ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

SCHOOL OF EARTH SCIENCES

Geospatial technology–AI (Machine learning algorithm) Application in Flood Inundation Frequency and Extent Analysis and Modeling in Becho plain, Western Central Ethiopia

A thesis submitted to

The school of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Masters of Science in Remote Sensing and Geo-informatics

By

Jifara Dabessa Gemechu

(GSR/1161/15)

Advisor: Dr. Tibebu Kasawmar

Co-advisor: Dr. Tilahun Az.

Addis Ababa University

June, 2024

Addis Ababa University

School of Graduate Studies

This is to certify that thesis prepared by Mr. Jifara Dabessa Gemechu, entitled: **"Geospatial technology- AI(Machine learning algorithm) Application in Flood Inundation Frequency and Extent Analysis and Modelling in Becho plain, Western Central Ethiopia"** and submitted in partial fulfillment of the requirements for the degree of Masters of Science in Remote Sensing and Geo-informatics complies with the regulations of the University and meets the accepted standards with respect to the originality and quality.

Signed by the Examining Committee:

Approved by the board of examiners:

_____ _____ _____

Chairman, School of Earth Sciences Signature Date

Examiners

1. Prof. Surya K. Signature_____ Date _____
2. Dr. Binyam Tesfaw Signature_____ Date _____
3. Dr. Tibebu Kasawmmar Signature  Date _____

   Advisor
4. Dr. Tilahun A. Signature_____ Date _____

   Co-advisor
5. Dr. Bayisa Signature_____ Date _____

   Chairperson

Table of Contents

Acknowledgments

First and for most, I would like to thank the “Almighty GOD” who made it possible and for strength and patience that he gave me to complete my study successfully within the given short period of time. Next to “GOD”, I express my heartfelt gratitude and indebtedness to my advisor Dr. Tibebu Kasawmar and co-advisor Dr. Tilahun Az., who helped me from the very beginning of the research to its advise and for sharing me their time for valuable discussions and giving me their kind helpful guidance throughout this study.

I would like to thank Ministry Water and Energy (MoWE) for allow me to pursue this MSc. And providing the necessary data and information of water resources and climate data.

Further, I would like to extend my thanks to Dr. Binyam Tesfaw and Professor Surya K. for their unlimited support and comment from the beginning to the end of this thesis work.

My thanks extend to all Earth Science Department staff members and Remote Sensing and Geo-informatics stream staff members for securing lab. Facility and unreserved support.

Finally, I would like to express my deep gratitude to all my family and friends who encouraged me in my educational pursuits.

List of Figures

List of Tables

List of Appendices

List of Acronyms

AI	Artificial intelligence
AAU	Addis Ababa University
CSA	Central statistical agency
DEM	Digital Elevation Model
DIP	Digital Image Processing
ESA	European Space Agency
FAO	Food and Agricultural Organization
GEV	Generalized extreme values
GEE	Google earth engine
GIS	Geographic Information Sciences
GPS	Global Positioning System
LST	Land Surface Temperature
LU/LC	Land-Use/Land-Cover
ML	Machine learning
MoWE	Ministry of water and energy
OLI	Operational Land Imager
SAGA GIS	System for Automated Geoscientific Analyses
SAR	Synthetic Aperture RADAR
SRTM	Shuttle Radar Topography Mission
TIRS	Thermal Infrared Sensor
TM	Thematic Mapper
UNEP	United Nations Environmental Program
USGS	United States Geological Survey
UTM	Universal Transverse Mercator

Geospatial technology-AI (Machine learning algorithm) Application in Flood Inundation Frequency and Extent Analysis and Modelling in Becho plain, Western Central Ethiopia

Jifara Dabessa Gemechu, MSc. Thesis

Addis Ababa University, June 2024

Abstract

Floods are one of the most devastating forces in nature. Flooding is a significant natural hazard affecting the Becho plain in western-central Ethiopia, causing extensive environmental, social, and economic damage. This study utilizes advanced geospatial technologies and artificial intelligence (AI) techniques, machine learning (ML) algorithms, to predict and analyze flood inundation frequency and extent in this region. The study integrates historical flood records from the Awash River (1995−2018), climate data (1990−2023), and SAR imagery (2015−2023) to examine the relationship between climatic variables, flood events, and the spatial extent of inundation. The analysis involves evaluating various flood frequency models such as Generalized Extreme Value (GEV), log-Pearson III, and Gumbel distributions, with the GEV distribution demonstrating the best fit for river flow data (K-S statistic of 0.17 and p-value of 0.39). Furthermore, machine learning models including Support Vector Regression (SVR), Random Forest (RF), and Linear Regression (LR) were applied to predict flood frequencies and extents. Linear Regression was identified as the most accurate model, achieving a Mean Absolute Error (MAE) of 17.71 and a Cross-Validated MAE of 20.63. The calculated flood frequency and river flow overall mean and rainy season, the trend analysis showed a general increasing trend in the mean flow, with an overall trend of 0.12 m^3/s per year and a rainy season trend of 0.56 m^3/s per year. Moreover, based on linear regression analysis of time series flood extent, the slope is positive that the area flooded is increasing per year. Furthermore, geospatial analysis revealed critical flood-prone areas and their temporal changes, providing valuable insights into flood dynamics. The results highlight that integrating geospatial and AI methodologies significantly enhances the prediction and modeling of flood-prone areas. This approach offers a robust framework for improving flood risk management and mitigation strategies, ultimately contributing to better preparedness and resilience in the Becho plain.

Keywords: Climate data, Flood inundation, Frequency analysis, Machine learning, Remote sensing

CHAPTER I

Introduction

1. 1 Background and Justification

Flood inundation is a significant environmental hazard that poses a major threat to lives, livelihoods, and infrastructure in many regions around the world, including Ethiopia (Hallegatte et al., 2013; Worqlul et al., 2018). Flood events can have devastating consequences, causing loss of life, property damage, and disruption to economic activities. Understanding the frequency, extent, and dynamics of flood inundation is crucial for effective flood risk management and disaster preparedness (Gebremichael et al., 2018; Komi et al., 2017).

1.2 Flood inundation

Flood inundation refers to the temporary covering of land by water overflowing from the normal confines of a water body, such as a river, stream, or lake (Teng et al., 2017). Flood inundation can be caused by various factors, including heavy rainfall, rapid snowmelt, dam or levee failures, and coastal storm surges (Merz et al., 2010; Zheng et al., 2018). The frequency, duration, and extent of flood inundation can vary widely depending on the specific hydrological and climatic conditions of a region (Gebremichael et al., 2018; Komi et al., 2017).

1.3 Flood types

Floods can be categorized into different types based on their causes and characteristics, such as fluvial floods, pluvial floods, and coastal floods (Merz et al., 2010). Fluvial floods occur when the water level in a river or stream exceeds the capacity of the channel, leading to the overflow of water onto the surrounding land (Teng et al., 2017). Pluvial floods, on the other hand, are caused by heavy rainfall that overwhelms the ability of the land surface to absorb or drain the water, resulting in the accumulation of water in low-lying areas (Zheng et al., 2018). Coastal floods are primarily caused by storm surges, high tides, and rising sea levels, which can inundate low-lying coastal areas (Hallegatte et al., 2013).

1.4 Geospatial-AI technology application

Geospatial-AI (Artificial Intelligence) technology, including remote sensing and Geographic Information Systems (GIS), has become an increasingly valuable tool for studying and monitoring flood inundation (Gebremichael et al., 2018; Komi et al., 2017). These technologies can provide

detailed spatial and temporal information on flood dynamics, which is essential for understanding the causes, patterns, and impacts of flood events (Teng et al., 2017; Zheng et al., 2018). Remote sensing data, such as satellite imagery, can be used to map the extent of flood inundation, while GIS can be utilized for spatial analysis, modeling, and decision support (Gebremichael et al., 2018; Komi et al., 2017). Furthermore, the integration of AI algorithms, such as machine learning and deep learning, can enhance the analysis and prediction of flood inundation patterns (Worqlul et al., 2018).

1.5 Problem statement

The Becho plain, located in western-central Ethiopia, is a region that is frequently affected by flood inundation (Gebremichael et al., 2018). Despite the significant impact of these flood events on the local population and infrastructure, there is a lack of comprehensive understanding of the spatial and temporal dynamics of flood inundation in the area. The application of geospatial-AI technology could provide valuable insights into the frequency, extent, and patterns of flood inundation, which could inform and support effective flood risk management strategies (Worqlul et al., 2018).

1.6 Objective of the study

The primary objective of this study is to analyze the flood inundation frequency and extent in the Becho plain, western-central Ethiopia, using a combination of geospatial and AI-based techniques. Specifically, the study aims to:

- Assess the spatial and temporal patterns of flood inundation in the Becho plain.
- Identify the critical variables that influence flood inundation in the study area.
- Develop a flood inundation model using geospatial and AI-based techniques to predict the extent and frequency of floods.
- Provide insights and recommendations for effective flood risk management and mitigation strategies in the Becho plain.

1.7 Research questions

The study will address the following research questions:

- What are the spatial and temporal patterns of flood inundation in the Becho plain?
- What are the critical variables that influence flood inundation in the study area?
- How can geospatial and AI-based techniques be utilized to develop a flood inundation model for the Becho plain?

- What are the implications of the study findings for flood risk management and mitigation strategies in the Becho plain?

1.8 Significance of the study

The findings of this study will contribute to a better understanding of flood inundation processes in the Becho plain, a region frequently affected by devastating floods. The use of geospatial and AI-based techniques will provide a more comprehensive and data-driven approach to analyzing flood patterns, identifying critical variables, and developing predictive models. The insights gained from this study will be valuable for:

- Enhancing flood risk assessment and management in the Becho plain.
- Informing the development of effective flood mitigation strategies, such as the implementation of early warning systems, infrastructure improvements, and land-use planning.
- Supporting decision-making processes related to disaster preparedness, emergency response, and post-flood recovery efforts.
- Expanding the knowledge base on flood inundation modeling and the application of geospatial and AI technologies in flood studies in Ethiopia and similar regions.

1.9 Scope of the study

The scope of this study is focused on the Becho plain in western central Ethiopia, which is prone to recurrent flood events. The research aims to analyze the flood inundation frequency and extent using geospatial-AI (ML) technology applications. Specifically, the study covers the following aspects:

- Analysis of critical variables influencing flood inundation, such as climate, topography, hydrology, and land use/land cover.
- Accessing and processing satellite imagery data to map flood inundation extents over time.
- Conducting flood frequency analysis using the historical flow data of Awash River.
- Modeling the flood inundation extent and frequency in the Becho plain using the analyzed data and geospatial-AI techniques.

2.0 Limitations of the study

- The study is limited to the Becho plain area and may not be directly applicable to other flood-prone regions without further validation and adaptation.

- The study relies on historical data, which may have limitations in terms of completeness, accuracy, and spatial/temporal coverage.
- The availability and quality of satellite imagery data can be a constraint, particularly in areas with persistent cloud cover or limited data archives.
- The application of geospatial-AI techniques in flood inundation modeling may be limited by the complexity of the underlying processes and the availability of sufficient training data.

2.1 Organization of the thesis

This thesis was organized into six chapters. Chapter one provides an introduction to flood inundation, types of floods, and the application of geospatial-AI technology. It also presents the problem statement, objectives, research questions, and significance of the study. Chapter two, Literature Review, discusses the concepts of flood inundation, recent flood studies in Ethiopia, and previous flood studies on the Becho flood plain. Chapter three is Materials and Methodology that describes the study area, including its location, topography, geology, hydrology, climate, vegetation, land use, and soil. It also outlines the data collection methods and the techniques used for flood frequency analysis and flood inundation extent mapping. Chapter four is the Analysis and Results, presents the analysis and results of the study. It covers the critical variables in the flood inundation study, satellite imagery data access and processing, climate variables (precipitation, temperature, humidity, and wind speed), the impact of topography and morphometry on the Becho plain, the hydrology impact (Awash river maximum flow analysis), and the flooded area calculation and time series analysis. Chapter five is discussion that provides a detailed discussion of the findings from the study. Chapter six conclusion and recommendations which presents the conclusions of the study, followed by recommendations for future research.

CHAPTER II

LITERATURE REVIEW

2.1 Concept of flood inundation

Flood inundation is a complex phenomenon that is influenced by various factors, including rainfall patterns, river discharge, topography, land use, and soil characteristics (Teng et al., 2017; Zheng et al., 2018). Understanding the underlying concepts and dynamics of flood inundation is crucial for effective flood risk management and disaster mitigation. Flood inundation can be characterized by its frequency, duration, depth, and areal extent (Merz et al., 2010). The frequency of flood events refers to the likelihood of a flood of a particular magnitude occurring within a given time period, such as annual or decadal (Gebremichael et al., 2018). The duration of a flood event is the length of time that the inundated area remains flooded, which can range from hours to days or even weeks (Komi et al., 2017). The depth of flood inundation is the vertical distance between the water surface and the ground, which can have significant impacts on the severity of the flood's consequences (Hallegatte et al., 2013). The areal extent of a flood event is the geographic area that is covered by the inundating water, which can vary depending on factors such as the size and shape of the water body, the topography of the surrounding land, and the presence of flood control structures (Teng et al., 2017).

2.2 Global trends

In the recent years, there has been a significant amount of research on flood inundation frequency, extent, and depth analysis, mapping, and modeling using geospatial and machine learning (ML) techniques (Li et al., 2020). One notable trend in recent studies is the use of high-resolution satellite imagery, LiDAR data, and unmanned aerial vehicles (UAVs) to map flood extent and depth (Smith et al., 2017; Wang et al., 2018). Another important trend is the development of ML-based flood frequency analysis and modeling, allowing for better understanding of flood patterns and trends (Liu et al., 2019). Additionally, there has been a growing interest in the integration of geospatial and ML techniques for flood risk mapping and assessment (Zhang et al., 2020; Li et al., 2021). Furthermore, recent studies have also explored the use of Synthetic Aperture Radar (SAR) data for flood mapping and modeling, demonstrating the potential of SAR data for accurate and timely flood mapping (Chen et al., 2018; Li et al., 2020).

2.3 Recent flood research in Ethiopia

Ethiopia has experienced numerous flood events in recent years, with significant impacts on the country's population, infrastructure, and economic development (Gebremichael et al., 2018; Worqlul et al., 2018). Several studies have investigated the causes, patterns, and impacts of flood inundation in different regions of Ethiopia. Gebremichael et al. (2018) conducted a study on the impact of climate change on flood events in the Awash Basin, a major river basin in Ethiopia. The study used remote sensing data, hydrological modeling, and climate projections to assess the frequency and magnitude of future flood events. The results showed that climate change is expected to increase the frequency and intensity of flood events in the basin, posing significant challenges for disaster management and infrastructure planning. Worqlul et al. (2018) evaluated the performance of different rainfall datasets, including satellite-based and ground-based measurements, for use in hydrological modeling in the upper Blue Nile Basin, a data-scarce region in Ethiopia. The study found that the use of satellite-based rainfall data, combined with ground-based observations, can improve the accuracy of hydrological models and support flood risk assessment and management in the region.

2.4 Previous Research on Becho floodplain

The Becho plain, located in western-central Ethiopia, is a region that is frequently affected by flood inundation (Gebremichael et al., 2018). Several studies have been conducted to investigate the characteristics and impacts of flood events in this area. The article "Flood hazard map of the Becho floodplain, Ethiopia, using nonstationary frequency model" by Tadesse et al. (2020) presents a study on flood hazard mapping in the Becho floodplain in Ethiopia. The study utilizes a nonstationary frequency model to assess flood hazards and create a flood hazard map for the area. The research begins by highlighting the increasing frequency and severity of floods in the region, which have caused significant damage to infrastructure and livelihoods. The authors emphasize the importance of accurate flood hazard mapping in order to effectively manage and mitigate the impact of floods.

The study by Birhanu et al. (2019), "Flood hazard assessment and mapping using GIS integrated with multi-criteria decision analysis in upper Awash River basin, Ethiopia," presents a similar approach for the upper Awash River basin. The study uses GIS integrated with multi-criteria decision analysis (MCDA) to identify and map flood-prone areas. The authors found that the upper Awash

River basin is highly prone to flooding, with the areas around the Awash River and its tributaries being the most vulnerable. The article "Flood Risk Assessment Using Multi-Criteria Evaluation and Flood Inundation Modelling: The Case of Upper Awash Basin, Ethiopia" by Reta Birhanu, published in the Journal of Flood Risk Management, aims to assess flood risk in the Upper Awash Basin using a combination of multi-criteria evaluation (MCE) and flood inundation modeling. The findings of the study indicate that the combination of MCE and flood inundation modeling provides a comprehensive and effective approach to flood risk assessment. Overall, these studies highlight the importance of considering nonstationary factors, such as climate change and land use changes, in flood hazard mapping and risk assessment. The integration of various techniques, including GIS, MCDA, and flood inundation modeling, provides a robust framework for identifying and prioritizing flood risk factors in the region.

2.5 Geospatial-AI technology for flood inundation mapping and analysis

2.5.1 Remote sensing for flood mapping

Remote sensing technologies, such as satellite imagery and aerial photography, have become increasingly valuable tools for mapping and monitoring flood inundation (Teng et al., 2017; Zheng et al., 2018). These technologies can provide detailed spatial and temporal information on the extent and dynamics of flood events, which is crucial for understanding their causes and impacts. Several studies have utilized remote sensing data to map flood inundation in different regions. For example, Gebremichael et al. (2018) used Landsat satellite imagery to delineate the extent of flood events in the Awash Basin, Ethiopia, and assess their relationship with climate change. Similarly, Komi et al. (2017) employed satellite-derived flood maps to model flood hazard in the Oti River basin in West Africa.

2.5.2 GIS for flood analysis and modeling

Geographic Information Systems (GIS) have become an essential tool for spatial analysis, modeling, and decision support in flood risk management (Teng et al., 2017; Worqlul et al., 2018). GIS can be used to integrate and analyze various spatial datasets, such as topography, land use, and hydrological data, to understand the complex dynamics of flood inundation. Several studies have utilized GIS for flood analysis and modeling. For instance, Merz et al. (2010) reviewed the use of GIS-based approaches for flood risk mapping at the local scale, highlighting the challenges and opportunities in incorporating spatial variability and uncertainty. Worqlul et al. (2018) employed GIS-based

hydrological modeling to evaluate the performance of different rainfall datasets for flood risk assessment in the upper Blue Nile Basin, Ethiopia.

2.5.3 Integration of AI with geospatial technologies

The integration of Artificial Intelligence (AI) algorithms, such as machine learning and deep learning, with geospatial technologies has the potential to enhance the analysis and prediction of flood inundation patterns (Gebremichael et al., 2018; Worqlul et al., 2018). AI-based techniques can be used to extract valuable information from remote sensing data, improve the accuracy of hydrological models, and support decision-making in flood risk management. Several studies have explored the application of AI-based methods for flood inundation mapping and prediction. For example, Gebremichael et al. (2018) used machine learning algorithms to integrate remote sensing data and climate projections for assessing the impact of climate change on flood events in the Awash Basin. Worqlul et al. (2018) incorporated AI-based techniques, such as support vector machines, to evaluate the performance of different rainfall datasets for hydrological modeling in the upper Blue Nile Basin.

2.6 Research gaps

The research gaps in the previous studies conducted on flood inundation analysis and modeling that could further improve the effectiveness of flood inundation analysis, modeling and risk management strategies in the Becho floodplain are:

- Limited use of advanced spatial and hydrological modeling techniques.
- Lack of comprehensive uncertainty analysis.
- Underutilization of remote sensing and geospatial data.
- Lack of machine learning and AI-based approaches.
- Limited validation and performance evaluation.

CHAPTER III

MATERIALS AND METHODS

3.1 Description of the study area

The study area is found in the upper Awash river basin in the Oromia Regional state of South west Shoa zone in Becho District, central western Ethiopia, around 60 km from Addis Ababa. The study area geographically situated between 8° 15′ 00″−9° 15′ 15″ E longitude and 38° 00′ 00″− 38° 40′ 00″ N latitude with areal coverage of 4117 km^2 (Fig.1).

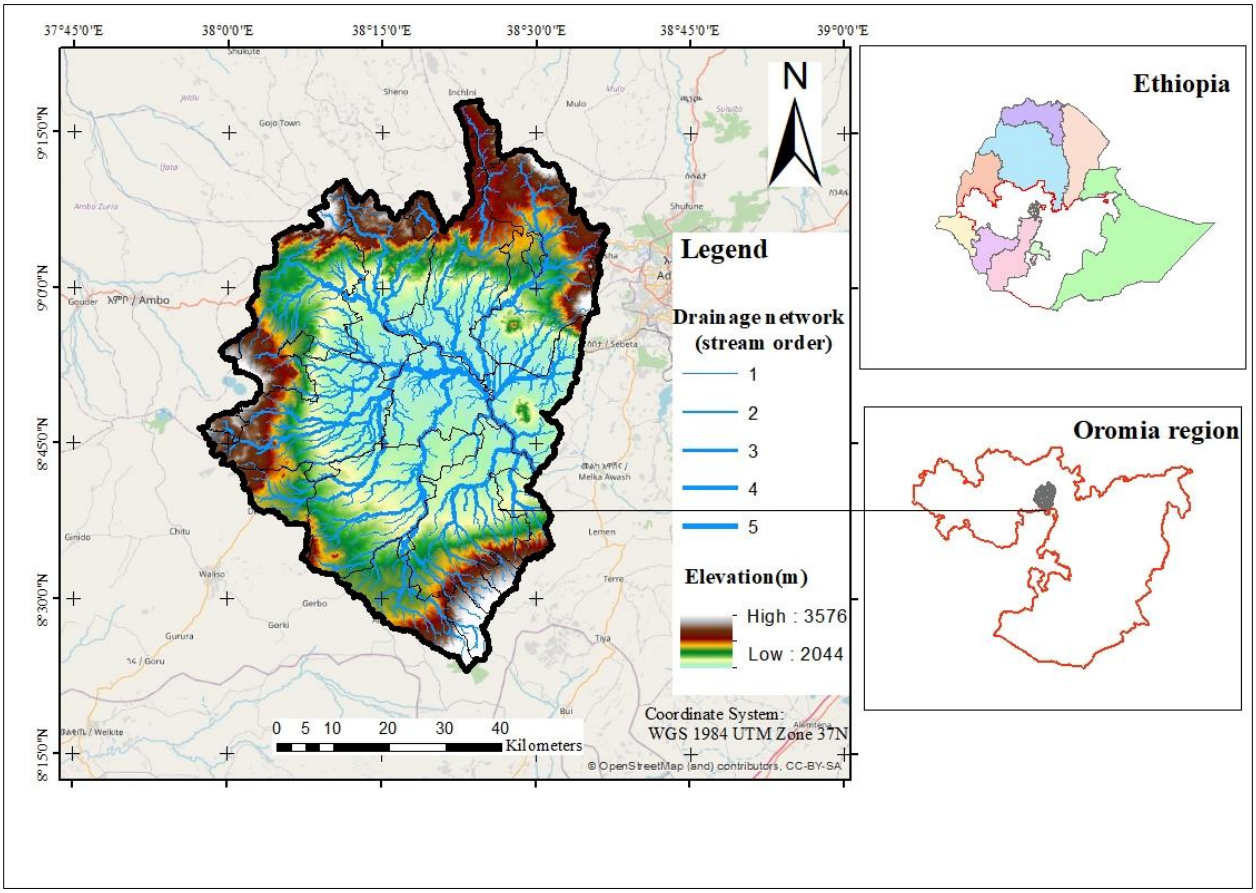


Fig. 1 Location map of the study area

3.2 Topography and Hydrology

The Becho plain is located within the expansive Awash River basin in central Ethiopia, situated at elevations ranging from approximately 800 to 1,000 meters above sea level (Kassa et al., 2019). The plain has a generally flat to gently undulating topography, with the meandering Awash River running

through the center of the area (Alemu et al., 2017). The Awash River floodplain extends for several kilometers on either side of the main channel, creating a broad, low-lying agricultural region (Teshome, 2016). The gentle slope of the plain facilitates the accumulation of water during the rainy seasons, leading to the formation of seasonal wetlands and flooding.

3.3 Geology and Soils

The geology of the Becho plain is dominated by alluvial and lacustrine deposits (Fig. 2) laid down over time by the Awash River and associated tributaries and lakes (Kassa et al., 2019). The soils are primarily classified as vertisols heavy shown in (Fig. 3), clay rich soils that exhibit significant shrinking and swelling behavior as they cycle through wet and dry conditions (Alemu et al., 2017). These Vertisol soils have high agricultural potential when properly managed, but also pose challenges due to their poor drainage characteristics and tendency to waterlog (Mokonen, 2021).

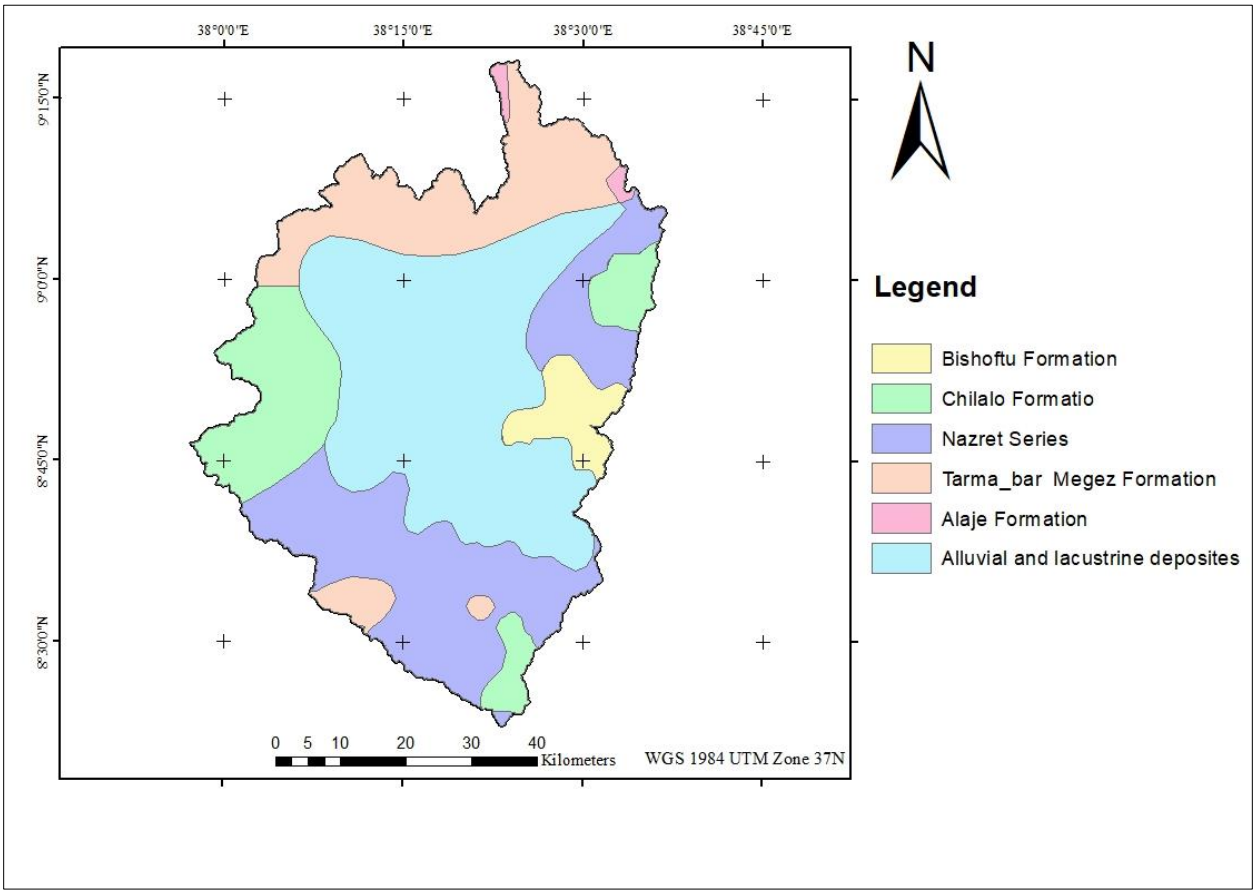


Fig. 2 Geology map

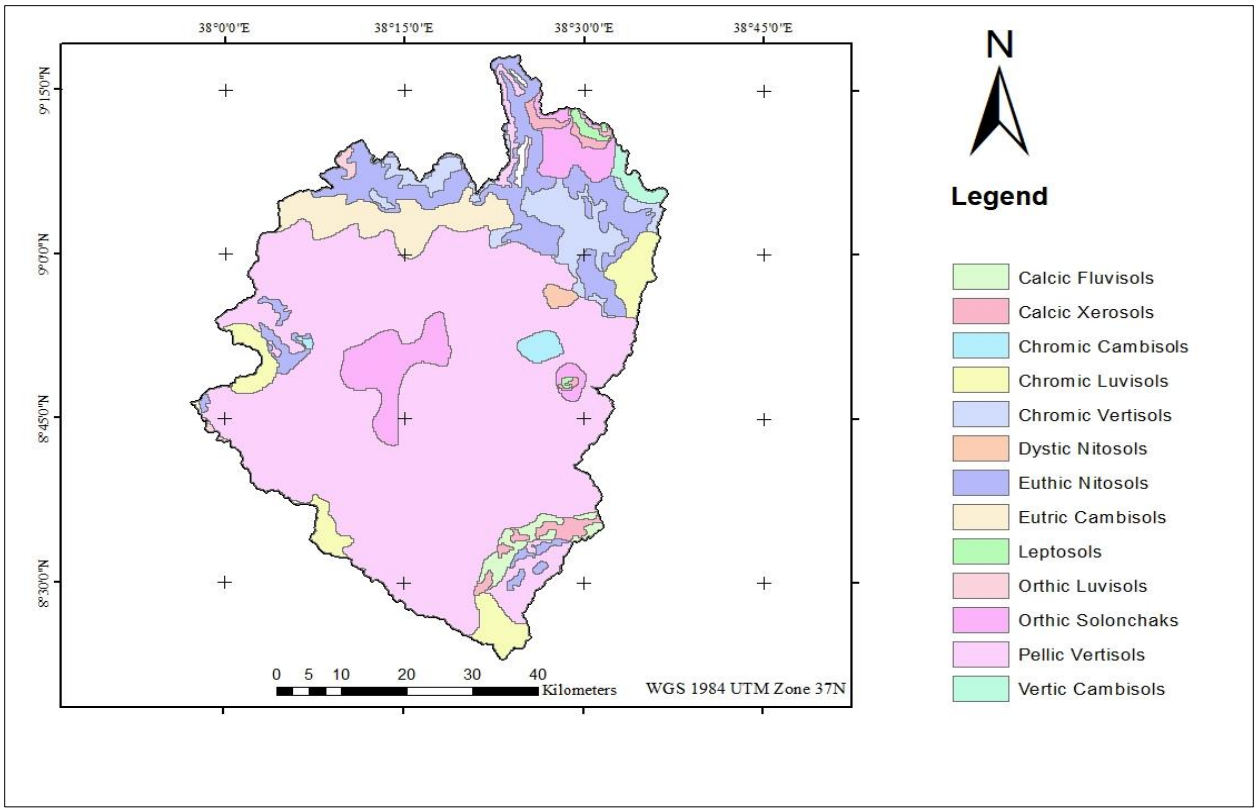


Fig. 3 Soil map

3.4 Climate and vegetation

3.4.1 Climate

The Becho plain experiences a semi-arid climate, characterized by a bimodal rainfall pattern (Mokonen, 2021). The main rainy season occurs from June through September, with average monthly rainfall during this period ranging from 150 to 250 mm (Kassa et al., 2019). A shorter rainy season takes place from March to May, contributing an additional 100 to 150 millimeters of average monthly precipitation (Alemu et al., 2017). The yearly total rainfall in the Becho plain typically falls between 1200 to 1500 mm (Mokonen, 2021). Temperatures in the region are consistently high, with an average annual temperature around 25°C and minimal seasonal variation (Teshome, 2016).

The vegetation cover in the area is predominantly characterized by grasslands and shrublands, with some patches of agricultural land. The seasonal flooding and waterlogging in the area have a significant influence on the distribution and composition of the vegetation. The vegetation in the Becho floodplain is predominantly composed of savanna grasslands, with scattered trees and shrubs. During the wet season, the floodplain becomes lush and green, with the growth of grasses and other

vegetation. However, during the dry season, the vegetation may become sparser, and some areas may experience drought conditions.

3.5 Agro-climate and flood risks

The Becho plain's semi-arid climate shown in (Fig. 4) and fertile Vertisol soils support both rainfed and irrigated agricultural production (Mokonen, 2021). However, the area is highly vulnerable to periodic flooding events along the Awash River, which can cause significant damage to crops, property, and infrastructure within the floodplain (Kassa et al., 2019). Effective flood risk management and mitigation strategies are critical for sustaining agricultural livelihoods and development in the Becho plain region (Alemu et al., 2017).

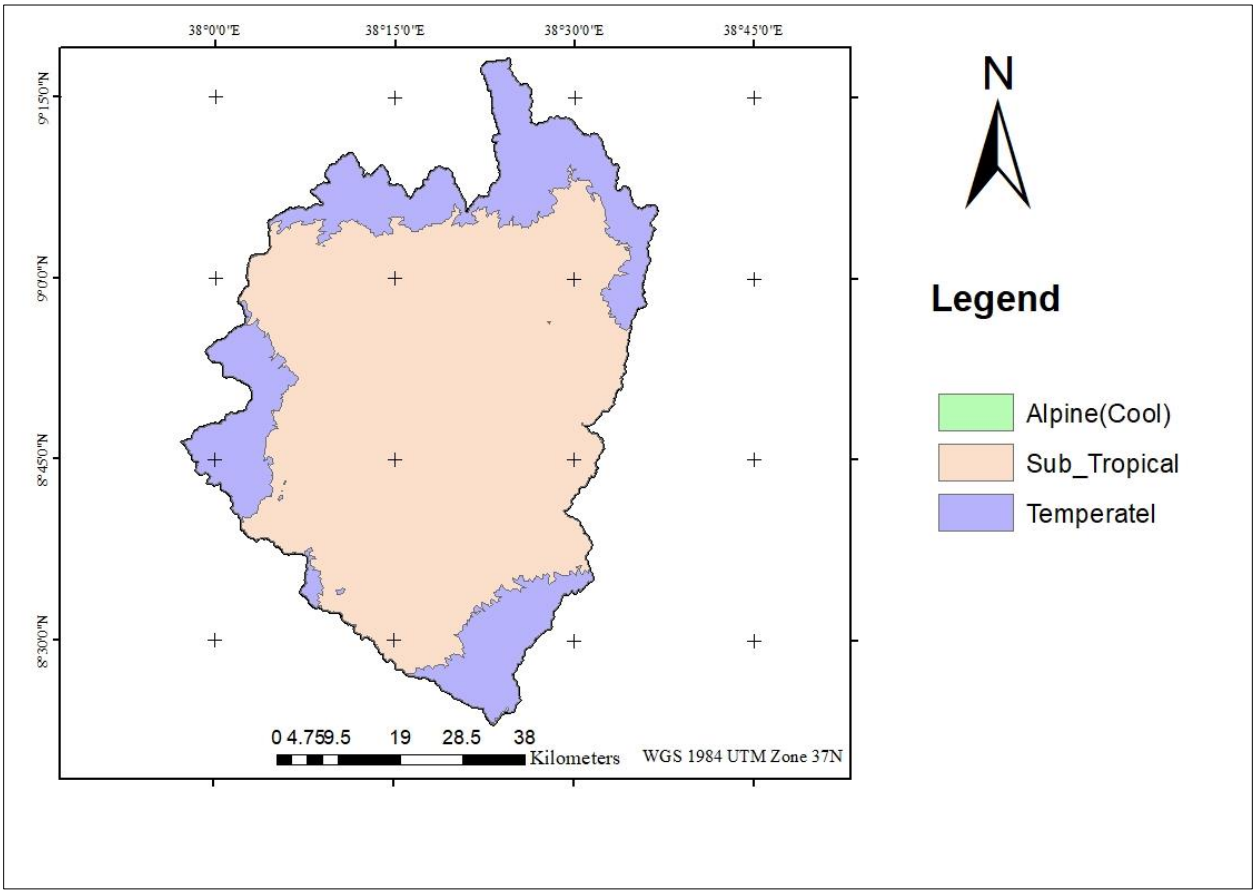


Fig. 4 Agro climate map

3.6 Land-use and land-cover

The Becho plain is an important agricultural region, with the primary crops including maize, sorghum, teff, and cotton (Teshome, 2016). In addition to the extensive farmlands, the plain also contains areas of natural grassland and scrublands, particularly along the fringes of the Awash River floodplain

(Kassa et al., 2019). Settlement and infrastructure development is concentrated along the main roads and near the Awash River, though the floodplain itself remains largely rural (Alemu et al., 2017) shown in (Fig. 5).

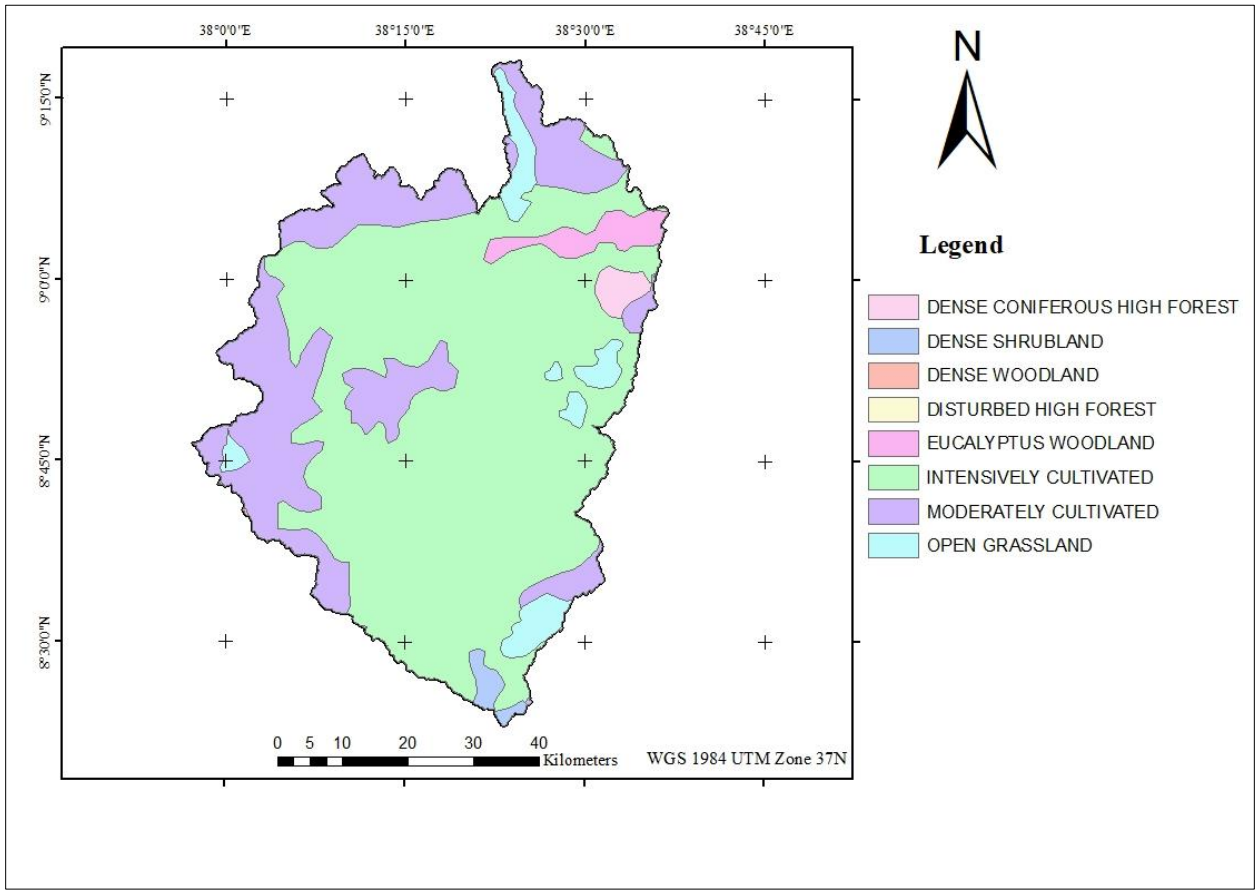


Fig. 5 Land-use and land-cover map

3.7 Annual rainfall

Rainfall play the most significant role in inducing flood (Pal and Samanta, 2011; Ali et al., 2019). Inverse Distance Weighted (IDW) algorithms (Othman et al., 2006; Chen and Liu, 2012) have been used in the ArcGIS environment to spatially map the rainfall. The rainfall map shows that the western portion of the research region experiences more rain than the eastern portion. The rainfall data for this study will be obtained from CRU TS (Climatic Research Unit gridded Time Series) which is a widely used climate dataset on a 0.5° latitude by 0.5° longitude grid over all land domains of the world except Antarctica. It is derived by the interpolation of monthly climate anomalies from extensive networks of weather station observations shown in (Fig. 6).

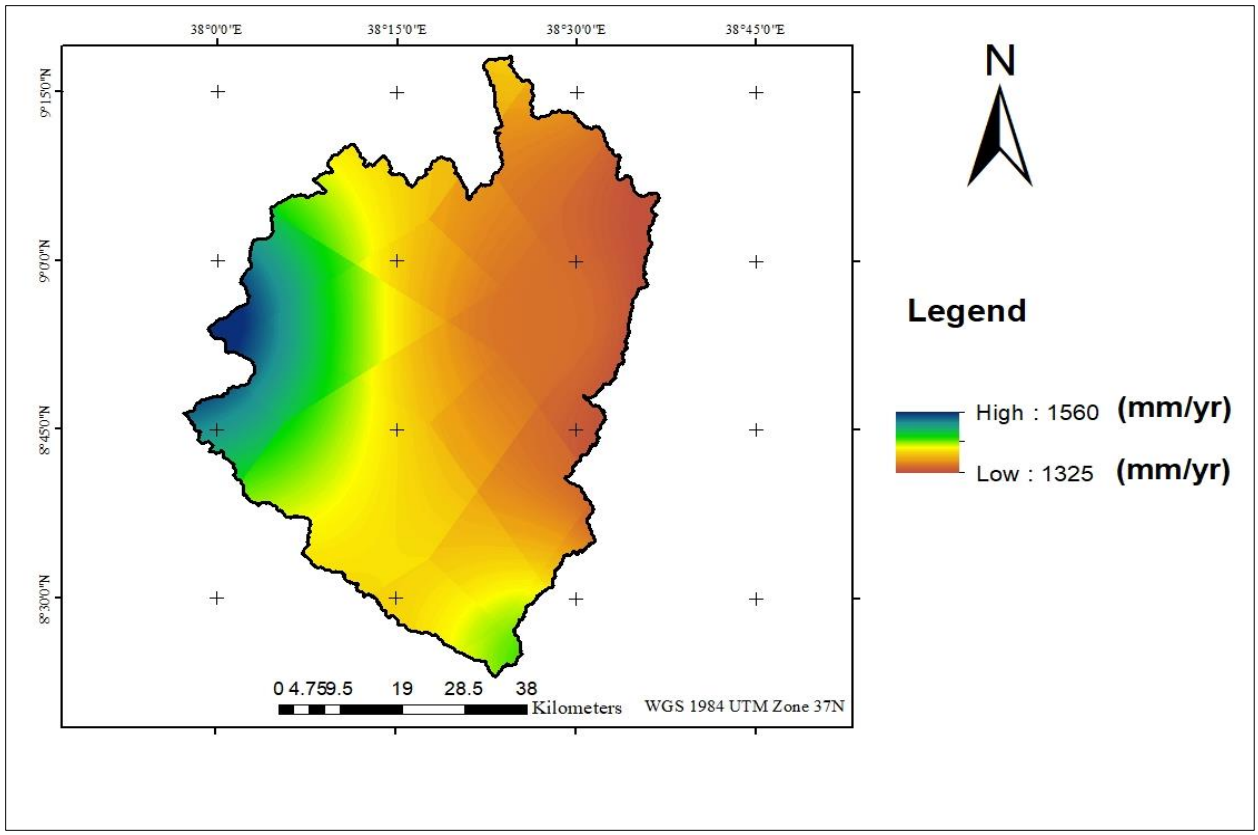


Fig. 6 Precipitation map

3.8 Population

The Becho plain is home to a significant rural population, with the majority of inhabitants living in small villages and farming communities scattered throughout the region (Alemu et al., 2017). According to the latest census data, the total population of the Becho plain and surrounding areas is estimated to be around 458,469 people (CSA, 2016). The population density is relatively high for a rural area, averaging around 150 people per square kilometer (Kassa et al., 2019). The ethnic composition is predominantly Oromo, with some minority groups also present (Mokonen, 2021).

3.9 Economic activities and Livelihoods

Agriculture is the backbone of the Becho plain's economy, with crop production and livestock rearing being the primary economic activities (Alemu et al., 2017). As mentioned previously, the main crops grown in the region include maize, sorghum, teff, and cotton (Teshome, 2016). These crops are cultivated through a combination of rain fed and irrigated farming practices, with the latter becoming increasingly important as water resource development projects are implemented (Kassa et al., 2019).

In addition to crop farming, many households also engage in livestock keeping, primarily raising cattle, sheep, and goats (Mokonen, 2021).

3.2 Data collection

3.2.1 Primary data

The primary data collected for this study included Geolocation of structures on the Awash River, measurements, and field observations. During field visits, the upstream and downstream of Awash River Bridge was conducted to gather information on the historical occurrence and impacts of flood events, as well as the coping strategies employed by the local communities. Measurements of water levels, bridge dimension, river width, and manmade dykes information were also collected support the analysis of flood inundation and its impact on the local environment (Table 1).

Table 1. Data and software

Item	Dataset	Source	Relevance
1	Digital Elevation Model (DEM)	SRTM(30m) USGS	To generate slope, elevation and different topographic morphology
2	Soil data	FAO	To differentiate different soil types
3	Climate data(Rainfall and Temperature)	CRU	To extract the soil types within the study Area
4	Geological data	EGA	To map geological type
5	Administrative shape files	CSA	For flood inundation boundary mapping
6	Land use land cover data		To determine different land classes
5	Population data	CSA	For deriving population density factor
6	SAR (RADAR) Imagery(2015-2023)	ESA	To extract water and no-water bodies
7	Multispectral image	USGS	To extract the different Land-use land cover data
8	Awash river flow data (1995-2020)	MoWE	To analysis of flood frequency and time series main river(Awash) flow analysis

3.2.2. Software and tools

The materials and Software that has been used in this study presented in Table 2.

Table 2. Dataset and sources

Item	Dataset	Source	Relevance
II	**Software and Tools**		
2.1	Python programming with GEE embedded(ML)	free	For,flood-analysis, producing statistical report generation
2.2	ENVI 5.3	free	Analysis and mapping of flood inundation
2.3	ArcGIS 10.7	ESRI	For analysis, modeling and mapping
2.4	SAGA GIS	free	For Terrain analysis, modeling and mapping
2**.5**	SNAP 6.0	free	For SAR image analysis
2**.6**	GPS	self	Geo-points collection
2**.7**	Measuring tape	self	To measure river channel width at selected points

3.3 Secondary data

In addition to the primary data, the study also utilized secondary data sources, including:

3.3.1Climate data

Historical climate data, such as rainfall and temperature records, were obtained from the and NASA Power project for Agro climatology and National Meteorological Agency of Ethiopia (for field verification). This data was collected from meteorological stations located within and around the Becho plain, covering the past 30 years (1990-2022). The climate data (Rainfall, temperature, Humidity and Wind speed) was used to analyze the long-term trends and variability in precipitation and temperature patterns, which can have a significant influence on the occurrence and severity of flood events.

3.3.2 River flood flow data

Hydrological data, including river discharge and water level records, were obtained from the Ministry of Water, Irrigation and Energy of Ethiopia. The data was collected from Awas-Belo gauging station along the Awash River and its major tributaries in the Becho plain, covering the past 20 years (1995−2018). This data was used to characterize the river flow regime, identify the timing and magnitude of maximum flood peaks, and conduct the flood frequency analysis.

3.3.3 Satellite imagery

Multi-temporal satellite imagery, including synthetic aperture radar (SAR) and optical imagery data, was obtained from free sources of European space agency(ESA) ,Sentinel−1 and from USGS, Landsat 8 respectively. The satellite data covered the study area and the time period of interest, providing information on land cover, vegetation, and the extent of flood inundation.

3.3.4 Other ancillary data

Additional data, such as topographic maps, soil types, and land use/land cover maps, were obtained from government agencies and research institutions to support the analysis and contextualize the findings. The integration of primary field data, climate and hydrological records, and remote sensing data allowed for a comprehensive assessment of the flood dynamics and their impacts in the Becho plain.

3.4 Methods

3.4.1 Flood frequency analysis

Flood frequency analysis was conducted to determine the probability and magnitude of flooding events in the Becho plain (Millington et al., 2011). This involved the use of historical hydrological data, such as river discharge records from the Awash River and its tributaries, to calculate the return periods and magnitudes of flood events (Millington et al., 2011). Statistical techniques for extreme values analysis, including the Generalized Extreme Value (GEV) Distribution, Gumbel distribution, and Log-Pearson Type III distribution, were employed to model the flood frequency curves and estimate the likelihood of floods of different magnitudes occurring (Millington et al., 2011).

The GEV distribution is a flexible distribution that combines the Gumbel, Fréchet, and Weibull distributions into a single framework (Millington et al., 2011). It is widely used in flood frequency analysis as it can accommodate a wide range of flood data distributions (Millington et al., 2011). The comparison of the performance of three different interpolation methods (linear, quadratic, and cubic) in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for the Awash Belo Maximum Flow dataset was analyzed (Millington et al., 2011). The interpolation method will be selected based on Root Mean Square Error (RMSE) (Millington et al., 2011).Here are some of the key flood frequency analysis methods, along with their parameters and formulas employed:

3.4.2 Log-Pearson Type III Distribution

Parameters: Location (μ), Scale (σ), and Shape (γ)

Formula: $Q = \bar{Q} + K \times \sigma$ (1)

where Q is the flood discharge, $\bar{Q}$ is the mean discharge, σ is the standard deviation, and K is a frequency factor dependent on the return period and the skewness coefficient.

3.4.3 Gumbel (Extreme Value Type I) Distribution:

Parameters: Location (μ) and Scale (α)

Formula: $Q = \mu - \alpha \, x \, ln(-ln(1 - 1/T))$ (2)

where Q is the flood discharge, μ is the location parameter, α is the scale parameter, and T is the return period.

3.4.4 Generalized Extreme Value (GEV) Distribution:

Parameters: Location (μ), Scale (σ), and Shape (ξ)

Formula: $Q = \mu - \sigma/\xi \times [1 - (-ln(1 - 1/T))^{\wedge}\xi]$ (3)

where Q is the flood discharge, μ is the location parameter, σ is the scale parameter, ξ is the shape parameter, and T is the return period.

The GEV distribution is a flexible distribution that combines the Gumbel, Fréchet, and Weibull distributions into a single framework. It is widely used in flood frequency analysis as it can accommodate a wide range of flood data distributions.

3.4.5 The key parameters of the GEV distribution are:

- Location parameter (μ): Determines the central tendency of the distribution.
- Scale parameter (σ): Determines the spread or variability of the distribution.
- Shape parameter (ξ): Determines the tail behavior of the distribution, i.e., whether it is light-tailed (Gumbel), heavy-tailed (Fréchet), or bounded (Weibull).

3.4.6 The Kolmogorov-Smirnov test

The Kolmogorov-Smirnov (K-S) test statistic is based on the greatest vertical distance between the empirical and theoretical Cumulative Density Functions (CDFs) (Massey, 1951). The hypothesis is rejected if the test statistic is greater than the critical value at a chosen significance level. For the significance level of α=0.05, the critical value calculated is 0.12555 (Massey, 1951). The test statistic (D) is calculated as:

$$D = \max_{1 \le i \le n} |F_n(x_i) - F(x_i)| \quad (4)$$

where $F_n(x_i)$ is the empirical CDF and $F(x_i)$ is the theoretical CDF (Massey, 1951).

3.4.6 Steps in model comparison and selection

- **Load the river flow data:** The code reads the 'AwashBelo_MaximumFlow.csv' file into a pandas DataFrame df.
- **Find the maximum flow for each year**: The code extracts the maximum flow value for each year from the monthly flow data in the DataFrame.
- **Display the original data**: The code prints the original annual maximum flow data, including the mean and standard deviation.
- **Interpolate the data using cubic spline**: The code uses the SciPy CubicSpline function to interpolate the annual maximum flow data.
- **Display the interpolated data:** The code prints the interpolated annual maximum flow data, including the mean and standard deviation.
- **Fit the Gumbel distribution**: The code fits a Gumbel distribution to the interpolated annual maximum flow data and calculates the cumulative distribution function (CDF).
- **Fit the Log-Pearson Type III distribution**: The code fits a Log-Pearson Type III distribution to the interpolated annual maximum flow data and calculates the CDF.
- **Fit the Generalized Extreme Value (GEV) distribution**: The code fits a GEV distribution to the interpolated annual maximum flow data and calculates the CDF.
- **Calculate the Kolmogorov-Smirnov (K-S) statistic**: The code computes the K-S statistic and p-value for each of the three fitted distributions.
- **Compare the empirical and theoretical CDFs**: The code plots the empirical CDF of the interpolated annual maximum flow data alongside the CDF of the Gumbel, Log-Pearson Type III, and GEV distributions.

3.5 Flood flow time series and trend analysis

For Awash river flow time series analysis and modeling, the model algorithms considered are Support Vector Regression (SVR), Random Forest, Linear Regression, and ARIMA (Auto Regressive Integrated Moving Average) (Millington et al., 2011). The validation will be made based on the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) values for different models used to predict the maximum flow of the Awash Belo River (Millington et al., 2011). The best and default model in Python for non-complex time series trend data analysis is linear regression

In general, the following steps are applied:

- Data loading and preprocessing
- Conversion of monthly records to numeric values
- Removal of unnamed column and handling of NaN values
- Interpolation of missing data
- Statistical analysis of monthly data
- Overall and rainy season time series a nalysis
- Trend analysis using linear regression
- Flood Frequency Analysis with Log-log Curve
- Peak Flow *vs* Return Period *vs* Probability Table
- Model Validation Techniques
- Statistical description and output results plot

3.5.1 Interpolation methods

The comparison of the performance of three different interpolation methods (linear, quadratic, and cubic) in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for the Awash Belo Maximum Flow dataset was analyzed (Millington et al., 2011). The interpolation method will be selected based on Root Mean Square Error (RMSE): RMSE is a commonly used metric to measure the difference between predicted and actual values. It provides a measure of how well the model's predictions fit the observed data.

3.5.2 Model selection and validation process for time series and trend analysis

For Awash river flow time series analysis and modeling the model algorithms considered are Support Vector Regression (SVR) Random Forest, Linear Regression, and ARIMA (Auto Regressive Integrated Moving Average). The validation will be made based on the Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) values for different models used to predict the maximum flow of the Awash Belo River. The best and default model in python in for non-complex time series trend data analysis is linear regression. In this study linear regression, Support Vector Regression (SVR) and Random Forest selected for analysis and comparison and to select the fit for Awash river maximum flow analysis. Over all methodological frame work is presented in Figure 7.

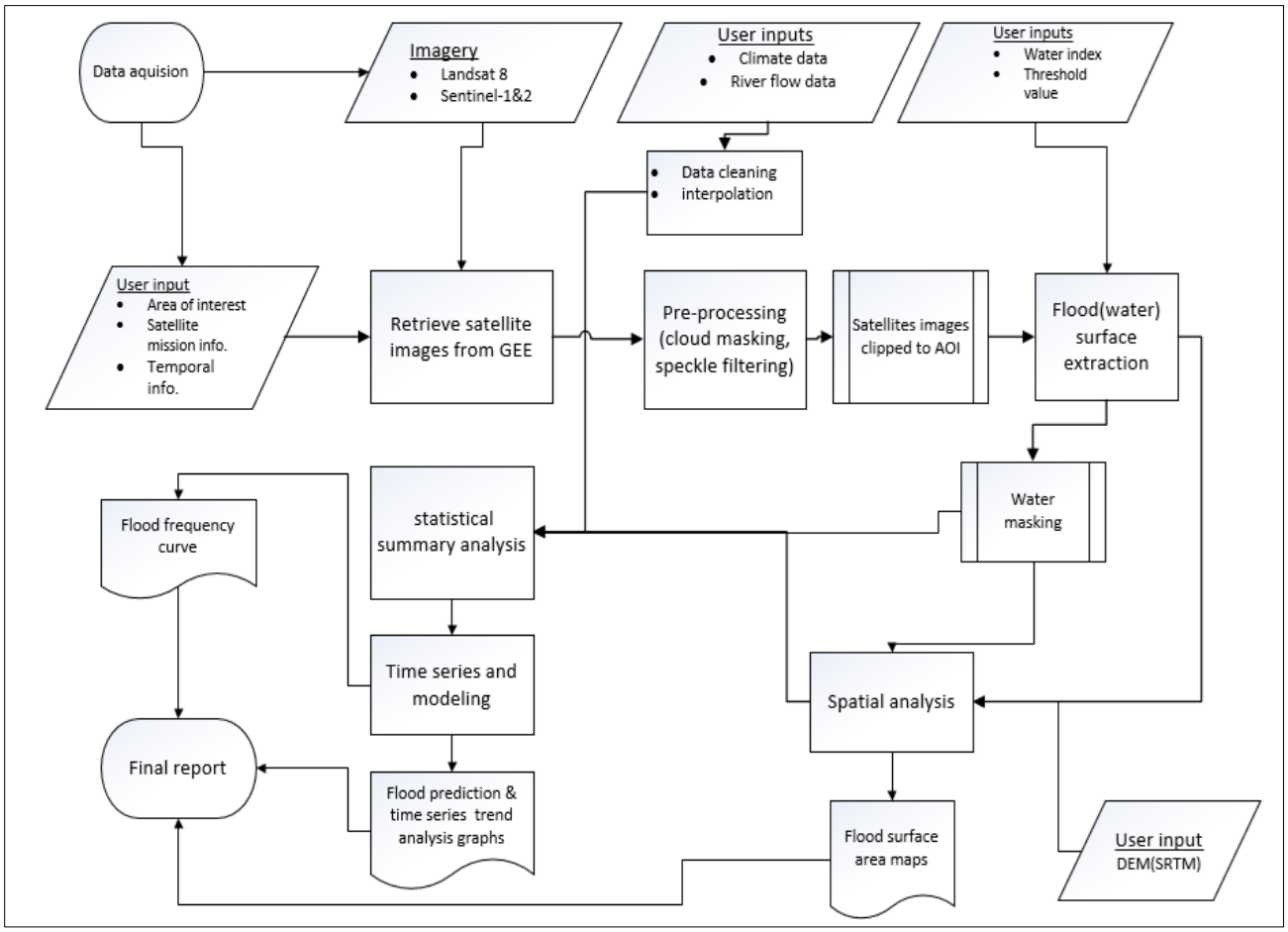


Fig. 7 Process flow diagram

CHAPTER IV

RESULTS

4.1 Identification of critical variables for flood inundation

The first step in the analysis was to identify the key variables that influence flood inundation in the Becho plain. These were determined to be precipitation, temperature, topography, river hydrology, and land-use and land-cover. The relationships and interactions between these variables were explored in detail to understand their impact on flood dynamics.

4.2 Climate variables

Climate variables such as temperature, rainfall, relative humidity, and wind speed data of time series (1990 to 2022) for thirty years (30) was downloaded and organized, converted into CSV format imported to Python notebook for further analysis.

4.2.1 Precipitation (rainfall)

The statistical analysis and spatial distribution of rainfall across the Becho plain is conducted to make time series analysis trend and identify rainfall hotspots and its impact on flood inundation. Time series plots were generated to examine precipitation trends and variability.

Table 3. Rainfall statistical summary

Months	count	mean	std	min	25%	50%	75%	max
January	33	9.61	13.1	0	0	0	15.82	47.46
February	33	22.52	31.92	0	0	10.55	36.91	142.38
March	33	40.49	30.99	0	21.09	36.91	58.01	137.11
April	33	74.54	47.62	5.27	36.91	68.55	94.92	210.94
May	33	67.6	46.09	0	26.37	68.55	102.42	168.75
June	33	138.21	54.59	47.46	100.2	137.11	163.48	284.77
July	33	267.53	78.94	116.02	237.3	258.4	321.68	474.61
August	33	268.68	78.09	131.84	221.48	290.04	321.68	469.34
September	33	139.35	72.43	42.19	84.38	137.11	174.02	400.78
October	33	36.89	39.8	0	5.27	26.37	47.46	168.75
November	33	8.16	13.15	0	0	5.27	10.55	63.28
December	33	5.64	8.5	0	0	0	5.27	31.64

4.2.2 Rainfall Characteristics

The mean annual rainfall is 138.21 mm, with a standard deviation of 54.59 mm (Rama and Moradkhani, 2016) show in (Table 3). The highly seasonal rainfall, with the lowest mean values in the winter months and the highest in the summer months, combined with the low-lying, flat terrain of the Becho plain, suggests increased susceptibility to flood inundation (Rango and Martinec, 1995).

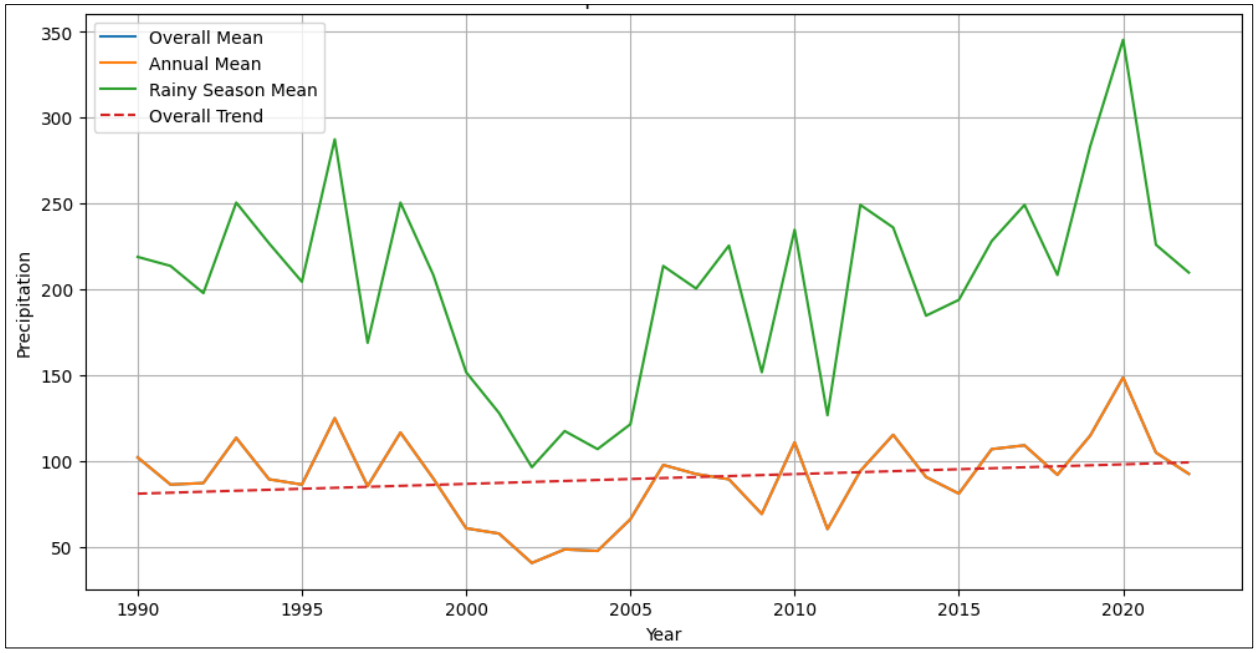


Fig. 8 Overall rainfall time series with trend

4.2.3 Time Series trend status

4.2.3.1 Overall rainfall period

The time series plot of overall precipitation shows the variation in precipitation levels over the years, and the linear regression trend line indicates an increasing trend in precipitation over time (Gocic and Trajkovic, 2013) indicated in (Fig. 8).

4.2.3.2 Rainy Season rainfall (June to September)

The mean precipitation during the rainy season is highest in July (267.53 mm) and August (268.68 mm), indicating the peak of the rainy season. The standard deviation for rainy season precipitation is relatively high, suggesting significant variability in rainfall patterns during these months (Gocic and Trajkovic, 2013). The minimum and maximum values for the rainy season indicate the range of

precipitation experienced during this period, and the quartiles provide information about the distribution of precipitation during the rainy season (Gocic and Trajkovic, 2013) shown in (Fig. 9).

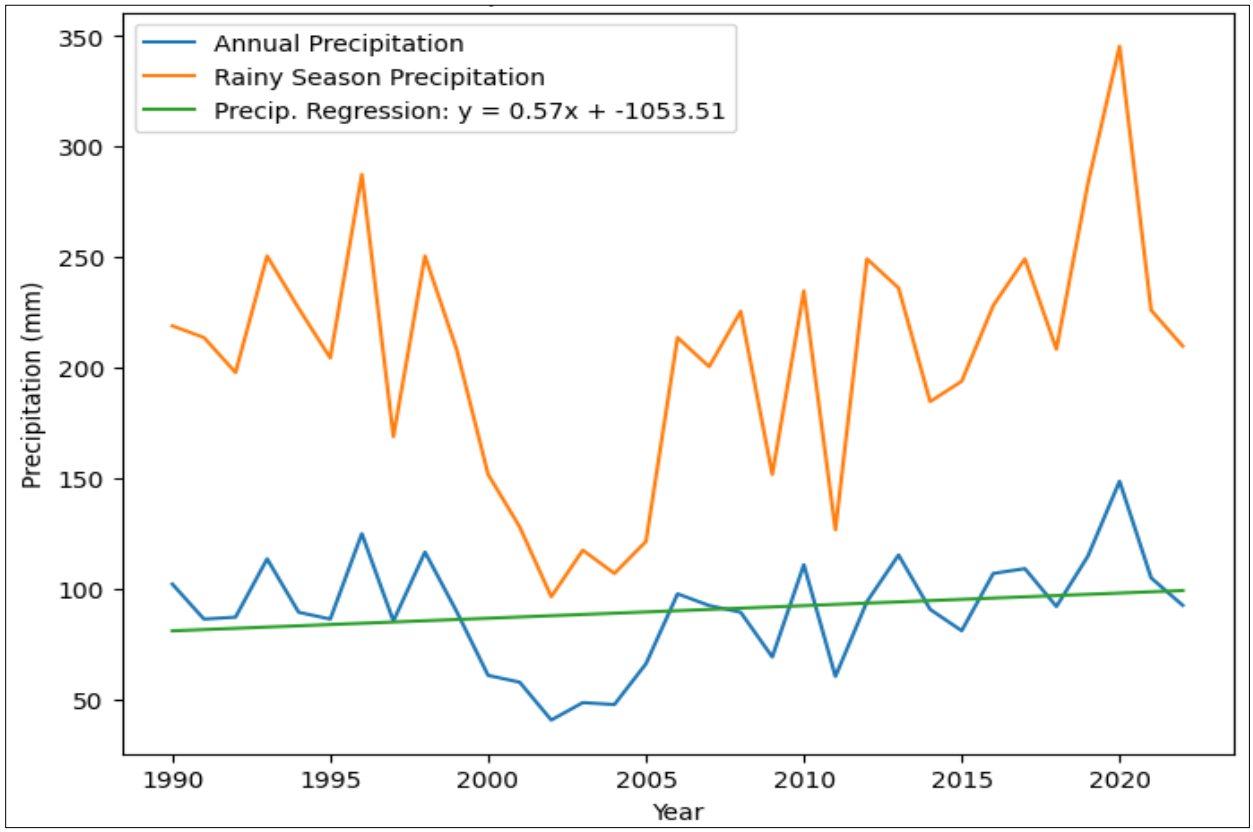


Fig. 9 Rainy season rainfall time series and trend

4.2.3.3 Temperature trend

The mean annual temperature is 16.77°C, with a standard deviation of 1.61°C. Temperatures follow a clear seasonal pattern, with the coolest months being December-February and the warmest being April-August (Yadav and Putnam, 2020). The highest mean temperature is in May at 18.67°C, while the lowest is in December at 14.44°C (Yadav and Putnam, 2020). The temperature range is relatively narrow, with a minimum of 12.21°C and a maximum of 20.69°C (Yadav and Putnam, 2020) shown in (Table 4). The R-squared value is 0.62, indicating a moderately strong positive correlation between year and temperature, suggesting a trend of slowly increasing temperatures over the time period (Cannon et al., 2015) shown in (Fig.10).

Table 4. Temperature Statistical summary

Stats	January	February	March	April	May	June	July	August	September	October	November	December
count	33.00	33.00	33.00	33.00	33.00	33.00	33.00	33.00	33.00	33.00	33.00	33.00
mean	15.74	17.39	18.44	18.50	18.67	17.42	15.33	15.29	15.61	14.79	14.32	14.44
std	0.86	0.75	0.65	0.71	0.99	0.70	0.50	0.34	0.35	0.83	1.08	1.11
min	13.68	16.01	16.60	17.37	16.80	16.04	14.27	14.40	14.89	13.32	12.33	12.21
25%	15.13	16.90	17.96	18.05	17.89	16.94	15.12	15.07	15.40	14.33	13.59	13.73
50%	15.78	17.32	18.51	18.49	18.73	17.28	15.30	15.31	15.54	14.69	14.08	14.42
75%	16.28	17.81	18.76	18.93	19.04	17.91	15.48	15.49	15.82	15.05	14.98	15.17
max	17.61	18.99	19.73	20.27	20.69	19.01	16.79	15.86	16.45	17.07	16.88	17.23

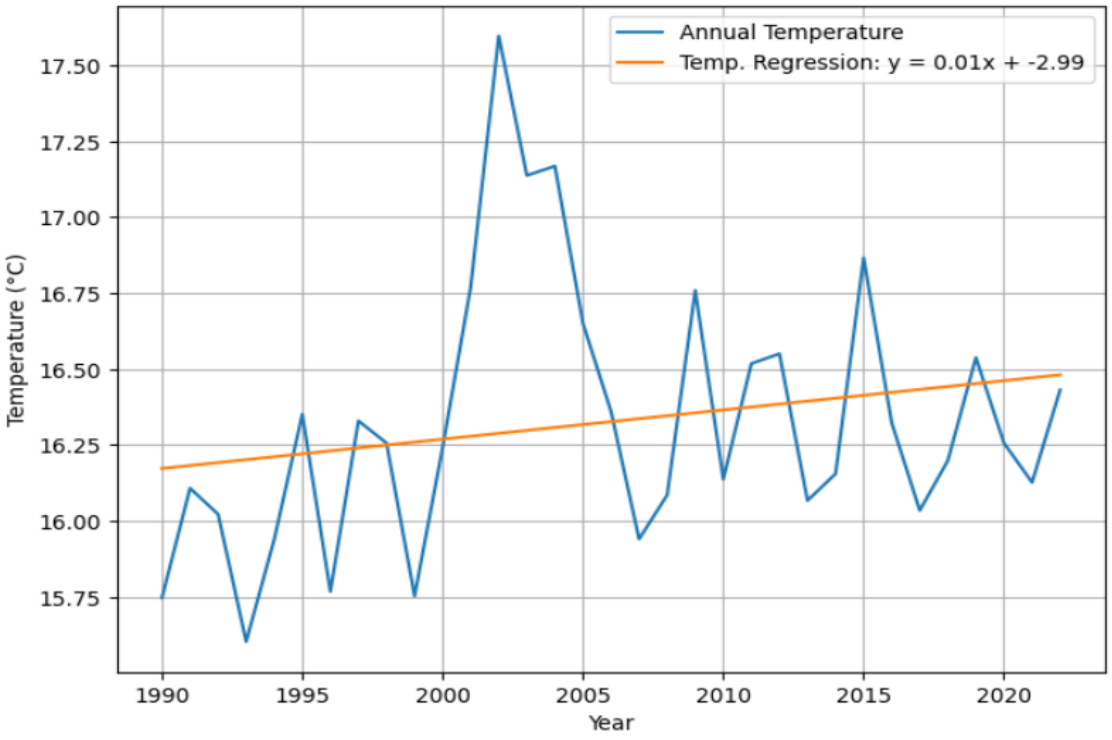


Fig. 10 Temperature time series with trend

4.3 Temperature-rainfall relationship

4.3.1 Correlation trend

Correlation coefficients range from −1 to 1, with −1 indicating a perfect negative correlation, 0 indicating no correlation, and 1 indicating a perfect positive correlation. A correlation coefficient of

−0.41 indicates a moderate negative correlation (Salkind, 2010), as stated in the conclusion. The calculated correlation coefficient between annual rainfall and temperature gives a value of 0.41. This indicates a moderate negative correlation as temperatures increase, (Fig. 11) rainfall tends to increases as shown in (Fig.12).

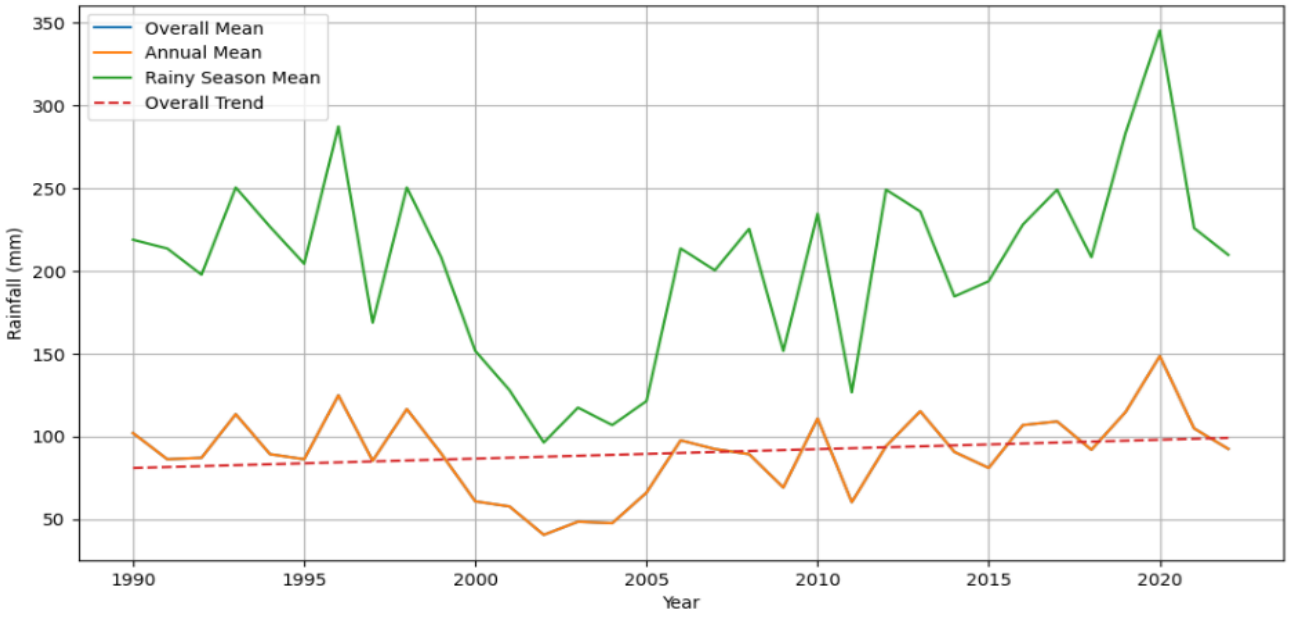


Fig. 11 Rainfall overall-annual mean and trend

4.3.2 Strength of Correlation

The strong negative correlation between annual precipitation and temperature ($r = -0.68$, p-value < 0.001) suggests a robust inverse relationship between the two variables. The weaker, but still significant, negative correlation during the rainy season ($r = -0.40$, p-value $= 0.02$) indicates that this relationship is also present within the wetter months of the year, though to a lesser degree (Cohen, 1988).

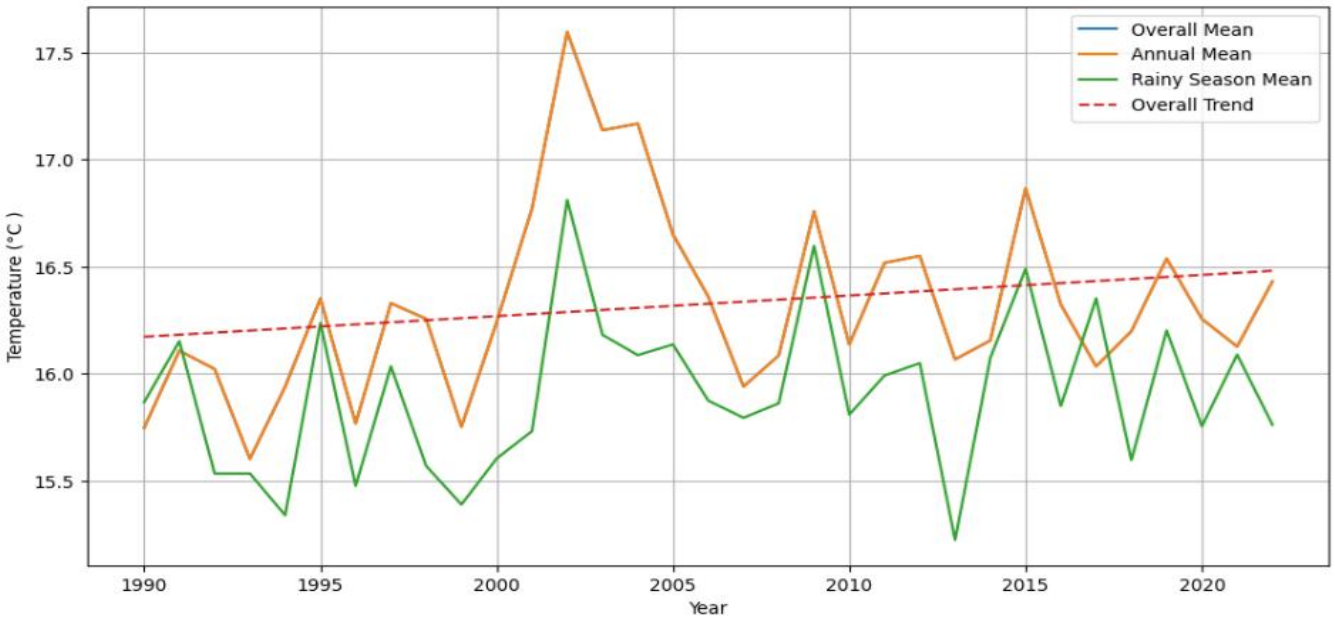


Fig. 12 Temperature overall-annual mean and trend

4.3.3 Temperature statistical significance

The p-value of less than 0.001 for the strong negative correlation ($r = -0.68$) indicates that the relationship is statistically significant, meaning it is unlikely to have occurred by chance. The p-value of 0.02 for the moderate negative correlation during the rainy season ($r = -0.40$) also indicates statistical significance **(**Wasserstein and Lazar, 2016**).**

4.4 Topographic and Morphometric factor trend

A detailed topographic and morphometric analysis was conducted using high-resolution digital elevation models (DEMs). Landform classifications were performed to identify areas prone to flooding based on factors such as slope, aspect etc. The impact of the complex terrain on flood propagation was evaluated.

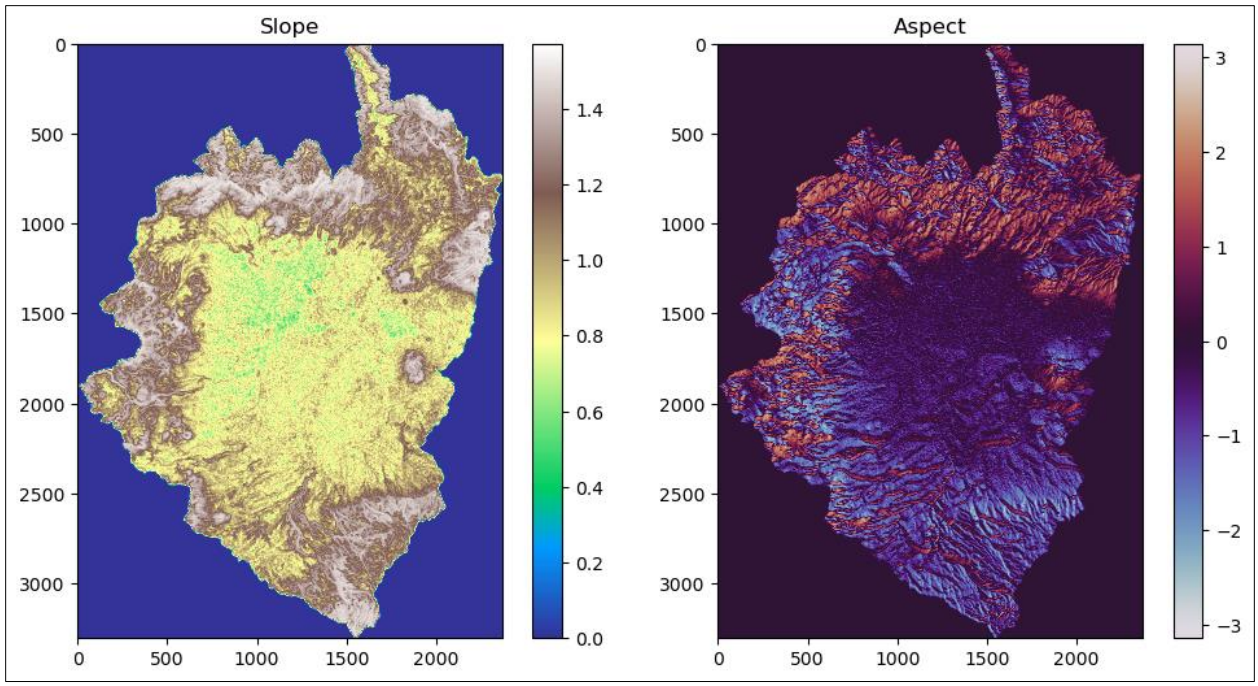


Fig. 13 Slope and aspect map

4.4.1 Slope

The slope range of 0.0 to 1.4 percent, with most values being less than 0.8, indicates a very gentle, relatively flat terrain. Most area of Becho plain slope is less than 0.8 are considered very gentle, often associated with areas such as plains, valleys, or gently rolling hills. The predominance of slopes less than 0.8 indicates a very flat, low-lying landscape in the study area, makes it more susceptible to flood inundation, as they lack the topographic variation to facilitate natural drainage (Wison and Gallant, 2000).

4.4.2 Aspect

Aspect represents the orientation of the slope with respect to the compass direction, and it is measured in degrees clockwise from the north. The Aspect values ranging from −3 to 3 in Python while working with DEM data, it is likely that these values have been normalized or scaled for easier interpretation, Aspect values close to 0 (between −0.8 and 0.8) typically represent slopes that are relatively flat or nearly flat with respect to the cardinal directions (North, South, East, West). Here is how aspect values close to 0 degrees can be interpreted:

The aspect range of −3 to 3, with most values being less than 0.8, indicates a terrain with a high degree of uniformity in slope orientation. Aspect values close to 0 (between −0.8 and 0.8) typically represent flat or gently sloping areas that do not have a pronounced directional orientation. The uniform aspect, also mostly less than 0.8, suggests a lack of significant directional orientation of the terrain. This lack of variation in slope direction can contribute to the accumulation of water during flood events, as there are fewer natural drainage pathways (Burrough and McDonnell, 1998).

4.4.3 Landforms

The terrain analysis (TPI-based landform classification) indicates that the Becho plain is predominantly a plain landform morphology, which is typically formed by the deposition of sediments carried by rivers (Weiss, 2001; Guida et al., 2016). Over time, rivers deposit fine particles such as silt, sand, and clay, creating flat or gently sloping areas along their banks that are prone to flooding during the rainy season (Burrough and McDonnell, 1998). Alluvial plains, such as the Becho plain, are often highly fertile and suitable for agriculture (Guida et al., 2016).

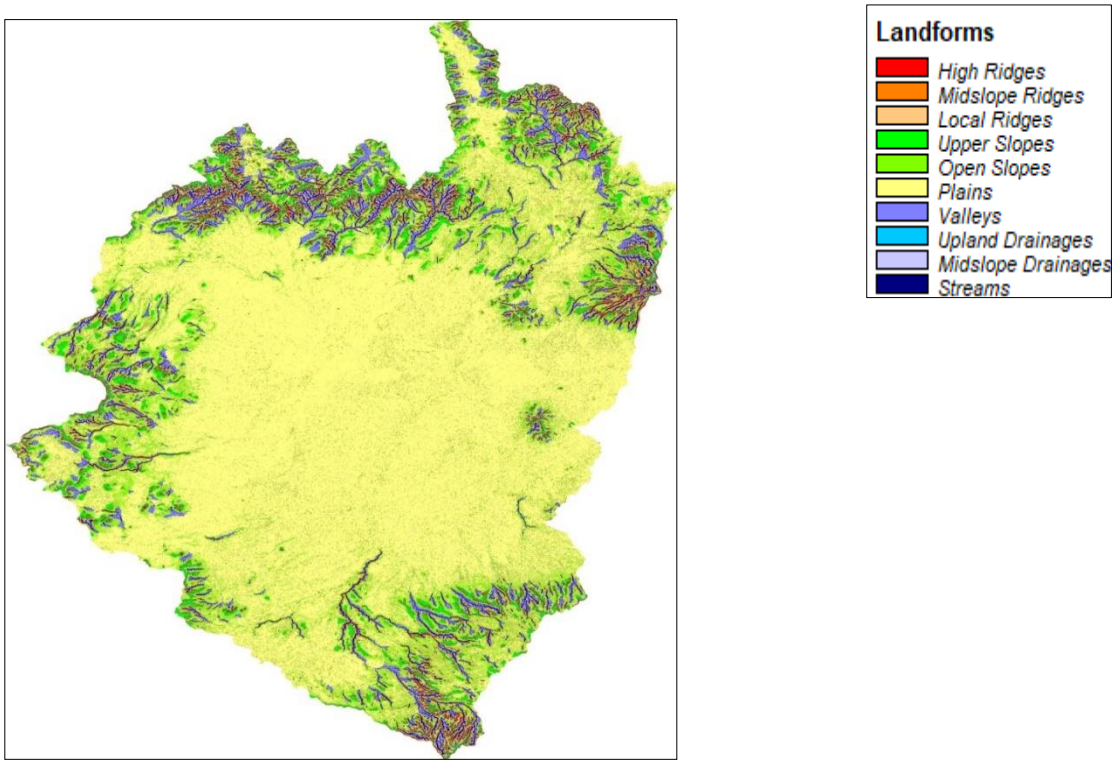


Fig. 14 Landforms map

4.5 Land-use and land-cover factor

The vegetation cover and type (such as forests, wetlands, agricultural lands) and soil characteristics (infiltration, permeability) are the main factors affecting flood inundation that were considered in this study (Balthazar et al., 2013; Amanambu et al., 2019). As described in section 3.1.5, the Becho plain land use is dominantly characterized by intensive farming, i.e., agricultural land cover (Alemayehu et al., 2009) shown in (Fig.15).

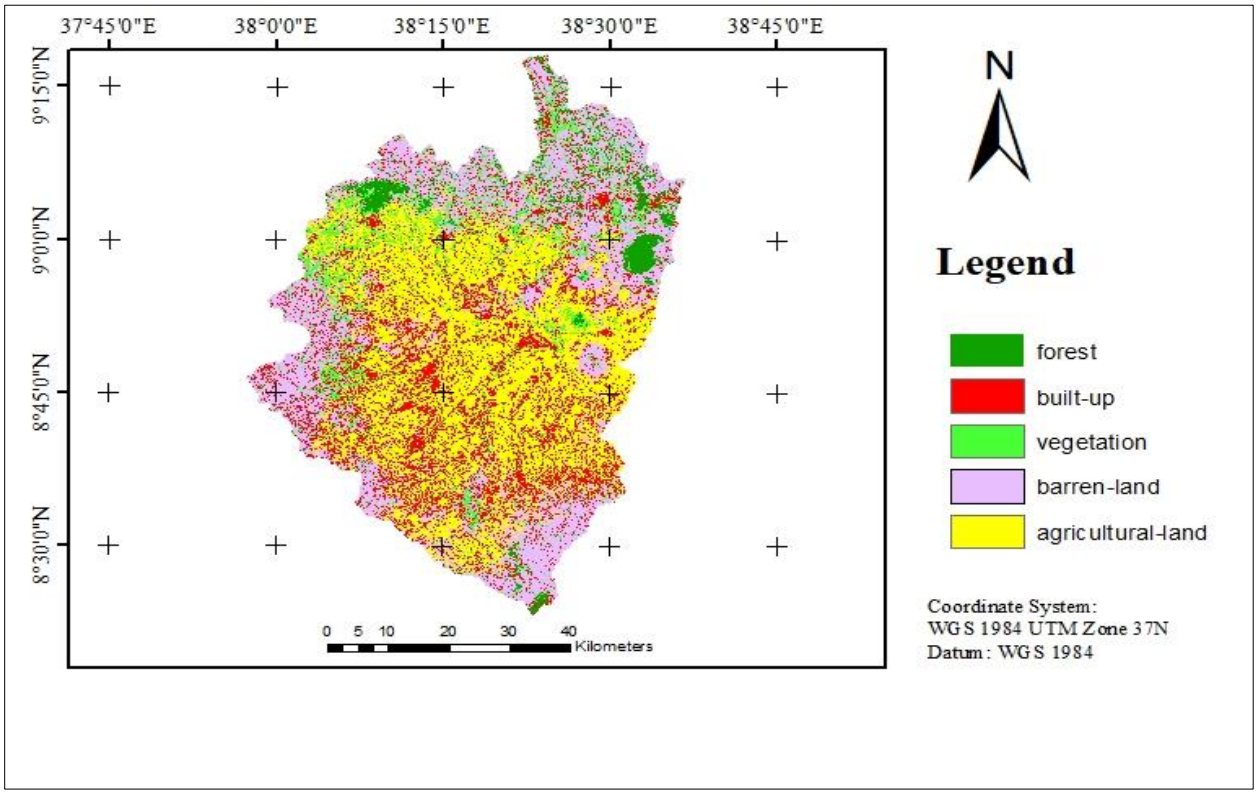


Fig. 15 Land-use and land-cover classification map

Table 5. Land-use and land-cover classes

Class name	Area (km^2)	Area (%)	Grid code
Forest	174.22	3.50	1
Vegetation	1187.57	23.88	2
Built-up land	547.37	11.01	3
Barren land	1467.55	29.52	4
Agricultural land	1595.45	32.09	5

4.5.1 Accuracy assessment

To assess the accuracy of the land cover classification, a random sample of 125 ground truth points were selected using ArcGIS software and Google Earth (Congalton and Green, 2019). These ground truth points were then compared with the classification results at the same locations. The overall accuracy of the land cover classification was calculated to be 84%, and the Kappa Coefficient was 78%, indicating a strong agreement between the classification and the ground truth data (Pontius and Millones, 2011) (Appedix 1).

4.6 Hydrological factor

4.6.1 Awash river maximum flood flow data analysis

Table 6. Awash river maximum flow statistics

	Jan	Feb	Mar	Apr	May	Jun	Jul	Augt	Sepr	Oct	Novr	Decr
count	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00	24.00
mean	0.94	1.08	2.25	4.91	4.69	14.65	36.61	32.60	34.80	15.87	2.71	1.04
std	0.84	1.05	2.59	4.32	4.78	10.69	13.40	20.10	17.90	13.00	3.02	0.74
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.37	0.50	0.72	1.30	1.32	8.02	33.48	26.37	31.88	7.91	1.24	0.63
50%	0.78	0.90	1.03	3.53	2.84	11.52	36.03	37.79	38.16	13.45	2.01	1.16
75%	1.20	1.27	2.67	7.55	5.78	21.45	48.15	49.07	47.59	21.83	2.89	1.44
max	2.93	4.55	8.17	13.06	16.31	43.84	50.00	52.72	56.58	44.34	13.52	2.97

4.6.2 Overall river statistical description

The overall flow data of the Awash River shows variation across different months. The mean flow ranges from 0.34 m^3/s in May to 43.84 m^3/s in June, with a relatively high standard deviation, indicating significant variability in flow rates (Kassa and Wale, 2018) show in (Table 6). The maximum flow occurs in June, while the minimum flow is observed in January. The flow data exhibits seasonal patterns, with higher flows during the rainy season (June to September) compared to other months (Kassa and Wale, 2018).

4.6.3 Flood frequency analysis models

The GEV (Generalized Extreme Value), Gumbel, and Log-Pearson Type III (LP3) distributions (Table 7) are commonly used in hydrological applications for estimating extreme values of given data sets (Millington et al., 2011). The GEV and LP3 distributions use three parameters (location, scale, and shape), while the Gumbel distribution uses two parameters (location and scale) (Millington et al., 2011).

Table 7. Flood frequency models

Sr .No	Distribution name	Parameters	Estimation method
1	Generalized Extreme Value(GEV)	Shape (k), scale (σ), location (μ)	Method of L-moments
2	Gumbel	max/EV1 Scale (σ), location (μ)	Method of moments
3	Log Pearson 3	Shape (α), scale (β), location (γ)	Method of moments

4.6.3.1 Goodness of Fit tests

To determine the best-fitting distribution, the Kolmogorov-Smirnov (K-S) goodness-of-fit test was conducted (Massey, 1951). The distribution with the lowest K-S statistic and the highest p-value (usually p-value > 0.05) is considered the best fit for the given data. The K-S test results showed that the GEV distribution had the lowest K-S statistic (0.1766) and the highest p-value (0.40), indicating that it was the best fit among the three distributions tested (Gumbel, LP3, and GEV) (Kassa and Wale, 2018). To select the best distribution, the K-S statistics and p-values for each distribution (Gumbel, Log-Pearson Type III, and GEV) were compared shown in (Table 8). The distribution with the lowest K-S statistic and the highest p-value (usually p-value > 0.05) is considered the best fit for the given data (Massey, 1951).

Table 8. K-S Test model comparison

Sr no.	Model	K-S statistic	p-value
1	Gumbel	0.2437	0.09
2	Log-Pearson Type III	0.1858	0.34
3	GEV	0.1766	0.40

Based on the K-S test results, the GEV distribution had the lowest K-S statistic (0.1766) and the highest p-value (0.40), indicating that it is the best fit among the three distributions tested (Gumbel, Log-Pearson Type III, and GEV) (Kassa and Wale, 2018). The GEV distribution is often used to model extreme value data, which is likely the case for this dataset (Fig. 16).

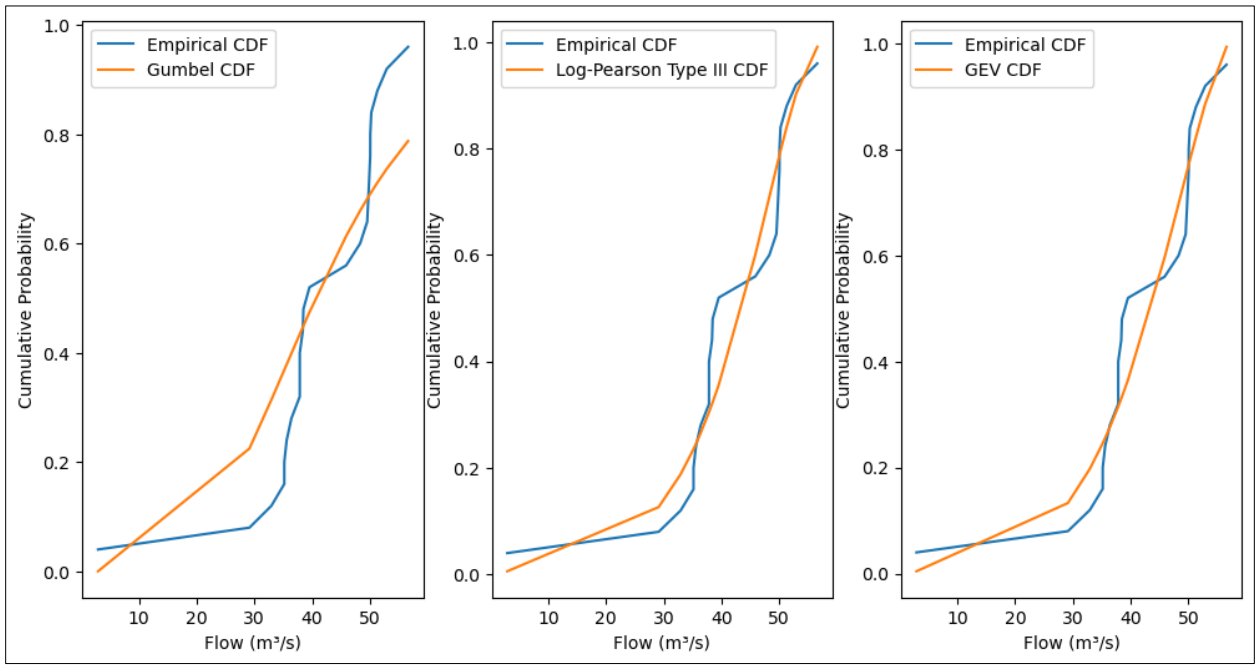


Fig. 16 Frequency models comparison graph

4.6.4 Flood frequency analysis using GEV distribution

The flood frequency analysis was conducted using the Generalized Extreme Value (GEV) distribution, which is often used to model extreme value data (Kassa and Wale, 2018). The overall

flood frequency was calculated as 127.98 events per year, while the rainy season flood frequency was 200.95 events per year (Kassa and Wale, 2018).

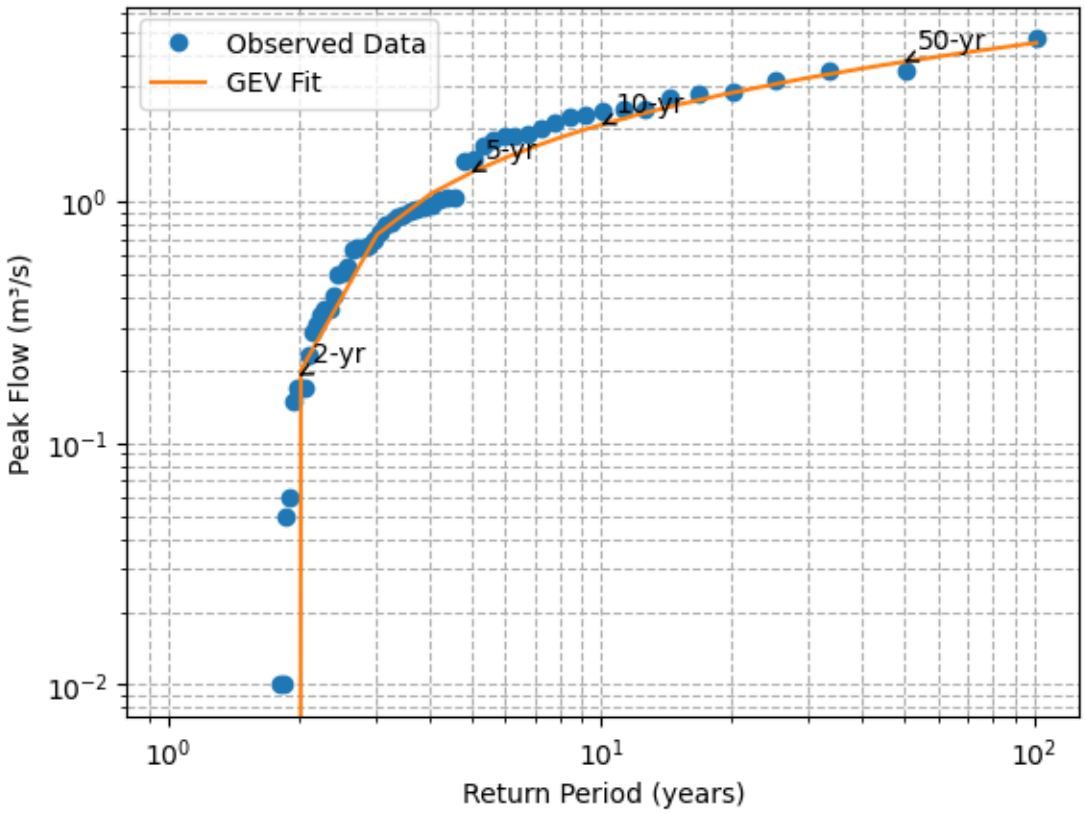


Fig. 17 Awash River flood frequency log-curve

Table 9. Return period(years) and Peak flow(m^3/s)

n-Year	Peak flow(m^3/s)	Probability
2	2.90	0.50
5	8.25	0.20
10	15.45	0.10
50	58.50	0.02

4.7 Flood frequency and Rainfall spatiotemporal relationship

4.7.1 Time series trend

The precipitation data shows a clear seasonal pattern, with higher rainfall in the July and August months compared to June and September (Kassa and Wale, 2018). The Awash River flow data also exhibits a similar seasonal pattern, with higher flows in the July, August, and September m onths compared to June (Kassa and Wale, 2018). This suggests a strong correlation between the precipitation and river flow data, indicating that the river flows are primarily driven by the seaso nal rainfall patterns in the region (Kassa and Wale, 2018) shown in (Fig.18).

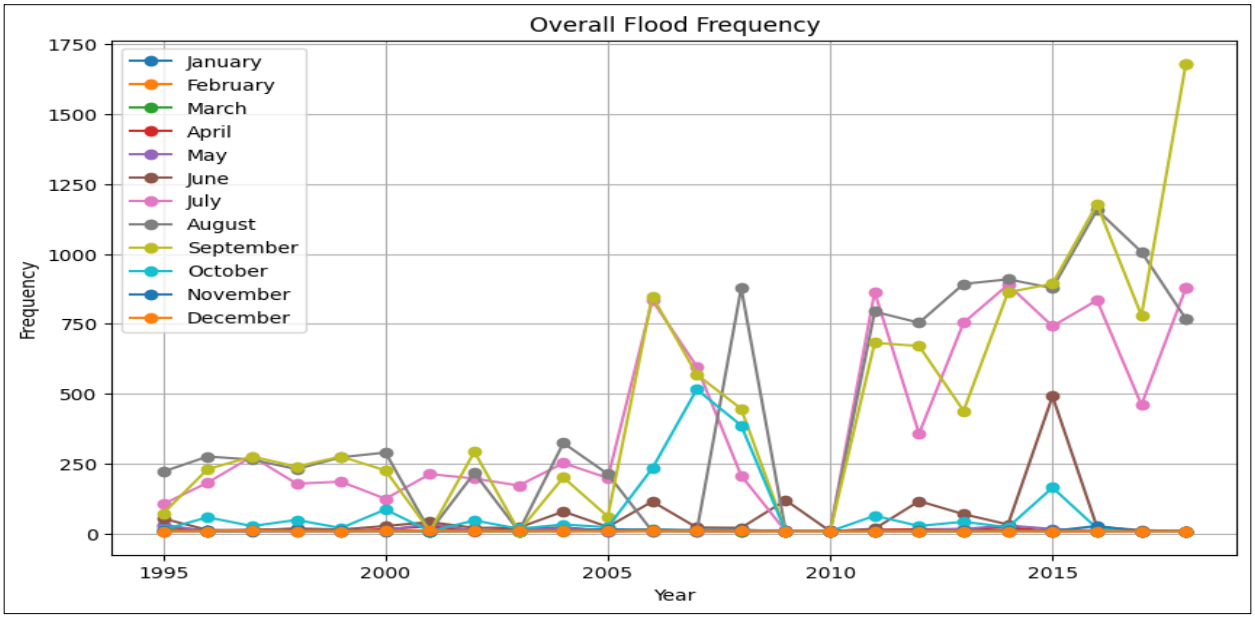


Fig. 18 Overall flood frequency

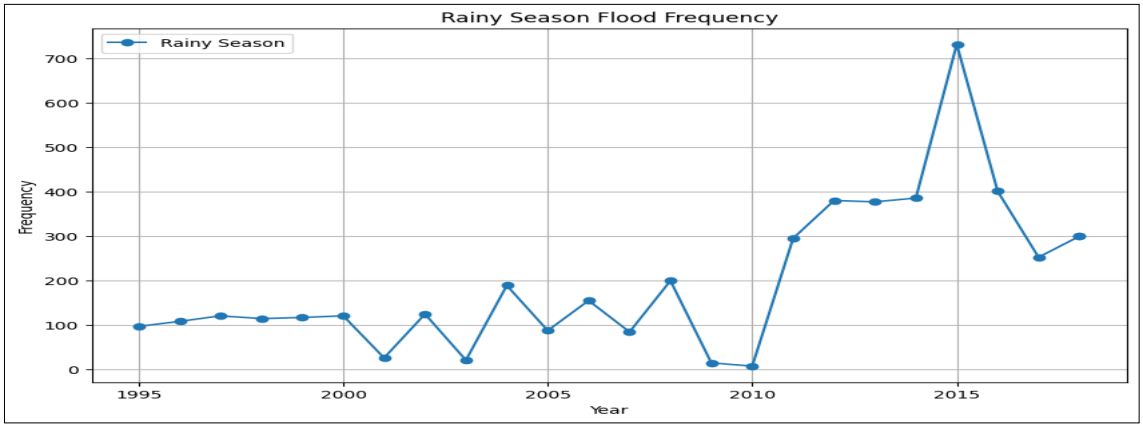


Fig. 19 Rainy Season flood frequency

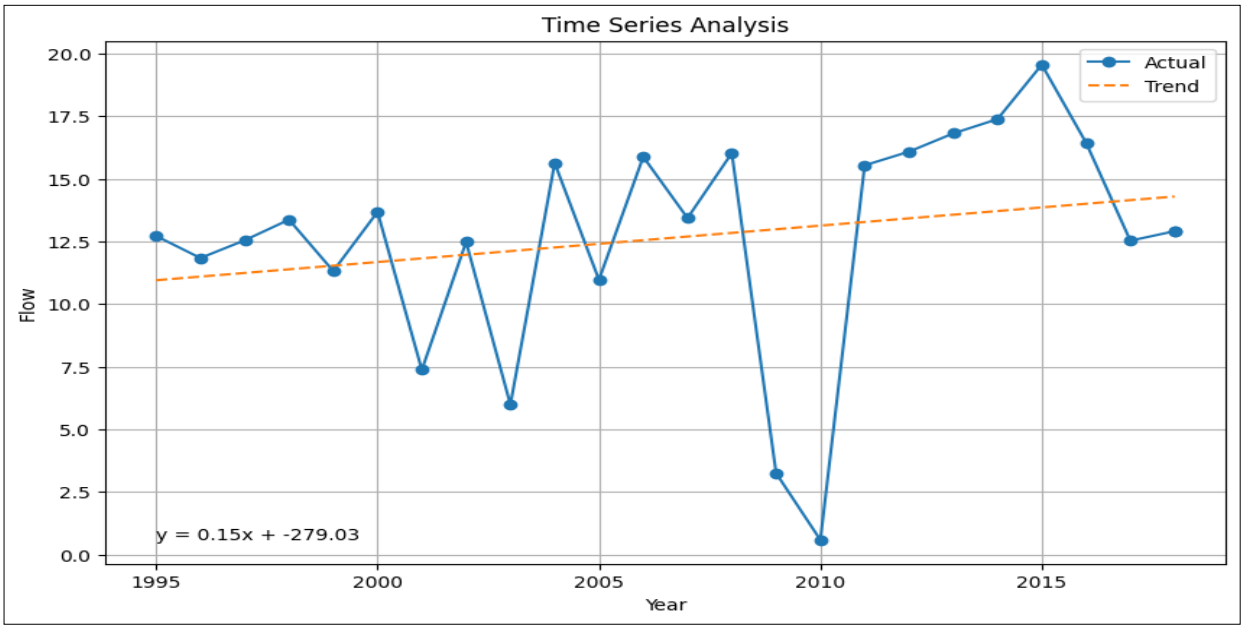


Fig. 20 Overall flood time series with trend

4.7.2 Flood trend status:

The trend analysis showed a general increasing trend in the mean flow, with an overall trend of 0.12 m^3/s per year (Fig. 20) and a rainy season trend of 0.56 m^3/s per year Kassa and Wale, 2018) shown in (Fig. 19).

Table 10. Rainy season (June–September) flow statistics

	June	July	August	September
count	22.00	22.00	18.00	20.00
mean	15.98	39.94	43.47	41.76
std	10.15	7.57	6.82	9.05
min	4.02	27.82	35.16	22.04
25%	9.50	33.88	37.50	36.29
50%	12.07	37.38	43.90	42.76
75%	22.78	48.24	49.86	48.85
max	43.84	50.00	52.72	56.58

4.7.3 Rainy season statistical description

The statistical description of the rainy season (June-September) flow in the Upper Awash River Basin is presented as opined by Kassa and Wale (2018). The mean flow during the rainy season ranged from 15.98 m³/s in June to 41.76 m³/s in September, with a relatively high standard deviation indicating variability in the flow rates (Kassa and Wale, 2018) show in (Table 11).

Table 11. Rainy season (June – September) precipitation statistics

	June	July	August	September
count	33.00	33.00	33.00	33.00
mean	138.21	267.53	268.68	139.35
std	54.59	78.94	78.09	72.43
min	47.46	116.02	131.84	42.19
25%	100.20	237.30	221.48	84.38
50%	137.11	258.40	290.04	137.11
75%	163.48	321.68	321.68	174.02
max	284.77	474.61	469.34	400.78

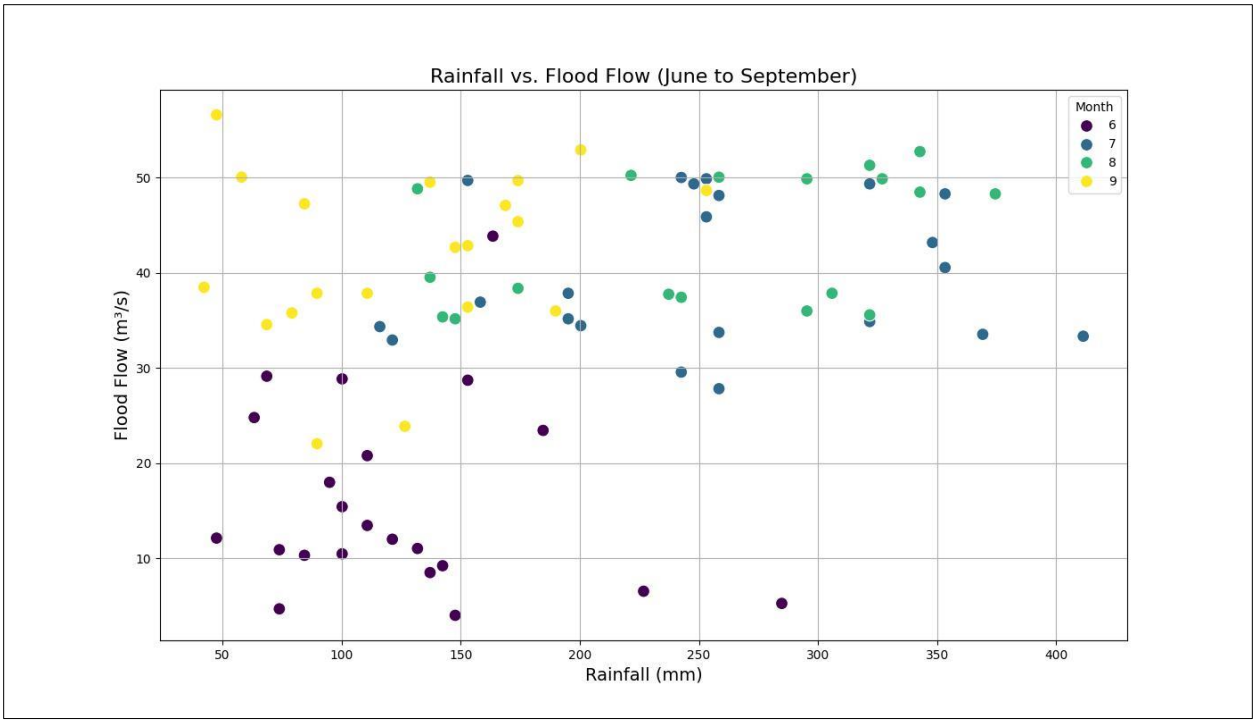


Fig. 21 Rainfall *vs* flood flow scatter plot

4.7.4 Correlation between rainfall and flood

The correlation analysis revealed a moderate positive correlation (r=0.39) between rainfall and f lood flow (Kassa and Wale, 2018). This suggests that as rainfall increases, the flood magnitude also tends to increase, but the relationship is not very strong (Kassa and Wale, 2018). The time s eries analysis showed that both precipitation and Awash river flow exhibit distinct seasonal patt erns, with higher values during the rainy season (June to September) and lower values during th e dry season (Kassa and Wale, 2018)

4.7.5 Rainfall and Flood flow trend pattern

The line plots illustrate the trends of mean rainfall and flood flow over time (Kassa and Wale, 2 018). Both variables show fluctuations, but there is a general increasing trend in mean rainfall a nd flood flow (Kassa and Wale, 2018). The precipitation data does not show any clear long-ter m trends, as the mean, standard deviation, and other statistical measures remain relatively consi stent across the years (Kassa and Wale, 2018). Similarly, the Awash River flow data does not e xhibit any significant long-term trends, suggesting that the river flow patterns are relatively stab le over the observed period (Kassa and Wale, 2018) shown in (Fig. 22).

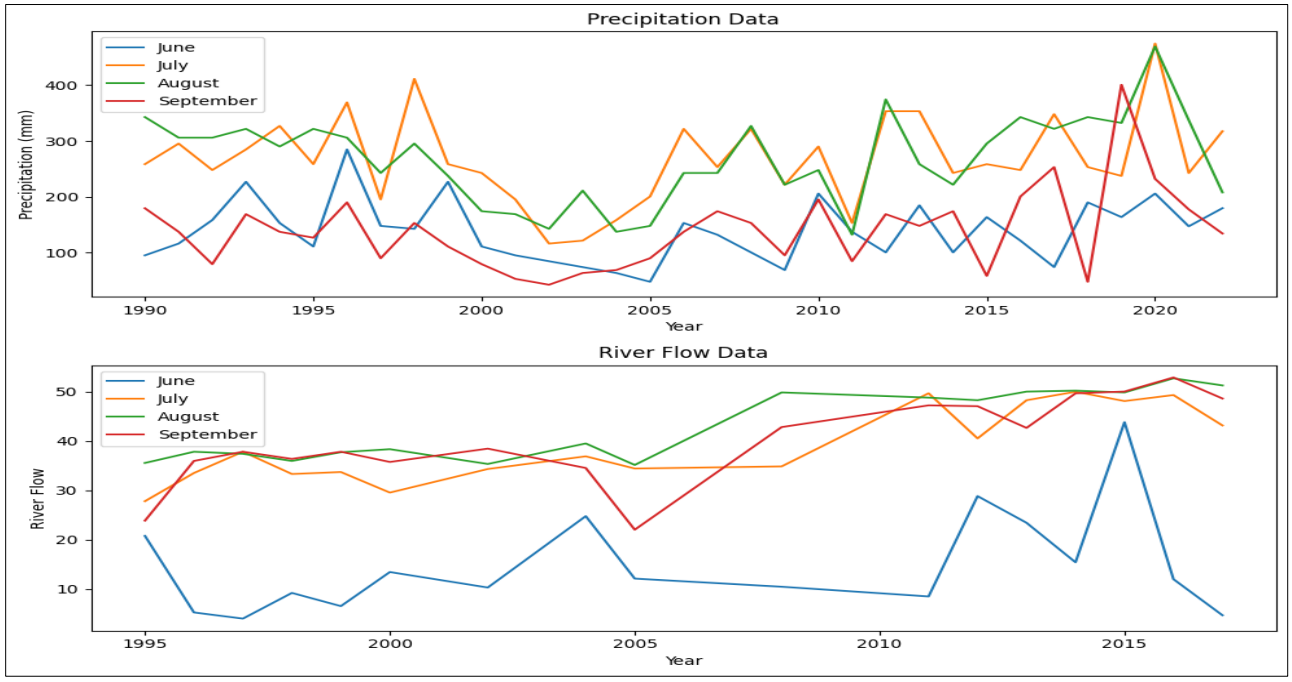


Fig. 22 Rainy season-flood flow and rainfall time series

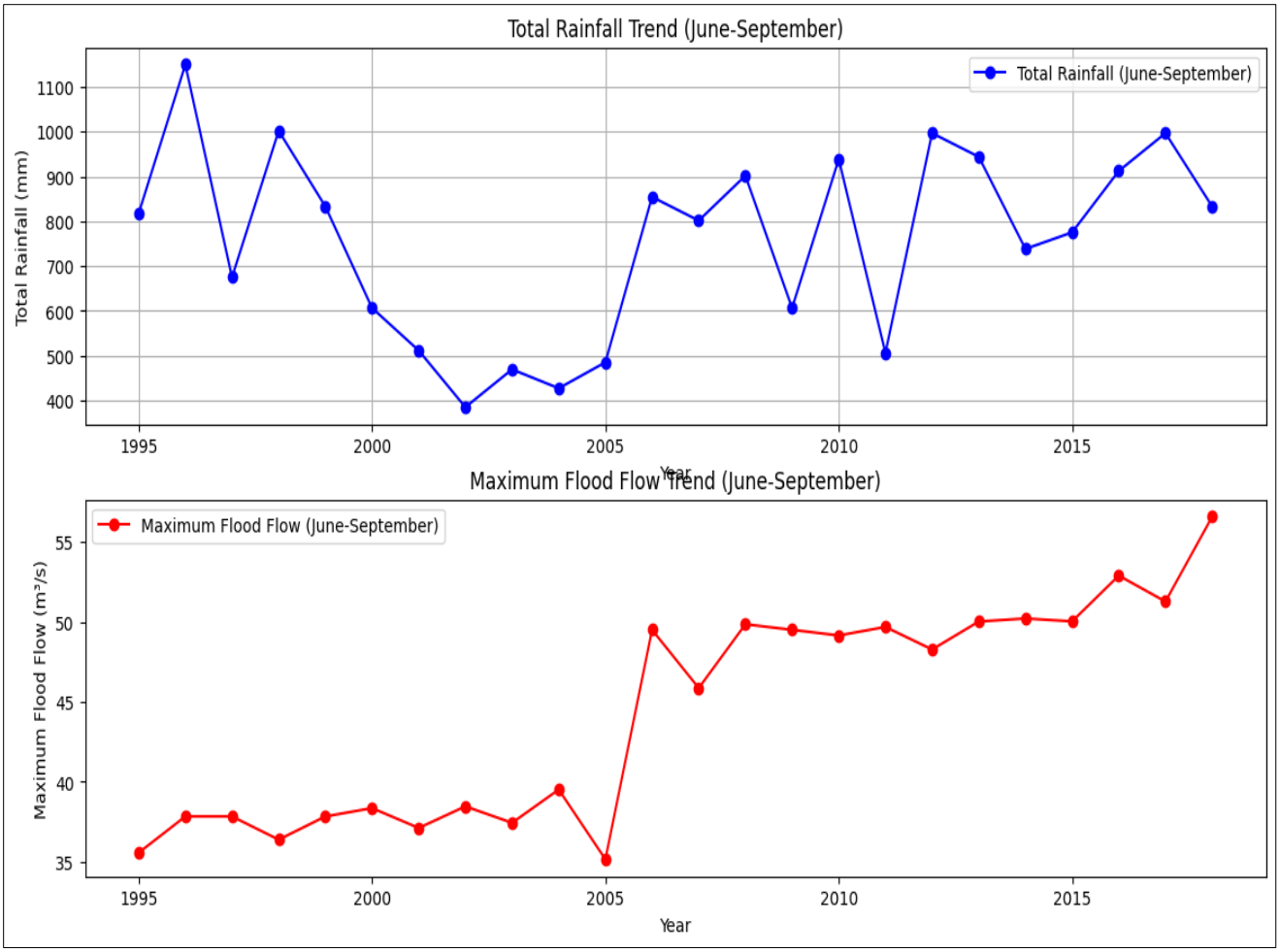


Fig. 23 Total rainfall and maximum flood flow trends

The correlation analysis revealed that overall rainfall and rainy season rainfall both exhibit a positive correlation with flood frequency, but the correlation appears to be stronger for rainy season rainfall (Kassa and Wale, 2018). The time series plots for both precipitation and Awash river flow show distinct seasonal patterns, with higher values during the rainy season (June to September) and lower values during the dry season (Kassa and Wale, 2018). The precipitation time series plot indicates that there are year-to-year variations in the precipitation levels, with some years experiencing higher or lower rainfall compared to the average (Kassa and Wale, 2018). The Awash river flow time series plot mirrors the precipitation patterns, demonstrating the strong influence of precipitation on the river's flow regime (Kassa and Wale, 2018) shown in (Fig. 23).

4.7.6 Recurrence interval and peak flow trend

The log-log curve allows for the estimation of the peak flow magnitude for different return periods (Kassa and Wale, 2018). By extrapolating the curve, the peak flow for rare events with higher return periods can be estimated (Kassa and Wale, 2018). This information is valuable for designing infrastructure, such as bridges and dams, to withstand floods of varying magnitudes (Kassa and Wale, 2018). The recurrence interval analysis shows the relationship between the annual maximum flows and the return periods, allowing for the estimation of flow magnitudes for various return periods (Kassa and Wale, 2018). Based on the recurrence interval analysis, the peak flow table provides the estimated flow magnitudes for different return periods (Kassa and Wale, 2018). This table can be useful for planning and design purposes, as it allows for the assessment of flood risks and the implementation of appropriate mitigation strategies (Kassa and Wale, 2018).

4.7.7 River peak flow

The peak flow *vs* return period log-log curve shows a linear relationship between the log of the return period and the log of the peak flow magnitudes, as expected from the Gumbel Extreme Value distribution (Kassa and Wale, 2018). The observed data points fit the GEV model well, indicating that the Gumbel distribution is a suitable choice for modeling the flood frequency in this region during the rainy season (Kassa and Wale, 2018). Table 12 presents the flood magnitudes for different periods of the study area.

Table 12. Flood magnitudes for different return periods

ID	Frequency(%)	n-year	Peak Flow (m^3/s)
1	50	2	42.15
2	20	5	53.42
3	10	10	60.88
4	5	20	68.03
5	2	50	77.30
6	1	100	84.24

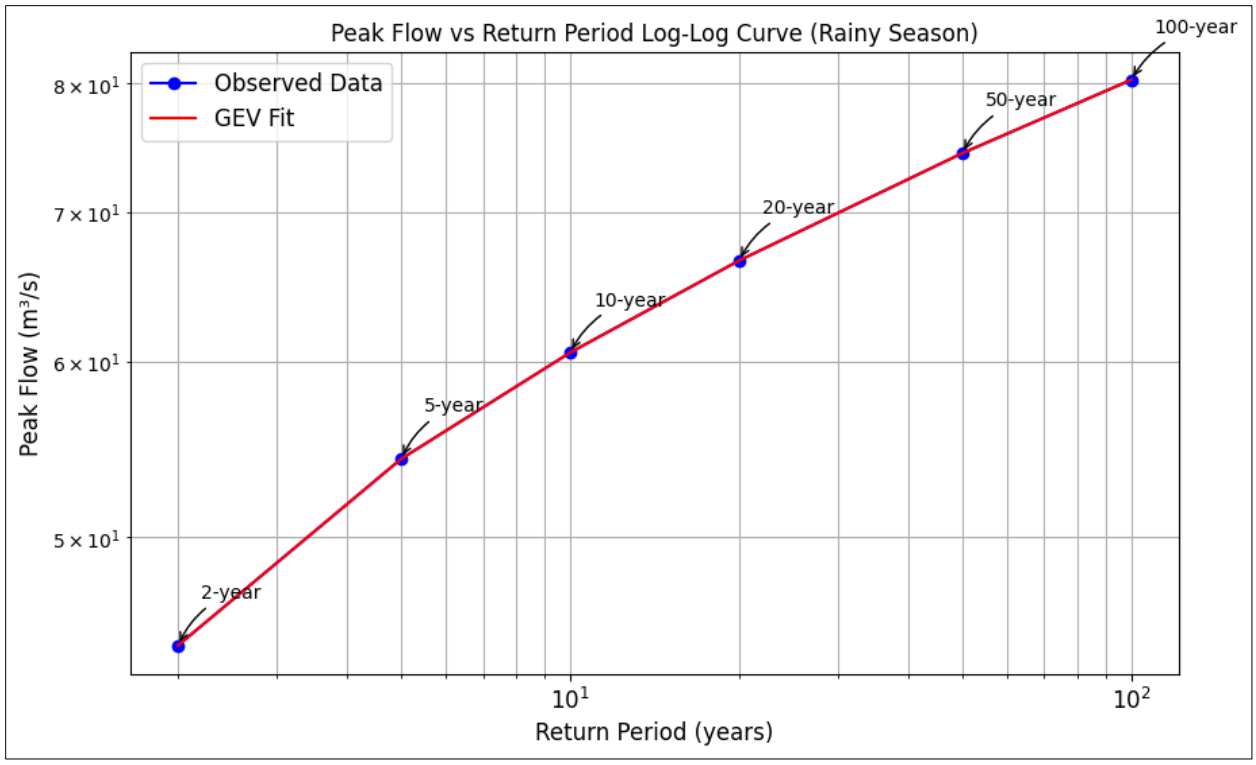


Fig. 24 Peak flow *vs* return period log-log curve

4.8 Flood Inundation extent (area) time series and trend pattern

The Python code that performs statistical analysis, trend analysis, time series analysis, and linear regression graph for the flood inundation frequency and extent study in Becho plain, upper Awash river basin, by importing pandas, matplotlib, and seaborn libraries (Kassa and Wale, 2018). The (Table 13) shows the flooded area in the Becho plain over the years (Kassa and Wale, 2018). Figure 25 and Figure 26 presents the flood area of the study area in year 2021.

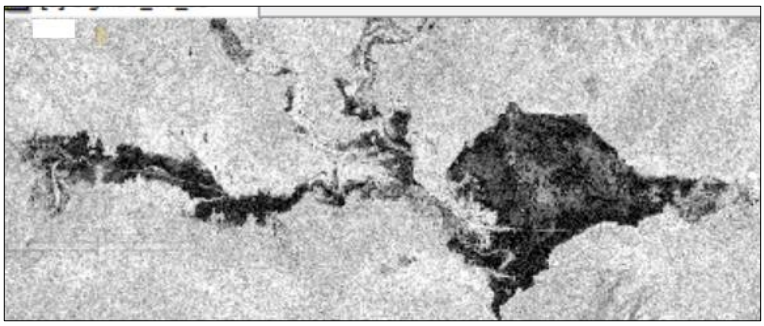


Fig. 25 Flood area image of year 2021

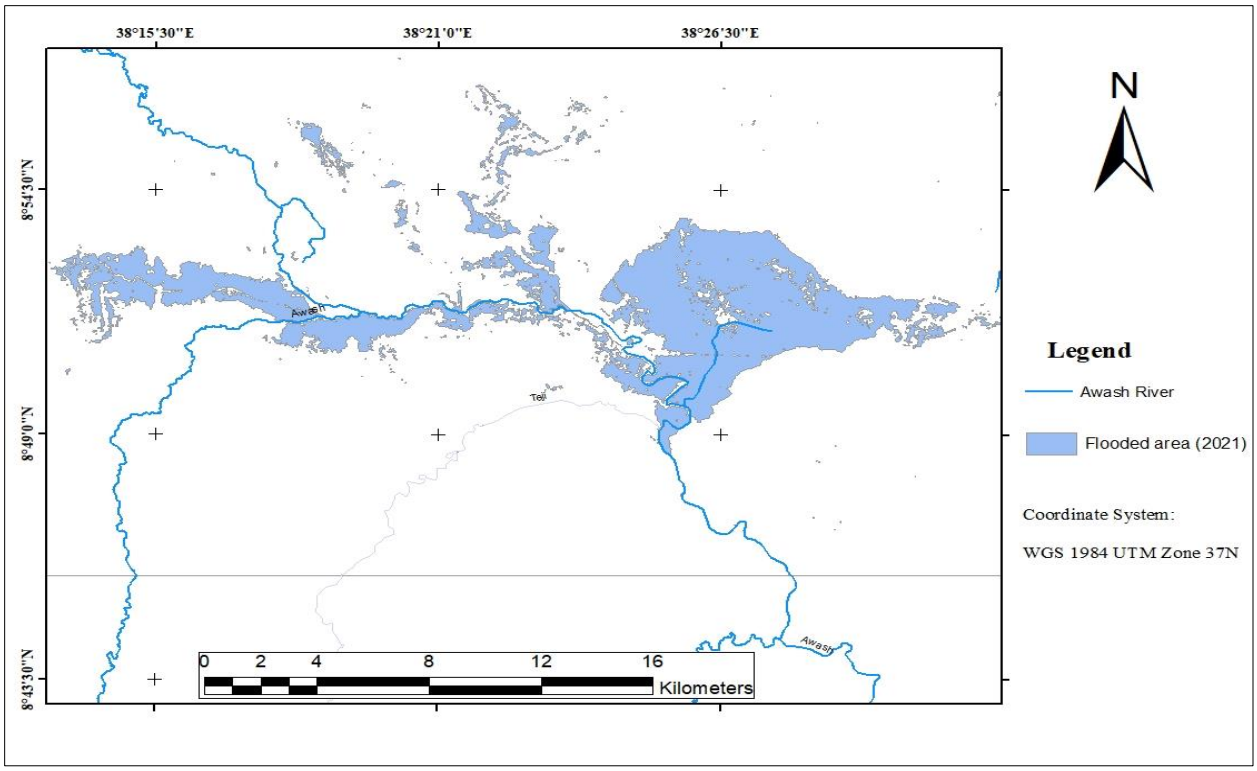


Fig. 26 Flood inundation area time series map

Table 13. Becho Plain flooded area

ID	Year	Area Flooded(km^2)
1	2015	20.03
2	2016	30.98
3	2017	75.92
4	2018	70.24
5	2019	66.52
6	2020	68.54
7	2021	72.66
8	2022	70.44
9	2023	50.47

Table 14. Flooded area statistics

	ID	**Year**	**Area (km^2)**
Count	9	9	9
Mean	5	2019	58.44
Std	2.74	2.74	20.23
Min	1	2015	20
25%	3	2017	50
50%	5	2019	69
75%	7	2021	70
Max	9	2023	76

4.8.1 Flood area time series pattern

The flood area time series pattern shows that the minimum flooded area recorded was 20.00 km^2, and the maximum was 76.00 km^2 (Kassa and Wale, 2018). Plotting these values against time can show the range of variation in flood extents over the years (Kassa and Wale, 2018). The quartiles (25^{th}, 50^{th}, and 75^{th} percentiles) at 50.00, 69.00, and 70.00 km^2, respectively, give insights into the distribution of flooded areas (Kassa and Wale, 2018). Analyzing how these quartiles change over time can reveal trends in inundation levels (Kassa and Wale, 2018) shown in (Table 15 and Figure 27).

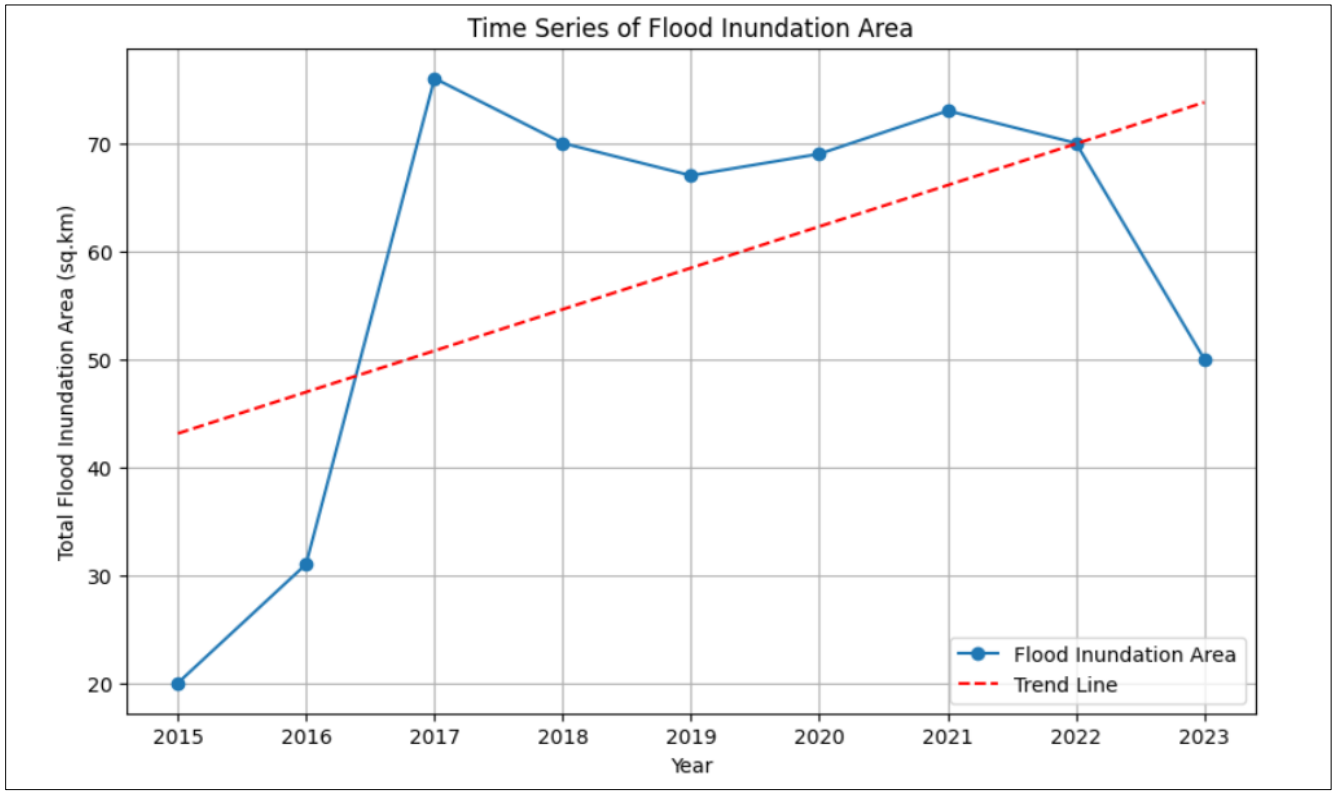


Fig. 27 Flood inundation area time series and trend

4.8.2 Flood inundation area time series trend

The flood inundation time series plot shows the trend of the flooded area over time, and the trend line indicates an increasing trend (Kassa and Wale, 2018).

4.9 Time series model algorithms validation techniques

Model validation is used to determine the best-fitting model for the study area in analysis modeling climate and river flow record, by exploring different models and evaluating their performance using appropriate metrics (Kassa and Wale, 2018). In this study, three models were considered for analysis and comparison: Support Vector Regression (SVR), Random Forest (RF), and Linear Regression (LR) (Kassa and Wale, 2018). These models were chosen because they can handle temporal dependencies and patterns in climate and river flow data, as well as non-linear relationships, interactions, and complex patterns in the data (Kassa and Wale, 2018). Table 16 shows the Mean Absolute Error (MAE) and Cross-Validated MAE values for the Awash Belo River maximum flow prediction from 1995 to 2018 for the three models (Kassa and Wale, 2018).

Table 15. Model types and validation

Sr. No	Model	MAE	MAE (CV)
1 .	Support vector regression (SVR)	25.19	27.00
2 .	Random Forest	20.14	23.08
3 .	Linear Regression	17.71	20.63

Based on the model comparison, the Linear Regression model performs the best, with a MAE of 17.71 and a Cross-Validated MAE of 20.63 (Kassa and Wale, 2018). This indicates that, on average, the predicted maximum flow values deviate by approximately 17.71 units from the actual values, and the model's performance is consistent across different subsets of the data (Kassa and Wale, 2018).

CHAPTER V

DISCUSSION

5.1 Climate variables (Rainfall and Temperature)

The precipitation analysis revealed an increasing trend over the study period, indicating a heightened risk of flood occurrence in the future (Gebremicael et al., 2013). The observed increase in precipitation could lead to higher surface runoff, enhanced river discharge, and more frequent extreme precipitation events all of which contribute to increased flood risk (Gebremicael et al., 2013). The temperature analysis also showed an increasing trend, which could further exacerbate the flood risk through enhanced evaporation and snowmelt processes (Asfaw et al., 2018). The strong positive correlation between precipitation and temperature suggests a complex interaction between these two climatic variables, necessitating a holistic approach in flood risk assessment and management (Asfaw et al., 2018).

5.2 Topography and morphology

The topographic and morphometric analysis highlighted the role of the Becho plain's terrain characteristics in shaping the flood dynamics (Jemberie and Gebrehiwot, 2021). The relatively flat and low-lying nature of the plain, with gentle slopes and poorly defined drainage channels, makes it particularly susceptible to inundation during high-rainfall events (Jemberie and Gebrehiwot, 2021). This topographic setting, combined with the expansion of agricultural and settlement areas, has likely modified the natural drainage patterns and increased surface runoff, leading to more frequent and extensive flood occurrences (Jemberie and Gebrehiwot, 2021).

5.3 Land use land cover

The land use and land cover changes emerged as a significant factor influencing flood inundation in the Becho plain (Gizaw and Gan, 2017). The expansion of agricultural and settlement areas has likely altered the natural infiltration and runoff processes, leading to increased surface water accumulation and heightened flood risk (Gizaw and Gan, 2017). This underscores the importance of considering land use planning and management as a crucial component of flood risk mitigation strategies (Gizaw and Gan, 2017).

5.4 Hydrology

The hydrology analysis, particularly the Awash River maximum flow, showed a clear link between high discharge events and flood occurrences in the Becho plain (Melesse et al., 2011). The flood frequency analysis further revealed an increasing trend in flood events and their magnitudes, which is consistent with the observed precipitation and temperature trends. This highlights the need for proactive flood risk mitigation measures to address the growing threat of flood hazards in the study area (Melesse et al., 2011).

5.5 Flood inundation extent

During the rainy season, the Awash River basin experiences increased water inflow from precipitation and tributary streams from the surrounding highlands (Jemberie and Gebrehiwot, 2021). Due to the flat plain characteristics of the Becho flood plain, the natural gravity flow of the river system can easily be disrupted, leading to flooding and inundation (Jemberie and Gebrehiwot, 2021). The Becho plain has a relatively flat or even below-zero slope in the river bed, which reduces the gravitational pull that would normally drive the water downstream, leading to stagnation and water level rise (Jemberie and Gebrehiwot, 2021) The narrow river channels at the Bridge and potential obstructions or constrictions create localized pressure flow effects (Melesse et al., 2011). As the water flow is constricted, the water pressure increases upstream, causing the water to back up and rise in level (Melesse et al., 2011) indicated in (Fig. 28).

Fig. 28 Awash River bridge at Tefki district

The combination of disrupted gravity flow and increased pressure flow due to the physical characteristics of the Becho plain leads to the accumulation and rise of water levels in the river basin (Jemberie and Gebrehiwot, 2021;Melesse et al., 2011). As the water levels rise and exceed the river banks, the surrounding Becho plain becomes inundated, leading to widespread flooding (Jemberie and Gebrehiwot, 2021; Melesse et al., 2011) (Fig. 29). The rainy season, which typically occurs from June to September, shows higher flow values compared to other months, indicating the influence of seasonal rainfall on the river flow (Jemberie and Gebrehiwot, 2021). The river flow exhibits significant variation throughout the year, with the highest flows typically recorded in July and August, and the lowest flows occurring in February (Jemberie and Gebrehiwot, 2021).

Fig. 29 Awash river channel upstream of the bridge

The overall flood frequency and the rainy season flood frequency provide insights into the frequency of flood events in the area, suggesting that the Becho plain is prone to flooding, particularly during the rainy season (Gebremicael et al., 2013; Kassa and Wale, 2018). The analysis also reveals a slight increasing trend in flow over the years, indicating potential changes in hydrological patterns (Gebremicael et al., 2013; Kassa and Wale, 2018).

CHAPTER VI

CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

This study utilized Geospatial technology and machine learning algorithms to predict flood inundation frequency and extent in the Becho plain, central western Ethiopia. By analyzing historical flood records of the Awash River from 1995 to 2018, climate variables from 1990 to 2023, and SAR imagery from 2015 to 2023, a model was developed to predict future flood occurrences and magnitudes. The results indicated that there is a significant correlation between climate variables (rainfall and temperature) and flood frequency and extent. The time series and trend analysis showed an increasing trend in flood events over the years, which suggests that the region is becoming more susceptible to flooding.

The statistical analysis of the Becho Plain reveals significant findings regarding precipitation and river flow patterns. The rainy season, particularly July and August, experiences the highest mean precipitation, indicating these months are more prone to heavy rainfall and potential flood events. The standard deviation of precipitation during the rainy season suggests significant variability, with periods of intense rainfall increasing flood risk and periods of dry conditions. Maximum precipitation values during the rainy season indicate the potential for extreme rainfall events, which can result in severe flooding and pose a significant threat to the Becho Plain area. Additionally, distinct seasonal temperature patterns were identified, with variations across different months, which can be utilized for planning activities like agriculture and reforestation.

The flow records and analysis highlight the importance of understanding flood inundation frequency and extent in the Becho Plain. The findings suggest frequent flood events, particularly during the rainy season. An increasing trend in flow indicates potential changes in hydrological conditions that may impact future flood risk. The river flow at the Awash-Belo station shows distinct seasonal patterns, with higher flows during the rainy season (June to September) and lower flows during the dry season (January to May and November to December). The highest river flows are typically observed in July and August, coinciding with the peak of the rainy season. This study emphasizes the need for effective flood management strategies to mitigate the impact of flooding on the Becho

Plain. The application of machine learning algorithms improved the accuracy of flood predictions, demonstrating their potential for effective flood management and mitigation.

The statistical analysis highlights that the rainy season, particularly July and August, has the highest mean precipitation, leading to a higher likelihood of heavy rainfall and potential flood events. The standard deviation shows significant variability, implying periods of intense rainfall and increased flood risk as well as dry conditions. The maximum precipitation values suggest the potential for extreme rainfall events, which can result in severe flooding, posing a threat to the Becho Plain area. Temperature patterns vary across months, providing valuable information for planning activities such as agriculture and forestation that are temperature-dependent.

Flow records emphasize the importance of understanding flood inundation frequency and extent in the Becho Plain. Flood events occur frequently during the rainy season, with an increasing trend in flow indicating potential future changes in hydrological conditions impacting flood risk. The river flow at the Awash-Belo station shows higher flows during the rainy season (June to September), especially in July and August, and lower flows during the dry season (January to May and November to December). The variability in river flow from year to year indicates climate variability and possible human activities' impact on the hydrology of the Becho Plain.

The flood inundation trend, especially during the rainy season, shows an increasing pattern over time, with June, July, and August consistently experiencing higher flood levels. Factors like climate change, agricultural land expansion, erratic rainfall patterns, high drainage density, and deforestation influence flood inundation trends.

Key insights:

The overall mean flow of the river shows an increasing trend of 0.12 m^3/s per year.

- The mean flow during the rainy season shows an increasing trend of 0.56 m^3/s per year.
- Rainfall during the rainy season positively correlates with flood frequency.
- The increasing trend in mean rainfall and flood flow supports the hypothesis of potential impacts of climate change, leading to increased rainfall and flood frequency.

Further investigation into other variables such as land use, drainage patterns, and local topography is necessary for a comprehensive understanding of factors influencing flood frequency. The GEV model is appropriate for estimating the frequency of extreme flood events, providing critical information for risk assessment and infrastructure planning. The average flood area is approximately

58.42 km², with a standard deviation of 20.16 km², a minimum of 20.03 km², and a maximum of 75.92 km². The trend analysis shows a positive trend in the flood area, increasing by approximately 3.86 km² per year.

Recommendations

- **Enhanced Monitoring and Data Collection:** Continuous and detailed monitoring of climate variables and river flow data is essential. Expanding the network of weather stations and utilizing remote sensing technologies can improve the accuracy and reliability of flood predictions.
- **Implementation of Early Warning Systems:** Based on the predictive models developed in this study, it is crucial to implement early warning systems that can alert communities in advance of potential flood events. This can significantly reduce the impact on lives and properties.
- **Infrastructure Development:** Constructing and upgrading flood control infrastructure such as levees, dams, and drainage systems can mitigate the effects of flooding. The government should invest in sustainable and resilient infrastructure that can withstand extreme weather events.
- **Community Awareness and Preparedness:** Educating local communities about flood risks and preparedness measures is vital. Training programs and awareness campaigns can empower residents to take proactive steps in protecting their properties and lives during flood events.
- **Integration of Geospatial and AI Technologies:** Future flood management strategies should integrate advanced geospatial and AI technologies to enhance the precision and efficiency of flood prediction models. Collaboration between governmental agencies, research institutions, and technology providers can facilitate the development of innovative solutions.

Future research potentials

The analysis results can serve as a foundation for further research on climate patterns, regional climate change, and its impact on the Becho flood plain, Upper Awash River basin;

- **Long-term climate impact studies:** Investigating the long-term impacts of climate change on flood patterns in the Becho plain and surrounding regions can provide deeper insights into future risks and inform adaptive strategies.
- **Enhanced machine learning models:** Developing more sophisticated machine learning models that incorporate additional environmental variables and higher-resolution data can improve flood prediction accuracy.
- **Socioeconomic impact analysis:** Conducting studies on the socioeconomic impacts of flooding, including the cost of damages and the effectiveness of mitigation measures, can help prioritize resources and efforts.

References

Alemu, A., Abera, D., and Lemma, T. (2017). Final Field Report on the Becho Plain Agro-Ecology. Regional Agricultural Research Institute.

Amanambu, A. C., Li, L., Esiobu, N. S., Singh, S. K., Uddin, S. K., Manipon, G. J. M., and Ahmed, I. (2019). Flood risk management through sustainable land use planning and flood forecasting using geospatial techniques. Geoscience Frontiers, 11(2), 589−603.

Alemayehu, F., Taha, N., Nyssen, J., Girma, A., Zenebe, A., Behailu, M., ... and Deckers, J. (2009). The impacts of watershed management on land use and land cover dynamics in Eastern Tigray (Ethiopia). Resources, Conservation and Recycling, 53(4), 192−198.

Ali, M., Khan, S., Sehgai, A. and Khan, S., 2019. Flood modeling and mapping using remote sensing and GIS: A case study of flooding in Peshawar city. Modeling Earth Systems and Environment, 5(1), 193−206.

Balthazar, V., Vanacker, V., Molina, A., and Lambin, E. F. (2013). Impacts of forest cover change on ecosystem services in high Andean mountains. Ecological Indicators, 31, 03−111.

Burrough, P. A., and McDonnell, R. A. (1998). Principles of Geographical Information Systems. Oxford University Press.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q. (2015). Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? Journal of Climate, 28(17), 6938−6959.

Chen, Y. and Liu, Y., 2012. A two-dimensional flood inundation modeling for the Wuhan urban area. The Scientific World Journal, 2012.

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Lawrence Erlbaum Associates.

Congalton, R. G., and Green, K. (2019). Assessing the Accuracy of Remotely Sensed Data: Principles and Practices (3rd ed.). CRC Press.

Gebremichael, E., Sibanda, M., and Nardini, A. (2018). Evaluating the impact of climate change on flood events in Awash Basin, Ethiopia. Catena, 164, 84-95.

Gebremicael, T. G., Mohamed, Y. A., Betrie, G. D., van der Zaag, P., and Teferi, E. (2013). Trend analysis of runoff and sediment fluxes in the Upper Blue Nile basin: A combined analysis of statistical tests, physically-based models and land use maps. Journal of Hydrology, 482, 57−68.

Gizaw, M. S., and Gan, T. Y. (2017). Impact of climate change and El Niño episodes on droughts in sub-Saharan Africa. Climate Dynamics, 49(1-2), 665−682.

Gocic, M., and Trajkovic, S. (2013). Analysis of changes in meteorological variables using Mann-Kendall and Sen's slope estimator statistical tests in Serbia. Global and Planetary Change, 100, 172−182.

Guida, D., Putorti, E., Siervo, V., Falcone, G., Tarantino, E., and Villani, P. (2016). Landform-soil-vegetation patterns analysis for ecosystem services assessment: An application for fire risk mitigation in a Mediterranean watershed. Procedia Environmental Sciences, 32, 265−276.

Hallegatte, S., Green, C., Nicholls, R. J., and Corfee-Morlot, J. (2013). Future flood losses in major coastal cities. Nature climate change, 3(9), 802-806.

Jemberie, A. L., and Gebrehiwot, S. G. (2021). Flood inundation mapping and hazard assessment in the Becho plain, upper Awash river basin, Ethiopia. Geomatics, Natural Hazards and Risk, 12(1), 2183−2207.

Kassa, A. K., and Wale, A. (2018). Trends and variability of stream flow in the Upper Awash River Basin, Ethiopia. Hydrology, 5(2), 23.

Kassa, H., Birhanu, K., Lemma, T., and Rashed, T. (2019). Flood Hazard Assessment and Mapping of Flood Inundation Area of the Awash River Basin in Ethiopia Using GIS and HEC-GeoRAS/HEC-RAS Model. Journal of Geographic Information System, 11(4), 363−377.

Komi, K., Neal, J., Trigg, M. A., and Diekkrüger, B. (2017). Modelling of flood hazard extent in data-sparse areas: a case study of the Oti River basin, West Africa. Journal of Hydrology: Regional Studies, 10, 122−132.

Massey, F. J., Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association, 46(253), 68−78.

Melesse, A. M., Abtew, W., Dessalegne, T., and Wang, X. (2011). Low and high flow analyses and wavelet application for characterization of the Blue Nile River system. Hydrological Processes, 25(21), 3277−3291.

Merz, B., Thieken, A. H., and Gocht, M. (2010). Flood risk mapping at the local scale: concepts and challenges. In Flood risk management in Europe (pp. 231-251). Springer, Dordrecht.

Millington N, Das S, Simonovic SP (2011) The Comparison of GEV, Log-Pearson Type 3 and Gumbel Distributions in the Upper Thames River Watershed under Global Climate Models. [Online]. http://ir.lib.uwo.ca/wrrr/40/. Accessed 10 feb 2017

Millington, N., Das, S., and Simonovic, S. P. (2011). The comparison of GEV, Log-Pearson Type 3 and Gumbel distributions in the upper Thames River watershed under future climate conditions. International Journal of Environment and Climate Change, 1(2), 37−58.

Mokonen, D. (2021). Socioeconomic Survey of the Becho Plain Region. Oromia Rural Development Bureau.

Othman, F., Sarudin, I., Yusop, Z. and Latif, Z.A., 2006. Flood hazard mapping by integrating hydrological, hydraulic, and GIS models for the Central Selangor Watershed, Malaysia. Hydrological Sciences Journal, 51(6), 1093−1110.

Pal, I. and Samanta, N., 2011. A hydrological modeling approach for flood management in a tropical river basin. Journal of Hydrology, 398(1-2), 239−247.

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., and Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. Environmental Modelling and Software, 90, 201−216.

Wasserstein, R. L., and Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. The American Statistician, 70(2), 129−133.

Weiss, A. (2001). Topographic position and landforms analysis (Poster presentation). ESRI User Conference, San Diego, CA.

Wilson, J. P., and Gallant, J. C. (2000). Terrain Analysis: Principles and Applications. John Wiley and Sons.

Worqlul, A. W., Yen, H., Collick, A. S., Tilahun, S. A., Langan, S., and Steenhuis, T. S. (2018). Evaluation of CFSR, TMPA 3B42 and ground-based rainfall data as input for hydrological models, in data-scarce regions: The upper Blue Nile Basin, Ethiopia. Catena, 167, 678−689.

Zheng, F., Westra, S., and Sisson, S. A. (2018). Quantifying the dependence between extreme rainfall and storm surge in the coastal zone. Journal of Hydrology, 561, 308−318.

Yadav, A. K., and Putnam, A. E. (2020). Long-term temperature and precipitation trends in the Upper Awash River Basin, Ethiopia. International Journal of Climatology, 40(7), 3413−3429.

Appendix. 1: Accuracy calculation

	Agricultural land	Barren land	Built up	Vegetation	Forest	Total User (total)
Agricultural land	36	3	0	0	0	39
Barren land	0	39	1	1	0	41
Built up	3	9	10	1	0	23
Vegetation	0	2	0	14		16
Forest	0	0	0	0	6	6
Producer (total)	39	53	11	16	6	125

User accuracy (%)		**Producer accuracy (%)**	
Agricultural land	92.31		
Barren land	95.12	Agricultural land	92.31
Built up	43.48	Barren land	73.58
Vegetation	87.50	Built up	90.91
Forest	100.00	Vegetation	87.50

Over all accuracy 84%

Kappa Coefficient 0.78

Appendix. 2: Precipitation data (mm)

Parameter	Year	January	February	March	April	My	June	July	August	September	October	November	December
Precipitation_SUM	1990	0.00	142.38	47.46	94.92	21.09	94.92	258.40	342.77	179.30	36.91	0.00	5.27
Precipitation_SUM	1991	0.00	21.09	73.83	47.46	36.91	116.02	295.31	305.86	137.11	0.00	0.00	0.00
Precipitation_SUM	1992	47.46	42.19	5.27	84.38	15.82	158.20	247.85	305.86	79.10	42.19	10.55	5.27
Precipitation_SUM	1993	26.37	21.09	21.09	163.48	84.38	226.76	284.77	321.68	168.75	36.91	0.00	5.27
Precipitation_SUM	1994	0.00	0.00	31.64	68.55	58.01	152.93	326.95	290.04	137.11	0.00	5.27	0.00
Precipitation_SUM	1995	0.00	52.73	31.64	89.65	26.37	110.74	258.40	321.68	126.56	5.27	0.00	10.55
Precipitation_SUM	1996	21.09	0.00	137.11	68.55	100.20	284.77	369.14	305.86	189.84	15.82	5.27	0.00
Precipitation_SUM	1997	15.82	0.00	36.91	94.92	73.83	147.66	195.12	242.58	89.65	116.02	10.55	0.00
Precipitation_SUM	1998	26.37	79.10	15.82	36.91	68.55	142.38	411.33	295.31	152.93	168.75	0.00	0.00
Precipitation_SUM	1999	15.82	0.00	26.37	10.55	68.55	226.76	258.40	237.30	110.74	121.29	0.00	0.00
Precipitation_SUM	2000	0.00	0.00	5.27	36.91	47.46	110.74	242.58	174.02	79.10	26.37	0.00	5.27
Precipitation_SUM	2001	0.00	10.55	84.38	15.82	63.28	94.92	195.12	168.75	52.73	5.27	0.00	0.00
Precipitation_SUM	2002	10.55	0.00	42.19	36.91	10.55	84.38	116.02	142.38	42.19	0.00	0.00	0.00
Precipitation_SUM	2003	5.27	5.27	36.91	36.91	0.00	73.83	121.29	210.94	63.28	0.00	0.00	26.37
Precipitation_SUM	2004	0.00	5.27	52.73	42.19	0.00	63.28	158.20	137.11	68.55	21.09	0.00	21.09
Precipitation_SUM	2005	10.55	0.00	63.28	94.92	110.74	47.46	200.39	147.66	89.65	26.37	0.00	0.00
Precipitation_SUM	2006	5.27	47.46	100.20	68.55	36.91	152.93	321.68	242.58	137.11	47.46	5.27	5.27
Precipitation_SUM	2007	5.27	36.91	68.55	47.46	131.84	131.84	253.12	242.58	174.02	15.82	0.00	0.00
Precipitation_SUM	2008	0.00	0.00	0.00	36.91	47.46	100.20	321.68	326.95	152.93	21.09	63.28	0.00
Precipitation_SUM	2009	42.19	0.00	21.09	31.64	36.91	68.55	221.48	221.48	94.92	58.01	0.00	31.64
Precipitation_SUM	2010	0.00	73.83	63.28	110.74	110.74	205.66	290.04	247.85	195.12	10.55	15.82	5.27
Precipitation_SUM	2011	0.00	10.55	58.01	42.19	79.10	137.11	152.93	131.84	84.38	5.27	21.09	0.00
Precipitation_SUM	2012	0.00	0.00	5.27	100.20	21.09	100.20	353.32	374.41	168.75	0.00	0.00	5.27
Precipitation_SUM	2013	15.82	0.00	68.55	163.48	94.92	184.57	353.32	258.40	147.66	79.10	15.82	0.00
Precipitation_SUM	2014	0.00	15.82	47.46	63.28	110.74	100.20	242.58	221.48	174.02	79.10	15.82	15.82
Precipitation_SUM	2015	0.00	10.55	10.55	5.27	142.38	163.48	258.40	295.31	58.01	5.27	10.55	10.55
Precipitation_SUM	2016	21.09	15.82	31.64	100.20	168.75	121.29	247.85	342.77	200.39	26.37	5.27	0.00
Precipitation_SUM	2017	0.00	68.55	42.19	21.09	142.38	73.83	348.05	321.68	253.12	31.64	5.27	0.00
Precipitation_SUM	2018	0.00	42.19	0.00	121.29	68.55	189.84	253.12	342.77	47.46	21.09	15.82	0.00
Precipitation_SUM	2019	0.00	0.00	47.46	131.84	21.09	163.48	237.30	332.23	400.78	21.09	21.09	0.00
Precipitation_SUM	2020	0.00	0.00	31.64	210.94	121.29	205.66	474.61	469.34	232.03	36.91	0.00	0.00
Precipitation_SUM	2021	21.09	15.82	0.00	88.02	102.42	146.92	242.37	337.16	177.25	73.49	36.33	16.28
Precipitation_SUM	2022	26.94	25.99	28.50	93.75	8.56	179.45	317.52	207.77	134.04	62.85	6.06	16.99

Appendix. 3: Temperature(°C) data

Parameter	Year	January	February	March	April	May	June	July	August	September	October	November	December
Temperature	1990	15.16	16.11	16.6	17.47	18.98	17.41	15.12	15.44	15.5	13.62	13.93	13.63
Temperature	1991	16.16	17.65	17.67	17.58	18.35	18.12	15.43	15.66	15.4	13.61	13.48	14.19
Temperature	1992	15.04	16.69	18.51	18.74	18.76	17.99	14.86	14.4	14.89	14.48	13.2	14.71
Temperature	1993	15.33	16.1	17.94	17.37	17.48	16.77	14.99	15.07	15.31	14.27	13.31	13.3
Temperature	1994	15.01	17.55	18.5	19.05	18.73	16.65	14.79	14.88	15.04	13.32	13.94	13.82
Temperature	1995	15.37	17.1	17.86	18.29	18.76	18.94	15.16	15.44	15.41	14.53	14.35	15.01
Temperature	1996	15.78	17.81	17.96	17.92	16.8	16.04	15.05	15.28	15.54	13.9	13.35	13.79
Temperature	1997	16.15	16.01	18.62	17.79	17.75	16.84	15.83	15.6	15.87	14.88	15.26	15.36
Temperature	1998	16.77	17.83	18.58	20.27	19.36	17.21	14.86	14.95	15.26	15.05	12.73	12.21
Temperature	1999	14.44	16.55	18.32	19.3	18.97	16.84	14.27	15.08	15.37	14.33	12.33	13.24
Temperature	2000	14.66	16.62	18.76	17.55	19.28	17.19	14.79	14.89	15.55	15.47	14.98	15.17
Temperature	2001	16.65	17.69	17.69	19.1	19.04	16.51	15.25	14.94	16.23	16.51	15.33	16.31
Temperature	2002	16.23	18.3	18.67	18.93	20.6	18.24	16.79	15.77	16.45	17.07	16.88	17.23
Temperature	2003	17.04	18.99	18.66	18.8	20.63	18.08	15.37	15.41	15.87	15.86	16.08	14.86
Temperature	2004	17.61	17.98	19.01	18.31	20.69	17.28	15.42	15.48	16.17	15.46	16.19	16.42
Temperature	2005	15.8	18.76	18.49	19.21	17.72	17.79	15.3	15.49	15.97	15.79	15.07	14.42
Temperature	2006	16.96	17.76	17.7	18	18.57	17.26	15.62	15.17	15.45	14.98	14.06	14.8
Temperature	2007	15.94	17.3	18.26	18.05	18.99	16.94	15.2	15.31	15.73	13.55	13.08	12.94
Temperature	2008	15.13	16.9	18.76	18.08	18.78	17.12	15.22	15.57	15.54	14.88	13.32	13.73
Temperature	2009	14.98	16.74	19.08	19.01	19.65	19.01	15.45	15.8	16.13	14.94	14.69	15.63
Temperature	2010	16.55	17.3	17.84	18.15	18.22	17.41	15.3	14.88	15.65	14.54	13.64	14.17
Temperature	2011	15.59	16.92	17.3	18.7	18.44	17.23	15.79	15.3	15.65	15.5	16.44	15.36
Temperature	2012	16.49	17.27	18.98	18.45	19.44	18.08	15.46	15.22	15.44	14.15	14.76	14.87
Temperature	2013	16.28	17.78	19.38	18.49	17.89	16.38	14.48	14.6	15.44	14.44	14.15	13.5
Temperature	2014	15.9	17.37	18.18	17.48	17.9	17.91	15.72	15.41	15.26	14.41	14.33	14
Temperature	2015	14.69	17.63	18.87	20.24	18.52	17.56	16.45	15.86	16.09	15.58	15.35	15.56
Temperature	2016	16.8	18.3	19.73	19.03	17.3	17.12	15.48	15.37	15.43	14.56	13.59	13.17
Temperature	2017	13.68	16.94	18.34	18.82	17.52	18.53	15.69	15.67	15.52	15.01	13.9	12.8
Temperature	2018	15.3	17.22	18.67	18.12	18.07	16.62	15.34	15.17	15.26	15.03	14.4	15.19
Temperature	2019	15.39	18.65	18.91	18.07	19.04	17.63	15.94	15.42	15.82	14.16	14.64	14.79
Temperature	2020	16.09	17.98	19.69	18.56	17.71	17.08	15.15	15.18	15.61	14.51	13.72	13.8
Temperature	2021	14.89	16.7	18.51	18.62	17.69	17.72	15.19	15.66	15.79	14.69	14.05	14.02
Temperature	2022	15.44	17.32	18.53	18.86	20.41	17.44	15.18	15.04	15.39	14.93	14.08	14.56

Appendix. 4: Awash river flow data (m^3/s)

Station	Year	January	February	March	April	May	June	July	August	September	October	November	December
Awash-Belo	1995	0.52	0.36	7.35	12.91	14.99	20.79	27.82	35.57	23.87	6.29	1.33	0.8
Awash-Belo	1996	0.62	0.95	0.65	1.28	0.62	5.27	33.53	37.84	35.98	21.59	2.5	1.2
Awash-Belo	1997	1.03	2.5	3.29	5.84	1.96	4.02	37.84	37.42	37.84	13.38	2.45	2.97
Awash-Belo	1998	1.12	4.55	7.75	7.48	1.41	9.23	33.33	35.98	36.39	19.73	2.25	1.2
Awash-Belo	1999	-	1.24	3.18	0.92	1.68	6.55	33.73	37.74	37.84	10.02	1.77	1.12
Awash-Belo	2000	1.03	2.86	1.77	7.48	3.62	13.46	29.56	38.36	35.77	25.73	2.97	1.54
Awash-Belo	2001	1.16	1.86	1.07	13.06	16.31	17.99	35.16	-	-	-	-	1.86
Awash-Belo	2002	1.33	0.92	1.28	1.59	2.5	10.32	34.34	35.36	38.47	19.38	2.86	1.41
Awash-Belo	2003	0.99	0.88	0.99	9.01	4.67	10.91	32.93	-	-	8.45	1.73	1.24
Awash-Belo	2004	2.45	0.76	8.17	11.21	9.88	24.79	36.91	39.52	34.55	15.48	2.4	1.41
Awash-Belo	2005	1.86	1.37	0.84	1.77	1.07	12.13	34.44	35.16	22.04	11.52	7.69	1.73
Awash-Belo	2006	2.66	0.62	6.96	4.77	5.04	28.71	49.33	-	49.51	36.1	4.97	1.87
Awash-Belo	2007	1.48	1.82	1.31	0.62	2.82	11.04	45.87	-	45.36	44.34	5.25	1.27
Awash-Belo	2008	0.74	0.54	0.36	1.31	2.77	10.48	34.87	49.86	42.84	41.35	5.6	1.6
Awash-Belo	2009	-	1.05	2.4	2.3	4.11	29.13	-	-	-	-	-	-
Awash-Belo	2010	2.93	0.94	0.91	1.43	0.78	-	-	-	-	-	-	-
Awash-Belo	2011	-	-	0.91	6.11	1.05	8.51	49.69	48.81	47.24	22.54	1.01	0.57
Awash-Belo	2012	0.65	0.54	0.74	7.11	3.62	28.85	40.54	48.28	47.07	13.52	1.39	0.68
Awash-Belo	2013	0.34	0	0.81	7.76	8.18	23.44	48.28	50.03	42.67	18.26	1.31	0.65
Awash-Belo	2014	0.38	1.16	2.5	11.14	14.15	15.43	50	50.22	49.68	11.42	2.5	-
Awash-Belo	2015	-	-	-	-	8.01	43.84	48.11	49.86	50.04	32.47	1.6	0.74
Awash-Belo	2016	0.54	0.36	0.41	2.2	2.87	12.01	49.33	52.72	52.9	9.39	13.52	1.01
Awash-Belo	2017	0.81	0.65	0.25	0.51	0.34	4.7	43.17	51.28	48.63	-	-	-
Awash-Belo	2018	-	-	-	-	-	-	49.86	48.46	56.58	-	-	-

Appendix.5: Field survey photos

Plate 1. Measuring the width and depth of the Awash River downstream of the main bridge.

Plate 2. Downstream of the bridge the main river flow gets stagnant and pressure flow non-gravity flow.

Appendix. 6: Awash river channel profile

Elevation profile: downstream of Awash river bridge (chainage 0-750m)

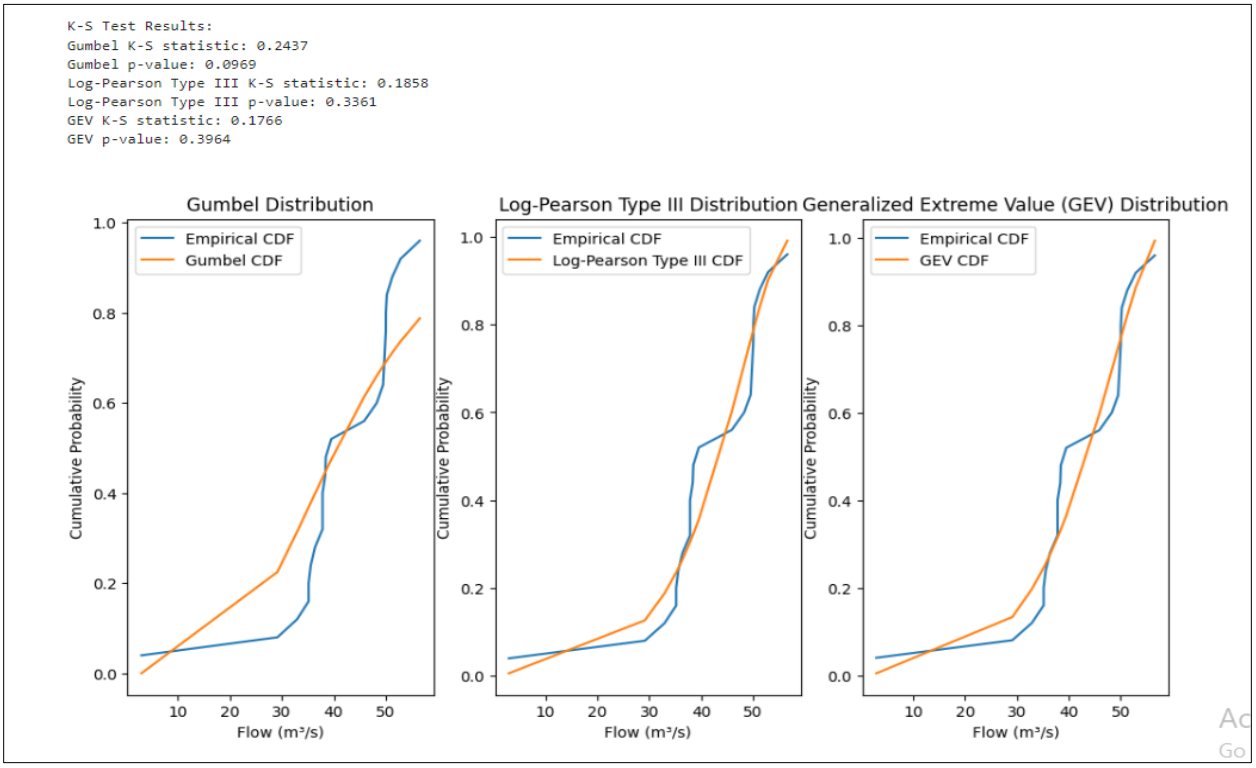
Appendix.7: Flooded area calculation (Python3 codes)

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Load the flood inundation area data from the CSV file
flood_area_data = pd.read_csv("C:/Users/Jifara/anaconda3/envs/SWToolbox/floodArea2015-2023.csv")

# Display the first few rows of the dataset to understand its structure
print(flood_area_data.head())

# Perform statistical analysis on the flood inundation data
mean_area = flood_area_data['AreaFlooded(sq.km)'].mean()
median_area = flood_area_data['AreaFlooded(sq.km)'].median()
max_area = flood_area_data['AreaFlooded(sq.km)'].max()
min_area = flood_area_data['AreaFlooded(sq.km)'].min()
std_area = flood_area_data['AreaFlooded(sq.km)'].std()

# Print the statistical measures
print("Mean flood area (sq.km):", mean_area)
print("Median flood area (sq.km):", median_area)
print("Maximum flood area (sq.km):", max_area)
print("Minimum flood area (sq.km):", min_area)
print("Standard deviation of flood area (sq.km):", std_area)

# Plot the time series of flood inundation area
plt.figure(figsize=(12, 6))
plt.plot(flood_area_data['Year'], flood_area_data['AreaFlooded(sq.km)'], marker='o', color='b')
plt.xlabel('Year')
plt.ylabel('Flood Inundation Area (sq.km)')
plt.title('Time Series of Flood Inundation Area')
plt.grid(True)
plt.show()
```

..continued

```
# Perform statistical calculations and analysis
correlation_matrix = df[months].corr()
print(correlation_matrix)

# Time series analysis and trend analysis
years = df['Year'].values.reshape(-1, 1)
temperature = df[months].values

# Overall time series analysis and trend analysis
overall_mean = np.mean(temperature, axis=1)
plt.plot(df['Year'], overall_mean)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Overall Mean Temperature Time Series')
plt.show()

regressor = LinearRegression()
regressor.fit(years, overall_mean)
trend_line = regressor.predict(years)
plt.plot(df['Year'], overall_mean)
plt.plot(df['Year'], trend_line)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Overall Mean Temperature Time Series with Linear Regression Trend Line')
plt.legend(['Actual', 'Trend Line'])
plt.show()

# Rainy season (June-September) time series and trend analysis
rainy_season_mean = np.mean(temperature[:, 5:9], axis=1)
plt.plot(df['Year'], rainy_season_mean)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Rainy Season Mean Temperature Time Series')
plt.show()

regressor.fit(years, rainy_season_mean)
trend_line = regressor.predict(years)
plt.plot(df['Year'], rainy_season_mean)
plt.plot(df['Year'], trend_line)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Rainy Season Mean Temperature Time Series with Linear Regression Trend Line')
plt.legend(['Actual', 'Trend Line'])
plt.show()
```

Appendix. 8: Flood frequency models comparison graph



Appendix. 9: SAR image analysis in Google earth engine

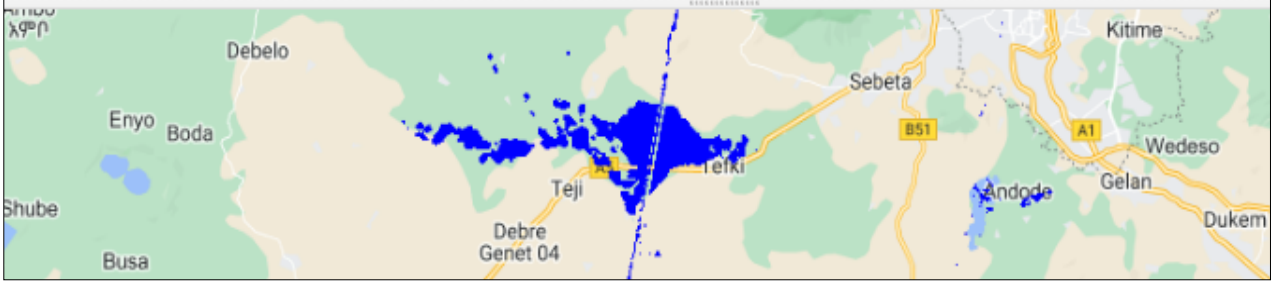
```
// Flooded area boundary

var cor = [
[38.19047064360, 8.90037829697],
[38.29848300590, 8.96033566369],
[38.38057900470, 8.96944308180],
[38.38057900470, 8.96944308180],
[38.59798361760, 8.87913614174],
[38.58451302750, 8.83879606366],
[38.45492076880, 8.80826208094],
[38.33283746000, 8.78964582144],
[38.18606968400, 8.85307699554]
];
crs: EPSG: 32637;

var roi = ee.Geometry.Polygon(cor)

Map.centerObject(roi);

var startDate = '2015-01-01', endDate = '2023-12-31'

function speckel(img){
  return img.focalMedian(100, 'square', 'meters')
  .copyProperties(img, img.propertyNames())
  }
var after = ee.ImageCollection("COPERNICUS/S1_GRD")
.filterBounds(roi)
.filterDate(startDate, endDate)
.filter(ee.Filter.calendarRange(7, 9, 'month'))
.filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VV'))
.filter(ee.Filter.eq('instrumentMode', 'IW'))
.select('VV')
.map(speckel)
.min();

Map.addLayer(after.clip(roi),[],'after', false);
```

..continued

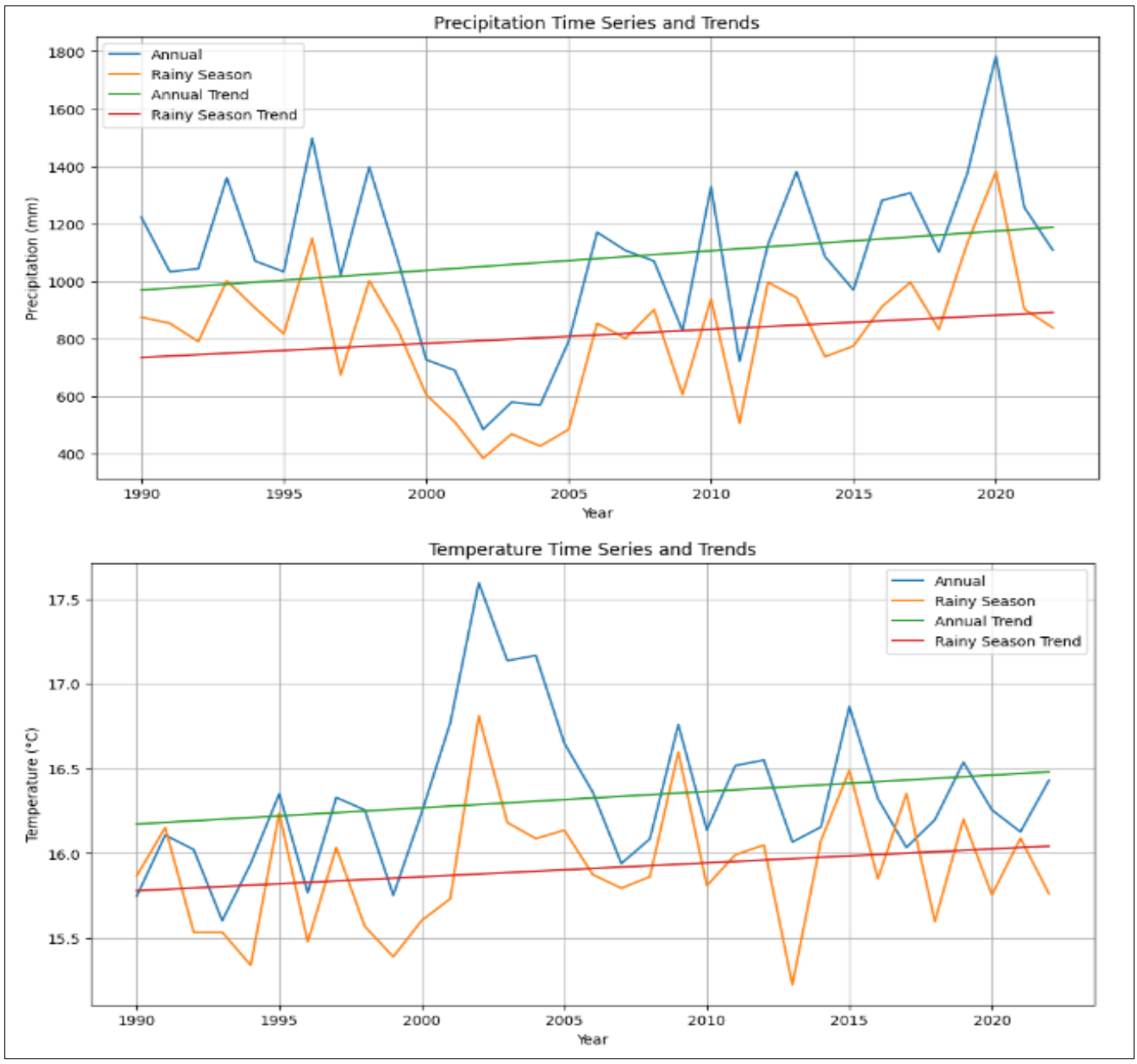
```
38
39  var before = ee.ImageCollection("COPERNICUS/S1_GRD")
40  .filterBounds(roi)
41  .filterDate(startDate, endDate)
42  .filter(ee.Filter.calendarRange(1,2,'month'))
43  .filter(ee.Filter.listContains('transmitterReceiverPolarisation', 'VV'))
44  .filter(ee.Filter.eq('instrumentMode', 'IW'))
45  .select('VV')
46  .map(speckel)
47  .min();
48
49  Map.addLayer(before.clip(roi),[],'before', false)
50
51  var change = before.subtract(after).rename('flood')
52
53  Map.addLayer(change.clip(roi),[],'flood', false)
54
55  // Generate histogram chart with labels and title
56  var histogramChart = ui.Chart.image.histogram(change, roi, 30)
57    .setOptions({
58      title: 'Histogram of Pixel Values (' + startDate + ' to ' + endDate + ')',
59      hAxis: {title: 'Pixel Value'},
60      vAxis: {title: 'Frequency'}
61    });
62
63  // Print the histogram chart
64  print(histogramChart);
65  //print(//  ui.Chart.image.histogram(change, roi, 30)   )
66
67  Map.addLayer(change.gt(7).clip(roi),[],'flood_thr', false)
68
69  var flood_thr = change.gt(7);
70  var flood_mask = flood_thr.updateMask(flood_thr);
71  var flood_area = flood_mask.multiply(ee.Image.pixelArea().divide(1e6));
72  var area_sum = flood_area.reduceRegion({
73    reducer: ee.Reducer.sum(), geometry: roi, scale: 30
74    }).values().get(0);
75  var centerPoint = ee.Geometry.Point(38.447, 8.87); // Define the center point coordinates
76  Map.centerObject(centerPoint, 9); // Zoom to the center point
77
78  Map.addLayer(flood_area, {palette: 'blue'}, 'Flooded Area');
79  // Create a feature collection with the flooded area and the attributes
```

..continued

```
// Create a feature with the flooded area geometry and the years 2015 and 2016 as attributes
var feature = ee.Feature(roi, {YearStart: startDate, YearEnd: endDate});

// Create a feature collection with the single feature
var featureCollection = ee.FeatureCollection([feature]);

// Print the feature collection to verify that it was created successfully
print('Feature Collection:', featureCollection);




var featureCollection = flood_area.reduceRegions({
  collection: featureCollection, // Define the region of interest
  reducer: ee.Reducer.sum(), // Calculate the sum of flooded area within each feature
  scale: 30 // Set the scale for the reduction
});

// Export the feature collection to a CSV file
Export.table.toDrive({
  collection: featureCollection,
  description: 'flooded_area_attributes',
  fileFormat: 'CSV'
});

print(ee.Number(area_sum).round())
```



Appendix. 10: Rainfall *vs* Temperature time series trends graph



Appendix.11: Python Jupyter notebook (machine learning language) coding

 # Create working directory and active for local serve…Time series analysis and modeling

```
activate SWToolbox
cd C:\Users\Jifara\anaconda3\envs\SWToolbox
jupyter notebook
```

 # Wrting python3 jupyter notebook

#.Importing the librarires and modules

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Load the temperature data from the CSV file
data_path = "C:/Users/Jifara/anaconda3/envs/SWToolbox/temperature_data.csv"
df = pd.read_csv(data_path)

# Interpolate missing values
df = df.interpolate()

# Remove unnamed column if present
df = df.loc[:, ~df.columns.str.contains('^Unnamed')]

# Statistical description and summary
description = df.describe().round(2)
print(description)

# Histogram plots
months = ['January', 'February', 'March', 'April', 'May', 'June', 'July', 'August', 'September', 'October',
'November', 'December']
overall = df[months].values.flatten()
plt.hist(overall, bins=20)
plt.xlabel('Temperature')
plt.ylabel('Frequency')
plt.title('Overall Temperature Distribution')
plt.show()

for month in months:
    plt.hist(df[month], bins=20)
    plt.xlabel('Temperature')
    plt.ylabel('Frequency')
    plt.title(f'{month} Temperature Distribution')
    plt.show()

# Perform statistical calculations and analysis
correlation_matrix = df[months].corr()
print(correlation_matrix)
```

```
# Time series analysis and trend analysis
years = df['Year'].values.reshape(-1, 1)
temperature = df[months].values

# Overall time series analysis and trend analysis
overall_mean = np.mean(temperature, axis=1)
plt.plot(df['Year'], overall_mean)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Overall Mean Temperature Time Series')
plt.show()

regressor = LinearRegression()
regressor.fit(years, overall_mean)
trend_line = regressor.predict(years)
plt.plot(df['Year'], overall_mean)
plt.plot(df['Year'], trend_line)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Overall Mean Temperature Time Series with Linear Regression Trend Line')
plt.legend(['Actual', 'Trend Line'])
plt.show()

# Rainy season (June-September) time series and trend analysis
rainy_season_mean = np.mean(temperature[:, 5:9], axis=1)
plt.plot(df['Year'], rainy_season_mean)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Rainy Season Mean Temperature Time Series')
plt.show()

regressor.fit(years, rainy_season_mean)
trend_line = regressor.predict(years)
plt.plot(df['Year'], rainy_season_mean)
plt.plot(df['Year'], trend_line)
plt.xlabel('Year')
plt.ylabel('Mean Temperature')
plt.title('Rainy Season Mean Temperature Time Series with Linear Regression Trend Line')
plt.legend(['Actual', 'Trend Line'])
plt.show()
```

Appendix.12: Dataset sources and its legacy

1. Soil map: Legacy maps and databases refer to data and maps compiled using field surveys backed up by remote sensing and other environmental data, expert opinion and laboratory analysis. The bulk of soil information was collected in this way. The technological advances in remote sensing, computers, terrain analysis, geostatistics, GIS data integration, and instrumentation have made it possible to achieve unprecedented reliability and utility in digital soil maps.
2. Land-use maps: Global Land Cover Facility This website, operated by the University of Maryland, provides free access to global land cover data and maps. https://glcf.umd.edu/data/.
3. Geological data: source; *Geological Survey of Ethiopia*
4. Climate Data: source: *National* Meteorology Agency for field verification.

D E C L A R A T I O N

I the undersigned declare that this thesis is my original work and has not been presented for a Degree in any other university and that all sources of materials used for the thesis have been duly acknowledged.

Jifara Dabessa Gemechu

Signature _____ Date_____

School of Earth Science

May, 2024

This thesis has been submitted for examination with my approval as university advisor.

Dr. Tibebu Kasawmar

Signature  Date_____