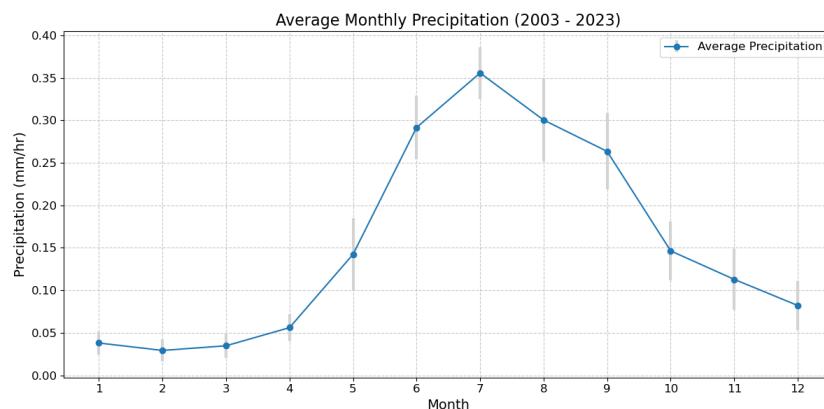
**Standard Deviation:** Indicates variability in monthly values. For instance, 2010 has a relatively lower standard deviation, suggesting lesser fluctuation in values (0.085965) throughout the year, whereas 2006 shows more significant variability.

**Extreme Values:** The rainfall values highlight a maximum value of 0.48 mm/hr in December 2011 and a minimum of 0.08 mm/hr in September 2015.

### **3.4 Average Monthly Precipitation Trends**

In this analysis, we've averaged 20 years of data to observe the typical monthly precipitation over two decades, aiming to determine if the observed rainfall aligns with the expected seasonal patterns of each region.

**India:** Figure 3.3 illustrates the significant impact of monsoon, with minimal rainfall from January to March and a peak in July. This annual pattern is consistent with the South-West monsoon dynamics with a decline from September to December.



*Figure 3.3: Average precipitation of Indian (2003 - 2023)*

**Indonesia:** Figure 3.4 reflects its tropical, monsoonal climate with high rainfall from January to March, decreasing through the dry season and picking up again in October, marking the onset of the wet season.

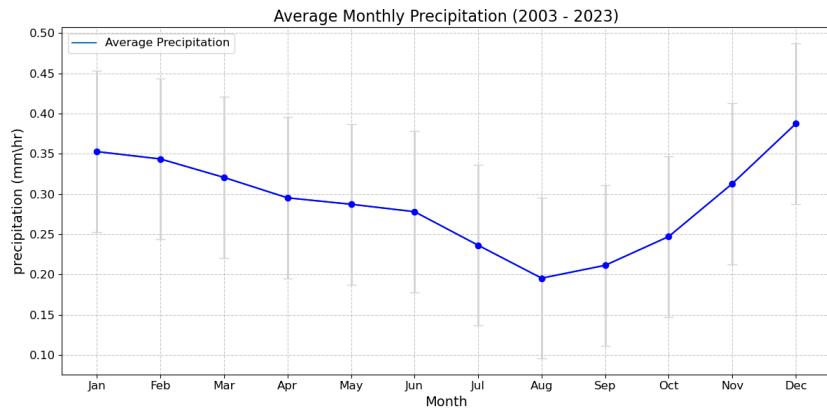


Figure 3.4: Average precipitation of Indonesia (2003 - 2023)

### 3.5 Regional Analysis within India

In this analysis, we compare the average rainfall patterns across different regions to determine whether the overall precipitation trends in India align with the regional averages.

Initially the data points are collected based on the latitude and longitude of each regions in India, compiling monthly data points from each year and aggregating them to visualize the average precipitation over 20 years.

Each region in India (Northwest is Rajasthan, Northeast is Arunachal Pradesh, Southwest is Kerala, and Southeast – Telangana ) is selected for precipitation trends. Graph 3.5 is scaled similarly on the y-axis to facilitate direct comparisons.

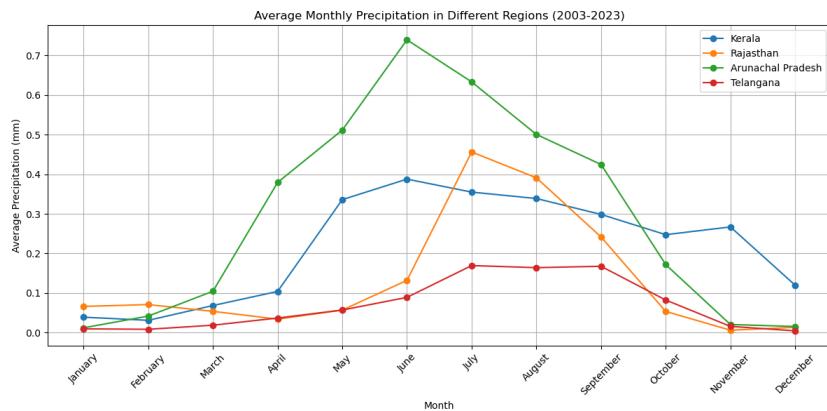


Figure 3.5: Average monthly precipitation in different region of India.

**Rajasthan:** Shows lower precipitation, aligning with its dry, low monsoon profile, where the point of latitude and longitude is 26.11, 72.483.

**Arunachal Pradesh:** Exhibits higher precipitation, consistent with its high monsoon impact. Where the point of latitude and longitude is 26.25, 91.549.

**Kerala and Telangana:** Represent the influence of both Southwest and Northeast monsoons, showing varied rainfall patterns. The point of latitude and longitude is 9.77, 76.36 (Kerala) and 16.33, 78.86 (Telangana). These both are in the costal region of Indian.

From this analysis it is evident that there is a distinct difference in the precipitation trends across each region compared to the overall average precipitation for India.

The figures from the data visualization do not clearly illustrate the precipitation trend on whether the period is dry or wet. To accurately understand the detailed trend, anomaly calculations are necessary. Converting the data into anomalies and applying smoothing techniques allows for a clearer identification of these trends and will be seen in chapter 4.

## Chapter 4

# Analysis of Rainfall Patterns and Climate Indices

In this chapter, we focus on analyzing rainfall patterns with an emphasis on understanding the influence of various oceanic climate indices on precipitation. Climate indices such as the Oceanic Niño Index (ONI), the Dipole Mode Index (DMI), and the Pacific Decadal Oscillation (PDO), represent large-scale fluctuations in ocean temperatures and atmospheric conditions. These indices are used to track ocean-atmosphere interactions that can influence climate patterns, including rainfall variability across different regions.

Previous research has looked closely at how climate indices, such as ENSO (El Niño-Southern Oscillation) are connected to climate anomalies including rainfall patterns, in different parts of the world. Studies have found strong links between these indices and rainfall, particularly in tropical and subtropical regions, showing how changes in the ocean can impact weather patterns (McPhaden et al. (2006)). Researchers have used techniques like correlation analysis and time-series modeling to measure the effects of these oceanic signals on rainfall. In this chapter, we build on this work by exploring the connections between smoothed rainfall data and climate indices, as well as how these relationships change with the seasons.

From section 4.2, we can see the comparison of smoothed rainfall data with selected climate indices. This comparison helps us identify patterns and similarities that may indicate a relationship between these indices and rainfall variations.

Following the pattern analysis, we conduct a detailed correlation analysis to quantify the strength of the relationship between rainfall and the climate indices. By identifying which indices are most strongly correlated with rainfall, we can gain insights into the underlying climatic drivers of precipitation patterns.

Finally, we explore the seasonal aspects of these relationships, examining how the influence of climate indices on rainfall varies throughout the year. Understanding this seasonality is critical for particular regions where seasonal rainfall is dominant.

Through this comprehensive analysis, we aim to identify the most influential climate indices

and provide a foundation for the forecasting models that will be developed in the subsequent chapter.

## 4.1 Calculating Monthly Average Precipitation

In this section, we focus on the transformation of raw monthly average precipitation data to better understand the relationship between rainfall patterns and climate indices, specifically for two key regions: India and Indonesia. These regions were chosen due to their significant seasonal rainfall patterns, which are highly influenced by various climatic factors, including climate indices.

The first step in this process involves converting the normal precipitation data into anomalies for both India and Indonesia. Here, anomalies represent the deviation of observed precipitation from the long-term average (2003 -2023), providing a clearer perspective on how current conditions differ from the expected norm. This transformation is crucial for highlighting unusual patterns in each region that may be driven by oceanic conditions.

$$\text{Anomaly} = \text{Precipitation} - \text{Average Precipitation}$$

### 4.1.1 India Raw Anomaly

Graph 4.1 displays the anomalies in monthly precipitation (2003 -2023) of India from the long-term average, with positive anomalies shaded in blue indicating wetter conditions and negative anomalies shaded in red indicating drier conditions.

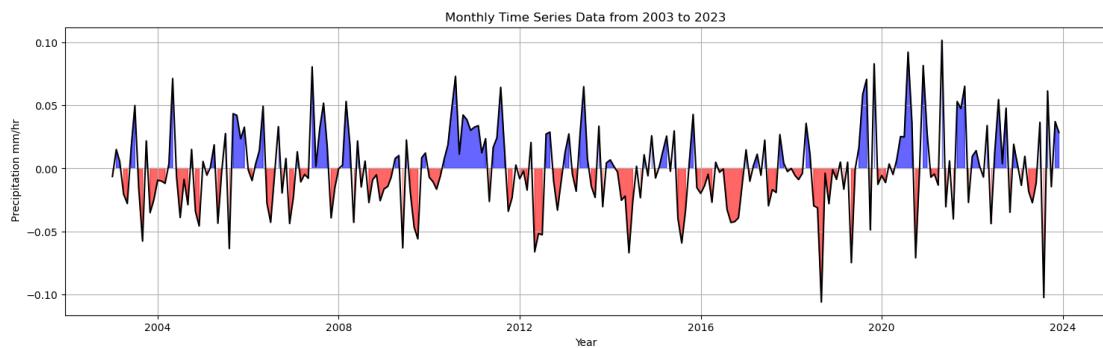
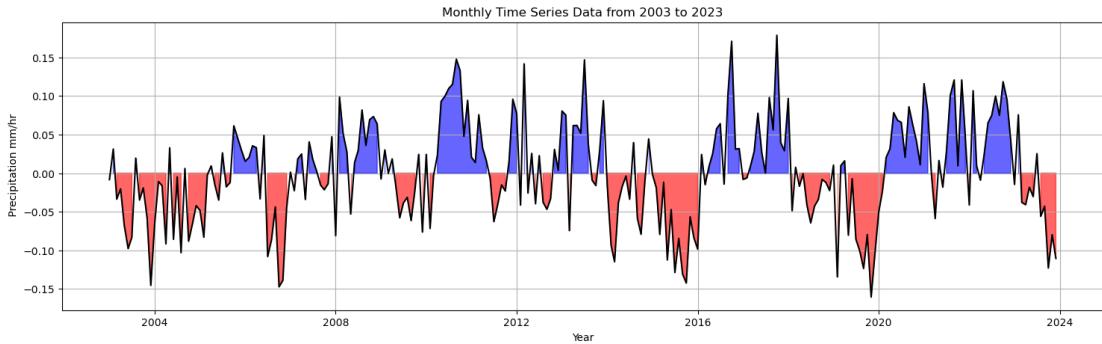


Figure 4.1: Monthly Precipitation Anomalies for India (2003-2023).

### 4.1.2 Indonesia Raw Anomaly

Graph 4.2 works the same as India anomaly, which displays the deviations in monthly precipitation (2003 -2023) of Indonesia from the long-term average.



*Figure 4.2: Monthly Precipitation Anomalies for Indonesia (2003-2023).*

## 4.2 Smoothing Techniques in Time Series Analysis

In this section, we explore the application of various smoothing techniques to test the effects of different parameter choices and filter types on precipitation anomalies. The filter we are going to use, Kolmogorov-Zurbenko (KZ) filter, triangular smoothing, Gaussian smoothing, and simple moving averages, are applied with varying window sizes to better understand how these choices influence the interpretation of rainfall data. The goal is to assess how different smoothing methods can reduce noise while preserving important trends and patterns in precipitation anomalies for both India and Indonesia.

Smoothing techniques are fundamental tools used in time series analysis to help reduce noise and reveal underlying trends and patterns. These techniques are especially useful in climatological studies, such as analyzing precipitation data, where they aid in highlighting significant variations and relationships that might be obscured by short-term fluctuations.

### 4.2.1 Kolmogorov-Zurbenko (KZ) Filter

The Kolmogorov-Zurbenko (KZ) filter is named after its developers, Andrey Kolmogorov, a prominent Russian mathematician, and Igor Zurbenko, a statistician who further developed the method for practical applications.

The KZ filter is a type of moving average filter that applies multiple iterations of a simple moving average (SMA) to the data. In a key study, Yang and Zurbenko (2010) used the KZ filter on 20 years of daily temperature data from the United States. The study showed how the KZ filter effectively separates short-term changes from long-term trend. This makes the KZ filter ideal for precipitation data.

#### Iterative Application in KZ Filter:

For the KZ filter, which applies the SMA iteratively, the formula is:

$$S_t^{(i)} = \frac{1}{k} \sum_{j=-(k-1)/2}^{(k-1)/2} S_{t+j}^{(i-1)}$$

$S_t^{(l)}$ : This is the smoothed value of the time series at time  $t$  after  $l=5$  iterations of the KZ filter.

$k$ : This is the window size (3, 5, 7) for the moving average. The window size determines how many surrounding data points are considered when smoothing each point.

$S_{t+j}^{(l-1)}$ : This refers to the smoothed value of the time series at time  $t + j$  after  $l - 1$  iterations of the KZ filter.

$j$ : This index runs from  $-\frac{(k-1)}{2}$  to  $\frac{(k-1)}{2}$ , ensuring that the moving average is symmetrically applied around the current data point  $t$ .

Here, the iteration is constant with the size of 5 over both regions.

### **India:**

In this section, we explore how varying window sizes and the number of iterations in the Kolmogorov-Zurbenko (KZ) filter affect smoothed precipitation anomalies and shows the impact of using different window sizes (3, 5, 7) on precipitation anomalies in India.

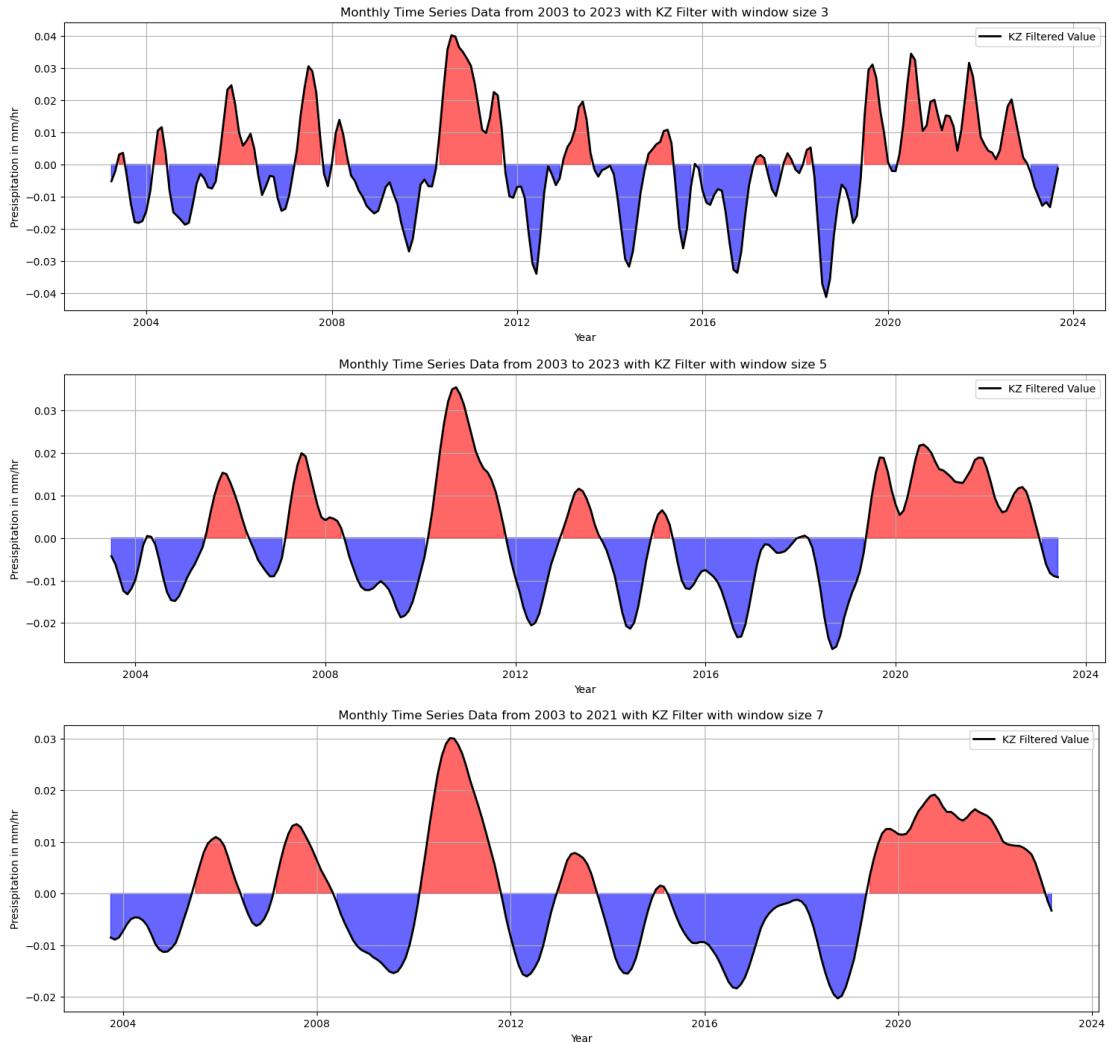


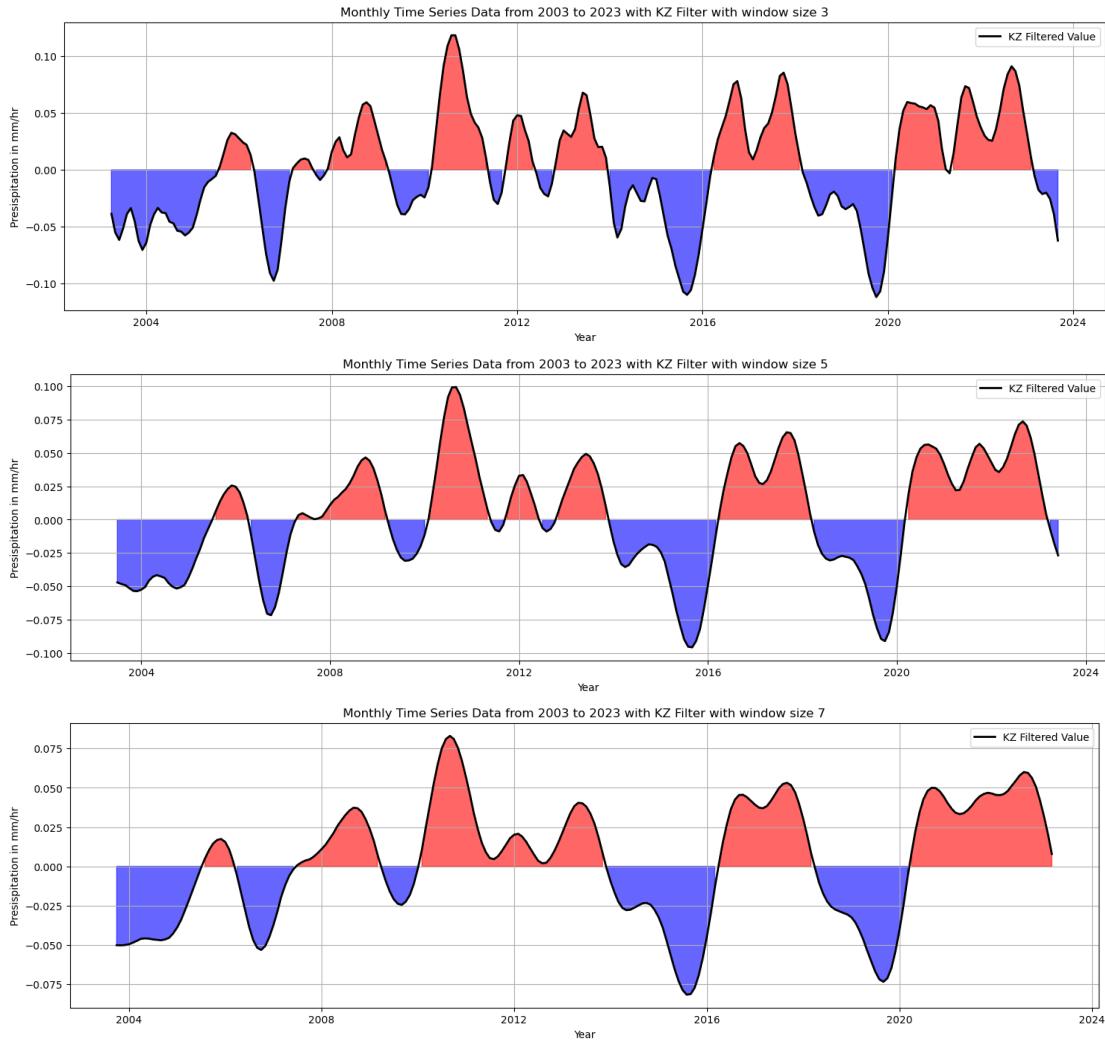
Figure 4.3: Smoothed Anomalies of India using a KZ Filter with a Window Size of (3,5,7).

As seen in Figure 4.3, smaller window sizes 3 highlight short-term variability with sharper peaks and troughs, which is useful for detecting sudden changes in precipitation. Increasing the window size to 5 results in smoother anomalies with less pronounced peaks, indicating more sustained weather patterns. With a window size of 7, the anomalies are even smoother, emphasizing long-term trends and filtering out short-term noise.

### Indonesia:

In this section, we also apply the Kolmogorov-Zurbenko (KZ) filter to analyze the effect of varying window sizes on smoothed precipitation anomalies in Indonesia, as shown in Figure 4.4.

Similar to the results for India, smaller window size 3 reveals sharper peaks, reflecting short-term variability in the precipitation data. As the window size increases to 5 and 7, the anomalies become progressively smoother, emphasizing longer-term trends and reducing the prominence



*Figure 4.4:* Smoothed Anomalies of Indonesia using a KZ Filter with a Window Size of (3,5,7).

of short-term fluctuations.

#### 4.2.2 Triangle Smoothing

Triangle Smoothing, or Triangular Moving Average (TMA), is a variation of the moving average technique used to smooth time series data. Unlike the simple moving average, TMA applies weights that decrease linearly from the center to the edges of the moving window. In technical analysis, traders use TMA to smooth price data and identify trends more accurately than with basic moving averages, making it valuable for analyzing stocks and commodities.

Given a time series  $Y$ , the triangular smoothed value  $S_t$  at time  $t$  using a window size  $n$  (preferably an odd number) can be calculated as follows:

$$S_t = \frac{2}{n(n+1)} \sum_{i=0}^{n-1} (n-i) Y_{t-n/2+i}$$

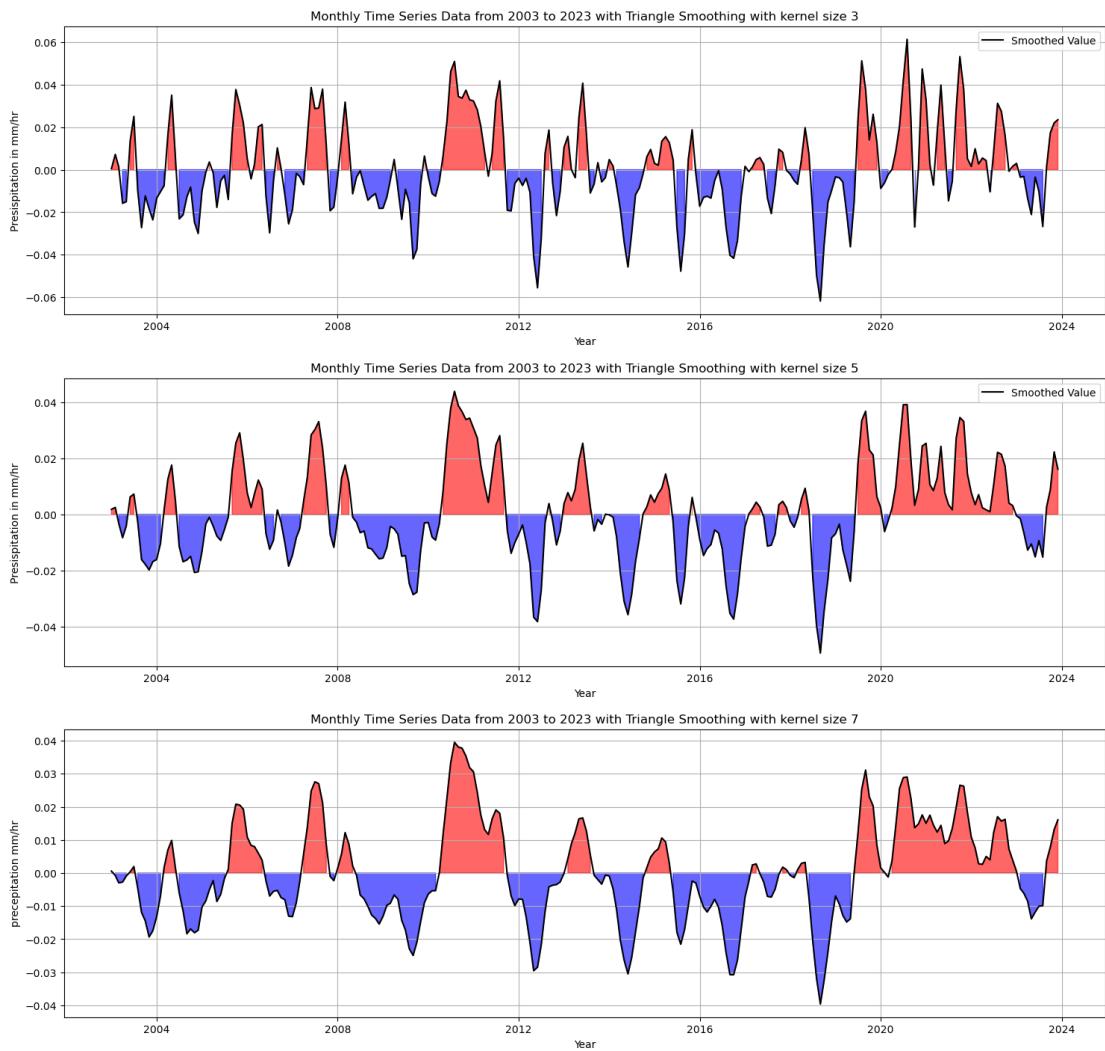
where  $n$  is the window size and  $i$  indexes each point within the window, extending from  $t - n/2$  to  $t + n/2$ . This implies that  $Y_{t-n/2+i}$  are the data points in the window, and this  $(n - i)$  represents the weights assigned to each data point, decreasing linearly from the center to the ends of the window.

**India:** Figure 4.5 shows how triangular smoothing affects precipitation anomalies in India when using different kernel sizes (3, 5, 7). Triangular smoothing gives more weight to data points in the center of the window and less to those at the edges, creating a smoother transition between points.

The top plot in figure 4.5 with kernel size 3 highlights short-term changes with more visible peaks and troughs, as the smaller kernel size closely follows the original data.

In the Middle Plot in figure 4.5 with kernel size 5 shows the smoothing more noticeable, with peaks and troughs becoming less sharp, representing more consistent weather patterns.

The bottom Plot in figure 4.5 with kernel size 7 shows the smoothest anomalies, with broad, subtle peaks, focusing more on long-term trends and filtering out short-term changes.

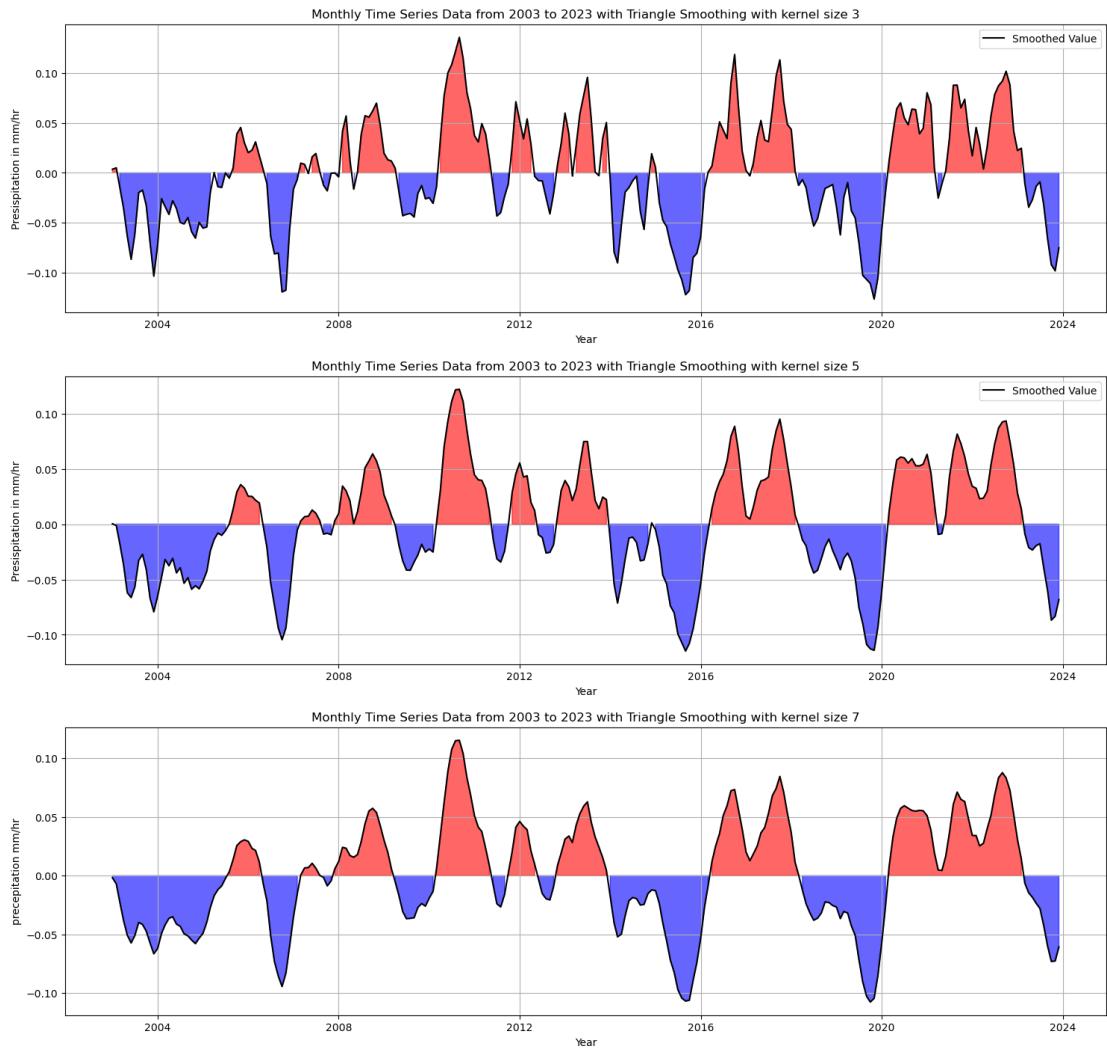


*Figure 4.5: Triangle Smoothed Anomalies of India using a kernel Size of (3,5,7).*

**Indonesia:** Figure 4.6 illustrates the effect of triangular smoothing on precipitation anomalies in Indonesia, using different kernel sizes (3, 5, 7). The overall trends of smoothing are similar to those observed in India.

In Indonesia, the precipitation anomalies, especially with smaller kernel size 3, have higher peaks and deeper troughs than in India. This means short-term rainfall variability is more pronounced in Indonesia.

As the kernel size increases from 3 to 7, the smoothing becomes more noticeable, similar to India.



*Figure 4.6: Triangle Smoothed Anomalies of Indonesia using a kernel Size of (3,5,7).*

#### 4.2.3 Gaussian Smoothing

This technique is used extensively in data processing and image processing (Gonzalez (2009)), particularly in the fields of signal processing and image processing, but also applicable to time series analysis like climatological data series. The key characteristic of Gaussian smoothing is the use of the Gaussian function, also known as the normal distribution curve, to smooth data. This method is especially valued for its properties of preserving the peaks and spatial locality of the original data, making it highly effective for applications where these attributes are critical.

The Gaussian function is defined as follows:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

where  $\sigma$  is the standard deviation of the distribution, controlling the spread of the bell curve.

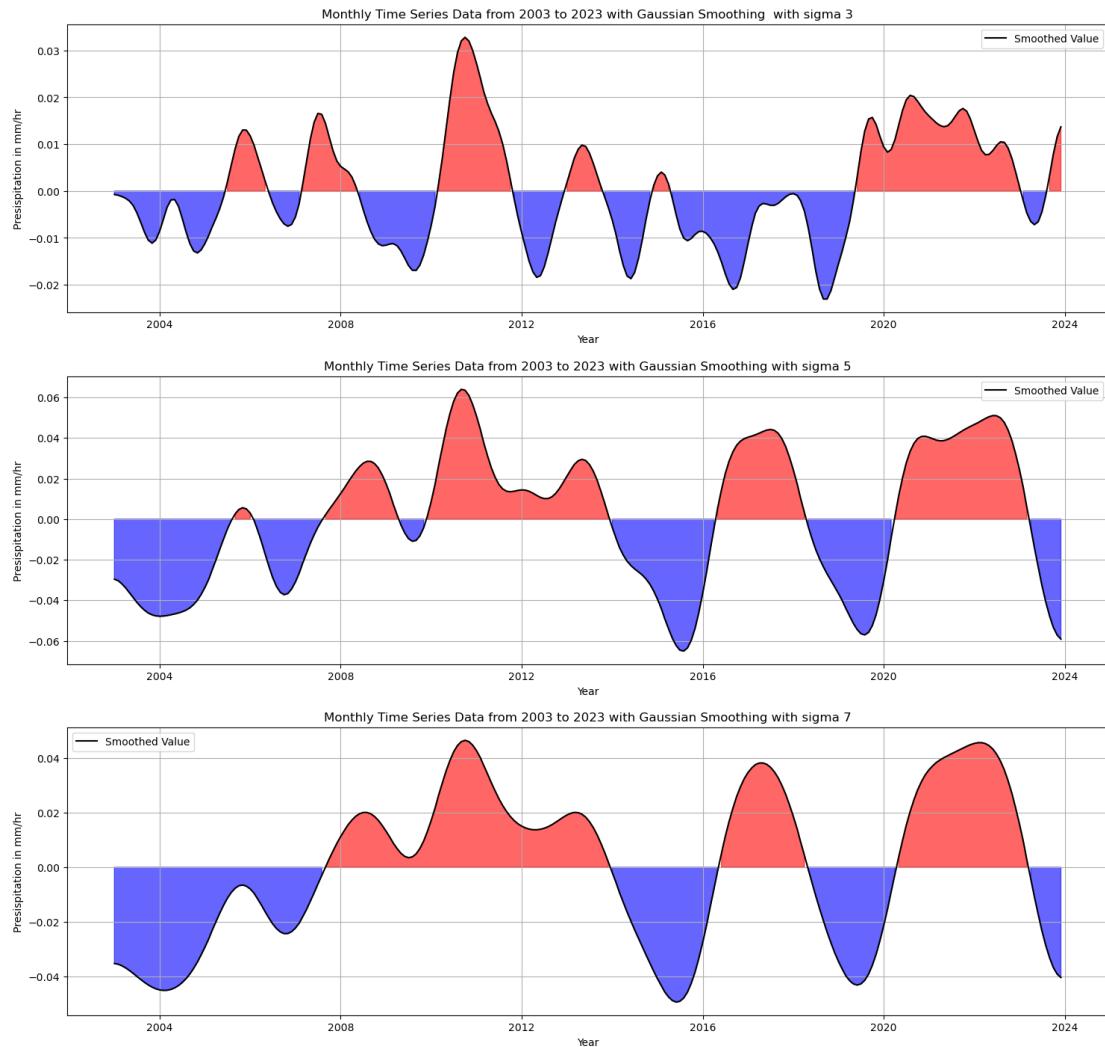
For Gaussian smoothing, the formula applied to data is:

$$S(t) = \sum_{i=-N}^{N} Y(t+i) \cdot G(i)$$

. Here,  $S(t)$  is the smoothed value at time  $t$ ,  $Y(t+i)$  are the data points around  $t$ , and  $G(i)$  are the Gaussian weights applied to each data point within the window defined by  $N$ , which typically extends several standard deviations on either side of  $t$  where  $i$  is an index that takes on values ranging from  $-N$  to  $N$ , where  $N$  is the number of time steps.

### India:

Figure 4.7 shows the effect of applying Gaussian smoothing to precipitation anomalies in India, using different sigma sizes (3, 5, and 7).

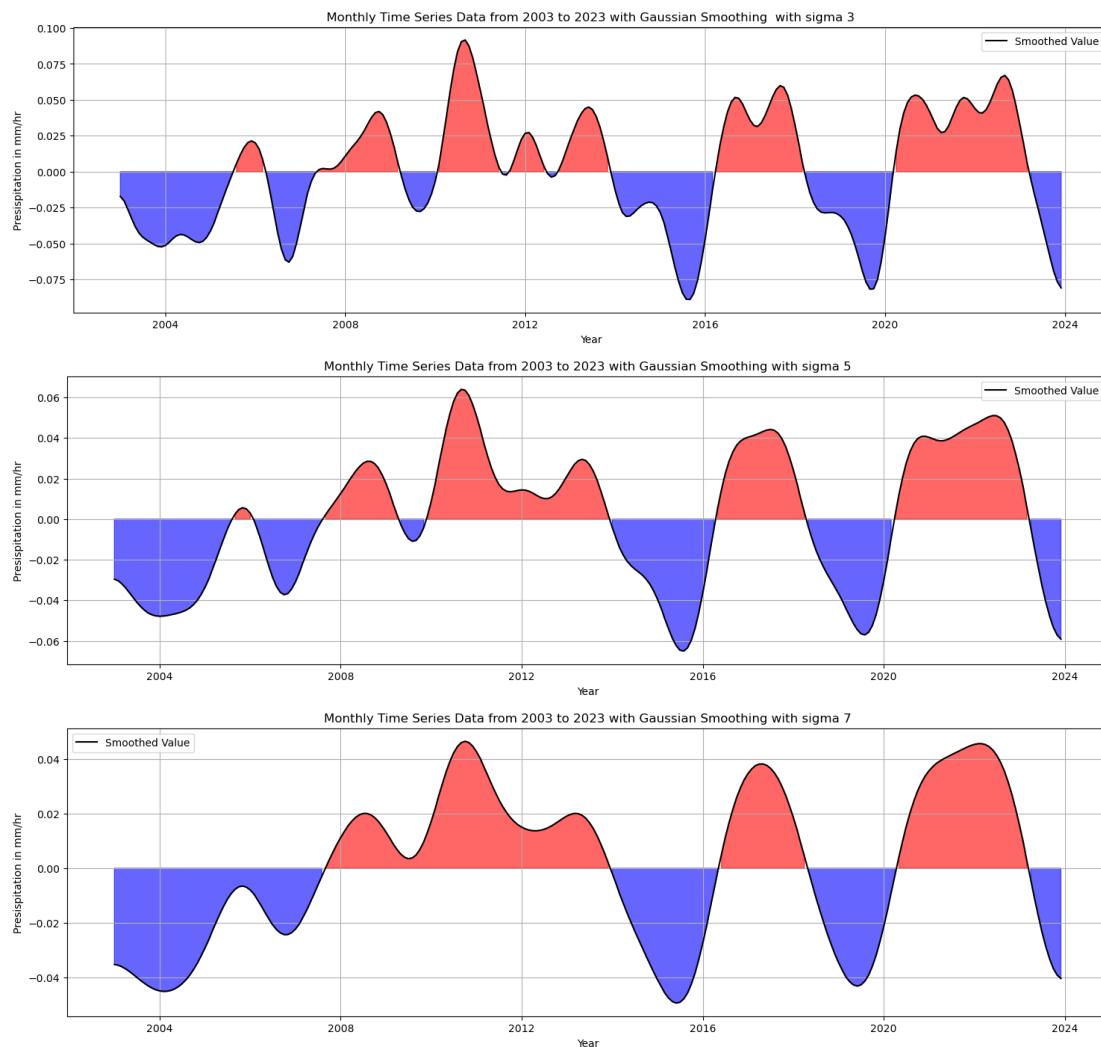


*Figure 4.7: Gaussian Smoothed Anomalies of India using a Sigma Size of (3,5,7).*

The sigma size of 3 offers good smoothing while preserving enough detail in the data. Sigma size 5 also works well for a more refined smoothing, while Sigma size 7 captures larger trends but risks over smoothing, which may obscure some of the short-term details. The choice of sigma size depends on the desired focus whether the goal is to observe short-term variations or broader, long-term trends.

### **Indonesia:**

Figure 4.8 presents the impact of Gaussian smoothing on precipitation anomalies in Indonesia, using different sigma sizes (3, 5, and 7). While the general behavior is similar to the Indian data,



*Figure 4.8: Gaussian Smoothed Anomalies of Indonesia using a Sigma Size of (3,5,7).*

here we are going to see the difference between the Gaussian filter and another filter. The Gaussian filter provides a more gradual smoothing effect compared to the KZ filter. The KZ filter, especially with larger window sizes, tends to flatten the data more and highlight long-term trends, making it better for long-term analysis. In contrast, the Gaussian filter retains more

short-term variability, particularly with smaller sigma sizes, offering more flexibility. The Gaussian filter creates smoother transitions between data points compared to triangular smoothing. Triangular smoothing, with its linearly decreasing weights, preserves a bit more short-term variability than the Gaussian filter at similar window or kernel sizes.

#### 4.2.4 Moving Average

The Moving Average method is introduced last because it is the simplest and most basic smoothing technique in time series analysis. Even though it's simple, it helps build a strong foundation for understanding the more advanced methods, like the KZ filter and Gaussian smoothing. By explaining the more complex filters first, and then ending with the Moving Average, we can better appreciate how the more sophisticated techniques expand on these basic principles.

Introducing the Moving Average at the end also helps reinforce the core ideas of smoothing, making it easier to compare its strengths and weaknesses with the more advanced methods that were discussed earlier. This way, the fundamental concepts are clear before diving into deeper comparisons.

Given a time series  $Y_t$ , the moving average  $S_t$  at time  $t$  with a window size  $n$  is calculated as follows:

$$S_t = \frac{1}{n} \sum_{i=0}^{n-1} Y_{t-i}$$

Where:

$S_t$  is the smoothed value at time  $t$ .

$Y_{t+i}$  are the data points within the window centered around  $t$ .

$n$  is the window size (3, 5, 7).

The sum ranges from  $i = -\frac{n-1}{2}$  to  $i = \frac{n-1}{2}$ , ensuring the window is centered around the point  $t$ .

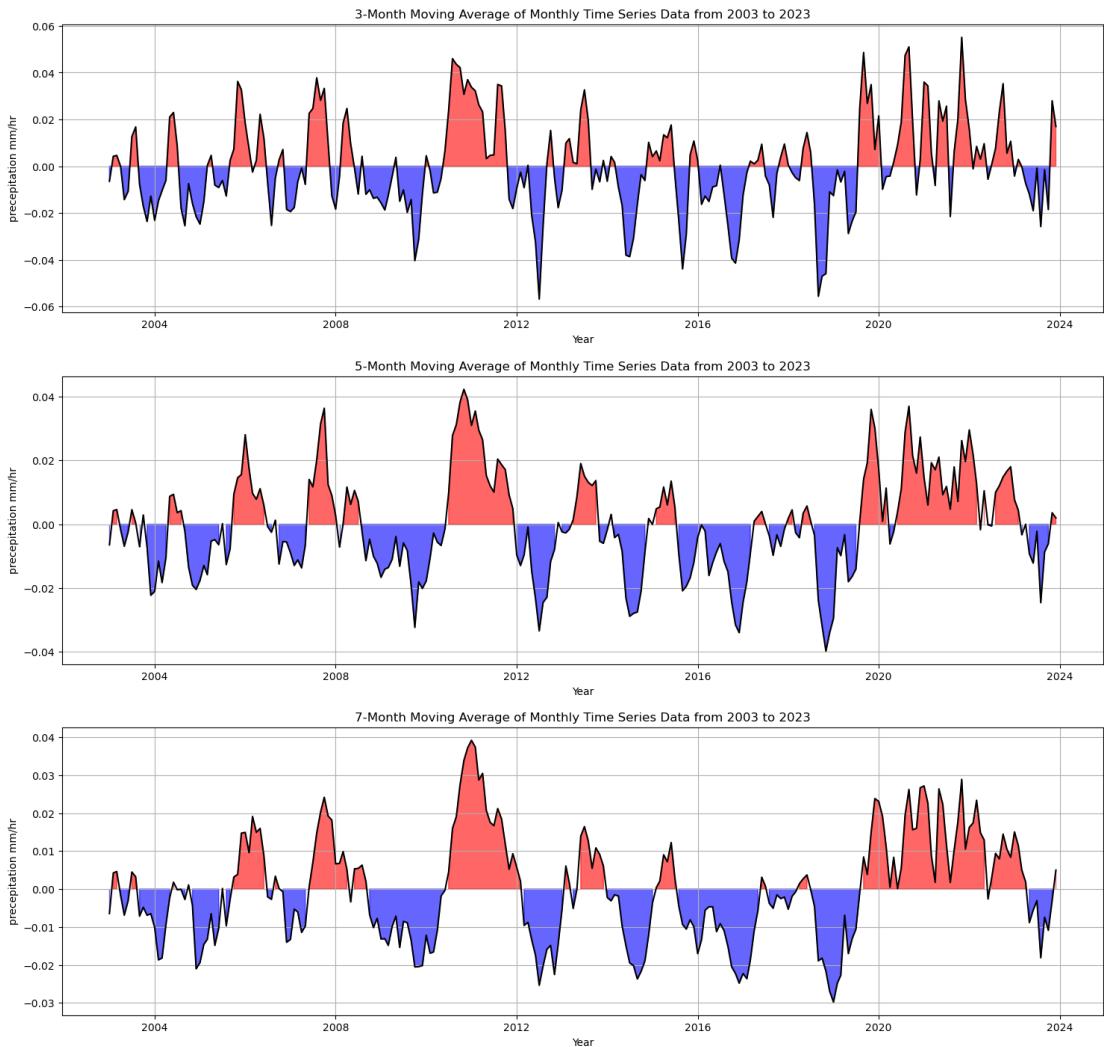
#### India:

Figure 4.9 displays the application of a 3-month, 5-month, and 7-month moving average to the precipitation anomalies in India.

The 3-month window provides subtle smoothing, allowing short-term fluctuations to still be visible. The peaks and troughs remain sharp, showing clear short-term variability.

The 5-month window smooths the data more than the 3-month window. Peaks and troughs are broader and less distinct, reducing short-term variability and highlighting more sustained patterns.

The 7-month window provides the strongest smoothing effect. Peaks and troughs are very broad, focusing on long-term trends by filtering out most short-term fluctuations.



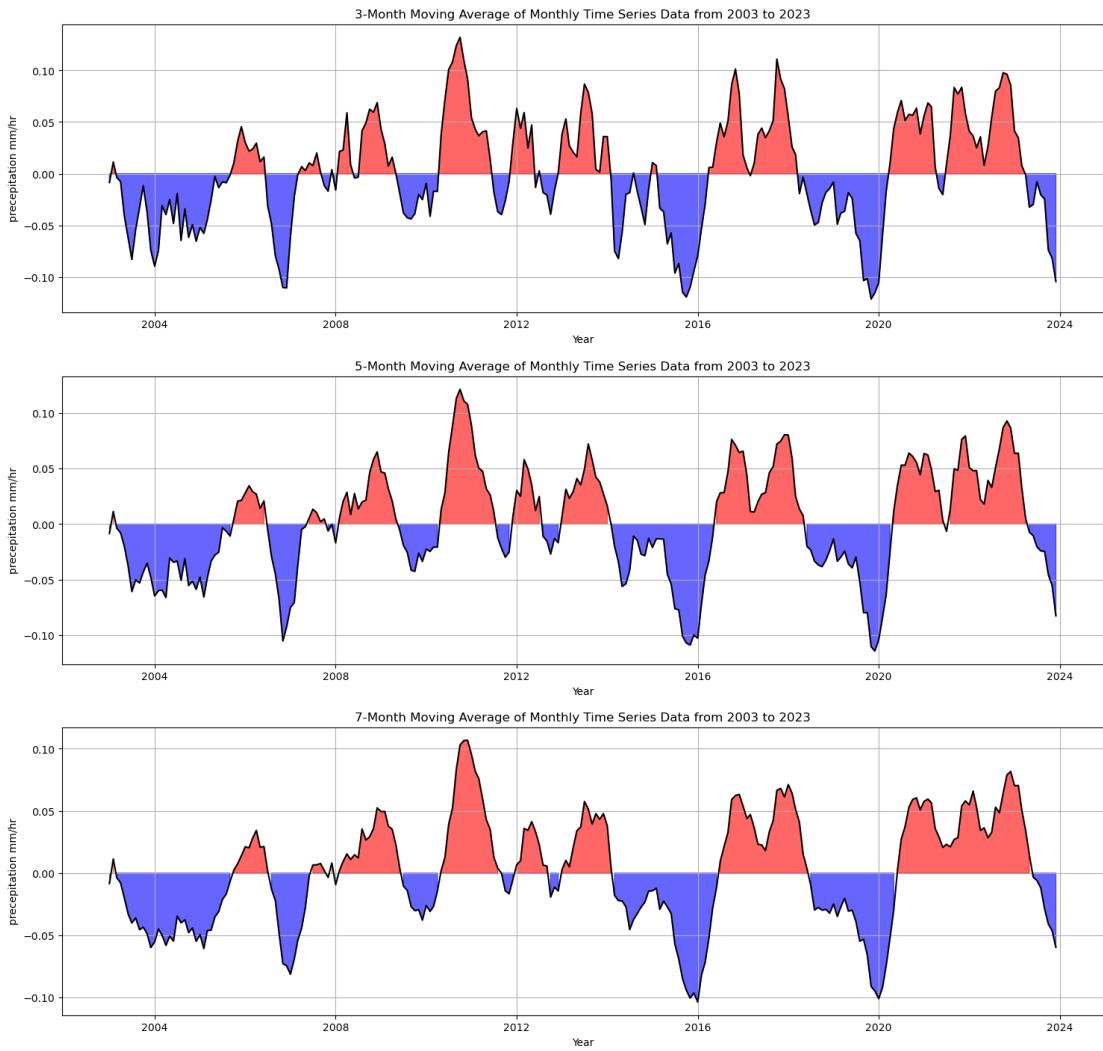
*Figure 4.9: (3,5,7)-Month Moving Average of India Anomaly.*

### **Indonesia:**

Figure 4.10 shows the application of 3-month, 5-month, and 7-month moving averages to precipitation anomalies in Indonesia.

The moving average is a reliable method for smoothing precipitation anomalies in Indonesia, particularly with larger window sizes that effectively highlight long-term trends. However, for more complex smoothing needs, such as preserving short-term variability while identifying trends, triangular and Gaussian smoothing provide more refined results from visual checks.

The KZ filter remains the best option when the goal is to isolate long-term trends with minimal influence from short-term noise. Each method has its own strengths, and the choice of filter should depend on the specific goals of the analysis and the nature of the data.



*Figure 4.10: (3,5,7)-Month Moving Average of Indonesia Anomaly.*

### 4.3 Impact of Window Size on Smoothing Precipitation Anomalies

A noticeable change that are common in all smoothing are is the effect of the filter at different window sizes (3, 5, and 7) on the precipitation anomalies for both India and Indonesia.

For India, with a window size of 3, sharp peaks and deep troughs are evident, indicating short-term variability, such as the pronounced peak around 2010, which reaches approximately 0.04 mm/hr. As the window size increases to 5, the peaks become less prominent, reflecting more sustained weather patterns. When the window size is expanded to 7, the peaks are even broader and less sharp, with values around 0.03 mm/hr, highlighting long-term trends while filtering out short-term fluctuations and revealing broader climatic patterns.

For Indonesia, a similar trend is observed. With a window size of 3, sharp peaks and troughs are present, showcasing short-term variability, such as the peak around 2006 reaching 0.10 mm hr.

As the window size increases to 5, the peaks become smoother and more representative of sustained weather patterns. With a window size of 7, the peaks are broader and less distinct, reaching around 0.075 mm/hr, thus emphasizing long-term climatic trends while minimizing short-term variability.

The anomaly plots for both regions show considerable short-term variability, with sharp peaks and troughs that can obscure long-term trends. This unpredictable nature underscores the necessity of applying smoothing techniques to reveal broader patterns and trends in precipitation anomalies. This leads us to the next section 4.4, where we analyze the relationship between these smoothed precipitation data and various climate indices.

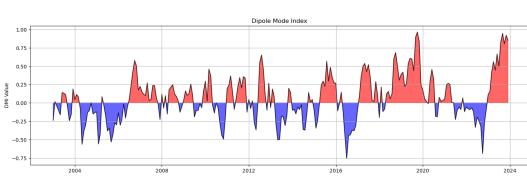
## 4.4 Pattern Resemblance Between Smoothed Precipitation and Climate Indices

In this section, we focus on analyzing the relationship between smoothed precipitation data and various Climate indices to understand how these large-scale climate phenomena influence regional rainfall patterns. The goal is to identify any patterns or correlations between these indices and the observed rainfall data after it has been smoothed using techniques like the Kolmogorov-Zurbenko (KZ) filter, triangular smoothing, Gaussian smoothing and monthly average.

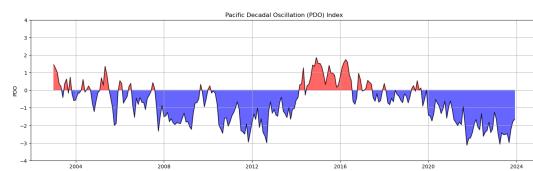
We generate time series plots of the smoothed precipitation data alongside the ONI, DMI, and PDO indices. By overlaying these time series, we can visually assess the alignment of trends, such as whether peaks and troughs in rainfall corresponding to specific phases of the indices.

### Visual Comparison Analysis

Visual comparison provides immediate insights into how well the patterns of the Climate indices match the smoothed precipitation data in India and Indonesia. Key observations from this comparison include:



*Figure 4.11: Dipole Mode Index (DMI).*



*Figure 4.12: Pacific Decadal Oscillation (PDO) Index.*

*Figure 4.13: Side-by-side comparison of DMI and PDO indices.*

**India:** The visual analysis indicates that there is no clear relationship between precipitation anomalies and either the DMI or PDO, as illustrated in Figure 4.13 there is no good visualization relation. The patterns in India's rainfall data do not align closely with the fluctuations observed in these indices.

**Indonesia:** Although the relationship is somewhat stronger than in India, it remains weak, and the alignment between the precipitation data and both the DMI and PDO, as seen in Figure 4.13, is still inconsistent.

**India:** The relationship with ENSO is less pronounced in India, although some correlation is still visible, particularly during strong ENSO events.

**Indonesia:** A clear visual alignment is observed between ENSO (represented by ONI) and the rainfall patterns in Indonesia. This strong correlation suggests that ENSO significantly influences rainfall in Indonesia.

### Graphical Comparison:

To support these observations, graphs in figure 4.14 comparing the smoothed precipitation data with the climate indices will be provided. These visual aids will highlight the degree of pattern alignment, making it easier to identify where the strongest relationships occur.

#### India:

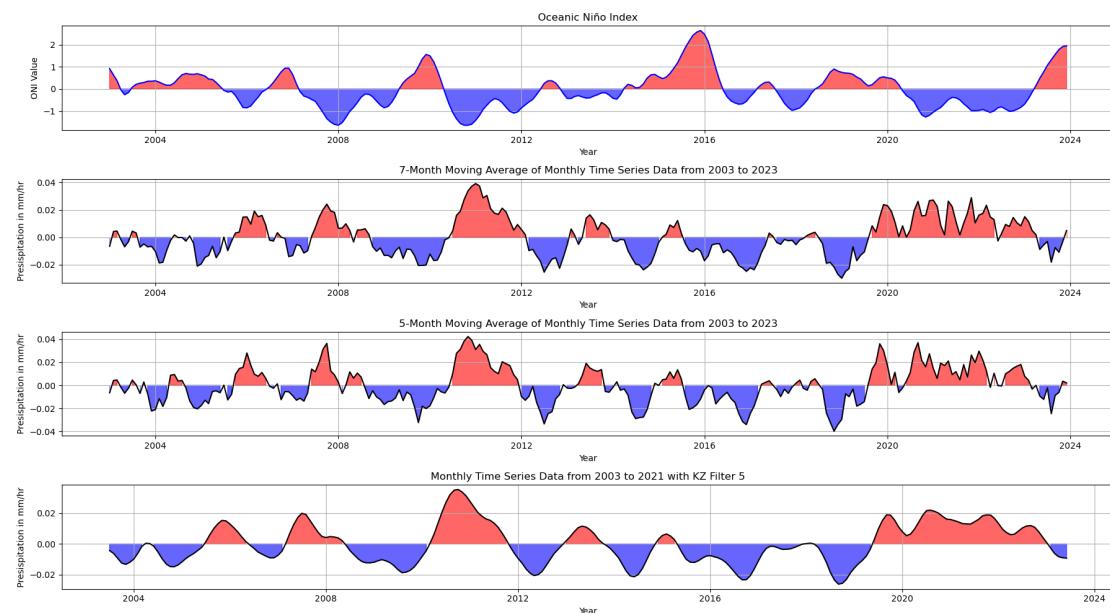


Figure 4.14: Comparing the ENSO indices with smoothing anomalies of India.

Figure 4.14 presents a comparison between the Oceanic Niño Index (ONI) or ENSO and the smoothed precipitation anomalies for India. The following observations are made based on the different smoothing techniques:

**7-Month Moving Average:** The precipitation anomaly for India smoothed with a 7-month moving average shows a relatively stable long-term pattern. The peaks corresponding to El Niño events, such as the one in 2015-2016, indicate drier conditions, though not as extreme as

in Indonesia. The 7-month filter provides insight into broader climatic shifts but smooths out some short-term variations

**5-Month Moving Average:** With the 5-month moving average, the precipitation anomaly reveals more distinct peaks and troughs, still reflecting the impact of ENSO events but with slightly more short-term fluctuations than the 7-month filter.

**KZ Filter (Window Size 5):** This filter produces even sharper and more defined peaks and troughs in the precipitation anomaly. It reflects both short-term fluctuations and long-term patterns, providing a more dynamic view of the anomalies in response to ENSO events. The filter captures significant variability, such as during the 2010 and 2015-2016 periods, highlighting the influence of ENSO on India's precipitation patterns.

### Indonesia:

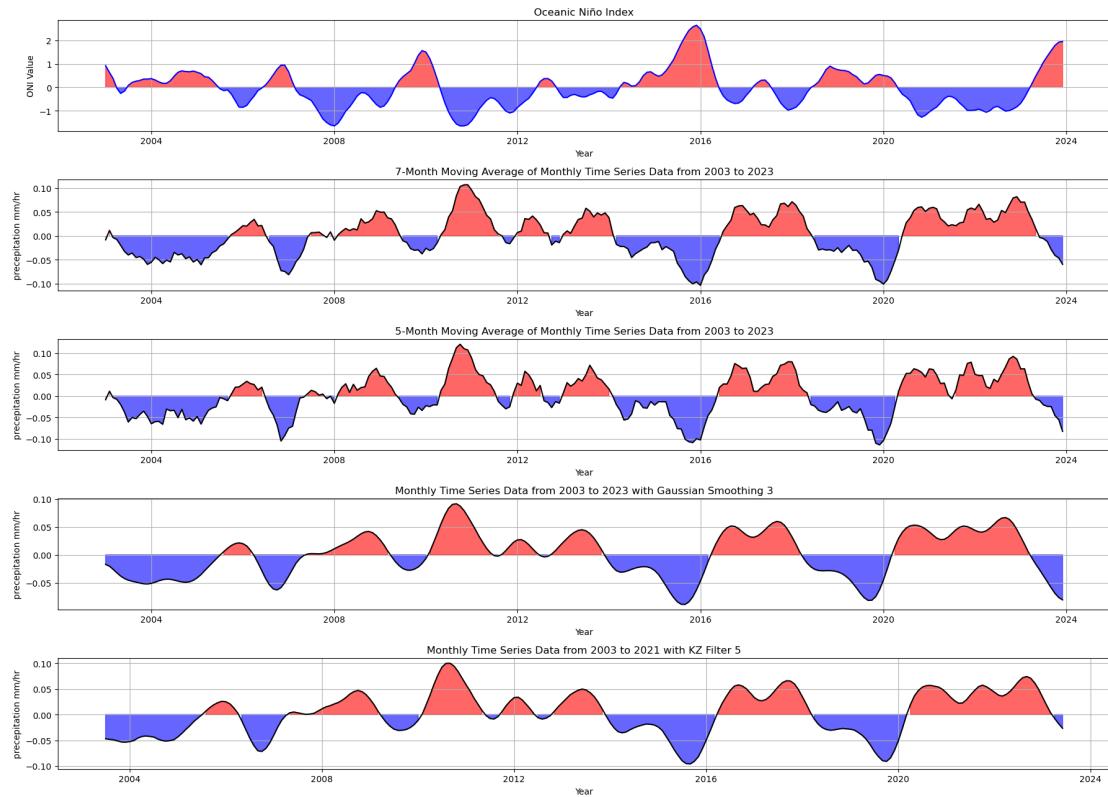


Figure 4.15: Comparing the ENSO indices with smoothed anomalies of Indonesia.

Figure 4.15 shows a similar comparison between the Oceanic Niño Index (ONI) or ENSO and the smoothed precipitation anomalies for Indonesia using various smoothing techniques.

**7-Month Moving Average:** The precipitation anomaly smoothed using a 7-month moving average for Indonesia shows broader and less sharp peaks compared to the ONI. Notably, during

the strong El Niño event around 2015-2016, the anomaly reflects a prolonged dry period. This filter effectively captures long-term precipitation patterns but lags in capturing sudden changes.

**5-Month Moving Average:** The 5-month moving average for Indonesia highlights more defined peaks and troughs compared to the 7-month average. It reflects the precipitation fluctuations more responsively to ENSO cycles, showing drier conditions during El Niño years and wetter conditions during La Niña years.

**Gaussian Smoothing (Sigma 3):** The Gaussian smoothing technique at a sigma size of 3 for Indonesia produces a smoothed anomaly that captures long-term climatic trends but with more sensitivity to short-term variations. This filter reveals the broader trends but still highlights important events like the significant dry periods during El Niño years.

**KZ Filter (Window Size 5):** The KZ filter with a window size of 5 produces sharper peaks in the anomaly. This highlights Indonesia's precipitation response to ENSO events with clear troughs and peaks, particularly during the 2010 and 2015-2016 El Niño events. The KZ filter allows for more precise tracking of these fluctuations compared to the broader moving averages. Before we dive into the correlation analysis in section 4.5, it is worth taking a moment to clarify what we are trying to achieve and why it matters. Our goal here is to see how well different smoothing methods can reveal the connections between precipitation anomalies and key climate indices (like ONI, PDO, and DMI) for both India and Indonesia.

## 4.5 Correlation Analysis

Here, we calculate correlation coefficients across various techniques—such as the Kolmogorov-Zurbenko (KZ) filter, Gaussian smoothing, triangular smoothing, and moving averages—we are trying to figure out which method does the best job of bringing out these relationships. Ultimately, this analysis helps us in identify the most effective filter for uncovering meaningful patterns in the climate data.

### 4.5.1 India

The correlation analysis for India shows the strength of the relationship between the climate indices and precipitation anomalies across different smoothing methods, which we can see from the table 4.1:

INDIA	PDO	ONI	DMI
KZ Filter 3	-0.26	-0.47	-0.05
KZ Filter 5	-0.31	-0.53	-0.06
KZ Filter 7	-0.37	<b>-0.54</b>	-0.04
Gaussian smoothing 3	-0.29	-0.45	-0.02
Gaussian smoothing 5	-0.41	-0.48	-0.01
Gaussian smoothing 7	-0.48	-0.47	-0.03
3_month (MAV)	-0.22	-0.38	-0.07
5_month (MAV)	-0.28	-0.50	-0.13
7_month (MAV)	-0.37	<b>-0.54</b>	-0.11
Triangular smoothing 3	-0.23	-0.32	-0.02
Triangular smoothing 5	-0.25	-0.39	-0.02
Triangular smoothing 7	-0.27	-0.44	-0.03

Table 4.1: Correlation Coefficients of Different Smoothing Methods on Climate Indices for India.

**ONI (ENSO):** The Oceanic Nino Index (ONI) consistently shows a strong negative correlation with precipitation across all smoothing methods, with values ranging from -0.38 to -0.54. The strongest correlation is observed with the 7-Month Moving Average and KZ Filter 7 (both at -0.54). This finding aligns with previous studies, such as those by Kumar et al. (2006), which documented the weakening or post Indian monsoon during El Nino events. La Nina phases (negative ONI values), on the other hand, typically enhance monsoon rainfall, though to a lesser extent compared to the drought-inducing effects of El Nino.

**PDO:** The Pacific Decadal Oscillation (PDO) shows a moderate negative correlation with Indian precipitation, noted in table 4.1, with the strongest correlation of -0.48 seen using the Gaussian Smoothing method with size 7. This finding is consistent with research by Singh et al. (2020), which indicates that PDO can modulate long-term climate patterns, including rainfall variability, though its influence on India's rainfall is less pronounced than ENSO's effects .

**DMI:** The correlation between the Dipole Mode Index (DMI) and Indian precipitation is notably weak, with values ranging from -0.01 to -0.13. Previous research by Ashok and Saji (2007). demonstrated that the DMI can have significant regional impacts focusing on the southwest and southeast, but its weaker correlations in this study suggest that in India, DMI's influence on rainfall may be overshadowed by ENSO, particularly during strong El Nino or La Nina phases.

#### 4.5.2 Indonesia

In contrast to India, the correlation results for Indonesia indicate a much stronger relationship between climate indices and precipitation anomalies, which we can see from the table 4.2:

<b>INDONESIA</b>	<b>PDO</b>	<b>ONI</b>	<b>DMI</b>
KZ Filter 3	-0.40	-0.71	-0.41
KZ Filter 5	-0.47	-0.74	-0.37
KZ Filter 7	<b>-0.53</b>	-0.74	-0.32
Gaussian smoothing 3	-0.42	-0.76	-0.43
Gaussian smoothing 5	-0.48	-0.73	-0.31
Gaussian smoothing 7	-0.53	-0.68	-0.21
3_month (MAV)	-0.35	-0.73	-0.42
5_month (MAV)	-0.40	-0.75	-0.37
7_month (MAV)	-0.44	<b>-0.77</b>	-0.28
Triangular smoothing 3	-0.32	-0.67	-0.42
Triangular smoothing 5	-0.36	-0.70	-0.43
Triangular smoothing 7	-0.38	-0.72	-0.43

Table 4.2: Correlation Coefficients of Different Smoothing Methods on Climate Indices for Indonesia.

**ONI (ENSO):** The ONI demonstrates very strong negative correlations with Indonesian rainfall, as noted in table 4.2, with values ranging from -0.67 to -0.77. The highest correlation is observed with the 7-Month Moving Average (-0.77), which aligns with previous studies highlighting ENSO's significant influence on Southeast Asian climates by Chang et al. (2020). El Niño events, characterized by positive ONI values, are strongly associated with droughts in Indonesia, while La Niña events tend to bring increased rainfall .

**PDO:** The PDO also shows moderate negative correlations with precipitation in Indonesia, as noted in table 4.2, with the strongest correlation of -0.53 found using the KZ Filter method size 7. The PDO's influence, while secondary to ENSO, has been shown to modulate long-term climate patterns in the Pacific, affecting rainfall variability across the region (Yulihastin et al. (2018)) .

**DMI:** The Dipole Mode Index (DMI) shows a stronger correlation with Indonesian rainfall compared to India, though still weaker than ONI. Correlation values range from -0.21 to -0.43, with the highest correlations seen using Triangular Smoothing methods.

One of the main insights from this analysis is that, for Indonesia, the correlation with ONI remains consistently strong across all the different smoothing methods we tested. This means that no matter which averaging technique is used, the relationship between ONI and Indonesian precipitation stays robust. It highlights how influential ONI is on rainfall patterns in the region, making it a dependable indicator for understanding and predicting precipitation in Indonesia.

#### Statistical Significant Testing:

We looked at how strongly the smoothed precipitation data (using 5-month and 7-month moving averages) correlates with the Oceanic Niño Index (ONI) for both India and Indonesia. In simple

terms, the p-value tells us whether the observed correlation is likely to be meaningful or just a coincidence. If the p-value is less than 0.05, it suggests that the correlation is probably real and not just by chance.

The statistical significance of correlations between the smoothed precipitation data (5-month and 7-month moving averages) and the Oceanic Nino Index (ONI) was assessed for India and Indonesia. In Indonesia, strong negative correlations were found:  $-0.77$  (5-month, p-value:  $5.37 \times 10^{-51}$ ) and  $-0.76$  (7-month, p-value:  $9.86 \times 10^{-49}$ ), showing a highly significant inverse relationship. In India, moderate correlations were observed:  $-0.50$  (5-month, p-value:  $1.93 \times 10^{-17}$ ) and  $-0.54$  (7-month, p-value:  $1.49 \times 10^{-20}$ ). From this we can say 7-month moving average has a better performance in relation with ENSO

#### 4.5.3 Analysis of Lagged Relationship Between ENSO and Precipitation

To further explore the relationship between ENSO and precipitation, a cross-correlation analysis was conducted using the 7-month moving average data for both India and Indonesia due to its correlation values. This analysis helps determine if there is any time-lag where the correlation between ENSO and precipitation is stronger, indicating a delayed response in precipitation to changes in ENSO conditions.

**India:** Figure 4.16 shows the cross-correlation between ENSO (ONI) and precipitation in India. The plot indicates that there is no significant lagged relationship, as the correlation is strongest at lag 0 (correlation of approximately  $-0.54$ ). This suggests that India responds almost immediately to changes in ENSO conditions, with no substantial delay in the impact.

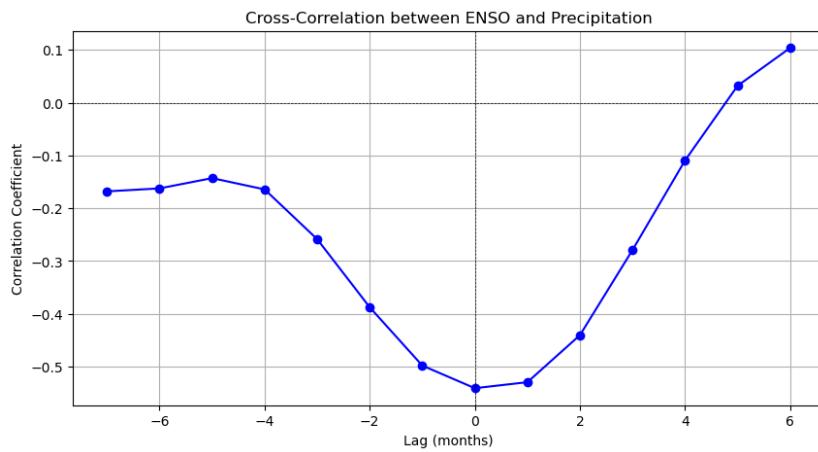


Figure 4.16: Lagged Correlation Trends between ENSO and India Precipitation Over Time

**Indonesia:** The second figure 4.17 illustrates the cross-correlation between ENSO (ONI) and precipitation in Indonesia. Although a minor increase in correlation is observed at a lag of 1 month, with a correlation coefficient of approximately  $-0.783$ , the best predictive relationship

remains at lag 0, where the correlation is only slightly lower (around -0.76 to -0.77). This suggests that ENSO's influence on rainfall in Indonesia is nearly immediate, with little practical benefit to considering the lagged relationship. Established research, such as Ropelewski and Halpert (1989), supports the conclusion that ENSO events tend to have their most pronounced effects on precipitation without significant delay. Therefore, while there is a slight increase in correlation at a 1-month lag, the normal (lag 0) relationship remains the most reliable and effective for predictive purposes.

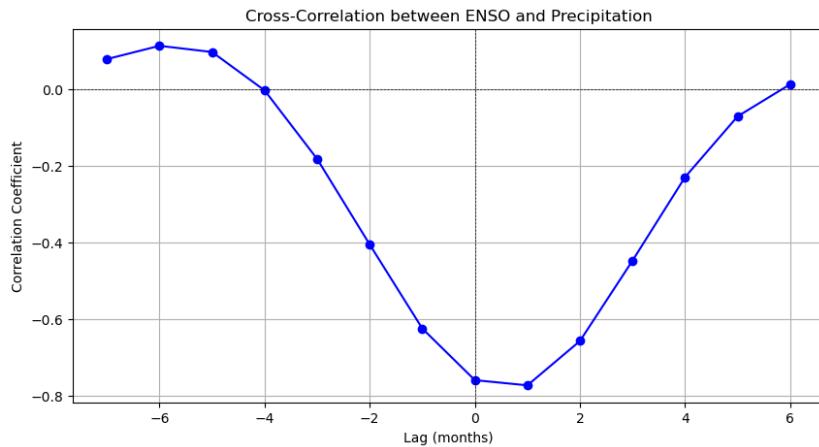


Figure 4.17: Lagged Correlation Trends between ENSO and Indonesia Precipitation Over Time

## 4.6 Seasonal Relationship Between Precipitation and ENSO

To better understand how ENSO influences rainfall in Indonesia and India, we analyzed data across four key meteorological seasons: DJF (December-February), MAM (March-May), JJA (June-August), and SON (September-November). This seasonal breakdown helps track the varying impacts of El Niño and La Niña throughout the year. Methodologically, we first applied a 7-month moving average to the precipitation data to smooth short-term fluctuations and highlight longer-term trends. We then segmented this smoothed data into the four seasons, allowing us to capture broader trends and more accurately assess how different ENSO phases impact seasonal rainfall patterns in these regions.

**Seasonal Correlation:** Table 4.3 highlights the correlation coefficients across different seasons for both India and Indonesia:

**DJF (December, January, February):** The correlation is higher for Indonesia (0.77) compared to India (0.55), indicating that ENSO's influence on precipitation is stronger in Indonesia during the wet season (DJF). For Indonesia, La Niña events tend to enhance rainfall, while El Niño suppresses it.

**MAM (March, April, May):** In this transition season, the correlation weakens slightly in both

Seasons	INDIA	INDONESIA
DJF	0.55	0.77
MAM	0.47	0.68
JJA	0.42	0.67
SON	0.61	0.81

*Table 4.3:* Seasonal Correlation between ENSO (ONI) and Precipitation in India and Indonesia Using the 7-Month Moving Average

regions. Indonesia still shows a stronger relationship (0.68) compared to India (0.47), suggesting that ENSO continues to affect precipitation but to a lesser degree as the seasons shift.

**JJA (June, July, August):** During the peak of the monsoon season in India, the correlation is lower (0.42), indicating that ENSO’s influence is present but less dominant. In Indonesia, which is experiencing its dry season, the correlation remains relatively stable at (0.67), highlighting ENSO’s continued impact, particularly during drought conditions brought by El Nino.

**SON (September, October, November):** ENSO’s influence increases again as the monsoon season retreats in India (0.61) and the wet season approaches in Indonesia (0.81). In Indonesia, this period marks the transition to the rainy season, where La Nina brings heavy rains and El Nino delays the onset of rainfall over Indonesia.

## 4.7 Conclusion

In this chapter, we explored the relationships between precipitation anomalies and key climate indices—ENSO (ONI), PDO, and IOD (DMI)—for India and Indonesia, applying different smoothing techniques across various seasons. Smoothing played a critical role in revealing these connections, with methods like the KZ filter and Gaussian smoothing demonstrating stronger correlations, particularly in Indonesia. Our analysis confirmed a strong negative correlation between ONI and precipitation in Indonesia (Wang et al. (2020)), consistent across all methods and seasons, emphasizing ENSO’s significant influence on Southeast Asian climates. In India, the correlation between ONI and precipitation was more moderate, aligning with the known impact of ENSO on the Indian monsoon (Roy et al. (2019)). We also found that seasonal correlations were notably stronger in Indonesia, suggesting that ENSO’s effects vary throughout the year. The weaker correlations observed with PDO and IOD are consistent with existing research, reinforcing their lesser influence on precipitation compared to ENSO (Mantua et al. (1997)). These findings highlight the importance of ENSO in regional rainfall variability and the need for careful selection of smoothing techniques to accurately uncover these patterns.

# Chapter 5

## Annual rainfall forecasting with SARIMAX

### 5.1 Introduction

Accurate rainfall prediction is crucial for sectors such as agriculture, water resource management, and disaster preparedness. Forecasts are needed on various timescales—ranging from predicting tomorrow’s weather to anticipating conditions over the next week, month, or even year. This is especially important in regions like India and Indonesia, where climatic extremes, such as droughts and floods, are common. Effective forecasting tools enable these sectors to prepare for and mitigate the impacts of such extremes, optimizing resource allocation.

To predict rainfall patterns over these timescales, particularly for the upcoming year, this chapter employs the Seasonal Autoregressive Integrated Moving Average with eXogenous variables (SARIMAX) model as it is well-suited for capturing complex seasonal patterns and external influences (Hyndman and Khandakar (2008)). In India, where rainfall patterns are relatively constant the model can achieve high predictive accuracy. In contrast, Indonesia’s precipitation is more varied in trend which makes the incorporation of El Niño Southern Oscillation (ENSO) data as an exogenous variable crucial for improving the model’s performance. This is due to the good relationship between ENSO and Indonesia precipitation, as shown in Chapter 4 of the correlation analysis.

This chapter will not only analyze and compare the SARIMAX model’s predictions focusing on forecasting monthly totals for the entire next year but also compare averaged precipitation of 20 years (climatology) values for both India and Indonesia, the previous year’s data to validate the model’s predictions for the following year. This approach provides a detailed assessment of the model’s effectiveness in forecasting annual rainfall totals.

## 5.2 Modeling

SARIMAX stands for Seasonal AutoRegressive Integrated Moving Average with eXogenous variables. It is an extension of the ARIMA model, which includes both seasonality and exogenous variables.

**Formula for SARIMA Model (Seasonality without Exogenous Variables):** This formula is used in prediction on India due to the seasonal pattern being smoother over the period. Here I have used SARIMAX Without Exogenous Variables, which works like a SARIMA model in SARIMAX.

$$\begin{aligned}y_t = & \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\& + \Phi_1 Y_{t-s} + \cdots + \Phi_P Y_{t-Ps} + \Theta_1 \epsilon_{t-s} + \cdots + \Theta_Q \epsilon_{t-Qs} \\& + \epsilon_t\end{aligned}$$

**Formula for SARIMAX Model (Seasonality with Exogenous Variables):** This formula is applied in predicting rainfall in Indonesia, where the ENSO variable is used as an exogenous factor. It is included due to the high fluctuations in the seasonal patterns, which the standard seasonal model alone cannot fully capture.

$$\begin{aligned}y_t = & \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \\& + \Phi_1 Y_{t-s} + \cdots + \Phi_P Y_{t-Ps} + \Theta_1 \epsilon_{t-s} + \cdots + \Theta_Q \epsilon_{t-Qs} \\& + \beta X_t + \epsilon_t\end{aligned}$$

### Explanation of both the formula:

$y_t$  is the observed value at time  $t$ .

$\phi_1, \dots, \phi_p$  are the non-seasonal autoregressive (AR) coefficients that model the relationship between current and previous values.

$\theta_1, \dots, \theta_q$  are the non-seasonal moving average (MA) coefficients that model the relationship between current and previous forecast errors.

$\Phi_1, \dots, \Phi_P$  and  $\Theta_1, \dots, \Theta_Q$  represent the seasonal autoregressive (SAR) and seasonal moving average (SMA) coefficients, capturing repeating seasonal patterns.

In the second formula,  $\beta X_t$  is the exogenous variable term, where  $X_t$  is the external predictor (ENSO in Indonesia) and  $\beta$  is its coefficient, used to adjust predictions for external influences.

### 5.2.1 India Model

Before applying the SARIMAX model, we have to do a preliminary check on whether the data is stationary or not, which means its statistical properties (mean, variance) remain constant over time. The Augmented Dickey-Fuller (ADF) Test is used to check for stationarity. A p-value below 0.05 confirms stationarity of our data has the p-value of 0.005287. If the data is non-

stationary, differencing (d) is applied to make it stationary before modelling.

**Diagnostic Plots:** To determine the appropriate order for the SARIMAX model, we use Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots, as shown in Figure 5.1. The ACF plot reveals correlations between the time series and its lagged values, which helps in identifying potential moving average (MA) components (denoted as q). The PACF plot, on the other hand, helps in isolating the direct effects of lagged terms, which is essential for identifying the AutoRegressive (AR) order (denoted as p) (Huffman et al. (2023)).

In setting the SARIMAX model parameters, p (the autoregressive order) which will be in range of 0 to 4 and for q (the moving average order) ranges same, later d (the number of differencing steps to make the series stationary) are carefully selected based on these diagnostic plots. From the ACF a seasonal model is noted by its fluctuation, the seasonal components (P, Q, D) are also determined, where D represents the seasonal differencing required to remove seasonal trends. The value of D is typically set to 1 if seasonal differencing is needed to stabilize the mean across seasons. The seasonal order of the model (P, Q) where it is also ranged from 0 to 4 is based on the seasonal spikes observed in the ACF and PACF plots, with the period of seasonality often set as 12 for monthly data to account for the annual cycle.

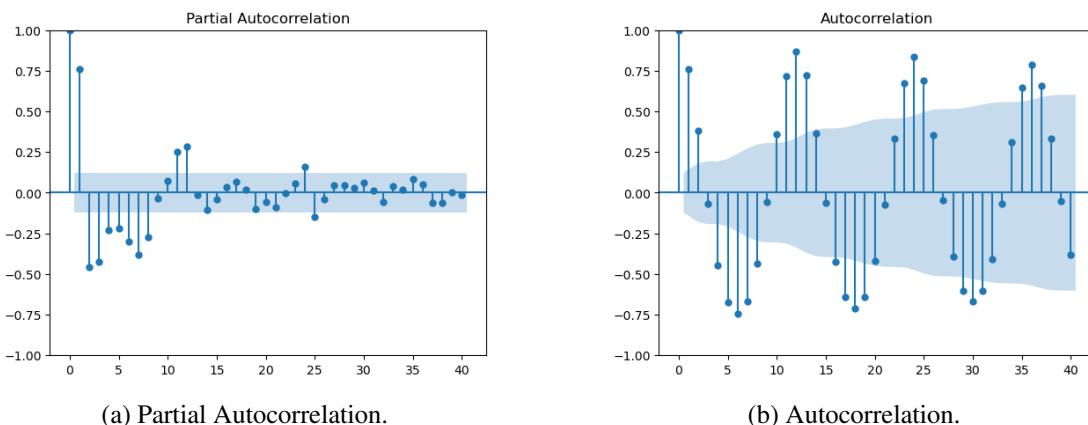


Figure 5.1: Partial Autocorrelation and Autocorrelation Plots for India.

We used the `auto_arima` function to automate model selection for each training dataset (Train =(2003-2020),(2003-2021),(2003-2022)), which was split year by year. The approach was as follows and gave the parameter value that has been visually picked from the diagnostic plot.

The best-fitting model was selected and is listed in Table 5.1 based on performance criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) which was explained by Hyndman and Khandakar (2008). These metrics help compare models. with lower values indicating better performance where we have chosen the lowest value by comparing with the other model using `auto_arima` (Smith et al. (2017–)).

TRAINING DATA SET	SARIMAX	AIC	BIC
Training set (2003–2020)	SARIMAX(1, 0, 0)(2, 1, 0)[12]	-762.237	-748.965
Training set (2003–2021)	SARIMAX(1, 0, 0)(2, 1, 0)[12]	-791.451	-777.950
Training set (2003–2022)	SARIMAX(1, 0, 0)(2, 1, 0)[12]	-833.194	-822.906

Table 5.1: Performance of SARIMAX models based on AIC and BIC for different training datasets for India.

### 5.2.2 Indonesia Model

Following the same process as with the Indian dataset where we first did Augmented Dickey-Fuller (ADF), it is used to check for stationarity where we get a p-value of 0.019765 which is stationary but it is close to 0.05 indicating a slight non-stationarity which implies variance in the data set.

Then, we initially used the `auto_arima` function to determine the best model parameters by analyzing the seasonality using autocorrelation (ACF) and partial autocorrelation (PACF) plots. These diagnostic tools highlighted the presence of seasonal patterns. However the model's performance in capturing the data was less optimal where the first three rows in table 5.2 use only SARIMAX without exogenous(SARIMAX).

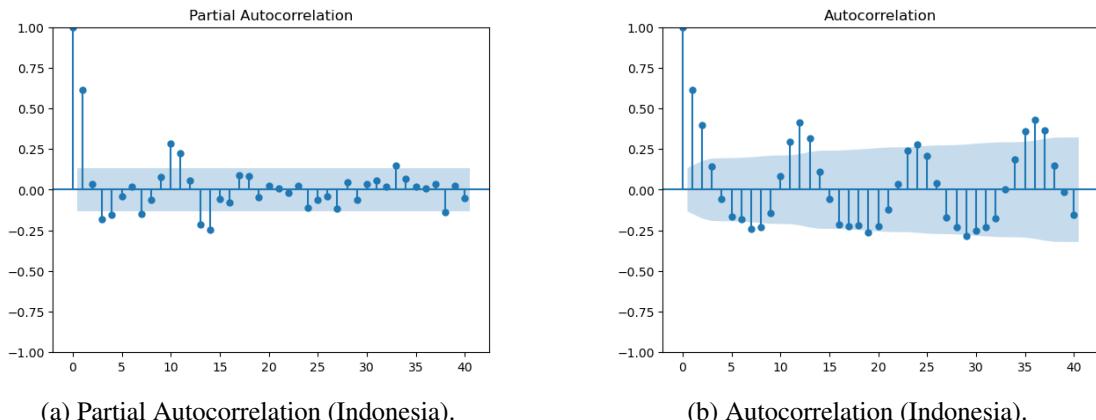


Figure 5.2: Autocorrelation and Partial Autocorrelation Plots for Indonesia.

To address this issue, we incorporated exogenous data related to ENSO (El Niño-Southern Oscillation), which is known to have significant impacts on climate and seasonal patterns in Indonesia from correlation. Including this exogenous variable helped resolve the performance issues observed in the initial model. As we can see from table 5.2, the performance of the exogenous model (SARIMAX) is better compared to the model that did not use exogenous variable (SARIMA).

The SARIMAX model that has been trained from 2003 to 2020 that includes, exogenous variables (ENSO data), has an AIC of -514.548 and a BIC of -489.971. These values are lower than those of the models without exogenous variables, indicating a better fit.

The SARIMAX model with parameters (2, 1, 1)(2, 1, 0)[12] and using data from 2003-2020

SARIMAX PARAMETER	TRAINING DATA SET	AIC	BIC
SARIMAX(2, 1, 1)(2, 1, 0)[12]	Training set (2003–2020)	-578.035	564.335
SARIMAX(0, 1, 1)(2, 1, 0)[12]	Training set (2003–2021)	-578.962	-564.480
SARIMAX(0, 1, 1)(2, 1, 0)[12]	Training set (2003–2022)	-547.962	-534.480
<b>SARIMAX(1, 1, 1)(2, 1, 1)[12]</b>	<b>(2003–2020 with Exogenous)</b>	<b>-514.545</b>	<b>-489.091</b>

Table 5.2: Model Comparison for Indonesia (Best Model in Bold, with Exogenous Variable).

has an AIC of -578.005 and a BIC of -564.835. The SARIMAX model using data from 2003–2022, without exogenous variables, shows an AIC of -547.982 and a BIC of -534.480, which is comparatively higher than the model that includes exogenous variables.

### 5.3 Comparison of Model Performance

We base our evaluation on the climatology of both regions, focusing on how well the predicted values align with the prior year's data. SARIMAX is used as the main model in this analysis. To assess its performance, we rely on three key metrics: R-squared ( $R^2$ ), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) which were explained by Polisetti and Ebenezer (2021) as to why we use these metrics .

**Climatology :** Climatology is the study of long-term weather patterns, usually averaged over 20 years. It helps define what the normal weather conditions are for a specific region. By comparing current weather data to these averages, we can see how current conditions differ from what is typically expected.

**Last year's observations:** Comparing predicted values to last year's observations allows us to assess how well the model can predict recent trends and short-term fluctuations in weather.

**The evaluation was conducted using the following metrics:**

**R-squared ( $R^2$ ):** This metric shows how well the model explains the variation in the data. A higher  $R^2$  value (closer to 1) indicates a better fit, while a lower  $R^2$  suggests that the model may require further tuning or additional predictors.

**Root Mean Square Error (RMSE):** Quantifies the average squared difference between predicted and actual values, emphasizing larger errors. A lower RMSE reflects greater accuracy, while a higher RMSE indicates a need for model improvement.

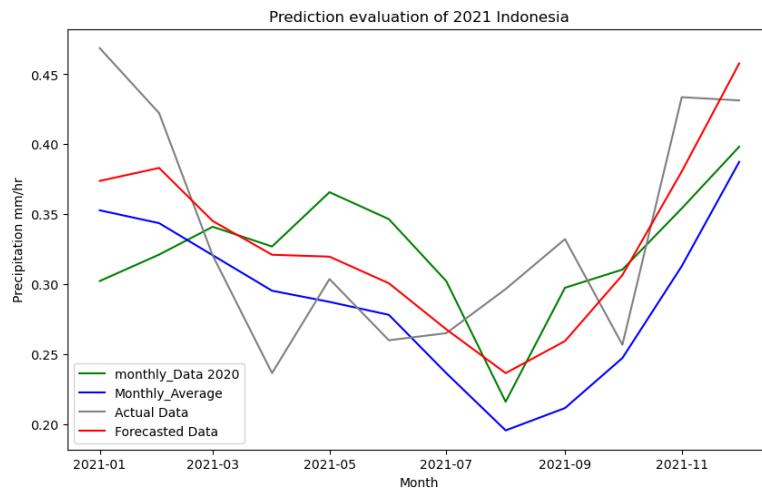
**Mean Absolute Percentage Error (MAPE):** Expresses errors as a percentage of actual values, making it easier to compare across models. Lower MAPE values indicate higher prediction accuracy.

### Prediction evaluation of 2021 for both regions:

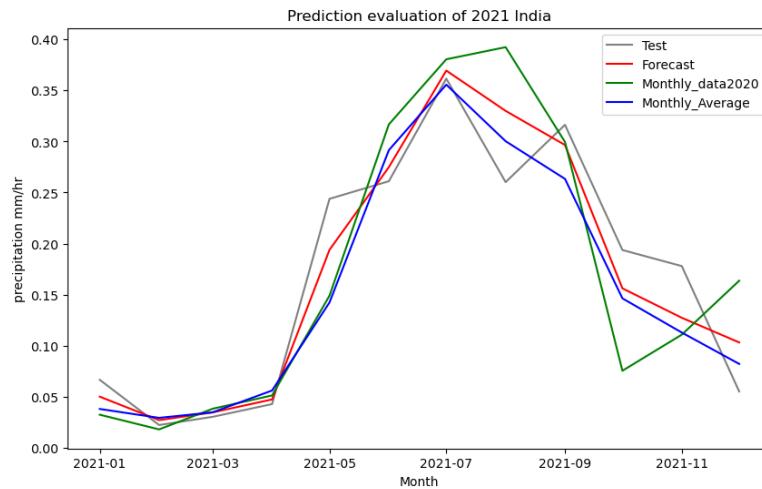
The two graphs in figure 5.3 display the SARIMAX model's performance in forecasting precipitation for Indonesia and India in 2021.

**Indonesia :** From the graph 5.3a, we can see all three predictors (SARIMAX model, Climatology, Prior year) are not able to access the pattern of testing data reliably, whereas the closest model that works better is the SARIMAX model with an exogenous (ENSO) variable.

**India:** As you can see from the graph 5.3b, all three predictors (SARIMAX model, Climatology, Proir year) are closely aligned with the test line. The SARIMAX model and Climatology (Monthly\_Average) perform well in predicting the test value .



(a) Prediction Evaluation for Indonesia (2021)



(b) Prediction Evaluation for India (2021)

Figure 5.3: Comparison of Prediction Evaluation for Indonesia and India for 2021.

From the table 5.3, we can see that overall, the SARIMAX model performs well in both regions compared to other predictors.

**Indonesia:** Here, the SARIMAX model with the ENSO variable performs decently well with  $R^2$  of 0.52 and MAPE of 0.42. The other predictors perform comparatively bad, indicating no dependence with  $R^2$  values, which are negative for test data (2021 year precipitation).

**India:** Here, SARIMAX model performs well with good  $R^2$  value of 0.85 and RMSE of 0.055 and MAPE with error rate of 0.29. In contrast to Indonesia, the other predictor also performs closely well with RMSE value for climatology (Monthly\_Average) at 0.04 and previous year data (2020) with  $R^2$  0.62.

PREDICTOR	R <sup>2</sup> (INDO)	RMSE(INDO)	MAPE(INDO)	R <sup>2</sup> (IND)	RMSE(IND)	MAPE(IND)
SARIMAX MODEL	0.52	0.038	0.424	0.85	0.055	0.29
MONTHLY_DATA(2020)	-2.43	0.118	0.459	0.62	0.071	0.44
MONTHLY AVERAGE	-0.76	0.064	0.227	0.78	0.04	0.26

Table 5.3: Comparison of Model Performance Metrics for Indonesia and India for 2021.

#### **Prediction evaluation of 2022 for both region:**

The two graphs in figure 5.4 illustrate how well the SARIMAX model predicted precipitation for Indonesia in 2022. Meanwhile, in India, the (monthly average) climatology performs well.

**Indonesia:** From the graph 5.4a, we can see the SARIMAX with ENSO model performs decently well, and the prior year has shown a close prediction for the test line.

**India:** Here in this graph 5.4b, we can see that both the climatology and SARIMAX models have performed presumably well.

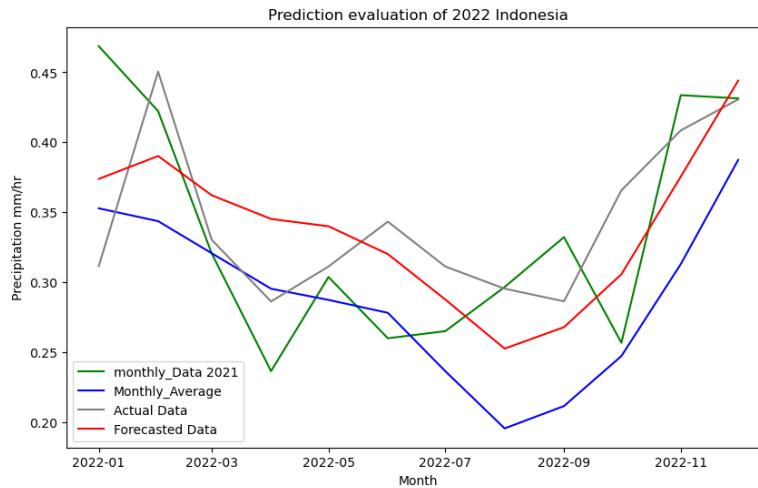
The table 5.4 shows that the SARIMAX model provided better predictions for 2022 in Indonesia. In India, the climatology performed surprisingly well.

**Indonesia:** In Indonesia, the main factor to note is that prior year data performs decently well, which partly matches the predictor data. With an  $R^2$  value of 0.29, which is less, but still, the pattern is quite similar to prediction data.

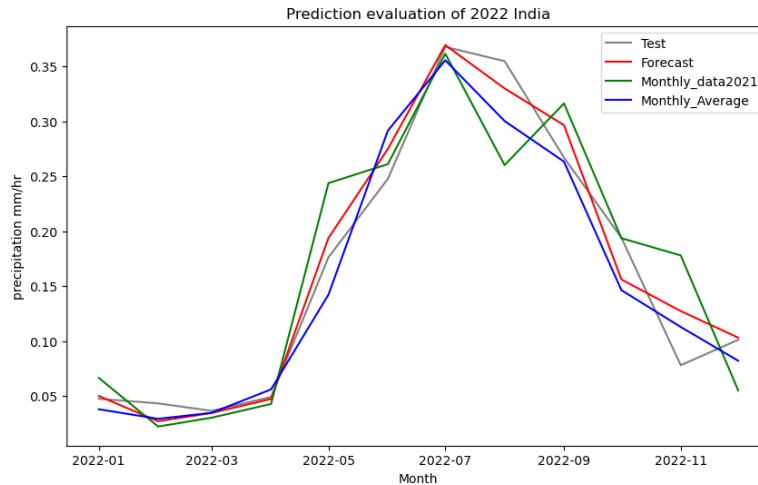
**India:** Here, even though the India SARIMAX model performs well with an  $R^2$  value of 0.89 compared to the climatology value, it has an  $R^2$  value of 0.87 where the climatology outperforms the other predictors with less RMSE of 0.02 and MAPE of 0.18.

PREDICTOR	R <sup>2</sup> (INDO)	RMSE(INDO)	MAPE(INDO)	R <sup>2</sup> (IND)	RMSE(IND)	MAPE(IND)
SARIMAX MODEL	0.42	0.038	0.467	0.89	0.04	0.25
MONTHLY_DATA(2021)	0.29	0.118	0.447	0.83	0.05	0.27
MONTHLY AVERAGE	-0.70	0.064	0.266	0.87	0.02	0.18

Table 5.4: Comparison of Model Performance Metrics for Indonesia and India for 2022.



(a) Prediction Evaluation for Indonesia (2022)



(b) Prediction Evaluation for India (2022)

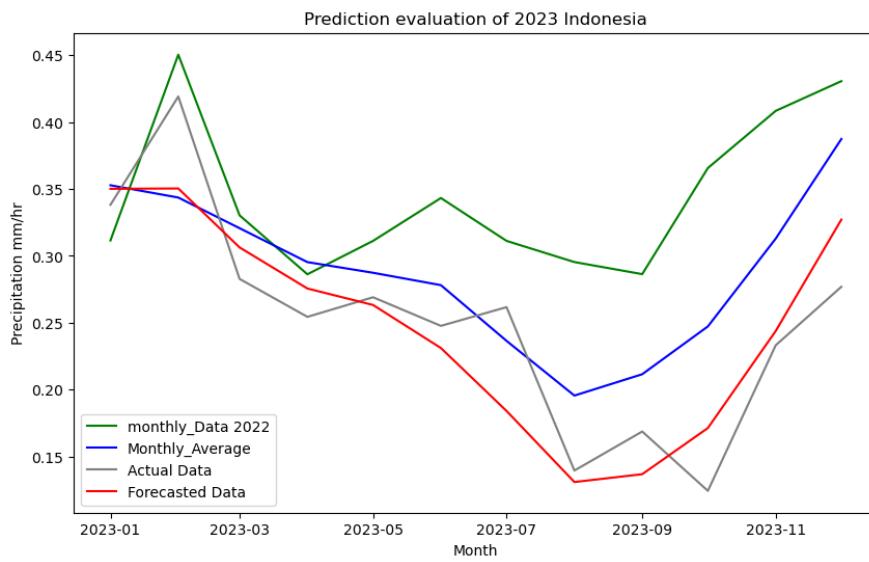
Figure 5.4: Comparison of Prediction Evaluations for Indonesia and India for 2022.

#### **Prediction evaluation of 2023 for both regions :**

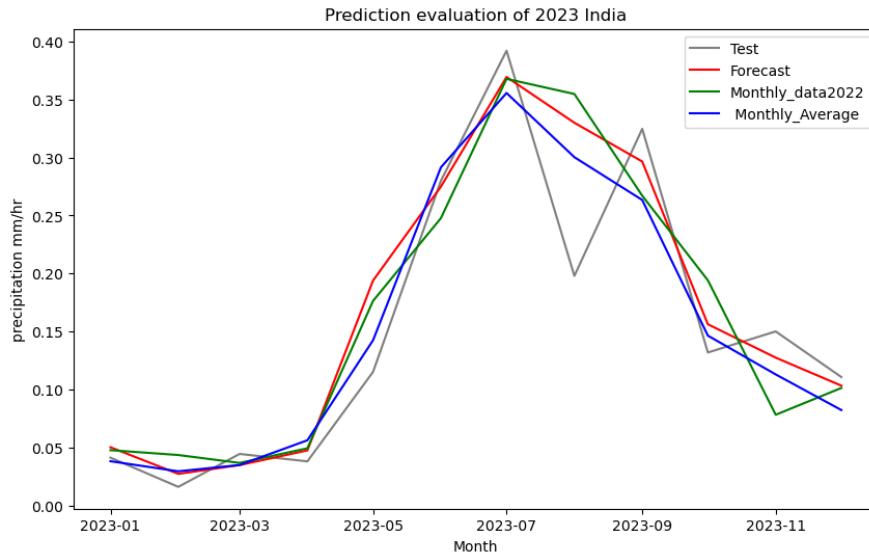
The graphs 5.5 show how the SARIMAX model predicted rainfall for Indonesia and India in 2023.

**Indonesia:** From the graph, we see that both the prior year and climatology follow the trend initially and at some point, the lines start to deviate from the test line. Meanwhile, the SARIMAX model with the ENSO variable performs really well in matching the test trend.

**India:** Here, the graph follows the same trend as in 2021, where the SARIMAX model performs well, while other predictors perform equally well.



(a) SARIMAX model predictions for Indonesia (2023)



(b) SARIMAX model predictions for India (2023)

*Figure 5.5:* Comparison of SARIMAX model predictions for Indonesia and India in 2023.

From the table 5.5, we can see that the SARIMAX model performed well in predicting rainfall for 2023.

**Indonesia:** The SARIMAX model with the ENSO variable performs really well compared to the other two predictions for 2021 and 2023 with a higher  $R^2$  value of 0.75 with a good error rate of MAPE (0.134) and RMSE (0.038). Meanwhile, the other predictors are still less predictive.

**India:** Here, climatology performs equivalently well with the SARIMAX model, where the SARIMAX model has a slight upper hand due to its  $R^2$  value of 0.88, whereas for climatology, it is around 0.83.

PREDICTOR	R <sup>2</sup> (INDO)	RMSE(INDO)	MAPE(INDO)	R <sup>2</sup> (IND)	RMSE(IND)	MAPE(IND)
SARIMAX MODEL	0.75	0.038	0.134	0.88	0.04	0.25
MONTHLY DATA(2020)	-3.71	0.118	0.520	0.70	0.06	0.29
MONTHLY AVERAGE	-0.31	0.064	0.266	0.83	0.03	0.27

Table 5.5: Comparison of Model Performance Metrics for Indonesia and India for (2023).

## 5.4 Overall Model Performance

The SARIMAX model generally outperforms the other models when analyzing monthly data from 2020 and average monthly data across various metrics. However, there is an exception in the 2022 predictions for India, where the Climatology (Monthly Average) model performed better than SARIMAX, achieving a Root Mean Square Error (RMSE) of 0.02 and a Mean Absolute Percentage Error (MAPE) of 0.18.

Overall, the data performance from the previous year is quite poor, suggesting a weak correlation for both regions when comparing data from the last year alone.

The monthly average model is more effective for India, while the performance of models for Indonesian data relies heavily on the ENSO value for enhanced results. When it comes to modeling, incorporating ENSO as an exogenous variable generally yields better outcomes for Indonesia. However, the results for India remain superior even without incorporating any exogenous variables, attributed to consistent seasonality throughout the period.

From the graphs, there is a noticeable difference in the performance of other predictors in Indonesia, whereas in India, the changes across predictors are relatively minor.

The SARIMAX model's superior performance indicates its effectiveness as a predictive tool for rainfall in both Indonesia and India. It not only captures the average conditions but also adapts well to the external influences that drive significant climate variations in these regions. This makes it an invaluable model for forecasters and climate scientists working in areas with complex, highly variable climate systems.

This detailed analysis provides a clearer understanding of how different models perform across varied metrics and regions, highlighting the importance of selecting appropriate models based on specific regional climatic characteristics and the type of prediction required.

# Chapter 6

## Conclusion and Discussion

### 6.1 Conclusion

In this study we explored the interesting relationships between regional rainfall patterns in India and Indonesia and key climate indices ENSO, PDO, and DMI. The study explores over two decades (2003-2023), employing various smoothing techniques to reveal correlations and applying the SARIMAX model for annual rainfall prediction.

#### **Influence of Climate Indices:**

**ENSO (ONI):** ENSO emerged as the most influential climate index affecting rainfall in both India and Indonesia. The impact was particularly dominant in Indonesia, with a strong negative correlation, confirming that ENSO plays a critical role in driving precipitation variability. In India, while the correlation was moderate, still ENSO significantly influenced the post-monsoon season.

**PDO and DMI:** These indices had a lesser effect on rainfall patterns, especially in India. While PDO showed some influence on long term trends in Indonesia, DMI's impact was minimal which suggests that ENSO remains the dominant force in these regions.

**Smoothing Techniques:** Here we applied various smoothing techniques, such as KZ filter, Gaussian smoothing, and moving averages, effectively revealing the underlying patterns in the rainfall data. Among them the 7 month moving average consistently provided the most reliable correlations with ENSO, particularly in Indonesia, where the effects of ENSO were more immediate and robust for all smoothing techniques. For India, it is moderately related with ENSO.

**Seasonal Variations:** The seasonal analysis showed that ENSO's influence varied throughout the year, with the strongest correlations observed during the SON (September-November) and DJF (December-February) seasons in Indonesia, where we have seen in EDA that the region has a strong monsoon period. In India, there is a good relation between SON (September-November) and ENSO, which indicates that ENSO has a relation with post monsoon period.

This highlights the importance of considering seasonal variability when checking climate indices and their impact on regional rainfall.

**Rainfall Prediction with SARIMAX:** The SARIMAX model proved to be an effective tool for predicting annual rainfall in both regions. In Indonesia, the inclusion of ENSO data significantly improved the model's accuracy, while in India, the model performed well even without exogenous inputs due to the consistent seasonality of the region's rainfall.

The comparison of SARIMAX with simpler models, such as climatology (20 year average) and previous year data, demonstrated the importance of SARIMAX in capturing both the average conditions and the external influences that drive significant climate variations. Whereas for India, climatology has a better performance equal to that of SARIMAX model.

From this study, we have an understanding of how global climate indices, particularly ENSO, influence regional rainfall patterns in India and Indonesia. It also helps in seasonal relations with ENSO and other regions where we have noted that for India, it is the post monsoon period, and for Indonesia, it is the main monsoon period. The findings specify the critical importance of incorporating these indices into predictive models for more accurate and reliable rainfall forecasts. Such models are important for sectors like agriculture, water resource management, and disaster preparedness, where accurate predictions can significantly mitigate the impacts of extreme weather events.

## 6.2 Discussion

In this study, we applied various smoothing techniques to the precipitation data with different filter sizes to better understand and reveal the underlying trends. The Gaussian filter particularly with a filter size of 3 has provided excellent smoothing results, effectively reducing noise while preserving important data trend. The filter size 5 also performed well by offering a balanced noise reduction in trend of data. However, filter size 7 provided a long term smoothing trend which tended to be unclear for short-term fluctuations making it more suitable for identifying broader and long-term trends rather than immediate and moderate changes where we could have tried a lesser filter size for observing better results.

The relationship between climate indices and precipitation revealed some interesting patterns, mainly in Indonesia where the correlation with the PDO index was moderately stronger than initially expected. This suggests that PDO could play a significant role in influencing rainfall in this region, and it presents a platform for future research.

Moreover, the comparison between DMI and PDO in relation to seasonal rainfall patterns indicates that these indices may have better seasonal relations than their normal relations, where the effects may tend to vary by season.

When it comes to modelling, the use of the SARIMAX model demonstrated strong predictive

capabilities for India and performs decently well when ENSO was included as an exogenous variable for Indonesia. It is essential to understand that the modelling could benefit from incorporating more extensive datasets and additional testing which would provide more robust evidence to determine which predictors work best across different metrics. where we have seen that the SARIMAX model generally outperformed other models. But the Climatology model, based on a 20-year average, performed exceptionally well for India in 2022 indicating the importance of considering different approaches depending on the region and the specific prediction needs.

In summary the findings presented here underline the importance of selecting appropriate smoothing techniques and models based on the specific characteristics of the region and the available data. While the SARIMAX model showed strong results in regions with normal seasonal patterns (India) but average for complex patterns (Indonesia), further research with more comprehensive data could refine these predictions and offer even greater accuracy for Indonesia by incorporating models like ETS (Exponential Smoothing State Space Model) where seasonal patters are inconsistent over time (like for Indonesia). Finally, we have seen the notable relationships observed between PDO and Indonesian rainfall suggesting that this area needs further investigation, potentially offering new insights into regional climate dynamics.

# Bibliography

Karumuri Ashok and NH Saji. On the impacts of enso and indian ocean dipole events on sub-regional indian summer monsoon rainfall. *Natural Hazards*, 42:273–285, 2007.

Chih-Pei Chang, Tim Li, and Song Yang. Seasonal prediction of boreal winter rainfall over the western maritime continent during enso. *Journal of Meteorological Research*, 34:294–303, 2020.

National Centers for Environmental Information. Pacific decadal oscillation (pdo). URL <https://www.ncei.noaa.gov/access/monitoring/pdo/>. 2024.

Rafael C Gonzalez. *Digital image processing*. Pearson education india, 2009.

Deborah E. Hanley, Mark A. Bourassa, James J. O'Brien, Shawn R. Smith, and Elizabeth R. Spade. A quantitative evaluation of enso indices. *Journal of Climate*, 16(8):1249 – 1258, 2003. doi: 10.1175/1520-0442(2003)16;1249:AQEOEI;2.0.CO;2.

George J Huffman, David T Bolvin, Eric J Nelkin, and Jackson Tan. Integrated multi-satellite retrievals for gpm (imerg) technical documentation. *Nasa/Gsfc Code*, 612(47):2019, 2015.

G.J. Huffman, E.F. Stocker, D.T. Bolvin, E.J. Nelkin, and Jackson Tan. Gpm imerg final precipitation l3 1 month 0.1 degree x 0.1 degree v07, 2023. URL <https://disc.gsfc.nasa.gov/datasets/GPM3IMERGM07/summary>. Accessed : 2024.

Rob J. Hyndman and Yeasmin Khandakar. Automatic time series forecasting: The forecast package for r. *Journal of Statistical Software*, 27(3):1–22, 2008. doi: 10.18637/jss.v027.i03. URL <https://www.jstatsoft.org/index.php/jss/article/view/v027i03>.

MR Jury. Representing the indian ocean dipole. *Physical Oceanography*, 29(4):417–432, 2022.

K Krishna Kumar, Balaji Rajagopalan, Martin Hoerling, Gary Bates, and Mark Cane. Unraveling the mystery of indian monsoon failure during el niño. *Science*, 314(5796):115–119, 2006.

NOAA Physical Sciences Laboratory. Download climate timeseries. URL  
<https://psl.noaa.gov/gcoswgsp/Timeseries/DMI/>. 2024.

Nathan J Mantua and Steven R Hare. The pacific decadal oscillation. *Journal of oceanography*, 58:35–44, 2002.

Nathan J Mantua, Steven R Hare, Yuan Zhang, John M Wallace, and Robert C Francis. A pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the american Meteorological Society*, 78(6):1069–1080, 1997.

Michael J. McPhaden, Stephen E. Zebiak, and Michael H. Glantz. Enso as an integrating concept in earth science. *Science*, 314(5806):1740–1745, 2006. doi: 10.1126/science.1132588. URL <https://www.science.org/doi/abs/10.1126/science.1132588>.

National Oceanic and Atmospheric Administration (NOAA). Oceanic niño index (oni) data, 2024. URL <https://www.cpc.ncep.noaa.gov/data/indices/>. Accessed:2024.

Kalpana Polisetty and Afrifa-Yamoah Ebenezer. An empirical study on rainfall patterns of monsoon season in the north-west india using time series models. *Journal of Statistics and Management Systems*, 24(3):559–572, 2021.

CF Ropelewski and MS Halpert. Precipitation patterns associated with the high index phase of the southern oscillation. *Journal of climate*, pages 268–284, 1989.

Indrani Roy, Renata G Tedeschi, and Matthew Collins. Enso teleconnections to the indian summer monsoon under changing climate. *International Journal of Climatology*, 39(6):3031–3042, 2019.

Divya Sardana, Prashant Kumar, Evan Weller, and Rajni. Seasonal extreme rainfall variability over india and its association with surface air temperature. *Theoretical and Applied Climatology*, 149(1):185–205, 2022.

Ashok Singh, Samir Thakur, and Nirab C Adhikary. Influence of climatic indices (amo, pdo, and enso) and temperature on rainfall in the northeast region of india. *SN Applied Sciences*, 2:1–15, 2020.

Taylor G. Smith et al. pmdarima: Arima estimators for Python, 2017–. URL <http://www.alkaline-ml.com/pmdarima>. [Online; accessed 1today].

Bin Wang, Xiao Luo, and Jian Liu. How robust is the asian precipitation–enso relationship during the industrial warming period (1901–2017)? *Journal of Climate*, 33(7):2779 – 2792, 2020. doi: 10.1175/JCLI-D-19-0630.1.

Wei Yang and Igor Zurbenko. Kolmogorov–zurbenko filters. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):340–351, 2010.

E Yulihastin, N Cholianawati, GA Nugroho, T Sinatra, and H Satyawardhana. Enso and pdo influence to climate variability in monsoon region of indonesia. In *IOP Conference Series: Earth and Environmental Science*, volume 166, page 012044. IOP Publishing, 2018.