NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

UNDERSTANDING THE DRIVERS OF EXTREME PRECIPITATION EVENTS IN PAKISTAN

by

Hassan Ilyas

December 2023

Thesis Advisor: Jefferson Huang
Second Reader: Mark A. Boothe

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** December 2023	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE UNDERSTANDING THE DRIVERS OF EXTREME PRECIPITATION EVENTS IN PAKISTAN			**5. FUNDING NUMBERS**
6. AUTHOR(S) Hassan Ilyas			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING /MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION /AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

Pakistan has been highly affected by the impacts of climate change, which have resulted in abnormal weather events. Over the past 14 years, Pakistan has witnessed two large-scale floodings (in 2010 and 2022) caused by abnormal precipitation events, which led to the loss of precious lives. The stark reality is that considering the geographical location of Pakistan, large-scale floodings caused by high precipitation events are likely to pose a threat in the future as well.

This study investigated the drivers for extreme precipitation events in Pakistan that led to these large-scale floodings. Data for 54 years of atmospheric variables from 1970–2023 were evaluated. The study selected a predictand region (28–34 N, 67–73 E) that had a wide representation of abnormal weather events in Pakistan, particularly the floods in 2010 and 2022. Subsequently, analysis of the atmospheric variables was done at long term means (LTM), including their composite anomalies, to select predictors linked to high precipitation. Later, correlation analysis between predictand and potential predictors led to our finalized set of predictors, including 850 hPa GPH over the Nepal and Red Sea at zero lead, SST over the South Central Indian Ocean and Southwest Indian Ocean (one month lead). Finally, statistical analysis using a logistic regression model was undertaken along with accuracy analysis using precipitation thresholds.

14. SUBJECT TERMS climate change, flooding, Pakistan, statistical analysis			**15. NUMBER OF PAGES** 105
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01-280-5500 Standard Form 298 (Rev. 2–89) Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

UNDERSTANDING THE DRIVERS OF EXTREME PRECIPITATION EVENTS IN PAKISTAN

Hassan Ilyas
Major, Pakistan Army
MMOAS, National Defense University, Islamabad, 2020

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
December 2023

Approved by: Jefferson Huang
Advisor

Mark A. Boothe
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Pakistan has been highly affected by the impacts of climate change, which have resulted in abnormal weather events. Over the past 14 years, Pakistan has witnessed two large-scale floodings (in 2010 and 2022) caused by abnormal precipitation events, which led to the loss of precious lives. The stark reality is that considering the geographical location of Pakistan, large-scale floodings caused by high precipitation events are likely to pose a threat in the future as well.

This study investigated the drivers for extreme precipitation events in Pakistan that led to these large-scale floodings. Data for 54 years of atmospheric variables from 1970–2023 were evaluated. The study selected a predictand region (28–34 N, 67–73 E) that had a wide representation of abnormal weather events in Pakistan, particularly the floods in 2010 and 2022. Subsequently, analysis of the atmospheric variables was done at long term means (LTM), including their composite anomalies, to select predictors linked to high precipitation. Later, correlation analysis between predictand and potential predictors led to our finalized set of predictors, including 850 hPa GPH over the Nepal and Red Sea at zero lead, SST over the South Central Indian Ocean and Southwest Indian Ocean (one month lead). Finally, statistical analysis using a logistic regression model was undertaken along with accuracy analysis using precipitation thresholds.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACE	Accumulated Cyclone Energy
AN	above normal
ANN	artificial neural networks
AUC	area under curve
BN	below normal
BOM	Bureau of Meteorology
BRNN	Bayesian regularization neural networks
CAF	Composite Analysis Forecast
CPC	Climate Prediction Center
DOD	Department of Defense
EB	European Blocking
ECMWF	European Centre for Medium-Range Weather Forecasts
EN	El Nino
ENLN	El Niño/La Niña
ENSO	El Niño-Southern Oscillation
ESRL	Earth System Research Laboratories
g/kg	grams per kilogram
GLOF	Glacier Lake Outburst Flood
GPH	Geopotential height
HOA	Horn of Africa
hPa	hecto-Pascal
IAHR	International Association for Hydro Environment Engineering and Research
IO	Indian Ocean
IOZM	Indian Ocean Zonal Mode
IRI	International Research Institute for Climate and Society
Jul–Aug	July–August

K2	Karakoram Two
LHFOL	latent heat fluxes over land
LN	La Nina
LRF	long-range forecast
LSCD	large-scale climate drivers
LSEF	large-scale environmental factors
LTM	long-term mean
m	meter
m/s	meters per second
MEI	Multivariate ENSO Index
mm/day	millimeters per day
MRA	multiple regression analysis
NAO	North Atlantic Oscillation
NCAR	National Center for Atmospheric Research
NCDC	National Climatic Data Center
NCEI	National Centers for Environmental Information
NCEP	National Centers for Environmental Prediction
NN	near normal
NOAA	National Oceanographic and Atmospheric Administration
NWS	National Weather Service
OLR	outgoing longwave radiation
Pa/s	Pascals per second
PCA	principal component analysis
PR	precipitation rate
PRA	precipitation rate anomaly
PSL	Physical Sciences Laboratory
SAH	South Asian high
SD	standard deviation
SHFOL	sensible heat fluxes over land

SLP	sea level pressure
SSH	surface-specific humidity
SST	sea surface temperature
SU	surface zonal velocity
SWA	Southwest Asia
TA850	850 hPa air temperature
TC	tropical cyclone
U500	500 hPa zonal velocity
UKMO	United Kingdom Meteorological Office
UN	United Nations
W/m^2	watts per square meter
WNP	Western North Pacific
Z850	850 hPa geopotential height

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

This thesis delves into the investigation of the drivers behind extreme precipitation events in Pakistan, a country profoundly impacted by climate change-induced abnormal weather events. The catalysts for this research are the devastating large-scale floods experienced in 2010 and 2022, resulting in significant loss of lives. The study aims to comprehensively understand and analyze the factors contributing to these extreme precipitation events.

The initial chapter sets the stage by providing an overview of the catastrophic floods in 2010 and 2022, emphasizing the severe implications on human lives and infrastructure. This chapter serves as a backdrop for the subsequent exploration of the underlying causes. The second chapter delves into existing studies conducted post the 2010 and 2022 floods in Pakistan and some other studies related to long-range forecasting for different other regions of the world. It synthesizes insights from various studies to build a comprehensive understanding of the complex interplay of factors leading to extreme precipitation events.

The third and fourth chapters outline the meticulous process of data collection and analysis followed by the results. Spanning 54 years (1970–2023), the study focuses on different atmospheric variables by using various advanced weather data sets primarily from NOAA. A specific predictand region (28–34 N, 67–73 E) was selected, showcasing a robust representation of abnormal weather events, particularly the floods in 2010 and 2022. The analysis involved long term means (LTM) and composite anomalies to identify predictors linked to high precipitation. Predictors were finalized using results obtained from the correlation analysis between the predictand and potential predictors. The study culminates in a finalized set of predictors, including 850 hPa geopotential height (GPH) which measures actual height of pressure surface above mean sea level over Nepal and the Red Sea at zero lead, Sea Surface Temperature (SST) over the south-central Indian Ocean at zero lead, and the Southwest Indian Ocean at a one-month lead. Logistic regression modeling and accuracy analysis revealed a notable 82% accuracy for precipitation thresholds at 3 mm/day.

The concluding chapter synthesizes the key findings and insights garnered throughout the study. It offers a concise summary of the identified predictors and their role in extreme precipitation events. The thesis concludes with a reflection on the study's significance, implications, and potential avenues for future research.

In essence, this research advances our understanding of extreme precipitation events in Pakistan, providing valuable insights for policymakers, climate scientists, and disaster management authorities. The study's nuanced approach and robust methodology contribute to the broader discourse on climate change impacts and adaptation strategies.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to the Institute of Pakistan Army for providing me with the invaluable opportunity to pursue advanced studies at the Naval Postgraduate School. The unwavering support and encouragement from the Pakistan Army played a pivotal role in shaping my academic journey.

I am profoundly grateful to my advisor, Prof. Jefferson Huang, for his guidance, mentorship, and expertise. His insightful feedback, encouragement, and commitment to excellence have significantly enriched my research and academic experience. I would also like to express my sincere thanks to my second reader, Mark Boothe, for his valuable contributions, constructive feedback, and insightful perspectives that have enhanced the depth and quality of my work. He was always available in case of any queries, which was a source of encouragement. I would also thank Prof. Tom Murphree, whose useful mentorship was valuable in keeping myself on the right track. I am also highly grateful to the National Oceanographic and Atmospheric Administration for providing requisite atmospheric datasets and tools to conduct the study.

I extend my deepest appreciation to my wife, Taliya Masood, whose unwavering support, understanding, and encouragement have been my pillars of strength throughout this academic endeavor. Her sacrifices and belief in my capabilities have been instrumental in overcoming challenges and achieving milestones.

My heartfelt thanks go to my parents, Muhammad Ilyas and Khalida Parveen, for their unending prayers, love, and support. Their belief in my abilities has been a driving force, and I am deeply appreciative of the foundation they provided for my education and personal growth.

This journey would not have been possible without the collective support of these individuals and institutions. I am grateful for the opportunities, guidance, and encouragement that have shaped my academic and personal development.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

This chapter gives an overview of the broad weather system of Pakistan, including its climatology, to provide a basic understanding of the weather system in the region and what are some major climate patterns.

A. CLIMATE CHANGE AND ITS GLOBAL IMPACTS

In recent years, the world has witnessed an unprecedented transformation of the global environment, which has primarily been driven by the persistent release of greenhouse gases into the atmosphere of Earth. These emissions, originating from a wide array of human activities, are the catalysts propelling us into an era of climatic change characterized by the acceleration of unprecedented phenomena across the atmosphere, oceans, cryosphere, and biosphere. This unrelenting anthropogenic influence has resulted in a trajectory that is reshaping the very fabric of our planet. With global surface temperatures exceeding 1.1ºC above pre-industrial levels during the period from 2011–2020, the reality of global warming has moved beyond conjecture to unequivocal fact (Calvin et al. 2023). Global greenhouse gas emissions, as projected by nationally determined contributions (NDCs) announced in Climate Change Report 2023 of October 2021, threaten to breach the 1.5ºC threshold during the 21st century, thus making the daunting task of limiting global warming to below 2ºC even more challenging (Calvin et al. 2023).

Because of this global transformation, climate change-induced alterations of weather have cascaded to impact every corner of the Earth. The burgeoning impacts, evident in the intensification of weather and climate extremes, have manifested in both natural ecosystems and human societies alike. The consequences of global warming are not just limited to climbing temperatures, but are more complex. This evolving phenomenon is producing heat, which is melting glaciers as well as sea ice, shifting the pattern of precipitation, and setting animals on the move. Moreover, use of terms like “global warming” and “climate change” are generally interpreted as identical or synonymous by the people; however, scientists like to use the term “climate change” when explaining the multifaceted changes affecting climate systems and weather of our planet. The term climate change is not limited to rising average temperatures but also includes abnormal or extreme weather events, increasing sea temperatures, and many other significant factors (National Geographic 2019). Climate change is thought to have played a substantial role in the floods that occurred in Pakistan in both 2010 and 2022.

B. RECENT HISTORY OF FLOODING IN PAKISTAN

The weight of this climate crisis is not borne evenly across the globe. Vulnerable communities, who historically have had the least contribution to the rise in global temperatures, find themselves disproportionately affected by the consequences of these activities (Joseph Hun-wei Lee 2022). With the bigger picture of climate change resulting in abnormal weather events in view, Pakistan has witnessed two massive floods during the last 13 years during which thousands of people lost their lives, while millions of others were affected or displaced by these large-scale catastrophic events. The floods in Pakistan during 2010 and 2022 stand as somber reminders of the country's vulnerability to extreme weather events. In 2010, Pakistan experienced the most fatality intensive floods in its history, when about 2,000 people lost their lives while millions were displaced. Moreover, vast tracts of agricultural land were inundated, while immense human suffering was seen due to various health concerns. Similarly, the floodwaters of 2022 were the most widespread in the history of the country and wreaked havoc across the nation, submerging communities, destroying infrastructure, and disrupting livelihoods. At one time about one third of the country was affected by flood in 2022 while around 1,700 people lost their lives. These catastrophic events serve as stark illustrations of the profound challenges posed by climate change and the urgent need for comprehensive disaster preparedness and mitigation strategies in Pakistan, a country that remains all too susceptible to the impacts of a changing climate.

It is believed by weather analysts that climate change had a substantial role in the 2010 and 2022 floods of Pakistan. Pakistan is listed amongst the top eight countries being affected by long-term effects of climate change (Zaidi 2022). The comparison of the damages caused by the floods in Pakistan in 2010 and 2022 (see Figure 1) underscores the evolving and increasingly severe nature of the country's vulnerability to climate-related disasters. Figure 1 illustrates the extent of flood-affected areas of Pakistan in 2010 with green (left) and 2022 with light blue color (right). As per different analysis carried out, 2010 floods affected about 20% of the total land area and 2022 floods affected almost 33% of the total land area of Pakistan. The 2010 floods, often termed as the worst in Pakistan's history, not only resulted in a staggering loss of life, economic devastation, and displacement of millions of people but also damaged infrastructure, disrupted agriculture, and brought significant humanitarian challenges. In contrast, the 2022 floods, though lesser than the 2010 disaster in terms of human losses, still inflicted substantial harm on communities and critical sectors. They caused widespread inundation, damage to homes, and interruptions in vital services, demonstrating that Pakistan's susceptibility to such events has not diminished. According to various statistics, the impacted regions witnessed 900mm of rainfall from June to August 2022, marking an increase of nearly 350 percent compared to the long-term average (Tillman 2023).

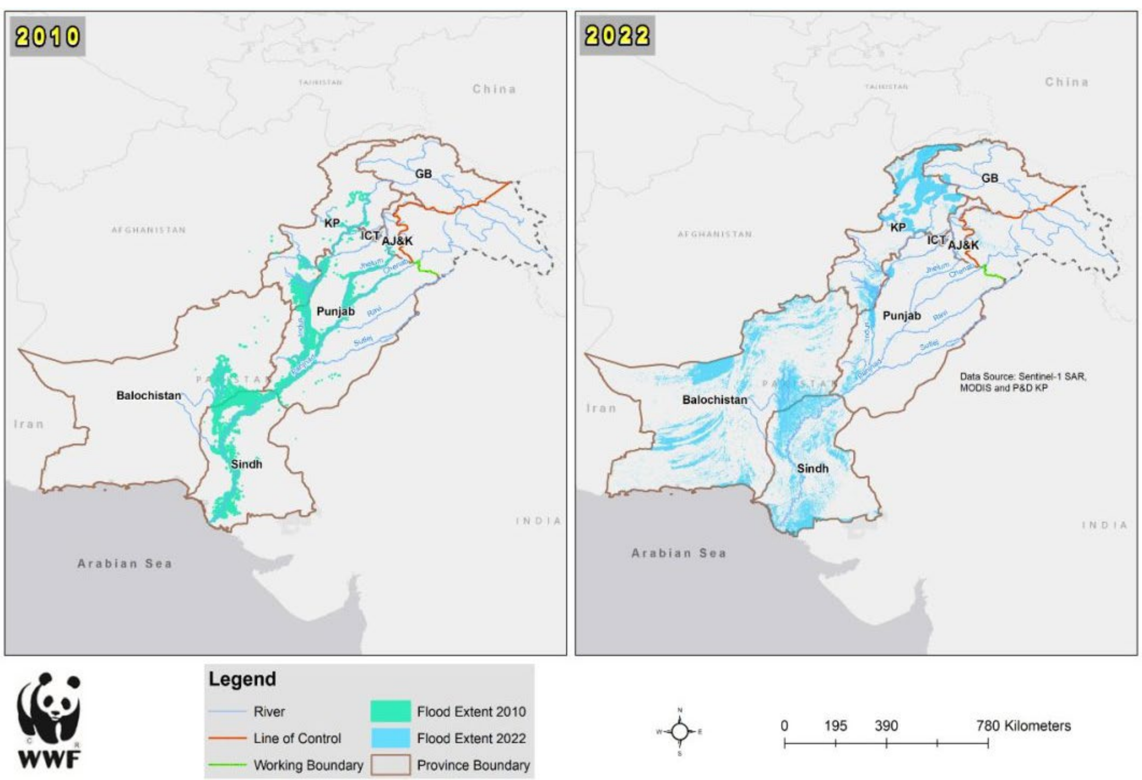

Figure 1. Flood affected areas. Source: Zaidi (2022).

C. MOTIVATION

These contrasting but interconnected events not only highlight the growing imperative of building resilience, improving disaster response, and addressing climate change but also call for better long-term forecasting to mitigate the impacts of future floods in Pakistan. These catastrophic events serve as stark illustrations of the profound challenges posed by climate change and the urgent need for comprehensive disaster preparedness and mitigation strategies in Pakistan. The primary motivating factor for the study was to understand the drivers of extreme precipitation events in Pakistan and to create a model to forecast extreme precipitation events.

1. Significance for the World

During September 2022, Secretary-General of United Nations Mr. António Guterres visited Pakistan to express support for the populace in the aftermath of extensive floods, characterizing the situation as a "climate catastrophe." He also asserted that "No country deserves this fate, but particularly not countries like Pakistan that have done almost nothing to contribute to global warming" (United Nations News 2023). There has been lots of discussion around the world regarding how climate change is impacting extreme weather events. A study by the team from International Association for Hydro Environment Engineering and Research (IAHR) examined 60 days of summer rainfall patterns in 2022 across Pakistan and the heaviest five-day period in Baluchistan and Sindh (2 provinces which witnessed maximum deviation from the average rainfall). The results indicate that climate change likely amplified the intensity of the five-day rainfall by as much as 75 percent, and it enhanced the intensity of the 60-day rain period by 50 percent. Furthermore, the likelihood of such substantial rainfall, referred to as a "1-in-100-year rain event," stands at approximately 1 percent today. However, in a scenario without climate change, this occurrence would have been even less probable (Joseph Hun-wei Lee 2022). Overall, the study is significant for the world in terms of understanding the factors behind abnormal weather events in any part of the world.

2. Significance for Pakistan

The unparalleled flooding in 2022, set off by intense monsoon rains, submerged approximately one-third of the nation and led to approximately 1,700 fatalities (United Nations News 2023). According to the UN Office in Pakistan, over 33 million individuals (about one in every seven Pakistani) experienced the impact of flooding, while almost eight million faced displacement and about 13,000 were reported injured. The floodwaters led to the loss of approximately one million livestock animals and inflicted damage to 2.2 million houses and about 4.4 million acres of agricultural land (United Nations News 2023). Vital infrastructure, such as schools, hospitals, roads, bridges, water and sanitation facilities, and government buildings, were left in shambles. If there are long-range forecasting mechanisms which are well developed and which incorporate the leading meteorological departments of the world, then there is all the likelihood that extent of damage can be reduced by timely sharing of the information about any unusual weather patterns.

D. OVERVIEW OF PAKISTAN'S CLIMATE

A brief overview in terms of climate of Pakistan and its contributing factors is given below.

1. Geography of Pakistan

Pakistan's geography is diverse, featuring a rich tapestry of landscapes that encompass a wide range of terrains and ecosystems. Located in South Asia, Pakistan lies between 23 degrees 35 minutes to 37 degrees 05 minutes north latitude and 60 degrees 50 minutes to 77 degrees 50 minutes east longitude (American Institute of Pakistan Studies 2023). Pakistan has an area of 796,096 sq. km. and shares its borders with India to the east, Afghanistan and Iran to the west, China to the north, and a lengthy coastline along the Arabian Sea to the south. The country's topographical diversity ranges from the towering peaks of the Himalayas in the north, including the world's second-highest mountain, K2, to the vast and arid plains of the Indus River basin, which dominate the country's center. Pakistan's western regions encompass rugged mountain ranges like the Hindu Kush and the Sulaiman Mountains, which give way to the vast deserts of

Balochistan. The country's southern coastline is characterized by the warm waters of the Arabian Sea. Figure 2 illustrates the considerable variations in elevation, ranging from the lofty peaks of the Karakoram and Hindu Kush mountains in the north to the Thar Desert in the southern side.

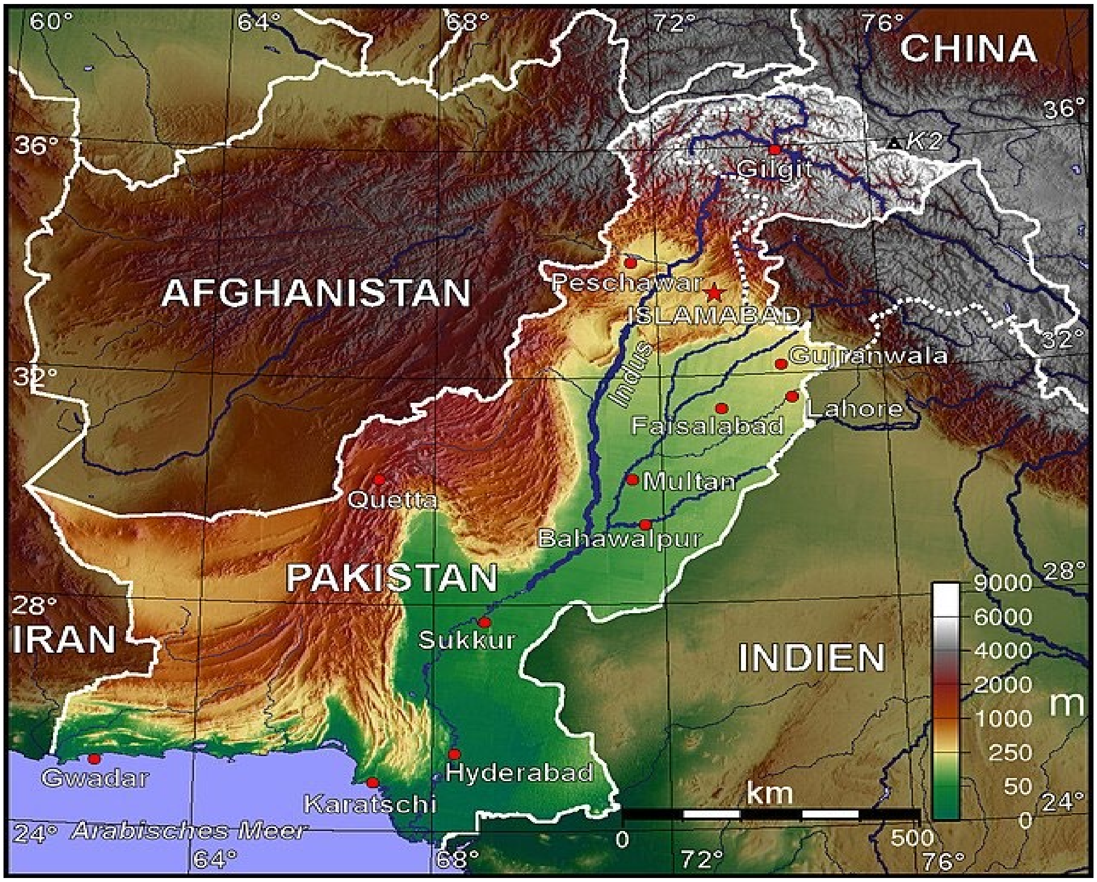

Figure 2. Physical relief map of Pakistan and its neighboring countries.
Source: Wikipedia (2019).

2. Climate Trends of Pakistan

Pakistan's climate is as varied as its geography, exhibiting a range of climatic zones from arid deserts to alpine snow-capped peaks. Pakistan's position in South Asia, with its extensive coastline along the Arabian Sea to the south and its proximity to high mountain ranges to the north, greatly affects its climate. The Himalayas and the Karakoram Range act as significant barriers, blocking the moisture-laden monsoon winds from entering certain regions. The country experiences four distinct seasons: a scorching summer, a mild and pleasant autumn, a relatively cold and dry winter, and a warm spring. The northern regions, including areas of Azad Kashmir and Gilgit-Baltistan, feature a subarctic and alpine climate, with harsh winters and mild summers. The central and southern plains, especially in Punjab and Sindh, endure sweltering summers with temperatures exceeding 40°C (104°F) while experiencing a wide temperature range throughout the year. Pakistan's monsoon season, which typically runs from June to September, brings heavy rainfall, especially to the eastern and northern regions, and can lead to flooding. In contrast, the southwestern province of Balochistan remains arid, with minimal rainfall and high temperatures. The country's climatic diversity, with its associated challenges and opportunities, plays a pivotal role in influencing agriculture, water resources, and overall living conditions for its population. Pakistan's varied topography, from high mountain ranges to low-lying plains, influences temperature and precipitation patterns. Highland areas in the north experience lower temperatures, while the Indus River plains in the central and southern regions tend to be hot and arid.

3. Key Contributors to Climate of Pakistan

The climate of Pakistan is influenced by a variety of factors, some of the significant ones are listed below. Understanding and monitoring these key causes and factors are essential for managing the challenges and opportunities associated with Pakistan's diverse climate.

a. Summer Monsoon

The monsoon is a pivotal climatic event in Pakistan, spanning from June to September. During this period, Pakistan experiences a pleasant change from the sweltering summer heat. Monsoon rains are known for their substantial intensity, which can sometimes result in significant flooding, especially when they interact with westerly pattern waves in the northern regions. The monsoon plays a crucial role in replenishing water resources and sustaining agriculture, making it a vital aspect of Pakistan's climate.

b. Western Disturbance

Western Disturbances, primarily active during the winter months, are a significant climate-altering factor in Pakistan. They bring light to moderate precipitation in the southern regions of Pakistan, while the northern areas witness moderate to heavy rainfall, often accompanied by heavy snowfall. These westerly waves tend to lose moisture as they traverse towards Pakistan, impacting the winter weather patterns and contributing to the country's water resources.

c. Fog

The winter season in Pakistan is marked by persistent fog, which can last for several weeks in upper parts of the country including Northern Sindh, central Khyber Pakhtunkhwa, and Punjab provinces.

d. Tropical Storms

Tropical storms, typically forming during end April to June and again from end September to November, primarily affect coastal areas of Pakistan. These heavy storms can often cause infrastructure damage.

e. Dust Storms

Violent dust storms occur during the summer season, with their intensity higher during the months of May and June. These dust storms often signal the arrival of the monsoons or the onset of winter in the autumn.

f. Continental Air

Air masses of continental origin form above landmasses, hence they are characterized by dry conditions. (Wikipedia 2023a). When continental air masses dominate the climate, the country experiences no precipitation.

Understanding and monitoring these key causes and factors are essential for managing the challenges and opportunities associated with Pakistan's diverse climate.

E. IMPORTANT CONCLUSIONS FROM CLIMATE OF PAKISTAN

To have some basic insight about this study on drivers of extreme precipitation events, the focus precipitation months were identified first by studying the month wise precipitation record of central Pakistan from 1970–2023 (Figure 3). The results were in line with the general concept that the summer monsoon during the three-month Jul-Sep period brings the highest amount of rainfall in Pakistan, with pronounced rainfall especially during Jul–Aug. Precipitation Rate (PR) during the months of Jul–Aug is the highest during the entire year both in terms of concentration as well as standard deviation (SD). The high amount of PR in Jul–Aug along with the variability makes this an important variable for prediction at long lead times. With these results in consideration, Jul– Aug PR in Pakistan was chosen as our primary target timeframe, or predictand time, for this study. Data for PR was obtained from a data set of global reanalysis (NOAA Monthly Mean Timeseries 2023).

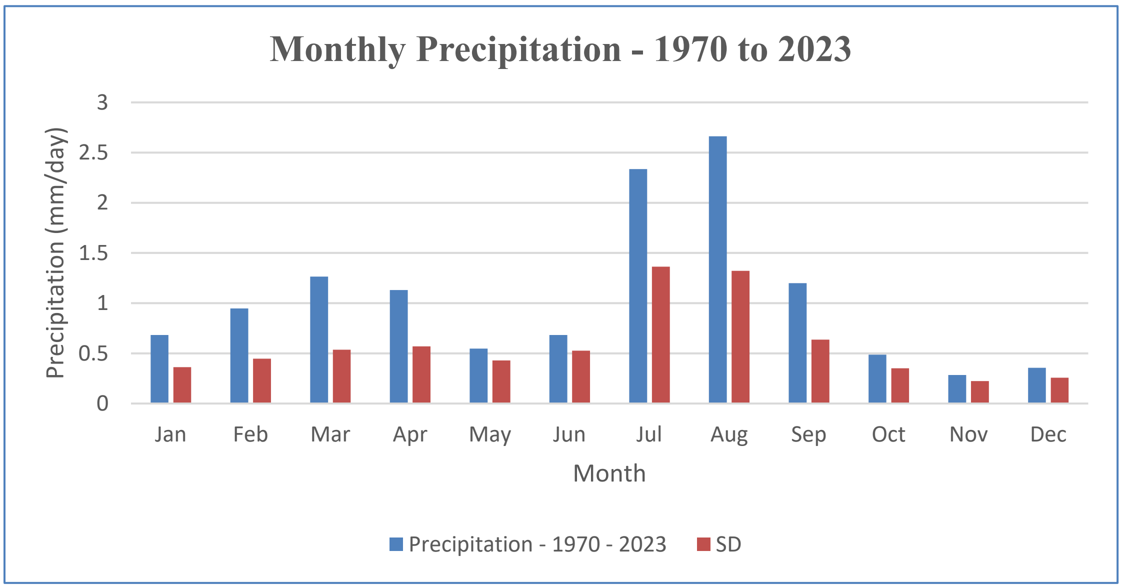

Figure 3. Month wise precipitation rate (PR) from 1970 to 2023 in mm/day (blue bars) and standard deviation (SD) of the PR in mm/day (red bars) for region of central Pakistan (28o–34oN, 67o–73oE). Adapted from NOAA Monthly Mean Timeseries (2023).

After identifying Jul–Aug as focus precipitation months (Figure 3), spatial distribution of PR was studied in Figure 4 during the focus period of Jul–Aug. It is evident that Pakistan separates the South Asian region (highly wet) from the Southwest Asia (SWA) region (mostly arid). We also see that although Jul–Aug is the wettest time of the year for most of Pakistan, this is particularly true for Northern Pakistan which also encompasses the south-facing mountains. Pakistan precipitation is: (1) highest in northern Pakistan with its high mountains; and (2) lesser PR than south Asian countries to the east but higher than countries in Southwest Asia (SWA). We can also conclude that the areas encompassing highest PR totals are generally along the foothills of these mountainous areas.

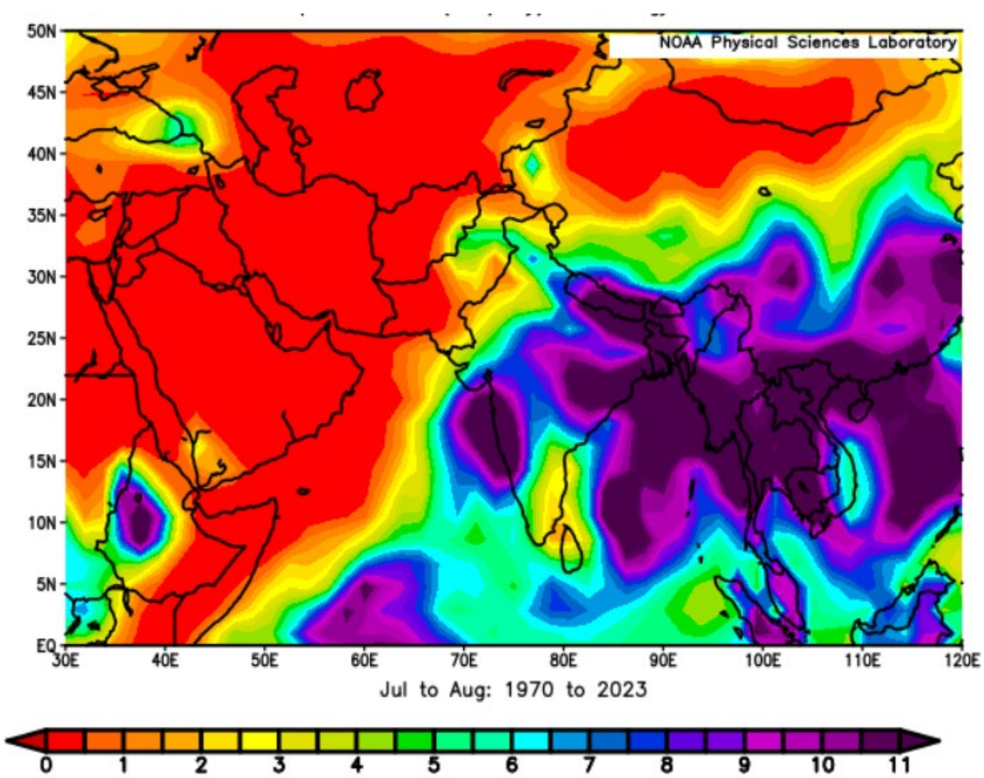

Figure 4. Long-Term Mean (LTM) composite of Jul–Aug surface PR (mm/day) for the period from 1970 – 2023. Adapted from NOAA Monthly/Seasonal Composites (2023).

As identified by DeHart (2011) as well, the arrival of moisture into Pakistan during the months of Jul–Aug is primarily because of the southwest monsoon. Figure 5a, depicted below, provides a visual representation of the circulation patterns in the lower troposphere responsible for driving this influx of moisture. In this system, the low-level pressure prevailing over South Asia forces the onshore flow of humid air from the Indian Ocean, ultimately leading to substantial rainfall. Across a vast expanse, stretching from Southwest Asia (SWA) to the southeastern regions of China, a broad region of low atmospheric pressure is discernible. Meanwhile, Figure 5b illustrates the wind vectors at the surface, offering a comprehensive portrayal of the resultant airflow patterns. Notably, a prominent feature within this system is the Somali Jet, which flows toward the northeast over the northwestern Indian Ocean (IO) and the Arabian Sea. The northern branch of the Somali Jet, represented by the blue arrow, transports warm, moist air into Pakistan.

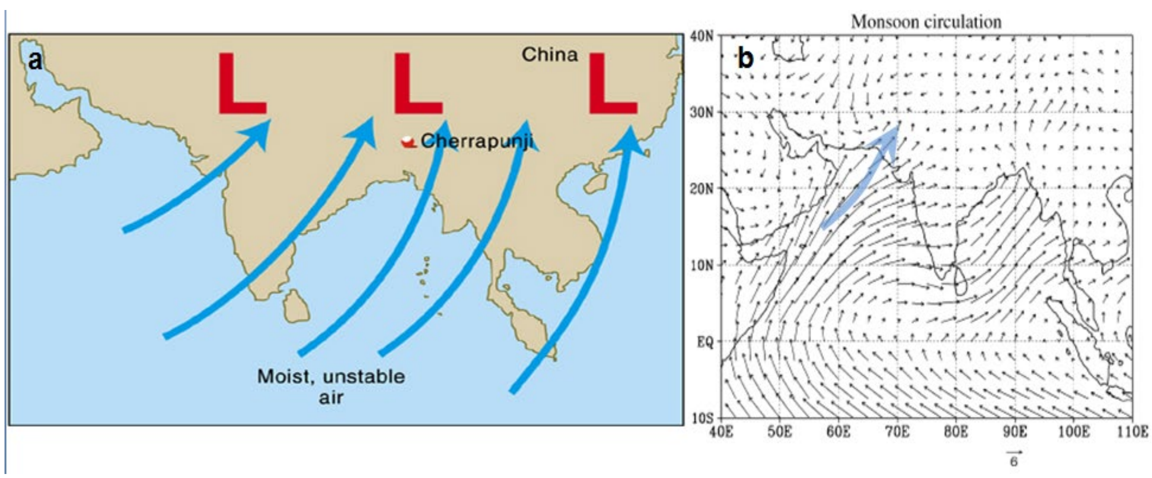

Figure 5. (a) Schematic low-level circulation from the summer southwest monsoon. Source: Danielson et al. (2003). (b) Surface vector winds (m/s) for June-September. Source: Kripalani et al. (2007).

Another very significant thing to analyze is teleconnections. Teleconnections represent noteworthy associations or connections between meteorological occurrences in geographically distant regions on Earth, typically encompassing climate patterns that extend over extensive distances. For instance, an elevation in atmospheric pressure in one area leads to a corresponding decline in pressure in a distant location, often separated by vast geographical expanses (Zavadoff and Arcodia 2022). Teleconnections stemming from major climate phenomena like El Niño/Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), and the Indian Ocean Zonal Mode (IOZM) can also be highly impactful when examining the weather patterns in a specific region (DeHart 2011).

Some past studies have shown that a notable common teleconnection factor in both the 2010 and 2022 calamities in Pakistan is the presence of La Niña. These years witnessed La Niña conditions during the northern summer months. In fact, from May to July, the years 2010 and 2022 stand out as the two most prominent instances of La Niña in the 21st century (Henson 2022). Nevertheless, it is evident that the magnitude of the flood impacts in Pakistan during these two years exceeds that of major flood events experienced in previous decades, suggesting that additional factors other than just the monsoon rains are at play where La Niña effect could be a significant contributor.

Another important conclusion drawn from the IAHR study done by Lee (2022) was that although Pakistan is home to more than 7200 glaciers, the floods of 2010 and 2022 were a result of extreme precipitation and not glacier melting. He further stated that Glacier Lake Outburst Flood (GLOF) occurrences would have typically been observed in May and June when heatwave was at its peak, however, the 2022 flood event happened during the month of July and August when there was extreme precipitation. Moreover, Lee also added that Glacier Lake outbursts are more of a localized incidents that pose challenges in spreading from mountainous regions to plains which are limited to only hundreds of kilometers downstream. Lee also deliberated that upstream of the Indus River, several reservoirs were also constructed, and it was not until August 2022 that these reservoirs were utilized for emergency releases, signifying their role in impeding upstream flooding. Furthermore, upon reaching the plains, the flood resulting from melting glaciers demonstrates a sawtooth flow pattern, characterized by heightened flow during the daytime due to elevated temperatures and a subsequent reduction in flow during the night as temperatures decline. This flow pattern stands in contrast to the peak shape associated with heavy rainfall (Joseph Hun-wei Lee 2022).

F. CLIMATE ANALYSIS RESOURCES USED FOR PAKISTAN

We have various climate resources available for carrying out our analysis. Some of the relevant resources are listed below.

1. U.S. Resources

We consulted several resources available from the U.S. government.

a. National Oceanographic and Atmospheric Administration (NOAA)

It is one of the primary weather analysis resources used for this study. It is Headquartered in Washington, DC, and its mission is "To understand and predict changes in climate, weather, oceans, and coasts, to share that knowledge and information with others, and to conserve and manage coastal and marine ecosystems and resources" (NOAA 2023). NOAA provides most of the U.S. operational climate products. NOAA's subagencies provide operational climate products for analyzing weather data and providing products which are listed below.

b. National Weather Service

NOAA's National Weather Service (NWS) delivers data on weather, water, and climate, along with forecasts, alerts, and decision support services that are grounded in the assessment of impacts. These services are dedicated to safeguarding lives and property while also contributing to the growth and resilience of the national economy. (US Department of Commerce 2023). NWS Headquarters is in Silver Spring, MD and its objective is to furnish forecast information in a manner that more effectively assists emergency coordinators, initial responders, governmental authorities, enterprises, and the general populace in swiftly and intelligently making decisions that result in the preservation of lives and assets, and the advancement of livelihoods. NOAA's Office of Oceanic and Atmospheric Research and the National Environmental Satellite, Data, and Information Service are facilitating the integration of cutting-edge science and technology into the operational procedures of the National Weather Service. This integration will enhance forecast accuracy and, consequently, bolster preparedness for varying weather conditions. (US Department of Commerce 2023).

c. *Climate Prediction Center (CPC)*

The offerings from the Climate Prediction Center (CPC) encompass operational climate forecasts, ongoing real-time climate monitoring with essential data repositories, and evaluations of the underlying factors contributing to significant climate irregularities. These products encompass various time spans, ranging from weekly to seasonal forecasts, extending into the foreseeable future, and encompass observations of terrestrial, marine, and atmospheric conditions, including the stratosphere, to the extent technologically achievable. Headquartered in Camp Springs, MD, its mission is "deliver real-time products and information that predict and describe climate variations on timescales from weeks to years thereby promoting effective management of climate risk and a climate-resilient society" (Climate Prediction Center 2023).

d. *National Climatic Data Center (NCDC)*

The National Climatic Data Center (NCDC) is an essential component of the National Oceanic and Atmospheric Administration (NOAA) in the United States. Situated in Asheville, North Carolina, the NCDC serves as a repository for a vast and comprehensive collection of climate and weather-related data. Its mission is to gather, archive, and provide access to a wealth of historical and contemporary climate information, including temperature records, precipitation data, and climate monitoring observations. The NCDC plays a crucial role in assisting researchers, meteorologists, policymakers, and the public in understanding climate trends, assessing climate change impacts, and making informed decisions related to environmental and climatic factors. Through its data management and dissemination efforts, the NCDC contributes to the advancement of climate science and the mitigation of climate-related risks.

e. *Earth System Research Laboratory (ESRL), Physical Sciences Division*

Earth System Research Laboratories (ESRL) is a vital component of the United States' National Oceanic and Atmospheric Administration (NOAA), dedicated to the study and understanding of Earth's complex systems. With its headquarters in Boulder, Colorado, ESRL conducts pioneering research focused on meteorology, atmospheric science, and the broader Earth system. Its work encompasses various facets of atmospheric and climate research, such as the study of greenhouse gas concentrations, air quality, weather patterns, and the impact of climate change. Through its efforts, ESRL plays a crucial role in advancing scientific knowledge, providing essential data for weather forecasts, and addressing critical environmental challenges in the United States and beyond. ESRL tools including Physical Sciences Laboratory (PSL) containing the NCEP/NCAR reanalysis dataset were extensively used in this study. Details of these will be discussed in Chapter III.

f. International Research Institute for Climate and Society (IRI)

IRI, situated in Palisades, NY, is an integral part of Columbia University. Its primary objective is to enhance the public's understanding of how climate affects developing nations and to offer scientific assistance in better predicting and handling these effects (IRI 2023). This encompasses the creation of seasonal climate predictions, which encompass products tailored for regions like SWA and south-central Asia.

2. International Resources

We also used several international weather resources.

a. Bureau of Meteorology (BOM)

The Bureau of Meteorology is Australia's premier agency responsible for meteorology, climate science, and weather forecasting. Headquartered in Melbourne, the Bureau plays a pivotal role in providing accurate and up-to-date weather information, warnings, and forecasts to the Australian public. It also conducts vital research on various aspects of the Earth's atmosphere, including climate patterns, oceanography, and environmental science. The Bureau of Meteorology employs advanced technology, such as radar systems and satellite imagery, to monitor weather conditions and deliver essential services for a wide range of sectors, including agriculture, emergency management, and aviation. In addition to its core forecasting duties, the Bureau is instrumental in advancing our understanding of climate change and its impacts, playing a crucial role in Australia's resilience to meteorological and environmental challenges.

b. European Centre for Medium-Range Weather Forecasts

The European Centre for Medium-Range Weather Forecasts (ECMWF) stands as a preeminent global institution in the realm of weather forecasting and climate research. Founded in 1975, ECMWF is headquartered in Reading, United Kingdom, and serves its member countries across Europe, providing invaluable meteorological services and climate insights. ECMWF specializes in medium-range weather predictions, covering a time frame that extends from a few days to several weeks, making it an essential resource for both short-term weather forecasting and longer-term climate monitoring (ECMWF 2023). The center leverages state-of-the-art computational models and observational data to produce highly accurate and reliable forecasts, contributing significantly to global meteorological advancements.

c. United Kingdom Meteorological Office

The United Kingdom Meteorological Office (UKMO) serves as the national weather service for the United Kingdom and offers support to the country's defense forces. Its primary function is to deliver weather and climate forecasts to aid decision-making, ensuring the safety, well-being, and prosperity of individuals (UK Met Office 2023).

G. SCOPE OF STUDY

The study objectives, questions and sequence are listed in the details given below.

(1) Research Objectives

This study focused on exploring the feasibility for usage of modern weather resources as well as means to skillfully forecast atmospheric conditions at intraseasonal to seasonal lead times (leads of 0–4 months) for enhancing the understanding of factors contributing to weather changes and improve the forecasting in case of extreme precipitation events during the Pakistan monsoon season.

(2) Research Questions

The study focused about examining the following questions:

- What are the major interannual variations in Pakistan summer precipitation that contribute to large flooding and drought events?
- What regional and global climate patterns characterize these variations? What climate processes contribute to these variations?
- What climate system variables might be used as potential predictors of Pakistan summer precipitation?
- What statistical methods might be useful in producing sub-seasonal to seasonal predictions of Pakistan summer precipitation? How skillful are those prediction methods?
- To what extent do the results of this research explain the most extreme precipitation events in Pakistan?
- How well would the prediction methods developed and tested in this research have predicted these most extreme events?

(3) Study Outline

Chapter II provides an overview of the relevant literature for extreme precipitation events; Chapter III focuses on the datasets and the methodology used for climate analysis. Chapter IV presents the results of our climate analysis and long-range forecast models. Chapter V contains a summary of our results, conclusions, and recommendations for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

As researchers continue to study intense, abnormal, and abrupt weather events, one such category of interest is extreme precipitation. Analyzing these unusual weather events leads us to the question of whether they can be forecasted. The study of extreme precipitation events in Pakistan has produced different viewpoints on factors affecting abnormal weather events. This chapter examines studies pertaining to extreme precipitation events in Pakistan i.e recent floods in Pakistan during 2010 and 2022. It also focuses on some other relevant weather-related literature to Pakistan as well as literature on forecasting/prediction and some pertinent weather patterns which could have an impact on high precipitation events related to Pakistan.

A. LITERATURE ON 2010 AND 2022 FLOODS

Galarneau et al. (2012) analyzed an extreme heat wave in Russia and simultaneous historic floods in northern Pakistan in July 2010. They linked the Russian heat wave to the floods through atmospheric patterns, where a blocking anticyclone intensified, causing recirculation of air and subsidence in Russia. The study explained that the active phase of the Madden-Julian oscillation played a role, creating an upper-level jet northwest of Pakistan. It concluded that, as a result, a southeasterly low-level jet formed over northern Pakistan that was associated with a monsoon depression. This jet, along with an easterly flow from an anticyclone over the Tibetan Plateau, brought tropical moisture and intense rainfall, which worsened the severity of the late July floods in northern Pakistan (Galarneau et al. 2012).

Di Capua et al. (2021) also investigated two simultaneous extreme weather events in summer of 2010 including a record-breaking heatwave in Russia and severe floods in Pakistan just like Galarneau studied above. They used a comprehensive ensemble climate model experiment to investigate the atmospheric wave event. Study revealed that the circulation in 2010 was indicative of a recurrent wave pattern connecting the heatwave and flooding incidents. Study demonstrated that three factors played a crucial role in the occurrence of this wave pattern which included: unusual sea surface temperature anomalies in 2010, increasing the likelihood of this wave pattern by a factor of 2-to-4 compared to the model's typical climate; secondly, lack of soil moisture in Russia during early summer increased the probability of local heatwaves and intensified rainfall extremes in Pakistan through an atmospheric wave response, and thirdly warming in high-latitude lands promoted the occurrence of the wave pattern, thereby leading to extremes in both rainfall and heat. These findings underscored the intricate and interdependent interactions among various drivers of precipitation (Di Capua et al. 2021).

Shaevitz et al. (2016) studied two major high precipitation events of July 2010 in northeast Pakistan and September 2014 in northern India which caused devastating floods. They analyzed flood events that shared very similar types of synoptic flow patterns, which resulted in moisture content that was exceptionally high in the affected areas. Shaevitz et al. utilized the quasi-geostrophic omega equation to analyze the elements of large-scale vertical motion, encompassing synoptic forcing, diabatic heating, and mechanically induced orographic ascent. The research uncovered diabatic heating as the predominant factor influencing large-scale vertical motion, indicating that the orographic forcing generated by airflow over the Himalayas played a crucial role in initiating the convection responsible for this heating. Shaevitz et al. concluded that eleven years of data demonstrated a strong correlation between extreme precipitation events in this region and the simultaneous presence of substantial orographic forcing and high column-total precipitable water vapor (Shaevitz et al. 2016).

Y. Ma et al. (2023) offered insights on extreme precipitation in Pakistan and provided essential information for policymakers to address the consequences of flooding. Authors highlighted that in the summer of 2022 Pakistan experienced an unprecedented and prolonged monsoon season, leading to severe flooding. The research investigated the sub-seasonal attributes and mechanisms underlying this unusually high precipitation event. It explained how excessive rainfall in July and August affected northern Pakistan and surpassed all historical records while displaying distinct spatial patterns. The author explained that a South Asian high-pressure area, in conjunction with an Iranian high-pressure region, were extended towards the east in July. This anomalous anticyclonic circulation to the northwest of Pakistan resulted in intensification of easterly winds to the south of the Tibetan Plateau. Meanwhile, a cyclonic system in the Arabian Sea region carried water vapors from the Bay of Bengal which enhanced precipitation over Pakistan. They explained that south Asian high (SAH) expanded towards the east in August while the subtropical high over western Pacific extended towards the west to the Tibetan Plateau. Moreover, the development of European blocking (EB) and presence of a deep trough in northwestern Pakistan altered atmospheric circulation. Authors concluded the study by claiming that extreme precipitation in the month of July was primarily the result of remarkably strong Indian monsoon, whereas high precipitation level in the month of August was outcome of a combination of the heavy Indian monsoon coupled with the role of EB (Y. Ma et al. 2023).

Q. Ma et al. (2023) focused on Pakistan's unprecedented extreme rainfall in July and August 2022, with particular attention to the associated mechanisms. According to this study, widespread flooding and landslides were triggered by consecutive extreme rainfall episodes which exceeded 500 mm in both northern and southern Pakistan. These events were highly linked to unusual snow cover in the western Tibetan Plateau during late spring. In July, reduced snow cover led to increased diabatic heating, resulting in the formation of the South Asian High (SAH) and strong southeasterly jet, coupled with cyclonic conditions in the Arabian Sea region, causing extreme precipitation with an estimated 50-years of repeat period. In August, anomalies, including an abnormally strong SAH and shifts in the West Pacific Subtropical High, influenced atmospheric dynamics. Cyclone anomalies further enhanced moisture transport from the Arabian Sea and the Bay of Bengal, culminating in extreme precipitation events with a 30-year return period (Q. Ma et al. 2023).

Nanditha et al. (2023) assessed increased flood risk in South Asia due to its vulnerability and exposure to changing climate conditions. August 2022 floods in Pakistan were explained by authors as a stark example of the potential devastation that could intensify as the climate warms. The event is the second deadliest flood on record, which led to displacement of approximately 33 million people. A comprehensive analysis of both observational data and climate projections was undertaken by these researchers to gain a deeper understanding of the causes and consequences of this catastrophic event.

According to them, the primary driver behind the August 2022 flood was a sustained period of intense precipitation, spanning approximately 15 days, which occurred on already water-saturated ground. This extreme precipitation event was a direct result of the passage of two atmospheric rivers over southern Pakistan. Utilizing multiple hydrological models, streamflow simulations further confirmed that this multiday extreme precipitation was pivotal in triggering the catastrophic floods. A noteworthy observation was that certain flood-impacted stations recorded significantly higher water flow levels compared to their upstream counterparts, underscoring the severity of the event (Nanditha et al. 2023).

Otto et al. (2023) examined the severe consequences of the historic 2022 summer flooding in Pakistan. They used a probabilistic event attribution methodology to understand climate change's role in this unprecedented event. Notably, state-of-the-art climate models struggled to accurately replicate the unique rainfall patterns observed. The observed trend in extreme rainfall exceeded model simulations. The study emphasized the challenge of quantifying the overall impact of human-induced climate change, suggesting unaccounted long-term variability and unmodeled processes might be significant contributors. Nevertheless, a substantial body of model projections and observations indicated a clear trend: rising temperatures in Pakistan led to more intense rainfall events, with some models indicating a potential 50% increase in rainfall intensity due to climate change (Otto et al. 2023).

B. LITERATURE ON SIGNIFICANT WEATHER PATTERNS STUDIED ON PAKISTAN

Imran (2013) focused on the temporal trends in peak monsoon rainfall events of Northeast Pakistan, which is a critical factor for the region's agricultural dynamics. He analyzed daily rainfall data from 1961 to 2010 (monthly and weekly trends) for selected stations in the northeast monsoon regime. He categorized extreme rainfall events into 50 mm, 100 mm, and 150 mm or more rainfall. Imran highlighted that the capital, Islamabad, experienced the highest frequency of 50 mm and 100 mm or heavier precipitation events, while Sialkot city led in 150 mm or more rainfall. Imran concluded that August emerged as the most frequent peak rainfall events month. His study found that the last week of July and first two weeks of August were the most likely times for high-intensity rainfall events. This conclusion could be useful in determining the predictand months for this study (Imran 2013).

Vellore et al. (2016) explored the connection between extreme Himalayan precipitation and monsoon-extratropical circulation. They analyzed 34 extreme precipitation events in the Western Himalayas from 1979 to 2013 using rainfall observations and reanalysis data. Study revealed a consistent large-scale circulation pattern linking the tropics and South Asian monsoon, creating extreme precipitation in the Western Himalayas. This pattern involved a southward-reaching westerly trough, a western Eurasian blocking high, and an intensifying Tibetan anticyclone. Key factors resulting in high precipitation instances in the region included mid-altitude Rossby wave breaking, low-pressure system spread during monsoon, anticyclone development in Tibetan, presence of ageostrophic waves over the Himalayan region, and detection of severe moist convection in the foothills of Himalayan range. High-fidelity simulations indicated that convective processes were intensified by diabatic heating and mesoscale ageostrophic effects. (Vellore et al. 2016).

Tandon et al. (2018) showed that climate model predictions of extreme precipitation levels varied significantly by region, as some areas witnessed substantial increase in precipitation, while others experienced decrease in extreme precipitation. Authors claimed that regional differences were closely linked to shifts in large-scale upward motion during extreme rainfall events, also referred as “extreme ascent.” They used data from a broad ensemble of the Canadian Earth System Model version 2 and highlighted those alterations in extreme ascent in subtropical regions resulted from variations in the horizontal extent of rising anomalies. Tandon et al. associated the changes with shifts in vertical stability and highlighted that those alterations in the average seasonal atmospheric circulation near equator played a significant role in extreme ascent (Tandon et al. 2018).

Bhatti et al. (2020) examined long-term precipitation changes in Pakistan to understand the effects of climate change. They investigated variations in extreme precipitation events in Pakistan between 1980 and 2016, using daily data from 51 weather stations. They employed non-parametric methods to assess trends in eight extreme precipitation indices and called them the simple daily intensity index. The study revealed an overall increase in the spatial distribution of different precipitation extreme indices, despite the monsoon and westerlies-influenced humid regions showing decreasing precipitation trends. Bhatti et al. identified 2011 as a significant year for increased trend changes. The investigation unveiled dynamic alterations in temporal trends for all precipitation extremes throughout the study period, highlighting an increasing inclination from 1980 to 2016. Bhatti et al. concluded that elevation was inversely related to most precipitation extremes, indicating reduced precipitation along the country's latitudinal extent. Authors also believed that spatiotemporal variations provided possible indications of climate change and variability affecting Pakistan's precipitation patterns (Bhatti et al. 2020).

Ullah et al. (2022) focused-on monitoring and assessment of long-term droughts in Pakistan which were highly significant for the management of meteorological disaster risks. Their analysis of atmospheric circulation patterns proposed that significant changes in wind speed, relative humidity, air temperature, and anomalies in geopotential height were likely contributors leading to droughts in the region. Therefore, its reversal was believed to be linked with high rainfall in a region. The research pinpointed Niño4, the multivariate El Niño-Southern Oscillation (ENSO4.0) Index, and sea surface temperature as the main influential factors contributing to periods of droughts in Pakistan (Ullah et al. 2022).

Ullah et al. (2023) conducted a comprehensive analysis of Pakistan's monsoon precipitation extremes from 1981 to 2018. It revealed a discernible increase in both the frequency and intensity of extreme precipitation events across the country. Notably, the most significant changes were observed in the northwestern, central, and eastern regions, while some areas in the north and southwest experienced a slight decrease. The statistical tools employed in this analysis underscore the robustness of these findings, confirming the trends in extreme precipitation. Per these, the rise in precipitation extremes was influenced by a complex interplay of meteorological factors, including intensified mid-latitude westerlies and easterlies, a reinforced monsoon trough, and the notable contrast in land-ocean temperatures. Research is also valuable reference for future studies aimed at mitigating the impact of extreme precipitation events and advancing sustainable development in the region (Ullah et al. 2023).

Abbas et al. (2023) investigated the impact of winter precipitation (December to March) on Pakistan's agricultural sustainability, water storage, and river flow. They focused on less-explored land-ocean and atmospheric variables. They observed a general decline in seasonal and monthly precipitation, except for February, which experienced an increase. One key precipitation pattern (EOF1) was significantly correlated with sea surface temperatures in the central Pacific and Indian Oceans, signifying the influence of ENSO-induced Hadley-Walker circulation changes. These alterations affected the strength of south-westerly jet streams, impacting water vapor transport and precipitation in Pakistan, with potential consequences for agriculture, cropping patterns, and river flow (Abbas et al. 2023).

Tajbar (2023) explained that Sindh province in Pakistan has a history of severe droughts, yet research is lacking on the impact of large-scale climate drivers (LSCDs) in this region. The study aimed to uncover the key LSCDs influencing monthly precipitation and improve forecasting. Utilizing techniques such as principal component analysis (PCA), artificial neural networks (ANN), Bayesian regularization neural networks (BRNN), and multiple regression analysis (MRA), along with a 12-month lagged association with LSCDs like oceanic temperature patterns, atmospheric pressures, and heat fluxes, the study pinpointed significant LSCDs. This included surface zonal velocity (SU), 500 hPa zonal velocity (U500), 2m air temperature (T2M), sea surface temperature (SST), sensible and latent heat fluxes over land (SHFOL and LHFOL), surface-specific humidity (SSH), and 850 hPa geopotential height (Z850), with a confidence level of 99%. During testing, the ANN and BRNN models outperformed multiple regression models, showing higher predictive capabilities with correlation coefficients ranging from 0.57 to 0.83 and 0.52 to 0.76, respectively. These results emphasize the effectiveness of ANN and BRNN models in forecasting monthly precipitation in Sindh, especially when considering lagged LSCDs (Tajbar et al. 2023).

C. LITERATURE RELATED TO FORECASTING/PREDICTION

There has also been a significant amount of literature generated on forecasting of weather patterns. Some studies related to forecasting in different parts of the world for prediction are listed below. These studies show different types of forecasting models used to understand weather patterns.

Whitaker (2021) examined the significant influence of the El Niño–Southern Oscillation (ENSO) index on global climate and hydrology patterns, particularly focusing on the Ganges River. This study aimed to develop a model for forecasting the Ganges River's flow with an extended lead time. Traditional rainfall-runoff models faced limitations due to short forecasting windows, especially in vast river basins like the Ganges. This pioneering research established a robust connection between the Ganges' annual flow and the ENSO index. The rate of change in the ENSO index was also statistically linked to Ganges flow. The study proposed a statistical model capable of forecasting the Ganges' annual flow one year in advance, along with quantifying forecasting uncertainty. One notable advantage of this model was its independence from upstream rainfall and stream flow data, making it applicable in data-scarce regions. With 45 years of data for development and calibration, and 15 years for validation, the model consistently provided accurate forecasts during both El Niño and La Niña events (Whitaker et al. 2001).

Meyer (2007) examined the connection between global warming and tropical cyclone (TC) activity in the Western North Pacific (WNP) by forecasting, with a focus on the impact of large-scale environmental factors (LSEFs). Meyer included sea surface temperatures (SST) above 26°C, weak horizontal wind shear, absolute positive vorticity at lower altitudes, upward motion of air and mid-level humidity as the factors. Global warming signals were identified in both SST and vertical wind shear data using least squares fit, showing variations at a 5°x5° scale. Logistic regression was employed to establish the relationship between LSEFs and TC formation probability, while linear regression assessed the link between LSEFs and Accumulated Cyclone Energy (ACE). Independent data were used for model validation, confirming the influence of LSEFs on

TC formation and ACE. His findings supported the hypothesis that global warming increased TC frequency and intensity in the WNP through LSEFs (Meyer 2007).

Crook (2009) assessed the feasibility of long-range dust storm forecasting in Iraq, with potential implications for Department of Defense (DOD) operations. He investigated two key dust storm variables including precipitation rates and surface winds. Anomalies in Iraq's precipitation leading up to dust storms and wind conditions during these events were determined. Significant correlations were established between Iraq-specific variables and with the remote climate variables. As a result, two long range predictors for dust favorable conditions were identified: (a) Indian Ocean Sea surface temperature and (b) an index measuring the sea level pressure difference between Tunisia and Kazakhstan, indicative of surface winds. These predictors were integrated into an adapted Composite Analysis Forecast (CAF) method to predict conditions conducive to dust storms in Iraq at one- and two-month lead times. The verification process confirmed the method's substantial potential for delivering accurate long range forecast of dust storms likelihood in Iraq, with authorship attributed to the U.S. Navy (Crook 2009).

Lemke (2010) studied that the Horn of Africa (HOA) faced recurring climate challenges, like droughts and floods, which necessitated effective mitigating/response strategies. These strategies included early warning systems, military commands, emphasizing the need for comprehensive knowledge of climate influences. Accurate long-range forecasts, especially for precipitation, became crucial for organizational planning. This study focused on long-range precipitation prediction during the October-November rainy season with multi-season lead times. Correlations between HOA precipitation rates (PR) and remote climate variables revealed the potential for skillful long-range PR forecasts. Methods included deterministic and probabilistic approaches. Results displayed improved forecast accuracy compared to mean-based forecasts, highlighting the forecasting techniques' potential for operational planning (Lemke 2010).

DeHart (2011) focused on Pakistan's summer precipitation forecasting, crucial for mitigating floods and droughts. The research developed methods for long-range predictions during July-August and unveiled 850 hecto-Pascal (hPa) geopotential heights (GPH) set in Pakistan as a likely predictor. Based on this predictor various LRF approaches were created, which included linear regression, optimal climate normal techniques and tercile matching. Independent hindcasts over a 41-year period (1970-2010) demonstrated the effectiveness of these approaches in predicting below and above normal events of precipitation. Furthermore, DeHart highlighted that using Sea Surface Temperatures (SSTs) as predictors for 850 hPa geopotential heights had the potential for skillful LRFs for July-August precipitation in Pakistan with lead times extending up to six months or more. The study highlights the need for further research to create operational long-range forecasting systems, emphasizing their value for strategic planning (DeHart 2011).

Gillies (2012) tested and developed a method for long-range forecasting system for Pakistan, which could predict environmental conditions on an intra-seasonal to seasonal level, spanning several weeks and seasons ahead. Multiple regression models were used to generate ensemble members using the connections established between atmospheric predictors, globally distributed oceanic and local forecast targets which were categorized into three tercile groups. Predictor selection depended on their long-lead correlations with the forecast target, while model choices were made on lagged average ensemble performance at various lead times, determined through cross-validated hindcasts which spanned over multiple decades. The key results comprised probabilistic forecasts at long-lead accompanied by evaluations of uncertainty quantitatively. The system underwent long-range forecasts testing for environmental situations in Pakistan, exhibiting superior skill in comparison to climatological forecasts. The integration of multiple forecast predictions in a multi-model, lagged ensemble average approach proved to be of greater accuracy than any individual forecast (Gillies 2012).

D. CONCLUSION

This thesis will extend the long-range forecasting work done by DeHart (2011) by examining the additional years 2011–2023, which include the 2022 major flood event. The addition of the recent 2022 flood allows comparison with the previous 2010 event. Data and Methods will be covered in Chapter III.

III. DATA AND METHODS

This chapter focuses on the datasets and methodologies employed in the research.

A. DURATION OF STUDY

This research spans a comprehensive 54-year period, from 1970 to 2023, a time frame chosen in consideration of data availability from the NCEP/NCAR reanalysis datasets. While data predating 1970 is available, our selection of this period was motivated by the desire to emphasize the benefits of satellite data and ensure data consistency within this study by excluding pre-satellite era years. Notably, as indicated in the works of Lemke (2010) and DeHart (2011), this timeframe encompasses various intraseasonal to interannual climate variations and provides some representation of decadal patterns. Our investigation focused on conducting seasonal climate analyses for the Pakistan region, specifically during the high precipitation months (as elaborated in Chapter 1), namely July and August. Among these years, 2010 and 2022 hold particular significance due to their association with substantial flooding events in Pakistan. Thus, their presence in the evaluation was crucial to yield a thorough insight of the observed patterns and impacts.

B. DATASETS AND SOURCES

The study sourced some of its datasets from the subsequent repositories.

1. National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) Atmospheric Reanalysis

The NOAA Physical Sciences Laboratory (PSL) carries out research with the purpose of improving observations, predictions and modeling of weather, water and climate extremes including their impacts. Among the available gridded climate data resources at NOAA PSL, we used NCEP/NCAR Reanalysis 1 (R1) as primary dataset for our study (Kalnay et al. 1996). Developed through a collective partnership amongst the National Center for Atmospheric Research (NCAR) and the National Centers for Environmental Prediction (NCEP), this dataset offers a consistent, global record of atmospheric conditions including comprehensive temporal and spatial coverage. The NCEP/NCAR reanalysis data is available on a consistent six hour temporal resolution and captures a horizontal grid which is uniform, boasting a 2.5° spatial resolution. This grid coverage encompasses all the standard levels of the troposphere and stratosphere, inclusive of sea surface temperature (Kalnay et al. 1996). This global coverage allowed the study to explore the intricate interactions between different regions and their contribution to the climate system.

Furthermore, the NOAA's PSL website (NOAA PSL 2023) offers a set of tools for data analysis and visualization, enabling graphical representations of our research findings. Our study primarily relied on the utilization of the R1 dataset in conjunction with the resources provided by the PSL website. While the R1 dataset has records dating back to 1948, we focused on the period from 1970 to 2023 for reasons explained above. Although dataset covers an impressive array of variables, our focus in the study was on key meteorological variables, including sea surface temperature (SST, °C), surface air temperature (°C), geopotential height (GPH, m) at different altitudes, 850 hPa vector winds (m/s), and precipitation rates (PR, mm/day). This selection was based on the recognition that these factors, as indicated by NCAS (National Center for Atmospheric Science) in 2023, exert a significant influence on weather outcomes (NCAS 2023).

2. Multivariate ENSO Index (MEI)

The Multivariate El Niño/Southern Oscillation (ENSO) Index (MEI.v2) synthesizes oceanic and atmospheric parameters into a singular composite index. This approach provides a comprehensive assessment of ENSO conditions (Wolter and Timlin 2011). It also serves as a real-time indicator of ENSO intensity and facilitates a meaningful comparative analysis of evolving climatic states. According to Zhang et al. (2019) "ENSO variability is evaluated using a revised MEI index (MEI.v2) that is based on five variables: sea level pressure (SLP), SSTs, 10-m zonal wind (U), 10-m meridional wind (V), and outgoing longwave radiation (OLR)." These calculations span the tropical Pacific basin (30°S-30°N and 100°E-70°W) region. According to Zhang et al. "a combined EOF analysis is conducted separately for each of the 12 partially overlapping two-month seasons” to account for ENSO’s seasonal variations while minimizing the impact of higher-frequency intraseasonal variability (Zhang et al. 2019). We analyzed this as a possible predictor given the significance between high (low) precipitation in Pakistan and ENSO.

3. North Atlantic Oscillation (NAO)

The North Atlantic Oscillation (NAO) index is determined by computing the contrast in sea-level pressure between the Subpolar Low and the Subtropical High (situated in the Azores). During the positive phase of NAO, it indicates that the pressure as well as heights in the North Atlantic’s high latitudes are lower than usual. Conversely, according to Hurrell and Deser the heights and pressure over the eastern United States, the central North Atlantic, and western Europe are higher than usual (Hurrell and Deser 2009). Similarly, an opposing pattern of pressure and height anomalies is represented by the negative cycle in those locations. As per the details on NAO given in National Centers for Environmental Information (NCEI) “Strong positive phases of the NAO tend to be associated with above-normal temperatures in the eastern United States and across northern Europe and below-normal temperatures in Greenland and oftentimes across southern Europe and the Middle East.” The NAO demonstrates notable inter-seasonal and interannual variability, and it’s common to observe prolonged durations, lasting spans of multiple months, characterized by both positive and negative phases of this pattern (North Atlantic Oscillation 2023). NOAA’s ESRL constructs an NAO index (NAOI), and for the purpose of this study, we tested NAO index as one of the potential predictors of Pakistan PR.

C. CLIMATE ANALYSIS METHODS AND FORECASTING TECHNIQUES

The methods used to carry out climate analysis for the study closely resembles those employed by DeHart (2011) in his examination of Long-Range Forecasting (LRFs) supporting operations in Pakistan, as well as Lemke (2010) in his investigation of LRFs for precipitation in the Horn of Africa (HOA). However, for prediction/forecasting we have used statistical modelling which was different from the ones used in the two above-mentioned studies.

1. Predictand Selection

In our study, we designated the predictand, or the forecast target, as the precipitation rate (PR) averaged over an area specific to a particular part of Pakistan throughout the months of July and August (Jul–Aug). For the rationale behind this choice of predictand and the chosen period, please refer to Chapter I. Opting for an area-averaged predictand was driven by the consideration that it represents a larger and more spatially comprehensive forecast target compared to a single point, thereby reducing vulnerability to small spatial as well as temporal changes. Additionally, area-averaged predictands offer several advantages, such as simplifying the forecast method development, enhancing predictability for extended lead times, while facilitating verification of forecast (van den Dool 2007). Nonetheless, it is crucial to recognize that choosing an area-average predictand without careful consideration can mask substantial spatial and temporal variability within the specified area of interest. The process of selecting the predictand variable itself was relatively simple, while the process of identifying the most suitable predictand region within Pakistan was a more involved endeavor. This involved a detailed analysis of long-term average July-August PR patterns within Pakistan (explained in Chapter IV), which allowed us to gain insights into the spatial distribution of precipitation in the country. Based on this analysis, we considered multiple potential predictand regions (details covered in Chapter IV) and ultimately chose one that aligned most closely with our research objectives. The primary criteria that informed our selection of the predictand region included following with criteria similar to DeHart's (2011) work:

- Examination of spatial patterns in long-term mean (LTM) PR within and in vicinity of the chosen area.
- Consideration of abnormal PR in the past two decades.
- Assessment of spatial arrangements (interannual to decadal level PR anomalies) within and in proximity to the selected region.

- Analysis of the repercussions of abnormal PR within the finalized predictand region to the other part of the country, including downstream effects such as impact of flooding.

- Identifying correlations among PR in the finalized predictand region and potential atmospheric predictors (e.g., geopotential heights and sea surface temperatures) at varying lead times (zero to four months).

- Investigating correlations among PR in the finalized predictand region and major factors involved in climate variations (e.g., MEI, NAO) at varying lead times (zero to four months).

- Capturing two high precipitation events of 2010 and 2022 floods to understand likely drivers of abnormal precipitation in these years.

2. Composites, Teleconnections, and Correlations

Upon the determination of selected predictand region, a time series of PR was compiled spanning over July to August, covering the years 1970 to 2023 within that specified area. We used this period (54 years) to identify 20% high (eleven) and 20% low (eleven) precipitation years to understand the factors driving these events. The time series was then used to determine the eleven wettest and eleven driest years of Pakistan since 1970 based on the data. Subsequently, we directed our attention to utilizing the mapping and analysis tools provided by NOAA PSL to generate seasonal mean and composites for July to August in terms of PR for the eleven wettest and eleven driest years (NOAA PSL 2023). The reason behind this process was to identify regions with abnormally high or low PR so that these could be narrowed down for the study and given us a general idea of the above normal (AN), near normal (NN), and below normal (BN) anomalies of PR for the selected region. After studying the changes in PR for predictand region using composites, we used composite means and anomalies to study selected environmental variables important for the study as also done by DeHart (2011) which mainly included regional PR (mm/day), GPH (m), global SST (°C), outgoing longwave radiation (OLR;

W/m2), 850 hPa vector winds (m/s), 700 hPa omega (Pa/s) and 850 hPa specific humidity (g/kg).

All anomalies considered in the study were computed using a reference base period spanning from 1991 to 2020 (ESRL 2023). Upon orientation with basic environmental factors, we studied various atmospheric variables with the purpose of choosing potential predictors for the predictand with emphasis on variables suitable at longer lead times. As a next step, we did correlation of Predictand (Pakistan PR timeseries) with some of the significant atmospheric variables (e.g., SST, air temperature, GPH) on a global and regional scale. The selected potential predictors led the Jul–Aug predictand at bi-monthly intervals from zero to four months. Consequently, we found teleconnections between the predictand and predictors of other atmospheric variables at different geographic regions that had significant correlations amongst each other (DeHart 2011). Then we went on to find the correlations that were statistically significant for our study. As per full study period spanning over 54 years (1970–2023), we adopted a pattern similar to DeHart (2011) where correlations greater than +/- 0.27 were considered statistically significant at a 95% confidence level based on the standard normal distribution of a two tailed t test (Wilks 2005). As we curtailed the dataset to 15 recent years (2009–2023) for the purpose of optimal climate normal analyses, correlations greater than +/- 0.51 were statistically significant at a 95% confidence level. Correlation analyses were done with the help of Microsoft Excel and the PSL website (PSL 2023).

3. Selection of Predictors

Potential predictors were identified as variables demonstrating substantial correlations at extended lead times and displaying plausible atmospheric connections with the predictand. Our inquiry delved into the feasibility of several candidate predictor variables, with specific attention directed towards Sea Surface Temperature (SST), Geopotential Height (GPH), and Air Temperature. Ultimately, our selection of the primary predictors focused on 850 hPa GPH and SST. This choice was based on several important considerations: (a) the most robust correlations involving high correlation values among the predictors examined; (b) consistent correlations across lead times ranging from zero to four months; and (c) their relationship with the predictand was conceptually sound. It is crucial to highlight that, while the relationship between predictor and predictand was relatively simple, the integration of predictors including 850 hPa GPH and SST from various locations was essential to grasp the complication.

4. Predictand and Predictor Time Series

Upon establishing the predictand for PR and its subsequent predictors, we embarked on an extensive examination of their time series data. This comprehensive analysis aimed to uncover patterns of variability spanning various timescales, ranging from intraseasonal to decadal, within both the predictors and predictands, while also assessing their interrelationships. Additionally, we used the DeHart (2011) pattern for the time series data to identify extreme events, which were subsequently used in conditional composite analyses, and to identify multi-year trends that could potentially influence the selection of long-range forecasting techniques. After the predictor's selection, we compared the time series data for predictors and predictand. This allowed us to evaluate the alignment of interannual variability and appraise the overall significance of their correlations. Moreover, this analytical procedure served multiple purposes, including (a) conducting a graphic control check on the predictor and predictand connection, (b) pinpointing stages as well as lead times when correlations displayed fluctuations in strength, and (c) identifying periods necessitating further examination through composite, correlation, and dynamical analyses.

5. Long-Range Forecast Method Development

Prior studies have evidenced the enhanced accuracy potential of forecasts that consider anomalies, or variations from the long-term mean (LTM), compared to those solely reliant on the LTM. Furthermore, long-range forecasts (LRFs) developed with custom-selected predictors tailored to specific forecast targets have shown increased skill when contrasted with those based on standardized climate variations like El Niño/La Niña (ENLN) events, as demonstrated in the research conducted by Lemke (2010) and Dehart (2011). To evaluate the efficacy of a particular predictor, we applied our statistical methods to establish the predictor-predictand relationship during the period from 1970 to

2023. Subsequently, we assessed the predictor's performance by scrutinizing its consistency with predetermined skill thresholds through the utilization of our statistical methods. The testing and result quantification were executed using primarily Microsoft Excel and R.

6. Statistical Analysis

To build our initial list of predictors, a single variable linear regression model was run between the predictors as well as between predictor-predictand at varying lead time times. To determine the fit of each regression, we found the correlation coefficient, R squared, adjusted R-squared, and significance values. To assess the viability of each predictor for single variable regression, we compared the correlation coefficients' values and discarded those predictor-predictor combinations that were highly correlated. The primary goal was to finalize our list of initial predictors so that we could build the logistic regression model that we could use to forecast PR in our predictand region at varying lead times from zero to four months.

7. Statistical Modelling

For our statistical analysis of this study, we used logistic regression modelling. The binary dependent variable takes on one of two possible outcomes, typically coded as 0 and 1. Logistic regression models the probability of the dependent variable being in one of the two categories as a function of the independent variables. It estimates the log-odds (logit) of the probability, which is then transformed into probabilities using the logistic function. After finalizing our initial list of predictors, we did logistic regression using R for our top four selected predictors to get the results for modelling. In R, we split the 54 years of data into test and train data. A logistic regression model was fit using the training data and evaluated using the test data. The results helped us to determine the model accuracy as well as Area Under Curve (AUC) to assess the performance of the model.

IV. RESULTS AND ANALYSIS

A. RESULTS FOR CLIMATE ANALYSIS

1. Predictand Selection

For the predictand selection, initially we chose our predictand variable (months) which we identified as July and August based upon the high precipitation rate in Pakistan during these months which was more significant than any other single month during the year (for details refer to Chapter I). However, as already explained in Chapter III, Section C, finalizing the process for predictand region containing Pakistan was a complicated process which required deliberated effort thus making it a lengthy process. While undertaking the task of identifying the predictand region, we carried out an analysis of interannual composite mean and anomalous PR during the focus period of 1970–2023 for the months of Jul–Aug which led to various patterns of PR and PR anomaly distribution. To begin our study, we focused on an area which was representative of the bulk of Pakistan and took its year-wise data to find out overall low and high PR years of Pakistan including their precipitation value. Some of the steps involved during the process of selecting the predictand region included:

- Finding PR of Pakistan by plotting the latitude/longitude with focus on high precipitation areas using seasonal time series analysis for July-August to record PR for each year from 1970 to 2023 (total 54 years).

- Experimenting with multiple possible rectangle-shaped geographical regions in Pakistan that experienced high PR and using seasonal time series analysis.

- Identifying 20% high PR (11 years) and 20% low PR (11 years) in Pakistan for the above-mentioned region and duration. Subsequently, plotting seasonal climate composite mean and anomalies separately for each of the 11 high and 11 low PR years.

- Identifying precipitation rate anomaly (PRA) for each of 11 high PR years by plotting a box over regions with +3 mm/day or more PRA for that specific year.

- Finding PRA for each of 11 low PR years by plotting a box over regions with -3 mm/day or lesser PRA.

- Selecting PRA boxes for each high (low) year and placing over one map to identify the common areas for all high (low) PRA regions.

- Identifying three potential regions for further analysis keeping in view their potential to be good representation of precipitation anomalies in Pakistan for all high and low PR years.

- Iteratively processing the time series analysis of three Predictand Boxes, including study of their time series analysis. The details of three boxes studied for finalizing the predictand are shown in Table 1.

Table 1. Potential predictand regions considered for the study. Displaying the list of predictand boxes including their latitude/longitude.

Predictand Box	**Latitude**	**Longitude**	**Remarks**
1	30-34 N	69-74 E	Not a good representation of high PR years particularly 2022 which has the highest PR in the history
2	29-33 N	68-73 E	Not a good representation of high PR years particularly 2010 being high flood year
3	28-34 N	67-73 E	Good representation of high as well as low PR years including focus years of 2010 and 2022

We focused on three different regions as potential predictand for our study on Pakistan PR while calling these as Predictand Box (depicted in Figure 6) and finalized one of these boxes as our main predictand box (Box 3 as depicted in Figure 6).

Predictand Box 3 (28º-34º N and 67º-73º E) was chosen as it captured the PRA of most of the high and low PR years and captured the potential focus years for study (i.e 2010 and 2022).

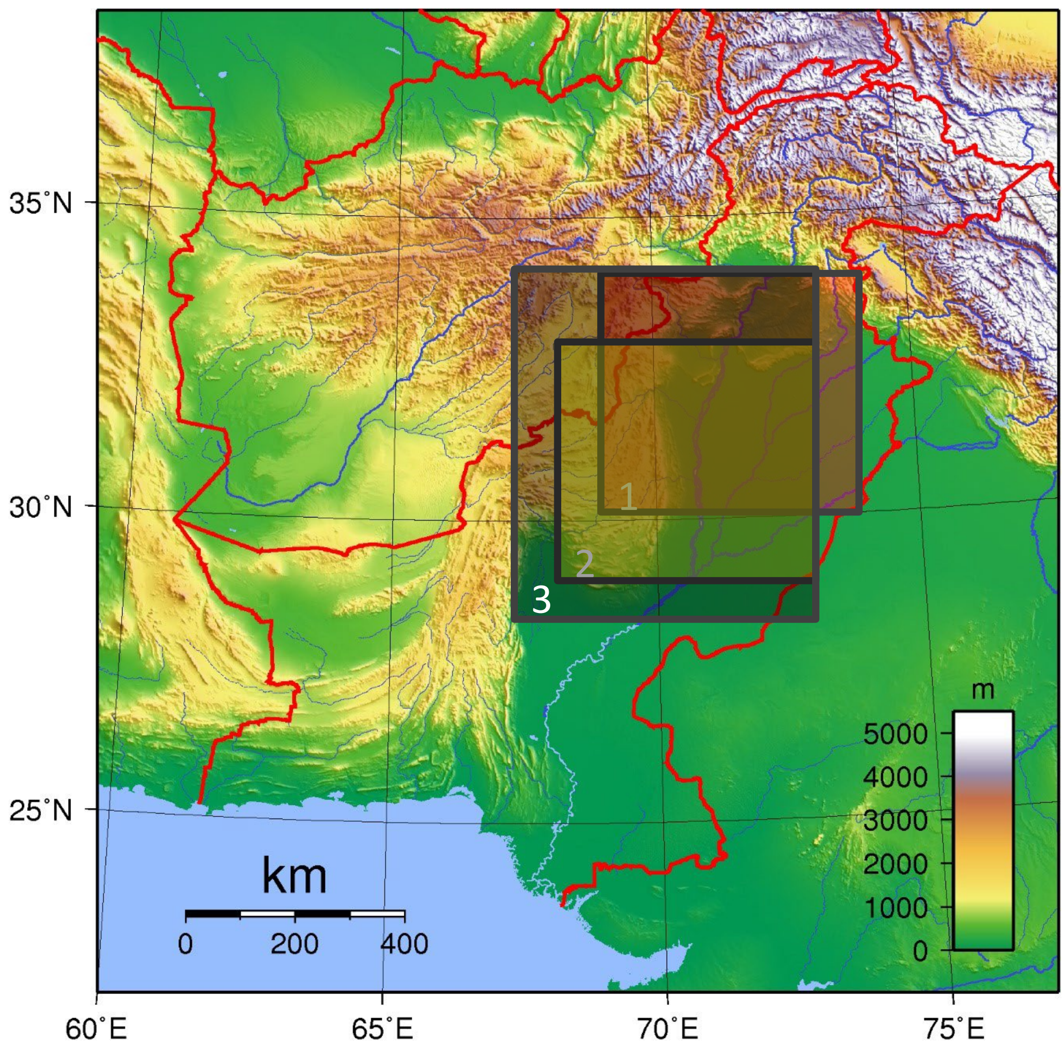

Figure 6. Pakistan physical relief map. Three predictand regions are plotted over the map and have been marked as such in the figure. Details of the three predictand regions are: (a) Predictand Box 1: 30o–34oN, 69o–74oE; (b) Predictand Box 2: 29o–33oN, 68o–73oE; and (c) Predictand Box 3: 28o–34oN, 67o–73oE. Source: Wikipedia (2023b).

Although the focus of study remained to analyze the abnormal precipitation events in Pakistan, the primary effort remained on flooding years of Pakistan (2010 and 2022) because of the extent of damage these floods caused in Pakistan. Therefore, it was obvious to focus on a predictand region which was a good representation of flooding in Pakistan. The significance of our selected Predictand Box 3 was that it showed a very good representation of the flood years while covering mostly the areas of central Pakistan. Another reason is Box 3 lies at a region where it captures most of the river flow in Pakistan; therefore, any abnormal precipitation event is likely to flood these rivers first. Both 2010 and 2022 saw heavy rain in Box 3 region which led to overflowing of rivers along with their banks particularly those flowing downstream. Some of the major reasons that DeHart (2011) also considered for choosing predictand box therefore remained:

- high LTM precipitation rate for the months of Jul–Aug.
- higher vulnerability to downstream flooding.

2. Predictand Time Series

Based on these criteria for choosing the predictand box, we then made a time series of the Predictand region from 1970–2023 showing year-wise precipitation. From this time series shown in Figure 7, we can see that 2010 and 2022 were amongst the high precipitation years in the predictand region. This time series indicates with orange dotted line the years over which there was above normal (AN) precipitation in the predictand region of Pakistan, whereas the yellow line indicates years below which were considered as below normal (BN) precipitation years in the predictand region. The years between the yellow and orange line were considered as near normal (NN) for the study of precipitation. This time series gave an easy reference for visually checking the eleven high and eleven low precipitation years in Pakistan. This identification of focus years helped in constructing composites related to the atmospheric conditions linked with these AN and BN precipitation years. Making of these composites was helpful in identifying potential predictors linked with PR variations and studying the key spatial and temporal patterns linked to climate variations in Pakistan PR.

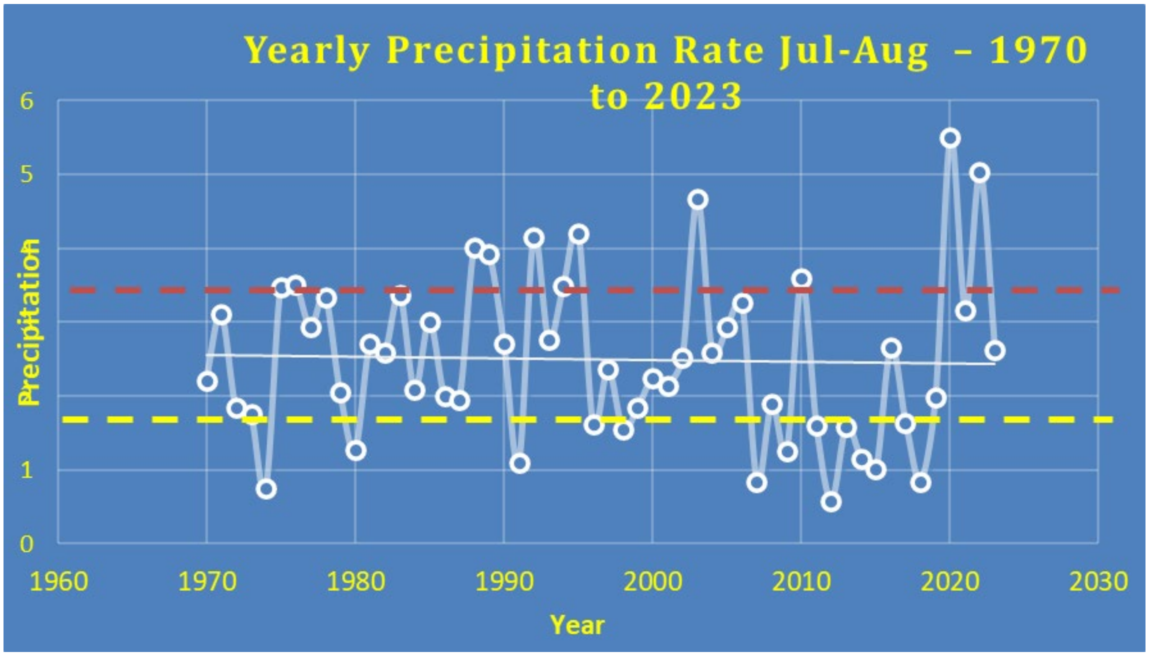

Figure 7. Time series of precipitation in predictand region. Adapted from NOAA Monthly Mean Timeseries (2023).

3. Composite Analysis

The first and foremost consideration for composite analysis was that it should be conducted along the lines of research performed by Dehart (2011) to draw comparisons or identify differences amongst his 41 years of study and my 54 years of study; the intent was to determine if there was any significant change in the atmospheric conditions. As already explained in Chapter I, the focus of our study was to comprehend the additional features that could have caused abnormal conditions while keeping DeHart work as the base.

a. Long Term Means (LTMs)

As part of composite analysis of the predictand and surrounding regions, we studied the long term mean (LTM) for different atmospheric conditions of Pakistan and surrounding regions for the month of Jul–Aug for the 1970–2023 period. This is a significant step to find the climatic conditions important in determining the processes that causes PR anomalies in Pakistan and helped in identifying potential predictors. The climatic variables considered were identical to DeHart (2011) to draw comparison of climatic behavior while including additional years of study from 2011 to 2023.

Figure 8 depicts different climatic conditions where Figure 8a is Jul–Aug LTM PR for South Asian Region (already explained in Chapter 1) whereas Figure 8b-d identifies corresponding atmospheric conditions. Figure 8b illustrates the formation of the Somali Jet stream along the coast of Somalia. Its primary direction of movement from this point of formation is towards the northeast where it enters the Arabian Sea. This whole process is the major contributor of bringing summertime high precipitation and moisture into Pakistan during the months of Jul–Aug, which is termed as the summer monsoon or the southwest monsoon period. Figure 8c depicts a strong upper-level ridge over a region of low 850 hPa geopotential heights in the lower troposphere as depicted in Figure 8d. As seen in Figure 8a, Pakistan is located on the border of an arid region found towards its west and high precipitation region located towards its east. Therefore, it displays a mix of both portions. One of the primary characteristics for such weather distribution as also highlighted by DeHart (2011) in his study that “the advection from the south of moist air into the lower tropospheric trough leads to high PR in south central Asia and arid conditions in the majority of SWA” (DeHart 2011). The LTM pattern for the climatic variables in Figure 8 was very similar to the results obtained by DeHart (2011).

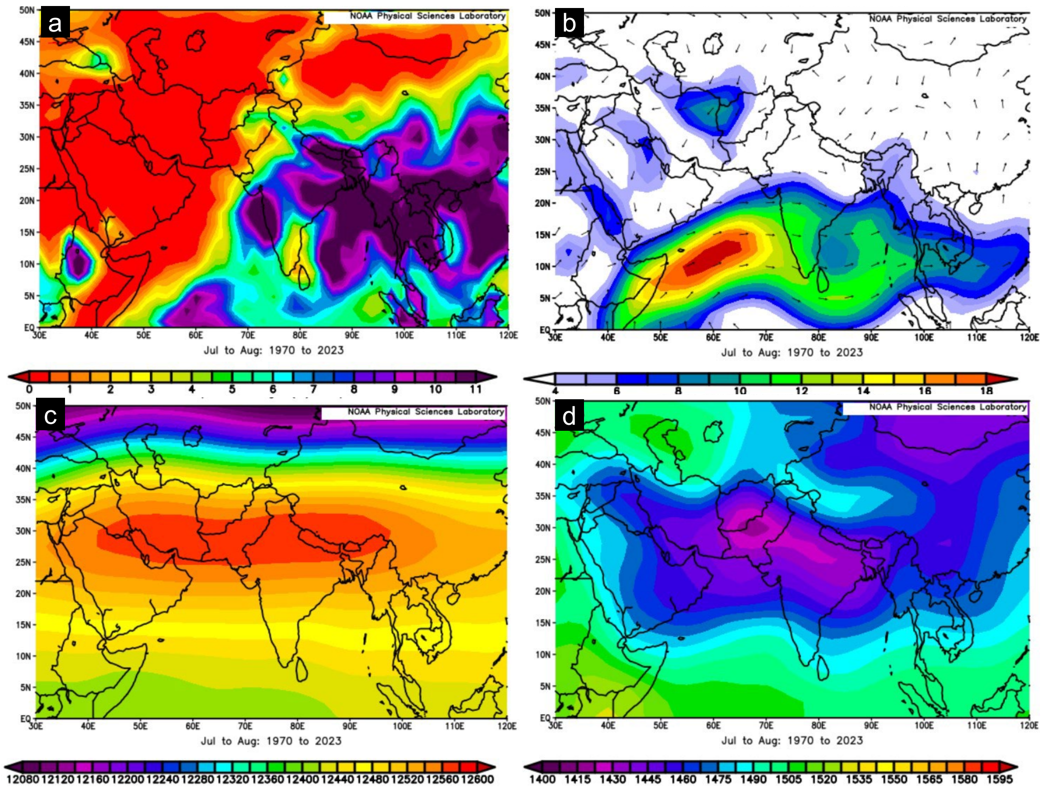

Figure 8. Different LTM atmospheric conditions. Figure 8a illustrates PR (mm/day) in the region, Figure 8b shows 850 hPa vector winds (m/s), Figure 8c depicts 200 hPa GPH (m) and Figure 8d examines 850 hPa GPH (m). Adapted from NOAA Monthly/Seasonal Composites (2023).

A few more atmospheric conditions are highlighted in Figure 9. Repeating Figure 8a, Figure 9a represents LTM PR for the months of Jul–Aug during 1970–2023 in the region with Pakistan at the junction point of high and low levels. Figure 9b is a composite LTM of outgoing longwave radiation (OLR), where low OLR values indicate deeper, higher convective clouds. We see slightly lower OLR values in the study towards the Bay of Bengal region. Figure 9c is a composite of specific humidity (850 hPa). In this case, the greater the value of specific humidity, the greater is the presence of moisture content in the air for precipitation. In the figure we can see higher specific humidity values over the Nepal region. As already explained, the position of Pakistan is on the border portion of high- and low-level precipitation regions. Thus, we can conclude that Pakistan is more likely to see fluctuations in its precipitation rate over the years and is likely to be sensitive to small changes in its surrounding areas. With this consideration, we can assume that for Pakistan above normal PR year can be attributed to the anomalies in advection from the positive side that is southeast of Pakistan, and it may be opposite for negative PR year in Pakistan. Omega is a measure of upward motion in which negative values of omega indicate upward motion that can produce clouds and possibly precipitation. Note in Figure 9d that, in addition to the area of strongest vertical motion over the Nepal and Bhutan region, a secondary region of upward motion over the mountains of western Pakistan indicates the persistent mountain-forced upslope flow during Jul–Aug. Another significant thing was the comparison of these LTM climate conditions with 41 years of work done by DeHart (2011), which revealed no impactful difference during the years 2011–2023.

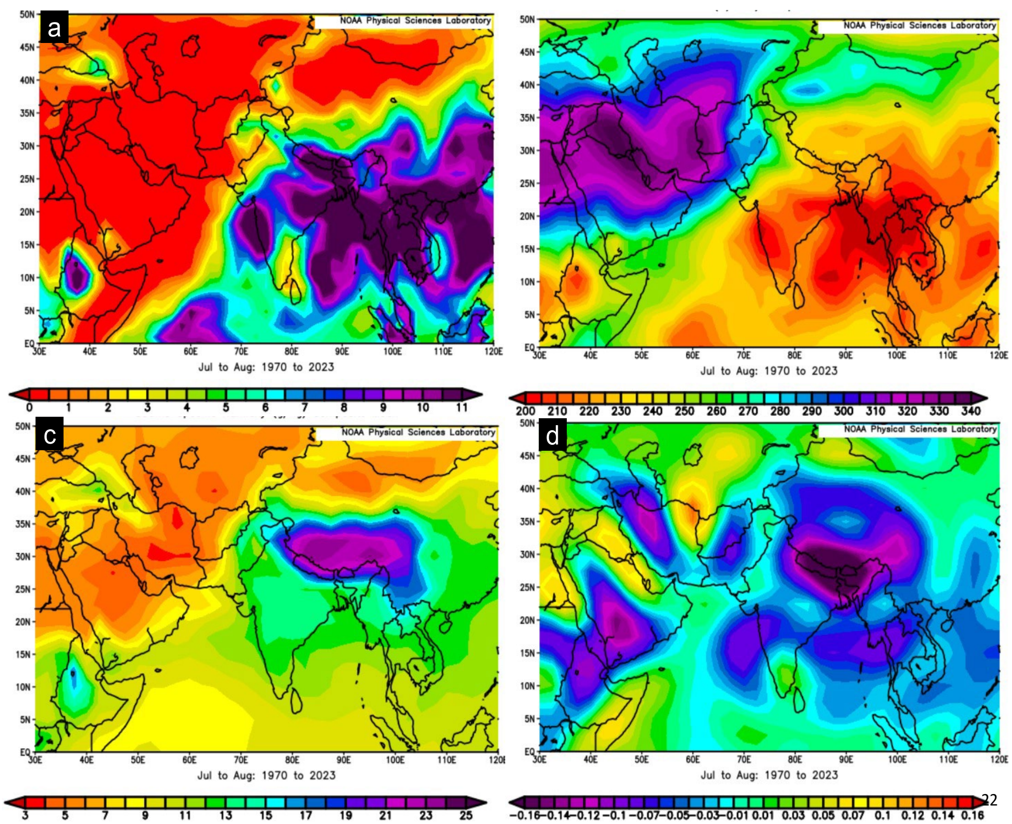

Figure 9. Different LTM atmospheric conditions. Figure 9a depicts PR (mm/day) in the region, Figure 9b examines OLR (W/m2), Figure 9c identifies 850 hPa specific humidity (g/kg) while Figure 9d shows 700 hPa omega (Pa/s). Adapted from NOAA Monthly/Seasonal Composites (2023).

b. Conditional Composites

After deliberate consideration of the basic atmospheric processes of Pakistan and the surrounding regions for the months of Jul–Aug PR, we were required to identify the likely predictors which could be useful for long lead times for PR in Pakistan. Then we followed the already obtained Jul–Aug time series for Pakistan as explained above to select eleven high and eleven low PR years. The high PR years for Predictand region included 1975, 1976, 1988, 1989, 1992, 1994, 1995, 2003, 2010, 2020 and 2022. The low PR years for the Predictand region included 1974, 1980, 1991, 1998, 2007, 2009, 2012, 2013, 2014, 2015 and 2018. The next step was to make conditional composites of anomalies for these high and low PR years which were termed as wet years and dry years composites for the purpose of this study. The main purpose behind preparation of these composite anomalies was to study the anomalies in regional as well as global level patterns of climate that can lead to either high or low PR in our selected predictand region.

For comparison of composite anomalies, we first studied the PR for wet and dry years at the regional level, and the same is also depicted in Figure 10. The red box in the figure for wet and dry years indicates the predicted region. The above normal (AN) precipitation in the predictand region during wet years occurred in conjunction with (1) AN precipitation in south and western India and (2) BN precipitation over Nepal/South China. Similarly, the BN precipitation in the predictand region for dry years can be associated with (1) BN precipitation over south and western India and (2) AN precipitation over Nepal/South China. The pattern of wet years anomalies for precipitation was different from the regions found by DeHart (2011) in his study, particularly over southern India, where we found AN precipitation as compared to BN precipitation found by him. From the plotted predictand region, we also ascertain that most of the PR anomalies within Pakistan are towards the North-Central region of the country, whereas PR towards north most and south most regions is generally consistent with overall BN or low PR. The predictand region has the greatest anomalies for both high and low years.

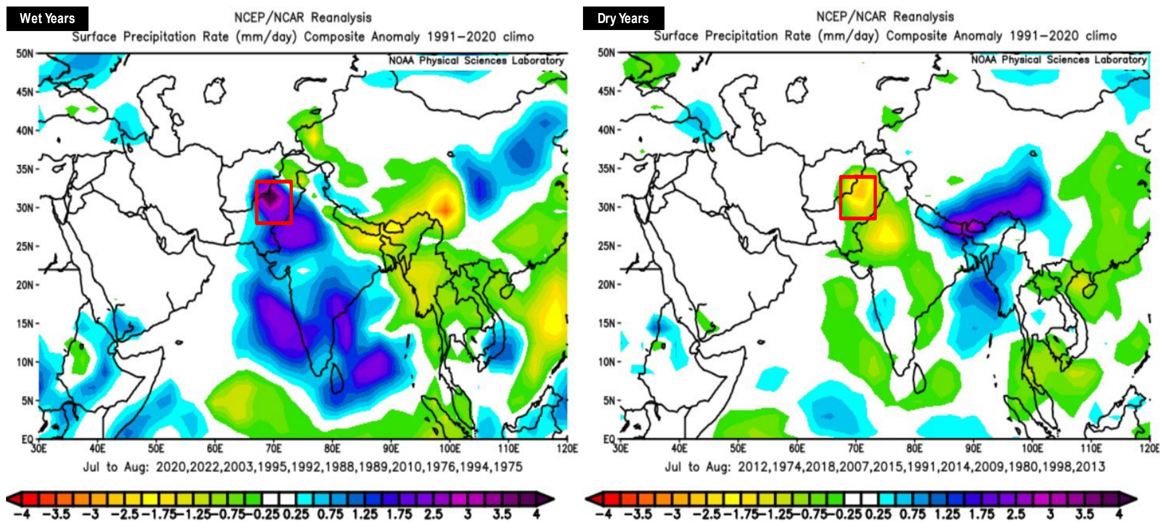

Figure 10. Regional PR anomalies for wet and dry years of the predictand region for the months of Jul–Aug and the duration from 1970 to 2023. Predictand box is shown in red color. Adapted from NOAA Monthly/ Seasonal Composites (2023).

We then studied the SST anomalies for both wet and dry years at the global level for our predictand box, and the same is also displayed in Figure 11. The conditional composite anomaly indicates that a portion of Arabian Sea adjacent to Pakistan was comparatively warmer (cooler) than normal, which could be the reason for higher (lower) precipitation/rainfall during wet (dry) years. Moreover, La Nina (El Nino) signals of lower (higher) SSTs than normal in the equatorial eastern Pacific can also be seen during wet (dry) years, which indicates warm (cool) winds moving towards Asia, resulting in higher (lesser) rainfall. The study carried out by DeHart (2011) did not identify any strong SST anomalies pattern; therefore, he did not include SST in his potential list of predictors. However, the same is not the case in our study as we identify strong SST anomalies for wet years, indicating La Nina signals.

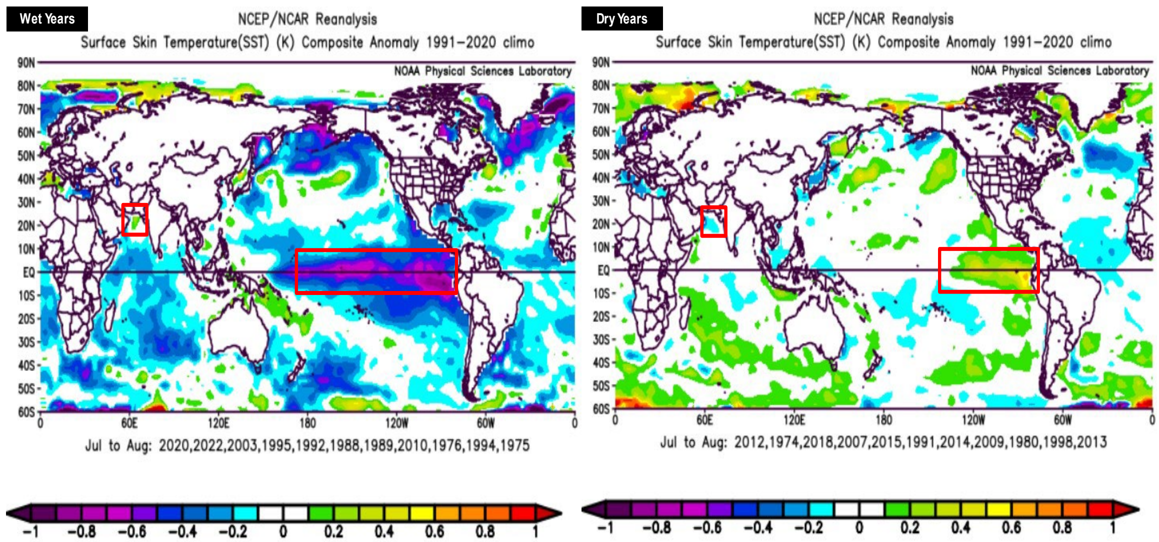

Figure 11. Composite anomalies of SST (°C) in case of eleven high or wet (left side) and eleven low or dry (right side) years of PR in Pakistan for the months of Jul–Aug and the duration from 1970 to 2023. Adapted from NOAA Monthly/Seasonal Composites (2023).

Next, we studied the 850 hPa GPH (m) anomalies for both wet and dry years at the regional level for our predictand box, and the same is also highlighted in Figure 12. Assuming flow is close to geostrophic at 850 hPa,GPH, anomalous wind blows clockwise (counterclockwise) around regions of anomalously high (low) 850 hPa GPH. Wet years have an anomalous flow from SE towards NW, which is different from dry years where it is flowing from NW towards SE. Arrows shown in the figure with red color represent the resultant anomalies of wind, which can likely impact Pakistan. Composite analysis for wet (dry) years highlights an anomalous high (low) over the region encompassing the Nepal and surroundings that generates a damp anomalous southeasterly (dry anomalous northwesterly) flow into Pakistan. This comparison was very close to that observed by DeHart (2011).

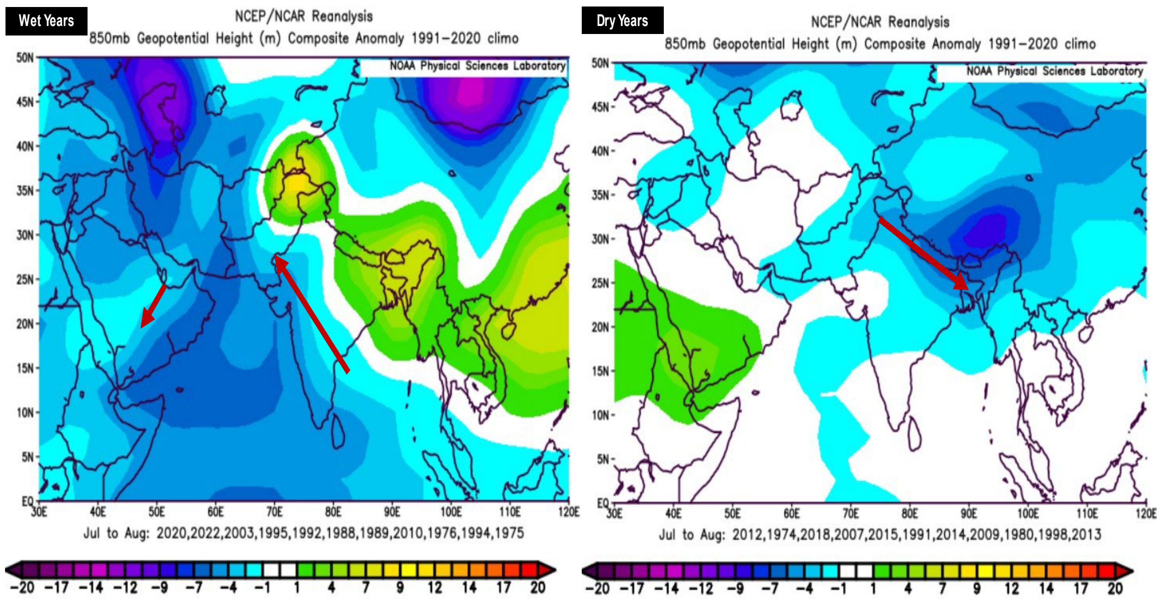

Figure 12. Composite anomalies of GPH (m) in case of eleven high or wet (left side) and eleven low or dry (right side) years of PR in Pakistan for the months of Jul–Aug and the duration from 1970 to 2023. Adapted from NOAA Monthly/Seasonal Composites (2023).

We also analyzed 850 hPa air temperature (°C) anomalies for the wet and dry years in Pakistan at a regional level for our predictand box, and the same is highlighted in Figure 13. Upon analyzing air temperature, we see that wet years result in about a 2 °C decrease in air temperature inside and near the predictand box, with an overall decrease in air temperature towards the Indian ocean region. Whereas dry years resulted in about 1 °C increase in air temperature over the predictand region where we also see a general increase in air temperature towards the Arabian sea/Indian Ocean region. There also appears to be a linkage between an increase (decrease) of air temperature in the Indian Ocean near the Equator with high (low) PR in Pakistan. Air temperature analysis was not done by DeHart (2011) in his study while analyzing composite anomalies.

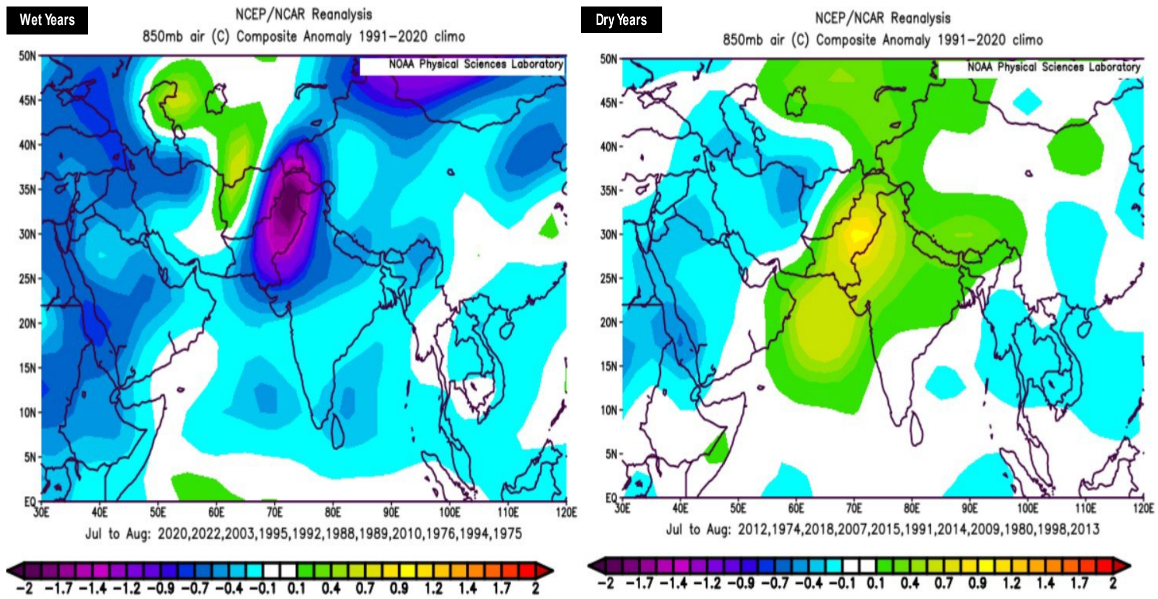

Figure 13. Composite anomalies for 850 hPa air temperature (°C) to include the eleven wet (left side) and eleven dry (right side) years of PR in Pakistan for the months of Jul–Aug and the duration from 1970 to 2023. Adapted from NOAA Monthly/Seasonal Composites (2023).

We also studied conditional composite anomalies for OLR (W/m^2) for the eleven highly wet and dry Pakistan PR years for the months of Jul–Aug for the period from 1970 to 2023, and the same is depicted in Figure 14. It is also known that high PR over Pakistan is associated with low OLR, whereas low PR is linked with high OLR. In Figure 14, we see low OLR values over the predictand region resulting in high PR in Pakistan, whereas we see high OLR values during low PR years. We also see some connection and signals over the Nepal region, which are linked with high or low OLR in Pakistan. When OLR is low over the predictand, then the Nepal region shows a slight increase in OLR which is opposite in the other case.

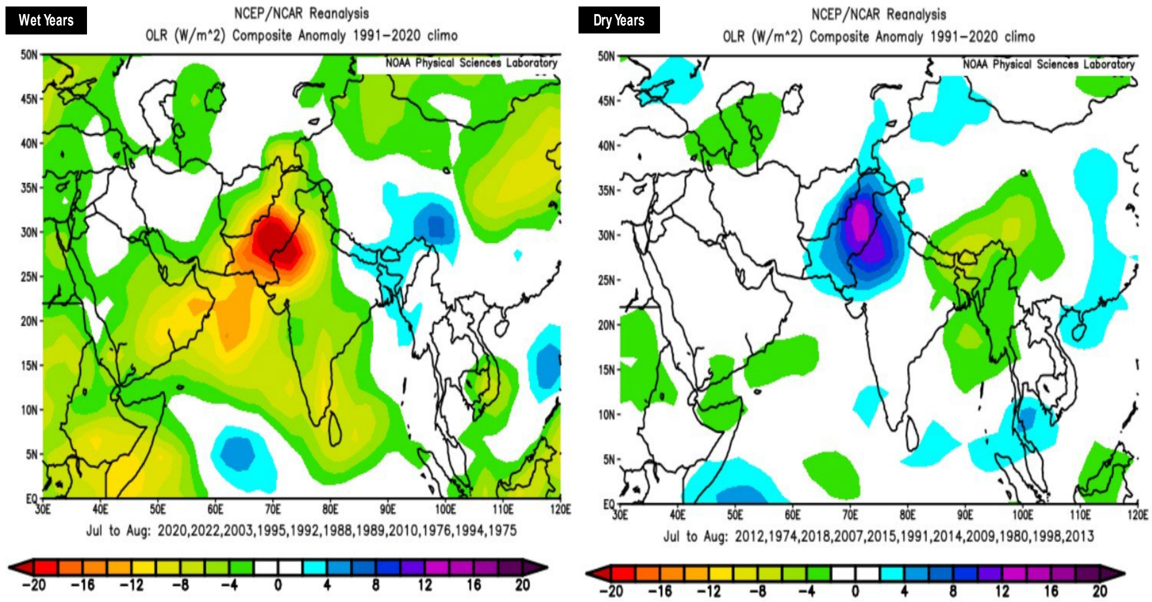

Figure 14. Composite anomalies of OLR (W/m2) for the eleven high or wet (left side) and eleven low or dry (right side) years of PR in Pakistan for the months of Jul–Aug and the duration from 1970 to 2023. Adapted from NOAA Monthly/Seasonal Composites (2023).

Conditional composite anomalies for 850 hPa specific humidity (g/kg) were studied for eleven wet and dry years for predictand PR for Jul–Aug from 1970 to 2023. For the dry years, the signal of anomalously low 850 hPa specific humidity over Pakistan is very similar, albeit a bit weaker, to what DeHart found in his work. However, for wet years the results seem very similar over Pakistan with much greater intensity than what Dehart had found. Signals over the Nepal region look somewhat opposite to over Pakistan when comparing results for wet and dry years.

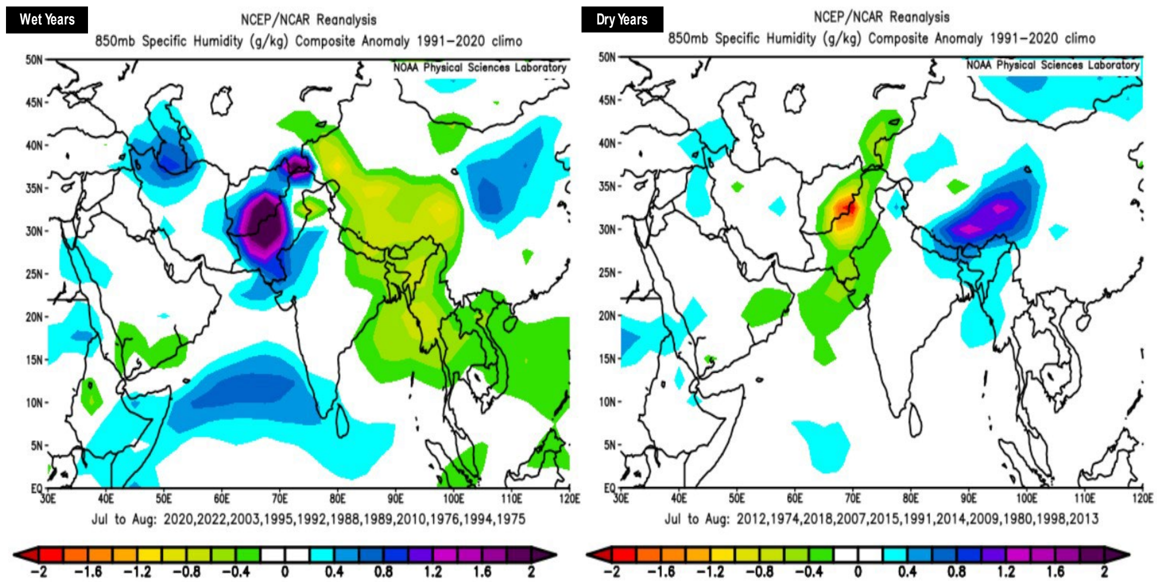

Figure 15. Composite anomalies of 850 hPa specific humidity (g/kg) for the eleven wet (left side) and eleven dry (right side) years of PR in Pakistan for the months of Jul–Aug while covering period from 1970 to 2023. Adapted from NOAA Monthly/Seasonal Composites (2023).

The last conditional composite anomalies of the study included 700 hPa omega (Pa/s) in the case of eleven most wet (dry) Pakistan PR years for the month of Jul–Aug from 1970 to 2023. In terms of omega, the negative anomalies over Pakistan indicate an increase in upward vertical motion during wet years whereas positive anomalies indicate a decrease in upward vertical motion during dry years. For low PR years, the results looked very similar to DeHart, but for high PR years, the northern most part of the country did not show the negative values as found by him, meaning lesser upward vertical motion as compared to southern half of the country where upward motion was high.

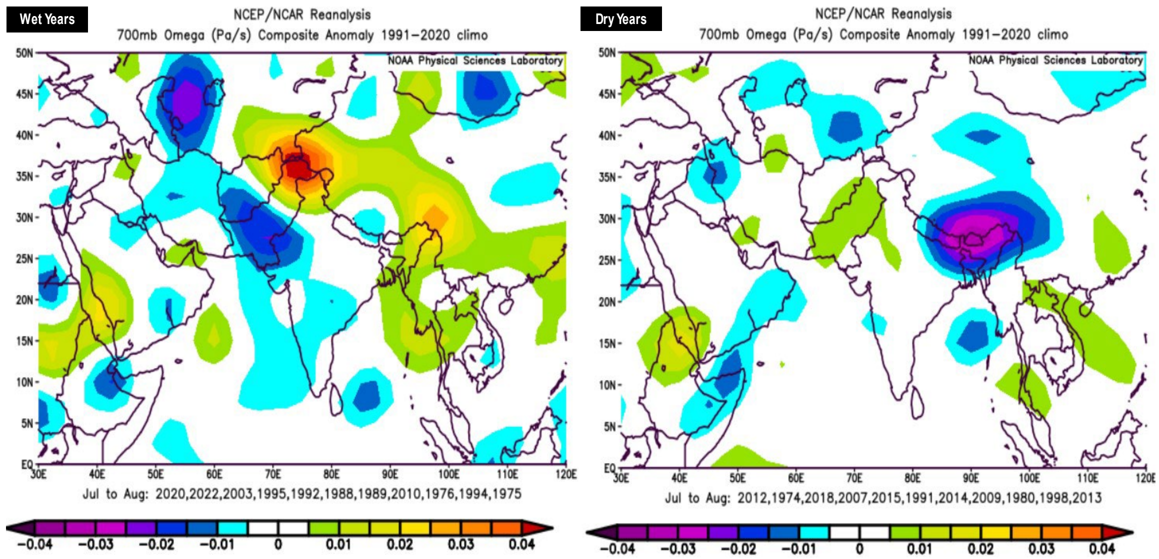

Figure 16. Composite anomalies of 700 hPa omega (Pa/s) in case of eleven wet (left side) and eleven dry (right side) years of PR in Pakistan for the months of Jul–Aug while covering period from 1970 to 2023. Adapted from NOAA Monthly/Seasonal Composites (2023).

4. Identifying Correlations and Studying Teleconnections

After carrying out composite analysis, the next step was to carry out correlations of potential predictors using time series analysis for PR in predictand between 1970–2023 with the predictand lagging the predictors from 0 to 4 months at bi-monthly level. As already explained in Chapter 2, correlations greater than +/- 0.27 were considered statistically significant at a 95% confidence level based on the standard normal distribution of a two tailed t test. The potential climate variables considered as predictors for the study included SST, 850 hPa GPH and 850 hPa air temperature. We restricted ourselves to these three climate variables to draw better comparison with DeHart results where he studied GPH and SST as potential predictors, however, air temperature was an addition in this study.

One of the main criteria that was considered for a potential predictor was for it to have substantial long-lead correlations, while showing some sort of credible teleconnections, with the predictand. To rule out chances of any outliers in terms of correlations and teleconnections, potential predictor boxes with at least 3x3 degrees of area were considered for GPH and 5x5 degrees of area were considered for SST and 850 hPa air temperature. The effectiveness and linkage between various potential predictor variables were considered; however, the emphasis remained on 850 hPa GPH and 850 hPa air temperature at regional scale, with SST at global level scale.

Firstly, we performed correlation of 850 hPa GPH (m) with predictand PR (mm/day) for Jul–Aug for a period from 1970–2023 with lead times of zero to four months at a bi-monthly level. Figure 17 shows the correlation between 850 hPa GPH and the predictand PR with no lead time; here we identify the strong positive correlation over the Nepal/Bhutan area, while negative correlation exists at the Gulf of Aden/Red Sea. To carry out identification of potential predictors, firstly the time series of seasonal mean for Predictand PR was done using the NCEP Reanalysis Dataset. Then, the time series of seasonal mean for 850 hPa GPH (commonly termed as Z850) was done using same NCEP Reanalysis data. Subsequently, correlations of both seasonally averaged variables were plotted using NCEP reanalysis which incorporates specified teleconnection and ocean index time-series. As already identified in Chapter III, regions with +/- 0.27 or higher correlation value were identified as statistically significant, and the same have been marked in Figure 17 for further study at additional lead time. Similarly, for Z850 at one month lead, areas of the Nepal/Bhutan region (positive correlation), Gulf of Aden/Red Sea (negative correlation) and Caspian Sea region (negative correlation) were picked as potential predictors (figure not attached). There was no statistically significant correlation for Z850 at lead times between 2 and 4 months.

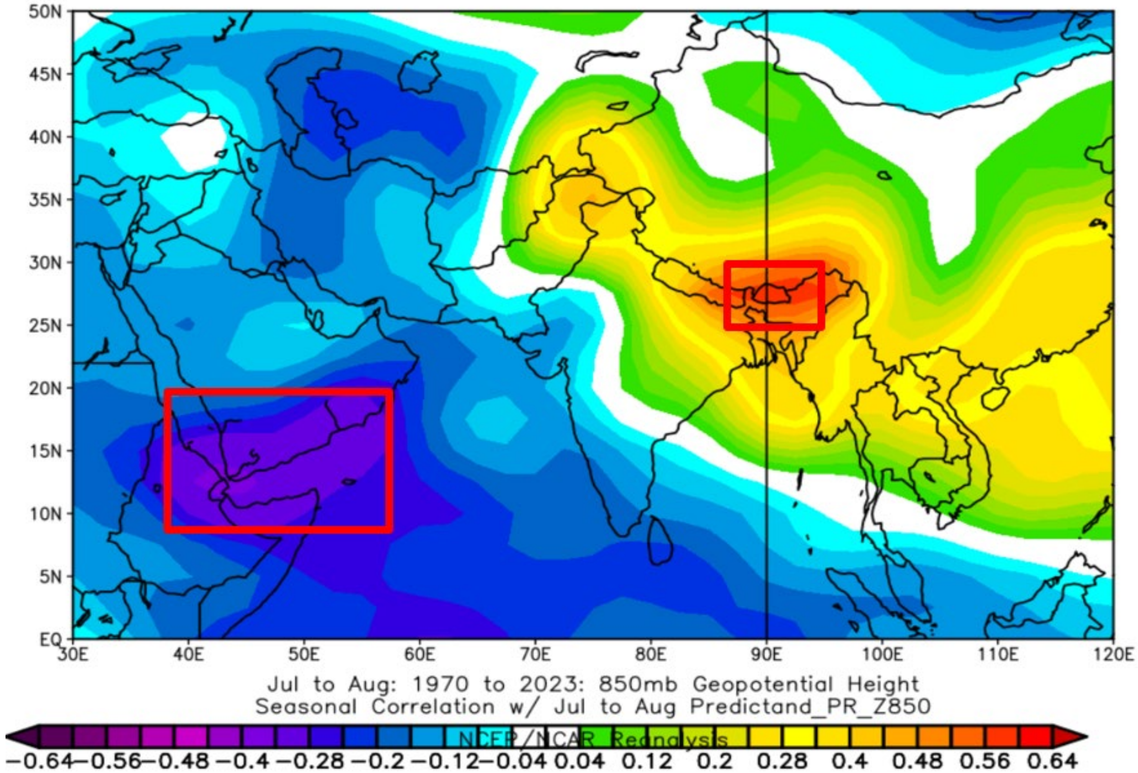

Figure 17. Correlation between Predictand PR (mm/day) with 850 hPa GPH (m) at zero lead. Adapted from NOAA Physical sciences Laboratory (2023).

Secondly, we performed a correlation of SST (°C) with Predictand PR (mm/day) for Jul–Aug for a period from 1970–2023 with lead times of zero to four months at bi-monthly level. Figure 18 shows the correlation of SST with Predictand PR at zero lead. Here we identify a strong negative correlation in the equatorial East Pacific, south central Indian Ocean, SW Indian Ocean, south of the Bering Sea and the African coast region. To carry out identification of potential predictors, firstly the time series of seasonal mean for Predictand PR was done using the NCEP Reanalysis Dataset. Then, the time series of seasonal mean for SST was done using same NCEP Reanalysis data. Subsequently, correlations of both seasonally averaged variables were plotted using NCEP reanalysis which incorporates specified teleconnection and ocean index time-series. Statistically significant correlations are marked in Figure 18 for further study at additional lead time.

Amongst all correlations, the South-Central Indian Ocean displayed the strongest correlation amongst the considered potential predictors.

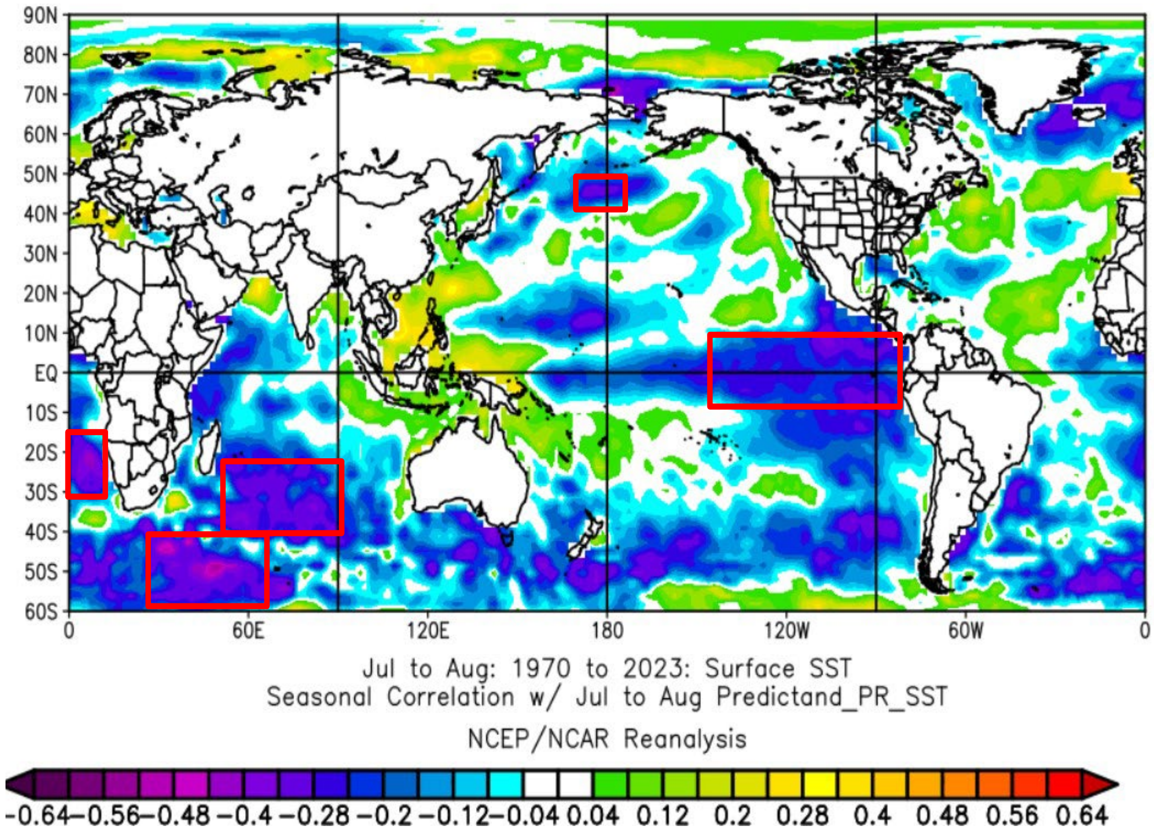

Figure 18. Correlation between Predictand PR (mm/day) with SST (C) at zero lead. Adapted from NOAA Physical sciences Laboratory (2023).

Then, we also performed correlation of SST (°C) with Predictand PR (mm/day) for Jul–Aug at one month lead as shown in Figure 19. Here again we identify strong negative correlation in the East Pacific, the South-Central Indian Ocean, SW Indian Ocean, south of the Bering Sea and the African coast region. Statistically significant correlations are marked in Figure 19 for further study at additional lead times. Amongst all correlations South Central and SW Indian Ocean regions displayed the strongest correlation, which were very consistent for lead times from 2 to 4 months as well and were the strongest potential predictors in SST.

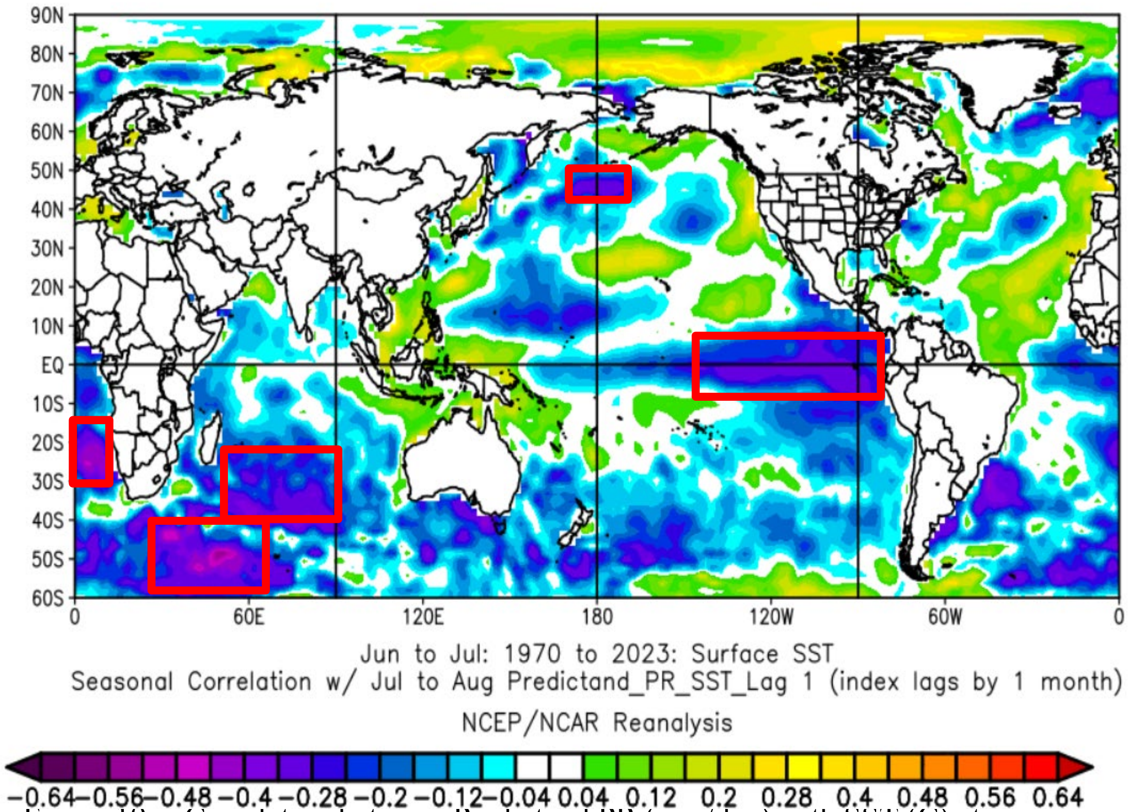

Figure 19. Correlation between Predictand PR (mm/day) with SST (C) at one month lead time. Adapted from NOAA Physical sciences Laboratory (2023).

5. Predictor Selection

For the selection of potential predictors, we initially worked on finalizing a list of initial predictors. The criteria on which this selection was based is that there should be significantly high correlation value between predictand and predictor. Secondly, potential predictors needed to have sufficiently high (statistically significant) long term correlations with the predictand PR. Following this criterion, the list of initial predictors included:

- Nepal Region Z850 (25-30 N and 85–98 E) at zero lead
- Red Sea/Gulf of Aden Z850 (11-14 N and 40–47 E) at zero lead

- Caspian Sea Z850 (39-42 N and 50–54 E) at one-month lead
- East Pacific SST (10 S-8 N and 220–275 E) at zero and one-month lead
- South central Indian Ocean SST (40-28 S and 50–90 E) at 0–4 months lead
- Southwest Indian Ocean SST (60-40 S and 25–70 E) at 0–4 months lead
- South of Bering Sea SST (44-50 N and 171–185 E) at 0–4 months lead
- African Coast SST (30-16 S and 0–11 E) at 0–3 months Lead
- Indian Ocean region TA850 (18-2 S and 70–105 E)

Details of initially investigated predictors to reach about mentioned predictors are also shown in Table 2.

Table 2. Potential predictors analysis. Identifies potential predictors including their correlation with predictand PR at lead times from zero to four months.

Predictor	**Location**	**Location Lat/Long**	**Correlation (zero lead)**	**Correlation (1 month lead)**	**Correlation (2 months lead)**	**Correlation (3 months lead)**	**Correlation (4 months lead)**
850 hPa GPH	Nepal Region	**25-30 N and 85-98 E**	**0.56**	**0.50**	0.21	0.12	0.10
"	Red Sea Region	11-14 N and 40-47 E	**-0.35**	-0.28	0.06	0.10	0.13
"	Caspian Region	**39-42 N and 50-54 E**	-0.21	-0.28	-0.18	0.01	0.02
Surface SST	East Pacific	10 S – 8 N and 220–275 E	-0.28	-0.29	-0.26	-0.17	-0.04
"	South Central Indian Ocean	40 – 28 S and 50 – 90 E	**-0.39**	**-0.36**	**-0.33**	**-0.34**	**-0.32**
"	SW Indian Ocean	60 – 40 S and 25 to 70 E	**-0.36**	**-0.47**	**-0.46**	**-0.39**	**-0.38**

Predictor	Location	Location Lat/Long	Correlation (zero lead)	Correlation (1 month lead)	Correlation (2 months lead)	Correlation (3 months lead)	Correlation (4 months lead)
"	South of Bering Sea	44-50 N and 171–185 E	**-0.28**	**-0.30**	**-0.34**	**-0.34**	**-0.38**
"	African Coast	30-16 S and 0-11 E	**-0.33**	**-0.37**	**-0.38**	**-0.34**	**-0.16**
850 hPa Air Temperature	Central Indian Ocean	18 – 2 S and 70-105 E	**0.35**	0.26	0.15	0.10	0.07

For the final list of potential predictors for statistical modelling, we limited lead times to 0 and 1 month to provide maximum accuracy for the data fed into the statistical model. After that, the preference was on selecting predictors having the highest correlation values during 0–1-month lead time. For SST, consistency of correlations from 0–4 months lead time was also given due significance. After adding these factors, our final 4 predictors for the statistical model included (see Figure 20):

- 850 hPa GPH over Nepal Region at zero lead
- 850 hPa GPH over Red Sea at zero lead
- SST in South Central Indian Ocean at zero lead
- SST in SW Indian Ocean at one month lead

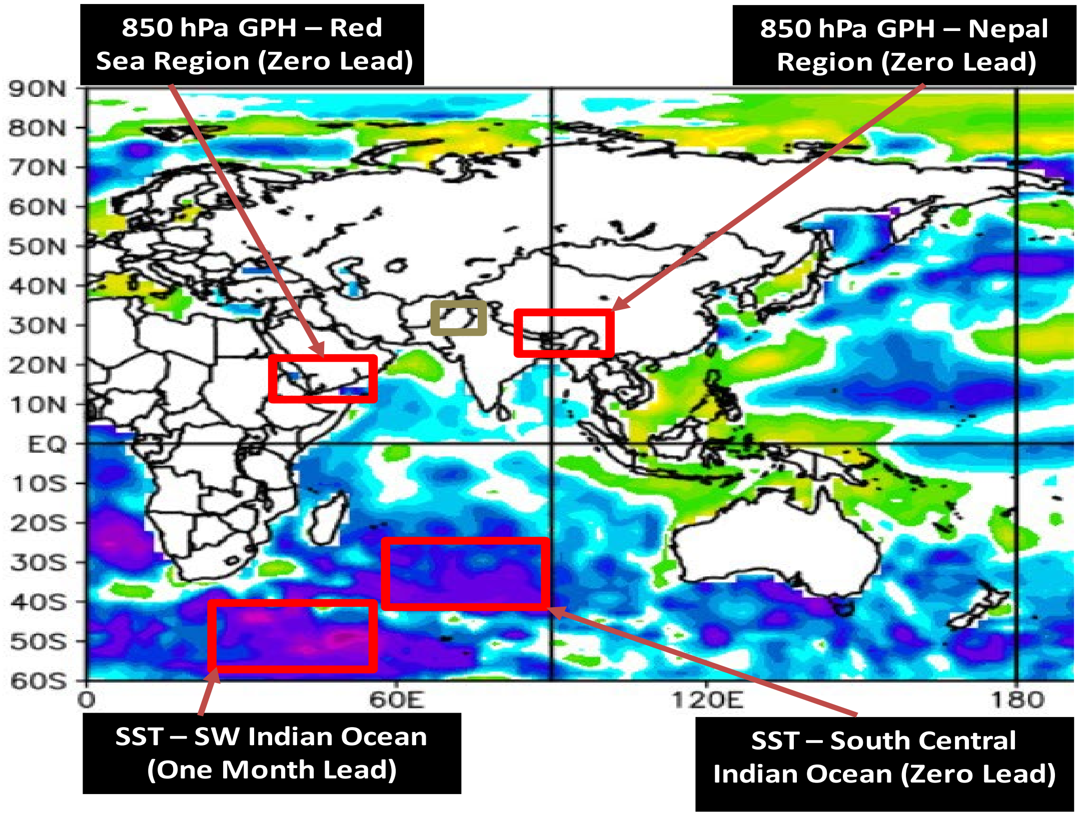

Figure 20. Graphical representation of final four predictors for the predictand region before conducting statistical modelling. Adapted from NOAA Physical sciences Laboratory (2023).

B. STATISTICAL ANALYSIS RESULTS

1. Predictand-Predictor Time Series

After doing correlation analysis, we plotted the time series for Predictand-Predictors to understand the trends of selected predictors with regards to high or low PR event in Predictand region. Figure 21 shows the time series analysis of Pakistan PR (depicted by dark blue color, mm/day) for Jul–Aug for the period from 1970 to 2023 with 850 hPa GPH for the Nepal region (light blue color, m) and the Red Sea region (orange color, m). As identified earlier the correlation with Nepal 850 hPa GPH is positive and very high at 0.56 whereas the correlation with the Red Sea region is negative at -0.35 (both at zero lead), both of which are statistically significant with > 95% of confidence level. The time series is helpful in providing a better depiction of long-term trends about

Pakistan PR and the interannual variations of PR. The results give an indication that 850 hPa GPH for the Nepal and Red Sea are good predictors, because from this time series we can see that most of the focus high and low PR years generally had high 850 hPa GPH for the Nepal region because of positive correlation and low 850 hPa GPH for the Red Sea region due to negative correlation.

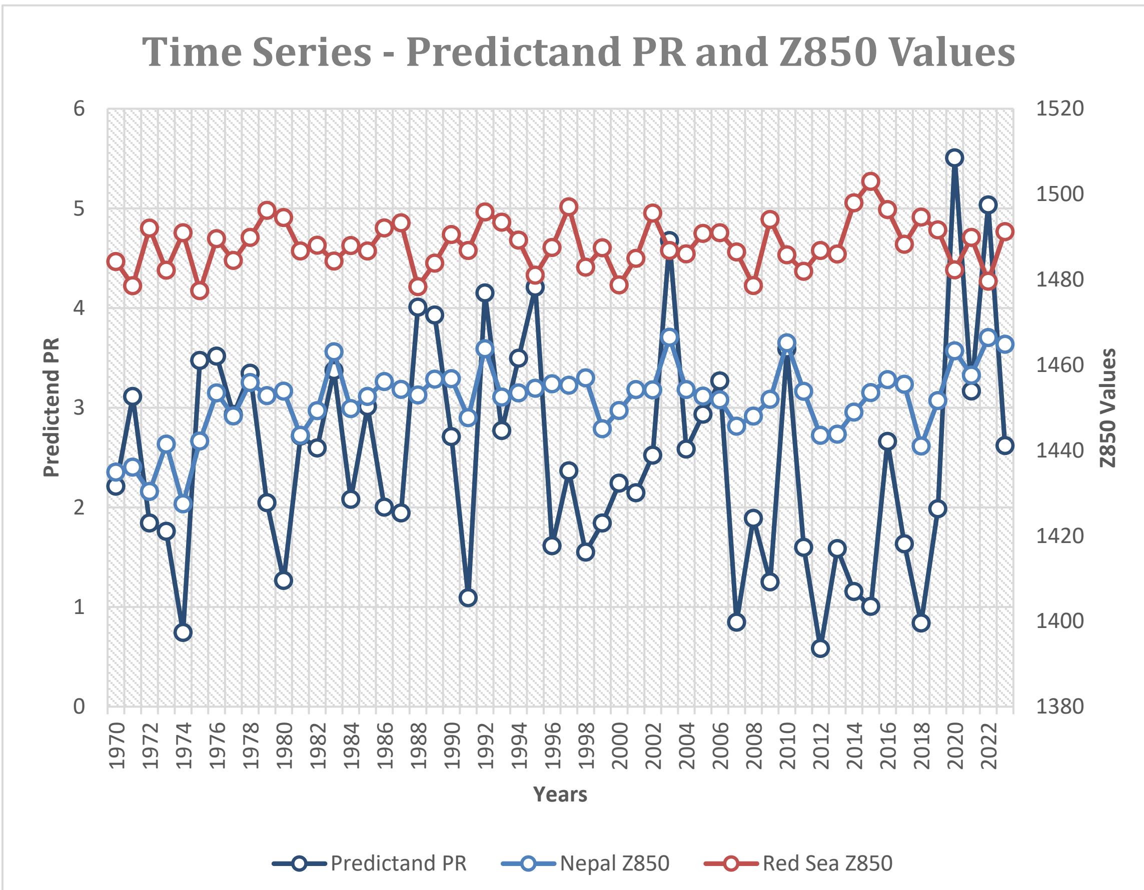

Figure 21. Time series analysis of Pakistan PR (depicted by dark blue color, mm/day) for Jul–Aug for the period from 1970 to 2023 with 850 hPa GPH for the Nepal region (light blue color, m) and Red Sea region (orange color, m). Adapted from (NOAA Monthly Mean Timeseries 2023).

Similarly, Figure 22 shows the time series analysis of Pakistan PR (depicted by dark blue color, mm/day) for Jul–Aug for the period from 1970 to 2023 with SST for the South Central Indian Ocean (orange color, C). As identified earlier the correlation with

SST for the South-Central Indian Ocean is negative at -0.39 at zero lead time which is statistically significant with > 95% of confidence level. The time series is helpful in providing a better depiction of long-term trends about Pakistan PR and the interannual variations of SST. The results give an indication that SST for the South-Central Indian Ocean is good predictor because from this time series we can see that most of the focus high (low) PR years generally had low (high) SST value for SST in the South Central Indian Ocean region due to negative correlation. Similarly, time series between Predictand PR and the Southwest Indian Ocean with a correlation of -0.49 displayed very good results (not shown) in terms of high and low PR years.

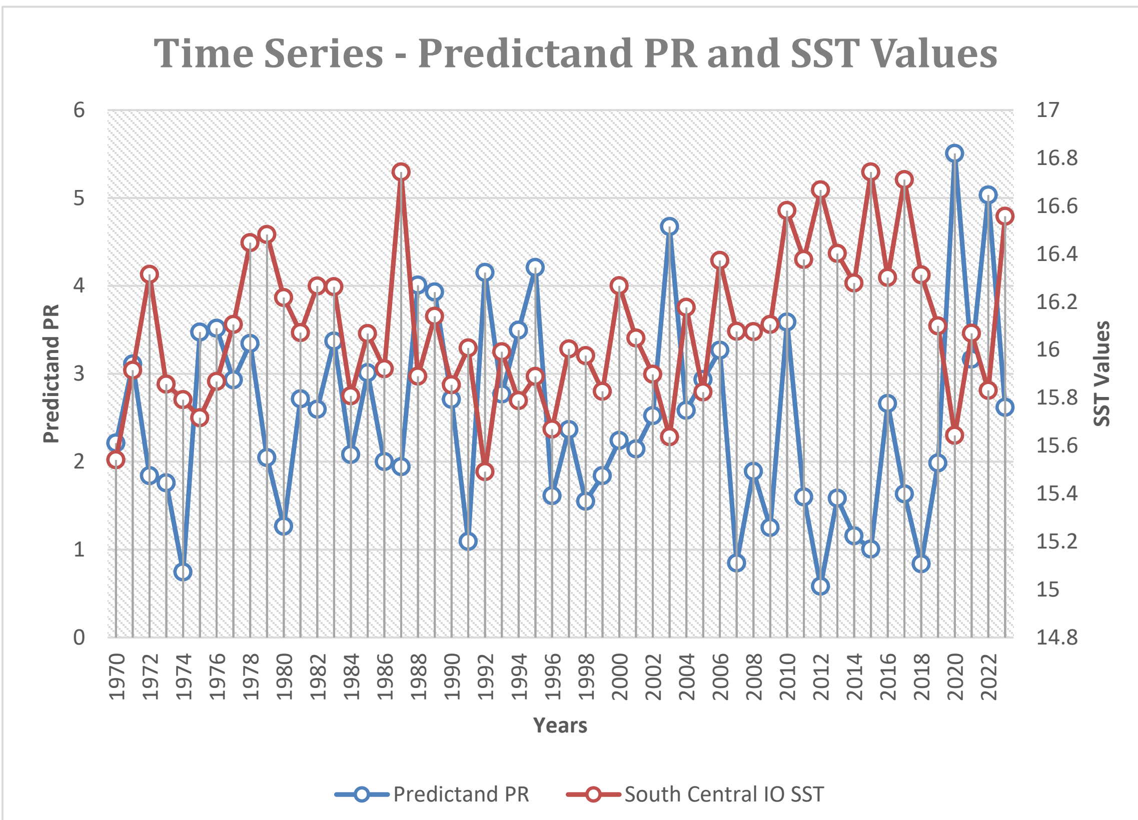

Figure 22. Time series analysis of Pakistan PR (depicted by dark blue color, mm/day) for Jul–Aug for the period from 1970 to 2023 with SST in the South-Central Indian Ocean (orange color, m). Adapted from (NOAA Monthly Mean Timeseries 2023).

2. Long-Range Forecast Method Development

For enhancing the long-range forecasting of the model, we studied lead times from 0 to 4 months. The purpose of studying lead time was to develop a model to enhance the forecasting potential at longer lead times. Initially we performed statistical analysis of models at different lead times and then upon finalizing of predictors, we carried out our statistical modelling. Our final predictors were focused until the lead time of one month. Some of the details of long-range forecasting are covered in subsequent paragraphs.

3. Statistical Analysis

As our Predictor selection method has already been highlighted above, including the finalized list of predictors, here the study will briefly describe the statistical analysis done to finalize the selected predictors. So firstly, we chose four predictors with the highest correlations and long lead time which also include Nepal Z850 for zero as well as one month lead. The list included:

- 850 hPa GPH over Nepal Region at zero lead
- 850 hPa GPH over Nepal Region at one-month lead
- SST in South Central Indian Ocean at zero lead
- SST in SW Indian Ocean at one-month lead

We did single variable linear regression to find R and R square values for each of the predictor-predictor combinations. The reason to perform this regression was that our selected predictors should not be correlated amongst themselves so that predictors do not misinterpret our prediction results due to biases.

Table 3. Statistical analysis of predictor-predictor regression including R and R Square values

Predictor 1	Predictor 2	R/R Square Value	Remarks
Nepal Z850 – Zero Lead	Nepal Z850 – 1 Month Lead	0.87/0.75	High correlation between Nepal Z850 at Zero- and One-month Lead.
"	South Central Indian Ocean SST – Zero lead	0.04/0.00	-
"	Southwest Indian Ocean SST – 1 Month Lead	0.37/0.14	-
Nepal Z850 – 1 Month Lead	South Central Indian Ocean SST – Zero lead	0.05/ 0.002	-
"	Southwest Indian Ocean SST – 1 Month Lead	0.39/0.15	-
Indian Ocean SST – Zero lead	Southwest Indian Ocean SST – 1 Month Lead	0.03/ 0.001	-

Upon checking correlations, we found out that the Nepal Z850 at zero lead and the Nepal Z850 at one month lead were highly correlated (R value of 0.87). We then replaced Nepal Z850 with the next high correlation value, which was selected as the Red Sea region (details already explained). Table 4 shows results of the Red Sea region when correlated with other predictors. The inclusion of Red Sea Z850 led to the final list of predictors.

Table 4. Updated predictor-predictor regression including R and R Square values.

Predictor 1	**Predictor 2**	**R/R Square Value**	**Remarks**
Red Sea Z850 – Zero Lead	Nepal Z850 – Zero Lead	0.08/0.006	New potential predictor with all other selected
"	South Central Indian Ocean SST – Zero lead	0.29/0.08	"
"	Southwest Indian Ocean SST – 1 Month Lead	0.05/0.003	"

4. Statistical Modelling

For statistical modelling for the study, we used the Logistic Regression model (James et al. 2013). The aim for using a classification model was to identify how well we can predict a particular year to be AN, BN or NN by fixing the precipitation rate at a certain threshold. The goal was to identify a particular year based on the four finalized predictors (listed in Chapter 3, Section A.5), and check if that will end up being an AN, BN or NN year or not. The logistic regression model for the study used the following criteria:

- Above Normal - < 3 (mm/day)
- Near Normal - 1.5 to 2.5 (mm/day)
- Below Normal - >1.5 (mm/day)

The procedure we used for our statistical modelling is described below:

- Develop R code to perform binary classification for the 54 years dataset based on a threshold value as AN, BN or NN.
- Identify threshold values of 1.5, 2.5 and 3 (mm/day) to decide for a particular year to be BN, NN or AN respectively.
- Create a new column in the data frame and filling it with binary values (0 or 1) based on whether precipitation exceeded certain threshold level.
- Divide the data into 70% train (37 years) and 30% test dataset (17 years) for the study.
- Compare the observations in the test data between predicted classifications and the actual classifications to create a confusion matrix for evaluating the performance of the classification.
- Create an ROC curve based on the predicted probabilities and actual classifications, and then plotting curve with the AUC printed.
- Evaluate performance of the logistic regression model using ROC and AUC.

5. Model – Results

a. Below Normal Threshold – 1.5 (mm/day)

When we kept below normal (BN) threshold, the model gave an accuracy of about 88% for our test data of 30% (17 years) regarding correct identification of the PR year.

However, the AUC value was about 69% of the model variations covered, which was not very high.

Table 5. Confusion matrix – Low threshold predictand PR.

		Actual Value	
		0	1
Model Prediction	0	1	0
	1	2	14

Figure 23 indicates the model fit at 1.5 mm/day threshold, from these results Nepal Z850 and Red Sea Z850 are statistically significant predictors at the 0.05 confidence level. ROC curve details are shown in Figure 24.

```
Coefficients:
                                       Estimate Std. Error z value Pr(>|z|)
(Intercept)                            319.9692   262.9395   1.217   0.2236
Nepal.Z850...Zero.Lead                   0.2096     0.0967   2.167   0.0302 *
Red.Sea.Z850...Zero.Lead                -0.4464     0.2264  -1.972   0.0486 *
South.Central.Indian.Ocean.SST...Zero.Lead  1.4956  3.2553   0.459   0.6459
Southwest.Indian.Ocean.SST...1.Month.Lead   4.5909  3.2017   1.434   0.1516
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 23. Model fit at below normal threshold of 1.5 mm/day.

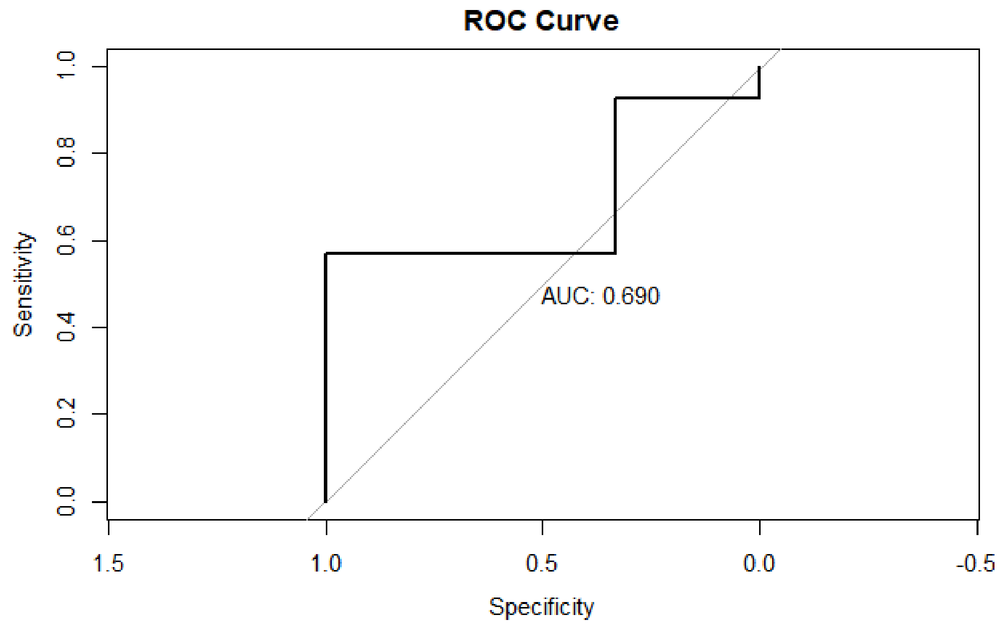

Figure 24. ROC curve for logistic regression model using low threshold.

b. Near Normal Threshold – 2.5 (mm/day)

When we kept the near normal (NN) threshold, the model gave an accuracy of about 76% for correct identification of the PR year. However, the AUC value was the highest in this case with about 90% of the model variations captured by the model.

Table 6. Confusion matrix – Near normal threshold for predictand PR.

		Actual Value	
		0	1
Model Prediction	0	7	2
	1	2	6

Figure 25 indicates the model fit at 2.5 mm/day threshold, from these results none of the predictors are statistically significant at the 0.05 confidence level. ROC curve details are shown below.

```
Coefficients:
                                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                                68.03022  129.08095   0.527   0.5982
Nepal.Z850...Zero.Lead                      0.09619    0.06678   1.440   0.1498
Red.Sea.Z850...Zero.Lead                   -0.11020    0.08424  -1.308   0.1908
South.Central.Indian.Ocean.SST...Zero.Lead -1.90365    1.50760  -1.263   0.2067
Southwest.Indian.Ocean.SST...1.Month.Lead  -2.94505    1.77639  -1.658   0.0973
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 25. Model fit at near normal precipitation threshold of 2.5 mm/day.

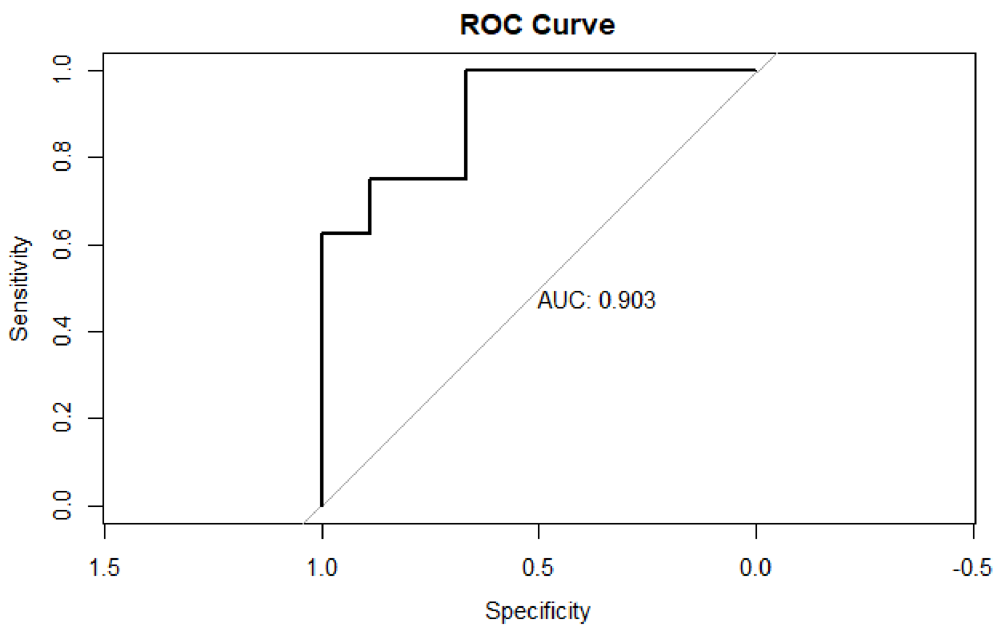

Figure 26. ROC curve for logistic regression model using near normal threshold.

c. *Above Normal Threshold – 3 (mm/day)*

When we kept the above normal (AN) threshold, the model gave an accuracy of about 82% in terms of correct identification of PR year, which was good keeping in view that we are more focused about results of high precipitation years. However, the AUC value was about 76% of the model variations captured by the model.

Table 7. Confusion matrix – Above normal threshold for predictand PR.

		Actual Value	
		0	1
Model Prediction	0	13	2
	1	1	1

Figure 27 indicates the model fit at 3 mm/day threshold, from these results Nepal Z850 is statistically significant predictors at the 0.05 confidence level. ROC curve details are shown in Figure 28.

```
Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                           44.89973  176.51928   0.254   0.7992
Nepal.Z850...Zero.Lead                 0.22328    0.09458   2.361   0.0182 *
Red.Sea.Z850...Zero.Lead              -0.20077    0.12678  -1.584   0.1133
South.Central.Indian.Ocean.SST...Zero.Lead -3.29593  2.02676  -1.626   0.1039
Southwest.Indian.Ocean.SST...1.Month.Lead -4.25019  2.27627  -1.867   0.0619 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 27. Model fit for above normal precipitation threshold of 3 mm/day.

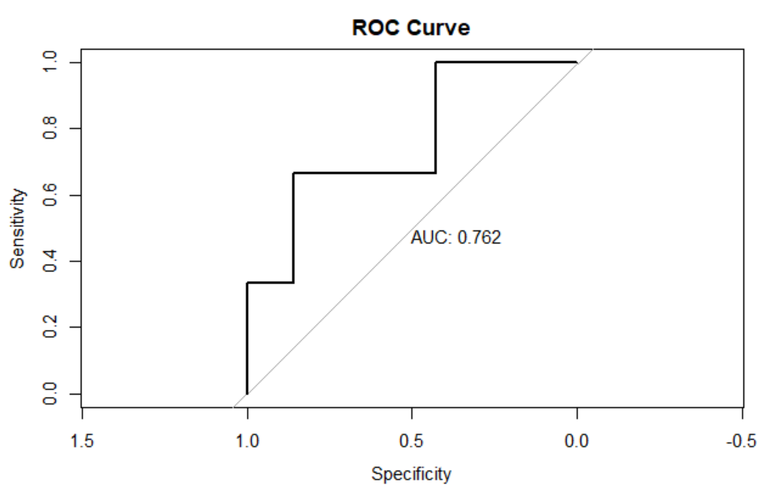

Figure 28. ROC curve for logistic regression model using above normal threshold.

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY, CONCLUSION AND RECOMMENDATIONS

A. SUMMARY OF METHODS AND RESULTS

The study undertaken focused on understanding the drivers for extreme precipitation events in Pakistan while also focusing on developing a model to predict abnormal precipitation along with the factors involved in it. The primary consideration behind the study was two abnormal precipitation events in Pakistan which caused large scale damage to both human lives and the infrastructure.

The study focused on choosing the predictand months and predictand region which could give us a good representation of abnormal precipitation. Using the monthly/ seasonal mean time series from NCEP Reanalysis dataset, we concluded that July and August have far greater amount of precipitation in Pakistan than any other single month. Similarly, we used composite climate analysis from NOAA to identify high and low precipitation years. This led to creating predictand boxes which captured the composite anomalies, and we finally selected a predictand box (28º–34ºN, 67º–73ºE) which we studied for abnormal precipitation.

After finalizing the predictand region, we worked on selecting the potential predictors associated with weather changes or patterns in Pakistan. We studied various atmospheric variables for potential predictors while selecting a basic criterion for predictors to avoid any outliers. We also used composite climate analysis including LTM and composite anomalies to analyze atmospheric variables which helped us to understand various weather patterns while leaving us with 850 hPa GPH, SST, and 850 hPa air temperature as variables to explore for our potential predictors.

Subsequently, we used linear correlations in atmospheric seasonal/monthly averages available at NOAA NCEP Reanalysis dataset. This helped us to identify correlations between the predictand PR and regions of potential atmospheric predictors. We use correlation greater than -/+ 0.27 as statistically significant at a 95% confidence level based on the standard normal distribution of a two tailed t test. We studied these correlations at zero to four months lead time to provide long range forecasting in terms of our selected predictors.

We then conducted statistical analysis including linear regression to verify that our potential predictors are not correlated with the aim to avoid any biases in the model. This led to our four finalized predictors to include 850 hPa GPH over the Nepal region at zero lead, 850 hPa GPH over the Red Sea region at zero lead, the South-Central Indian Ocean SST at zero lead and the Southwest Indian Ocean region at one month lead. The primary criteria to shortlist predictors remained higher correlation values and consistency over longer lead times.

We then used logistic linear regression model to predict the data set for a specific year based on 54 years of data available. After distributing 70% of data as train and remaining 30% as test data, we evaluate the performance of the model using confusion matrix for different thresholds.

Results gave us a good fit for each type of precipitation threshold (1.5, 2, 3 mm/day) that we used. As evident from the results, we concluded that logistic regression model displayed greater accuracy when it analyzed 3 mm/day and 1.5 mm/day as a precipitation threshold, which is useful in predicting any abnormal event on the bases of these identified predictors for both high and low precipitation years.

Another significance of greater accuracy for the model at higher threshold is that any unusual precipitation is likely to have precipitation threshold of at least 3 mm/day or more as seen in this study. Model accuracy can be further enhanced by using a different set of predictors or trying an even higher precipitation threshold; however, its relevance with the precipitation levels in predictand region needs to be carefully observed.

B. RECOMMENDATIONS FOR FUTURE RESEARCH

With ever evolving nature of climate around the world, particularly due to the pronounced effects of climate change, the field of meteorological study has avenues to explore. Statistical analysis of these climate changes can help in understanding various key phenomena that play a part in significant weather changes. The topic I undertook for my study also has an immense scope for exploration to contribute towards better understanding of the subject. Some of the recommendations for the future study on this topic are highlighted below.

1. Future research may include linking the effects of abnormal precipitation events with climate change to get the knowledge about significant predictors that contribute towards climatic changes. This would also give a better comprehension about which predictors are important to analyze the effects of climate change around the world.

2. Research work related to Pakistan precipitation can also be done for a different predictand region and comparison can be drawn about predictors relevant to both the regions while also understanding the major differences in term of climatic behavior.

3. Exploration of some other climatic variables like surface winds and specific humidity can be incorporated in future work to understand their effects and patterns to access them as potential predictors while studying precipitation events in Pakistan. Similarly, an effort can be made to enhance the lead times from 0 to 6 months to see which predictors are still relevant at longer leads.

4. Future work might also incorporate a deeper analysis of the dynamics that drive the individual 850 hPa geopotential height boxes and SST boxes, including their relationship with predictand. Some of the other potential predictors in the study can also be used as final predictors, and their effects can also be studied.

5. Study other statistical analysis tools to understand the accuracy of those models, as in my study I resorted to logistic regression modelling. Similarly, other classification models like random forest can be applied to same predictand region to understand the difference in accuracy levels and the behavior of these models.

6. Pakistan lies in the region where effects of climate change are pronounced in the form of abnormal weather events; therefore, any future abnormal precipitation event in Pakistan can also be studied in the light of the two events (large scale flooding of 2010 and 2022) that were analyzed in this paper. This will be highly relevant for the selection of any future predictand region in Pakistan.

LIST OF REFERENCES

Abbas A, Ullah S, Ullah W, Zhao C, Karim A, Waseem M, Bhatti AS, Ali G, Jan MA, Ali A (2023) Characteristics of Winter Precipitation over Pakistan and Possible Causes during 1981–2018. *Water* 15(13):2420.

American Institute of Pakistan Studies (2023) Geography | American Institute of Pakistan Studies. Retrieved October 24, 2023, https://www.pakistanstudies-aips.org/pakistan/geography.

Bhatti AS, Wang G, Ullah W, Ullah S, Fiifi Tawia Hagan D, Kwesi Nooni I, Lou D, Ullah I (2020) Trend in Extreme Precipitation Indices Based on Long Term In Situ Precipitation Records over Pakistan. *Water* 12(3):797.

Calvin K, Dasgupta D, Krinner G, Mukherji A, Thorne PW, Trisos C, Romero J et al. (2023) *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.* First. (Intergovernmental Panel on Climate Change (IPCC)).

Climate Prediction Center (2023) Climate prediction center – Vision & mission. Retrieved November 21, 2023, https://www.cpc.ncep.noaa.gov/information/who_we_are/mission/.

Crook JC (2009) Climate analysis and long range forecasting of dust storms in Iraq. Master’s thesis, Department of Meteorology, Monterey, California, Naval Postgraduate School.

DeHart JA (2011) Long-range forecasting in support of operations in Pakistan. Master’s thesis, Department of Meteorology, Monterey, California, Naval Postgraduate School.

Di Capua G, Sparrow S, Kornhuber K, Rousi E, Osprey S, Wallom D, van den Hurk B, Coumou D (2021) Drivers behind the summer 2010 wave train leading to Russian heatwave and Pakistan flooding. *NPJ Clim Atmos Sci* 4(1):1–14.

van den Dool H (2007) *Empirical Methods in Short-Term Climate Prediction* (Oxford University Press, United Kingdom).

ECMWF (2023) *ECMWF*. Retrieved October 26, 2023, https://www.ecmwf.int/.

Eric W. Danielson, James Levin, Elliot Abrams (2003) *Meteorology* 2nd ed. (McGraw Hill).

ESRL (2023) Composite Monthly Mean Plot: NOAA Physical Sciences Laboratory. Retrieved November 6, 2023, https://psl.noaa.gov/cgi-bin/data/composites/comp.pl?var=Precipitation+Rate&level=Surface&mon1=6&mon2=7&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&iy=&ipos%5B1%5D=1970&ipos%5B2%5D=2023&ineg%5B1%5D=&ineg%5B2%5D=&timefile0=&tstype=0&timefile1=&value=&typeval=1&compval=1&lag=0&labelcolor=Color&labelshaded=Shaded&type=3&scale=100&contourlabel=1&switch=0&cint=0.5&lowr=0&highr=11&proj=Custom&xlat1=0&xlat2=50&xlon1=30&xlon2=120&custproj=Cylindrical+Equidistant&level1=1000mb&level2=10mb&Submit=Create+Plot.

Galarneau TJ, Hamill TM, Dole RM, Perlwitz J (2012) A Multiscale Analysis of the Extreme Weather Events over Western Russia and Northern Pakistan during July 2010. *Monthly Weather Review* 140(5):1639–1664.

Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani (2013) *An Introduction to Statistical Learning with Applications in R* 7th ed. (Springer Science+Business Media New York).

Gillies SD (2012) *A Statistical Multimodel Ensemble Approach to Improving Long-Range Forecasting in Pakistan*. (Monterey, California. Naval Postgraduate School).

Henson B (2022) Cruel echoes of a 2010 disaster in Pakistan's catastrophic 2022 floods » Yale Climate Connections. *Yale Climate Connections*. Retrieved October 26, 2023, http://yaleclimateconnections.org/2022/08/cruel-echoes-of-a-2010-disaster-in-pakistans-catastrophic-2022-floods/.

Hurrell JW, Deser C (2009) North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems* 78(1):28–41.

Imran A (2013) Temporal trends in the peak monsoonal precipitation events over northeast Pakistan. *Pakistan Journal of Meteorology* 10.

IRI (2023) IRI – International Research Institute for Climate and Society | Climate.

Joseph Hun-wei Lee (2022) Pakistan floods 2022 – An interview by the IAHR president to share expert insights. Retrieved October 26, 2023, https://www.iahr.org/index/detail/776.

Kalnay et al. (1996) NCEP-NCAR Reanalysis 1: NOAA Physical Sciences Laboratory NCEP-NCAR reanalysis 1. Retrieved October 31, 2023, https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html.

Kripalani RH, Oh JH, Kulkarni A, Sabade SS, Chaudhari HS (2007) South Asian summer monsoon precipitation variability: Coupled climate model simulations and projections under IPCC AR4. *Theoretical and Applied Climatology* 90(3–4):133–159.

Lemke BD (2010) Long-range forecasting in support of operations in the Horn of Africa. Master's thesis, Department of Meteorology, Monterey, California, Naval Postgraduate School.

Ma Q, Lei H, Feng T, Hu R, Niu M, Hu Z, Feng G (2023) Impact of spring Tibetan Plateau snow cover on extreme precipitation in Pakistan in July and August 2022. *Atmospheric Research* 295:107007.

Ma Y, Hu X, Chen Y, Hu Z, Feng T, Feng G (2023) Different characteristics and drivers of the extraordinary Pakistan rainfall in July and August 2022. *Remote Sensing* 15(9):2311.

Meyer DW (2007) Relationships between global warming and tropical cyclone activity in the Western North Pacific. Master's thesis, Department of Operations Research, Naval Postgraduate School.

Nanditha JS, Kushwaha AP, Singh R, Malik I, Solanki H, Chuphal DS, Dangar S, Mahto SS, Vegad U, Mishra V (2023) The Pakistan Flood of August 2022: Causes and Implications. *Earth's Future* 11(3):e2022EF003230.

National Geographic (2019) Global Warming Effects. *Environment*. Retrieved October 22, 2023, https://www.nationalgeographic.com/environment/article/global-warming-effects.

NCAS (2023) What causes weather? *NCAS*. Retrieved November 21, 2023, https://ncas.ac.uk/learn/what-causes-weather/.

NOAA (2023) About our agency | National Oceanic and Atmospheric Administration. Retrieved October 26, 2023, https://www.noaa.gov/about-our-agency.

NOAA Monthly Mean Timeseries (2023) Monthly mean timeseries: NOAA Physical Sciences Laboratory. Retrieved November 21, 2023, https://psl.noaa.gov/cgi-bin/data/timeseries/timeseries1.pl.

NOAA Monthly/Seasonal Composites (2023) Monthly/seasonal composites: NOAA Physical Sciences Laboratory. Retrieved November 21, 2023, https://psl.noaa.gov/cgi-bin/data/composites/printpage.pl?var=Air%20Temperature;level=850mb;mon1=6;mon2=7;iy=2012;iy=1974;iy=2018;iy=2007;iy=2015;iy=1991;iy=2014;iy=2009;iy=1980;iy=1998;iy=2013;iy=;iy=;iy=;iy=;iy=;iy=;iy=;iy=;iy=;ipos%5B1%5D=;ipos%5B2%5D=;ineg%5B1%5D=;ineg%5B2%5D=;timefile0=;tstype=0;timefile1=;value=;typeval=1;compval=1;lag=0;labelcolor=Color;labelshaded=Shaded;type=2;scale=100;contourlabel=1;switch=0;cint=0.1;lowr=-2;highr=2;proj=Custom;xlat1=-60;xlat2=90;xlon1=0;xlon2=360;custproj=Cylindrical%20Equidistant;level1=1000mb;level2=10mb;Submit=Create%20Plot.

NOAA Physical sciences Laboratory (2023) Linear Correlations in Atmospheric Seasonal/Monthly Averages: NOAA Physical Sciences Laboratory. Retrieved November 27, 2023, https://psl.noaa.gov/data/correlation/.

NOAA PSL (2023) PSL Climate/Weather Products: NOAA Physical Sciences Laboratory. Retrieved November 21, 2023, https://psl.noaa.gov/products/.

NOAA PSL Monthly Mean Timeseries: NOAA Physical Sciences Laboratory. Retrieved November 6, 2023, https://psl.noaa.gov/cgi-bin/data/timeseries/timeseries1.pl.

North Atlantic Oscillation (2023) North Atlantic Oscillation (NAO) | National Centers for Environmental Information (NCEI). Retrieved November 6, 2023, https://www.ncei.noaa.gov/access/monitoring/nao/.

Otto FEL, Zachariah M, Saeed F, Siddiqi A, Kamil S, Mushtaq H, Arulalan T et al. (2023) Climate change increased extreme monsoon rainfall, flooding highly vulnerable communities in Pakistan. *Environ. Res.: Climate* 2(2):025001.

Shaevitz DA, Nie J, Sobel AH (2016) The 2010 and 2014 floods in India and Pakistan: dynamical influences on vertical motion and precipitation. (March 3) http://arxiv.org/abs/1603.01317.

Statstest (2023) Simple logistic regression. *StatsTest.com*.

Tajbar S, Khorshiddoust AM, Asl SJ (2023) Impacts of large scale climate drivers on precipitation in Sindh, Pakistan using machine learning techniques. *IDŐJÁRÁS = Quarterly Journal of the Hungarian Meteorological Service* 127(3):321–346.

Tandon NF, Zhang X, Sobel AH (2018) Understanding the dynamics of future changes in extreme precipitation intensity. *Geophysical Research Letters* 45(6):2870–2878.

Tillman C (2023) The interdependence of climate security and good governance: A case study from Pakistan. *The Center for Climate & Security*. Retrieved October 26, 2023, https://climateandsecurity.org/2023/10/the-interdependence-of-climate-security-and-good-governance-a-case-study-from-pakistan/.

UK Met Office (2023) Services. *Met Office*. Retrieved November 21, 2023, https://www.metoffice.gov.uk/services.

Ullah I, Ma X, Yin J, Saleem F, Syed S, Omer A, Habtemicheal BA, Liu M, Arshad M (2022) Observed changes in seasonal drought characteristics and their possible potential drivers over Pakistan. *International Journal of Climatology* 42(3):1576–1596.

Ullah W, Karim A, Ullah Sami, Rehman AU, Bibi T, Wang G, Ullah Safi et al. (2023) An increasing trend in daily monsoon precipitation extreme indices over Pakistan and its relationship with atmospheric circulations.

United Nations News (2023) UN continues to support Pakistan flood response | UN News. Retrieved October 23, 2023, https://news.un.org/en/story/2023/03/1134302.

U.S. Department of Commerce (2023) About the NWS. Retrieved October 26, 2023, https://www.weather.gov/about/.

Vellore RK, Kaplan ML, Krishnan R, Lewis JM, Sabade S, Deshpande N, Singh BB, Madhura RK, Rama Rao MVS (2016) Monsoon-extratropical circulation interactions in Himalayan extreme rainfall. *Clim Dyn* 46(11):3517–3546.

Whitaker DW, Wasimi SA, Islam S (2001) The El Niño southern oscillation and long-range forecasting of flows in the Ganges. *International journal of climatology* 21(1):77–87.

Wikipedia (2019) *Pakistan – Relief, Map, PopulationData.net*. PopulationData.net.

Wikipedia (2023a) *Climate of Pakistan*. https://en.wikipedia.org/wiki/Climate_of_Pakistan.

Wikipedia (2023b) *Topography of Pakistan*. https://en.wikipedia.org/wiki/Topography_of_Pakistan.

Wilks DS (2005) *Statistical Methods in the Atmospheric Sciences* (Elsevier Science & Technology, San Diego, CA).

Wolter K, Timlin MS (2011) El Niño/Southern Oscillation behaviour since 1871 as diagnosed in an extended multivariate ENSO index (MEI.ext). *Intl Journal of Climatology* 31(7):1074–1087.

Zaidi UF (2022) WWF reports differences between 2010 and 2022 floods. *The Diplomatic Insight*.

Zavadoff, Breanna, and Arcodia, Marybeth (2022) What are teleconnections? Connecting Earth's climate patterns via global information superhighways | NOAA Climate.gov. Retrieved October 26, 2023, http://www.climate.gov/news-features/blogs/enso/what-are-teleconnections-connecting-earths-climate-patterns-global.

Zhang T, Hoell A, Perlwitz J, Eischeid J, Murray D, Hoerling M, Hamill TM (2019) Towards probabilistic multivariate ENSO monitoring. *Geophysical Research Letters* 46(17–18):10532–10540.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Fort Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California