ADDIS ABABA UNIVERSITY

ADDIS ABABA INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

SPATIAL FLASH FLOOD SUSCEPTIBILITY PREDICTION IN SEMI ARID AREA (CASE STUDY IN JIGJIGA JERER UPSTREAM WATERSHED, ETHIOPIA)

A THESIS IN HYDRAULIC ENGINEERING

BY: WONDIMU WALIE (GSR/3231/10)

OCTOBER 1, 2019

ADDIS ABABA

A THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

The undersigned have examined the thesis entitled "**Spatial Flash Flood Susceptibility Prediction in Semiarid Area"** presented by **Wondimu Walie**, a candidate for the degree of **Master of Science** and hereby certify that it is worthy of acceptance.

Fiseha Behulu (PhD.)		
Advisor	Signature	Date
Internal Examiner	Signature	Date
External Examiner	Signature	Date
Chair person	Signature	Date

UNDERTAKING

I certify that research work titled “Spatial Flash Flood Susceptibility Prediction in Semiarid Area” is my own work. The work has not been presented elsewhere for assessment. Where material has been used from other sources it has been properly acknowledged / referred.

Signature

Wondimu Walie

This is to certify that the above certification made by the candidate is correct to the best of my knowledge

Signature

Fiseha Behulu (PhD)

ABSTRACT

The aim of this research was to identify areas that are most susceptible to spatial flash floods. The study applied an integration of ArcGIS based spatial data analysis and logistic regression model for producing areal flash flood susceptibility prediction map from flash flood conditioning factors and historical inventory maps. The flash flood inventory map consisting a total of equal 560 flood and non flood affected locations, were extracted from areal photographs and field observations. The flood inventory data were randomly divided into two groups where 70% (392) were used for training, and the remaining 30% (168) for validation respectively for Logistic Regression Model. Eleven flash flood conditional factors such as lithology, slope, topographic wetness index (TWI), stream power index (SPI), landuse-landcover (LULC), rainfall, normalized difference vegetation index (NDVI), distance from drainage, morphometric hazard index (MHI), drainage density, and soil type, were selected. These factors were mainly derived from Shuttle Radar Topographic Mission digital elevation model (SRTM DEM), and Landsat 8 OLI/TIRS imagery using spatial analysis tools (i.e. ArcGIS 10.4 and ERDAS 2014). Finally, the flash flood susceptibility index model was produed summing up all factors multiplying with their respective weights computed in logistic regression model. Furthermore, the spatial flash flood susceptibility probability map was produced. The accuracy of the final model for spatial flash flood sucesptibility prediction was evaluated by calculating the model relative operating characteristic (ROC) curve using R studio. For validation, success and prediction rate curves were produced using area under the curve (AUC) method. The predictive capability of the model was determined from the area of under the cuve relative operating characteristic (ROC) which is found to be 0.96 (i.e. highest prediction rate of 96%). Finally, the results of spatial flash flood susceptibility prediction map were classified into five susceptiblity classes such as very low, low, moderate, high and very high clsses. The spatial flash flood susceptibility map obtained from this study could therefore assist city planners and engineers for development, landuse planning and watershed management.

Keywords: Flash flood susceptibiliy; Logistic regression model; Flash flood conditioning factors; Receiver operating characteristic curve; GIS

ACKNOWLEDGMENTS

Above all, I would like to thank the Almighty God for giving and make everything possible for me in my entire life, especially gave me healthy, strength and patience to complete this work.

I would like to express my heartfelt thanks and genuine gratitude to my advisor Fiseha Behulu (PhD) at University of Addis Ababa, Addis Ababa Institute of Technology (AAiT) for his guidance and support, and for the freedom he gave me to pursue this research from starting to the end of my thesis work. Without his insight, encouragement, critical comments, fast responses for my queries and invaluable supports, I would not have been able to complete this research; thus I was lucky working my thesis with him.

I aslo gratefully acknowledge Addis Ababa Institute of Technology (AAiT) for helping me in various cases, Ethiopian Ministry of Water, Irrigation and Energy providing me data I requested, Ethiopian National Meteorological Services Agency for providing rainfall data, Ethiopian Geological Survey providing for Geological classes.

Finally, and most importantly, I would like to extend my deepest gratitude to my family, friends for their words of encouragement and various support and to all staff members of Civil and Environmental Engineering specially Hydraulic Engineering stream during my study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AHP	Analytic Hierarchy Process
AUC	The area under the curve
DEM	Digital elevation model
ERDAS	Earth Resources Data Analysis System
FDRENDRMCEWERD	Federal Democratic Republic of Ethiopia National Disaster Risk Management Commission, Early Warning, and Emergency Response Directorate
GIS	Geographic information system
GLM	Generalized linear model
GPS	Global positioning systems
Hec-GeoHMS	Hydrologic Engineering Centre Geographical Hydrologic Modelling System
IPCC	Intergovernmental Panel on Climate Change
LANDSAT	Land satellite
LR	Logistic regression
LULC	Land use land cover
NDVI	Normalized Difference Vegetation Index
OCHA	Office for Coordination of Humanitarian Affairs
OLI	Operational Land Imager
RADAR SAT	Radar Satellite
RDPPB	Somali Regional Disaster Prevention and Preparedness Bureau
ROC	Receiver operating characteristics curve
RS	Remote sensing
SPOT	Système Pour l'Observation de la Terre
SRDPPB	Somali Regional Disaster Prevention and Preparedness Bureau
TIRS	Thermal Infrared Sensor
TRMM	Tropical Rainfall Measuring Mission
TWI	Topographic wetness index
UN	United Nation
UTM	Universal Transverse Mercator

CHAPTER 1 INTRODUCTION

1.1 General introduction

Floods are one of the most common and devastating hydro-meteorological hazards in Ethiopia

and worldwide, causing both economic losses, infrastructure and human fatalities [1], [2]. Flash floods can develop rapidly in the catchment without warning leaving many deaths and lost many millions of dollars [3]. Flash floods are formed from excess rain falling on upstream watersheds, flow downstream with massive concentration, high speed, and typically occur suddenly. Heavy downpours in mountainous highlands can lead to surges of water that turn dry river beds or flood plains into raging torrents in minutes. Therefore, the damages of such floods become particularly pronounced and devastating when they pass across or along with human settlements and infrastructures. It is one of the most common, frequent and costly natural disasters and devastating hydro-meteorological hazards causing considerable damage to highways, settlement, agriculture, and livelihood and many infrastructures in different parts of Ethiopia. Usually, flash floods in the country are caused by heavy rainfall falling on upstream river catchments to downstream with huge concentration, speed, and force.

Flash floods are characterized by their violence and the rapidity of their occurrence. Because they don't occur frequently and are so unpredictable; their anticipation is still subject to research. Damages caused by flash floods can be dramatic. Such an event occurred in the Eastern part of Ethiopia in different consecutive years. Heavy localized rains in April/early May have caused flooding in belg rain-receiving areas, resulting in loss of lives and livelihoods and the destruction of infrastructures. They occur in small headwater catchments, less than a few hundred square kilometers, where these basins respond quickly to excessive rainfall amounts that fall in the short time periods characterized by flash flood-producing events [4]. Unfortunately, these small basins can also be located in urban areas where the effects of flash flood takes a short time for cities to collapse under flooding and residual solids loading where its magnitude and severity is determined by a number of natural and human-influenced factors (unplanned rapid settlement development, uncontrolled construction of urban infrastructures) and other factors such as : rainfall duration and intensity, antecedent soil moisture conditions, land cover and soil type, watershed or geomorphic characteristics, land use, drainage, engineering structures, and climate. Generally,

it is caused by complex interactions of topographical, geological, geomorphological, and hydrological parameters in the watershed. According to the Somali Regional State Government, frequent floods displaced more than 7,000 households in Fafan, Jarar, Korahe, and Shabelle zones. As the report of the United Nation Office for Coordination of Humanitarian Affairs (OCHA 2013) and particularly the flood killed more than 15 people in Jigjiga only in 2018. According to the April 2018 recently, thirty-five displacement incidents were reported during April alone displacing 170,760 people nationwide, the majority of which are due to flooding in Somali region. The Somali Regional Disaster Prevention and Preparedness Bureau (RDPPB) has also reported that, the flooding in Somali region has affected 43,887 families/households (263,322 people), of which, 25,238 households/151,428 people were displaced. The report also indicated that the floods destroyed 12,911 hectares of farmland and damaged 76 health facilities, mostly health posts. At least 123 schools were affected, interrupting schooling. The report also states that more than 15,643 houses were destroyed, requiring emergency shelter interventions as the report of the UN Office for Coordination of Humanitarian Affairs (OCHA 2013) indicates the issue.

An integrated approach using remote sensing, geographic information systems, and logistic regression are commonly applied to study spatial flash flood susceptibility prediction. Such approaches add an integrated multifactor (remotely sensed, topographic, geologic, morphometric) analysis for a comprehensive spatial flash flood susceptibility areas prediction. Many researchers around the world have integrated GIS and RS technologies with statistical, data-driven and machine learning models to study flood susceptibility with successful and promising results. GIS is a standard technology to analyze, manipulate, and integrate the factors contributing to flash flood susceptibility with great accuracy and efficiency [6]. In this study, the causative factors for flash floods will be constructed, analyzed, and integrated within the same geo-referencing outline using GIS, remote sensing data and then logistic regression will be applied so as to reveal the flash flood susceptibility index based on the factors relative contribution to flash flood susceptibility. Several GIS models including logistic regression [7], [8], bivariate statistical analysis, and analytical hierarchy process (AHP) [2], [6], [9], in addition hydrologic/hydraulic modeling to understand flood extend and peak discharge which later related to flood risk have been used in different applications around the world considering different factors.

The most challenging task in flash food susceptibility assessment is the limited availability of adequate data for flood hazard studies, which require the fusion of hydrologic observations, estimates of surface roughness and boundary conditions. Moreover, topographical surveys showing all-natural and manmade features for the channel of interest and the adjacent areas at high resolution is rarely available. From hydrologic study points of views, unfortunately, gauging station networks are not widespread throughout most the study areas and this scarcity is generally most pronounced in most catchments areas like Jigjiga watershed in Ethiopia. However, there are also other different alternatives in the determination of flash flood susceptibility assessment in data-scarce environments, which is a research topic of increasing interest in ungauged sites.

1.2 Statement of the problem

Ethiopia experienced to have two types of floods which are flash floods and river floods. Flash floods are the ones formed from excess and intensive rains falling on upstream watersheds and gush downstream with massive concentration, speed and force. Often, they are sudden and appear unnoticed. In the recent years, Jigjiga city and it's surrounding watershed is witnessing devastating flash floods more often due to heavy rainfall and infrequent extreme rainfall events, developments and settlements near major drainage channel, encroachment of water level in the drainage, inability of major canals to carry heavy rains and poor land use land cover of the area, rapid and unplanned urbanization, climate change, topography (low-lying area and almost like a pancake), inadequate and poor drainage systems, solid waste disposal in the drainage systems and etc, have exacerbated flash flood risk which has caused devastating loss of human life, livestock, damage to the property in Jigjiga, Ethiopia. Flood hazard management of the area is not well organized and established and hence every year flood risks occurred and create a social crisis. In 2016, significant and valuable work has been done by the city administration trying to widen the drainage channel which passes through the city though it was not enough to accommodate when a maximum flood discharge comes from the upstream catchment in the city.

According to the reports of Federal Democratic Republic of Ethiopia National Disaster Risk Management Commission, Early Warning and Emergency Response Directorate (2018); Flash floods occur in lowland areas when excessive rains fall in adjacent highland areas in Ethiopia; among these areas, Jigjiga Town in Somali region is most at risk of flooding. This

type of flood is characterized by sudden onset with little lead time for early warning and often resulting in considerable damage on lives, livelihoods, and property in the stud area.

Several flash floods have occurred in the city and its surrounding watershed in the past years. The last consecutive years most recently and notably in 2016, 2017 and 2018 were destructive due to huge flash flood through the river passing in the middle; which leading to huge damage to infrastructure, streets, schools, inundating many Kebele villages and threatening life and property. As the report of UN Office for Coordination of Humanitarian Affairs (OCHA, 2018), flash flood incidences since the second week of April have left hundreds of thousands of people of Somali (7 zones) regions. In Somali region, more than 27,000 households (165,000 persons) were flood-affected different woredas (districts) of After, Fafan (Jigjiga), Liben, Nogob, Siti, Shebele and warder Zones. Several Kebeles have submerged and farmlands were either flooded or washed away at the flowering stage. Many people's houses/shelters and livestock have reportedly been washed away, leaving people displaced and homeless. Consequently, a detailed analysis shall be carried out to reveal flash flood susceptibility in the study area. It is argued that more robust and scientific approaches to flood risk reduction such as: flood modeling and vulnerability assessment were lacking in the area so that the present study is initiated to evaluate flash flood susceptibility assessment/mapping in this study area. Such type of study are believed to be important so as to enhance the level of early awareness of probability of flood susceptible areas and reduce catastrophic outcomes. Moreover, the present study is found to be unprecedented in its nature as the scarcity of observed data in arid/ semi-arid regions (e.g Jigjiga) are common issue. The following images were collected to illustrate it more about the flooded areas.

Figure 1-1: Flash flood susceptible areas

1.3 Research Objectives

1.3.1 General Objectives

The overall objective of the present study is spatial flash flood susceptibility identification providing the possibility of flash flood suscepitible areas in the upper watersheds of Jigjiga.

1.3.2 Specific objectives

- To explore a unique Flash flood Susceptibility index model integrating multi-factor analysis
- To develop a good probable flash flood susceptible informative spatial map (categorizing areas as very low, low, moderate, high and very high hazard) which can create awareness for public and city planners on probable flash flood-prone areas
- To assess spatial relationship between flash flood-occurrence and factors influencing flooded area
- To provide an alternative approach for overcoming the limitations of extensive gauged flow data required for flash flood susceptibility areas identification in hydrologic and hydraulic models in semi-area of Jijiga Jerer upstream watershed.

1.4 Outline of the thesis

The thesis was compiled by five chaptes. The first chapter clearly explains an introduction about the background and nature of flash flood susceptibility; in addition the statement of the problem, general and specific objectives of the study were stated. The Second chapter sections reviews literatures that are relating the flash flood susceptibility. In the third chapter describes the study area followed by the material and methods of generating, extracting, manipulating and processing of all the input data from different sources, furtherly, the fourth chapter briefly presents results and discussions. In the final chapter, the conclusions and recommendations of the study were delivered according to the aforementioned sections.

CHAPTER 2 LITERATURE REVIEW

2.1 General

Problems related to flash flooding have greatly increased, and there is a need for effective modeling to understand the problem and mitigate its disastrous effects. Multicriteria analysis pairwise comparisons in the AHP method in different studies [2], [11], [12], used to identify areas likely to be inundated by future flooding. AHP method is suitable for regional studies, however; it has he main problem of AHP method is the inability to determinate the uncertainty in spatial decision-making methodology may occur from selection, comparison and ranking of multiple criteria and the biggest disadvantage of this method is the need for expert knowledge in assigning weights, which can be a source of bias.

Moreover, flood hazard and risk analysis is usually performed using hydrologic and hydraulic model which can simulate flood inundation extent, water depth and velocity through onedimensional (1D) or two-dimensional (2D) hydraulic model [13], [14] and [15] such as 1D Saint Venant model for flood routing of the rivers and 2D for modelling of region flood situation. However, this kind of method is typically applied to small scale, due to requirements for data of high quality, lack of macroscopic consideration for large-scale watershed and limited availability of adequate data for flood hazard studies, and it requires the fusion of hydrologic observations, historical information about past flood events, estimates of surface roughness and boundary conditions, and topographical surveys showing all-natural and manmade features for the channel of interest and the adjacent areas. Unfortunately, gauging station networks are not widespread throughout the study area and this scarcity is generally most likely happen in most catchments area like Jigjiga-Jerer upstream watershed in Ethiopia. Therefore, the development of an effective flood susceptibility assessment approach for data scarce watershed is crucial.

Logistic regression allows one to form a multivariate regression relationship between a dependent variable and several independent variables which is useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables. The logistic regression model has also been applied in different natural hazard modeling such as landslide hazard and susceptibility mapping [16], [17]. Flood susceptibility mapping using GIS and neural network methods have been applied in various case studies [8], [18], but it is complex and requirements of high computing capacity and time-consuming for

large numbers of variables. Radar remote sensing data have been extensively used for flood monitoring across the globe, and many of these studies have applied using probabilistic methods (logistic regression) [19]–[21]. The logistic regression model in GIS processing in this study is frequently used because of its straightforward and understandable concepts. In addition, Logistic regression can explain the role of factors and it shows a strong prediction ability when compared to other machine-learning techniques. Frequently many studies are preceded by flood risk study use hydrologic/hydraulic modeling to understand flood extent and peak discharge which later related to flood risk. Unfortunately, the lack of appropriate data availability (type and resolution) in some areas prevents the application of standard engineering flood models which is a common problem in developing countries [22]. In particular Komi et al. [22], and Fekadu [2] analyzed morphometric properties for assessing hazards of flash floods to solve data-scarce problems, and they suggested understanding a basin's response to extreme rainfall based on geomorphological indices can be valuable when studying flood hazard in ungauged catchments.

2.2 Flash floods in Jerer upstream watershed, Jigjiga

One of the most challenging problem in Jerer upstream watershed is flash flood damages and the susceptibility to flash flooding among the major flood disasters in Ethiopia; is more aggravated in the last some years which is due to extreme climatic events in upper catchment, poor drainage system, using inappropriate construction alongside the drainage line through Jigjiga city, rapid urbanization, population increase, and some communities are settled in flash flood-prone areas. Flash floods frequently occurred in areas Jerer upstream watershed when excessive rains fall in adjacent highland areas particularly in Chinaksen, Lefeissa, Eastern Oromia and in uphill side of Fafen zone mountain, which is characterized by sudden onset with little lead time for early warning and often resulting in considerable damage on lives, livelihoods, and property in Jigjiga town. According to FDRE National Disaster Risk Management Commission Early Warning and Response Directorate Flood Alert report in April 2016 flash flood is anticipated in the north-eastern, central and eastern parts of the country. Taking into consideration the presence of El Nino effect with anticipated above-normal rains in the *belg* season, the risk of flash and river flood is high in different parts of the country like Somali region Jigjiga town. The reports of Benard Juma (April 2016) the worst of the floods occurred in Jijiga (also known as Jigjiga), in the Somali Region of the

country. Reports claim at least 23 people have been killed and a further 80 or more injured. According to Tsegaye Tadesse reports to Reuter (May 31, 2008) Nineteen of 25 people killed and around 400 severely damaged by flash floods in Ethiopia's eastern town of Jijiga that week were children, a regional official said on Saturday. On that year, scores of people died after weeks of heavy rain. "The force of the flash flood that hit Jijiga town on Thursday night swept away 200 houses, killing 25 people of whom 19 were children," Nur Abdi Mohammed, spokesman for the Somali region, told Reuters. Forty-five people were receiving hospital treatment, he said. Nur said the flood occurred without warning after heavy rain around Chinaksen, 50 km (31 miles) north of Jigjiga, and there had been no rain in the town itself on the day. About 250 people whose houses were destroyed were receiving emergency assistance, according to a spokesman. Recently, according to the reports of Federal Democratic Republic of Ethiopia National Disaster Risk Management Commission, Early Warning and Emergency Response Directorate (2018); flash floods occur in lowland areas when excessive rains fall in adjacent highland areas in Ethiopia and the report shows areas affected and likely to be affected by flood as shown in figure below; among these areas Jigjiga in Somali region at risk of flooding. The problem of flash floods in the semiarid area of Jigjiga is not new, but their frequency has significantly increased in the past some years. This type of flood is characterized by sudden onset with little lead time for early warning and often resulting in considerable damage to lives, livelihoods, and property. The area was hit by a flash flood in that year as a result infrastructure damaged, livestock died and many people were displaced and affected.

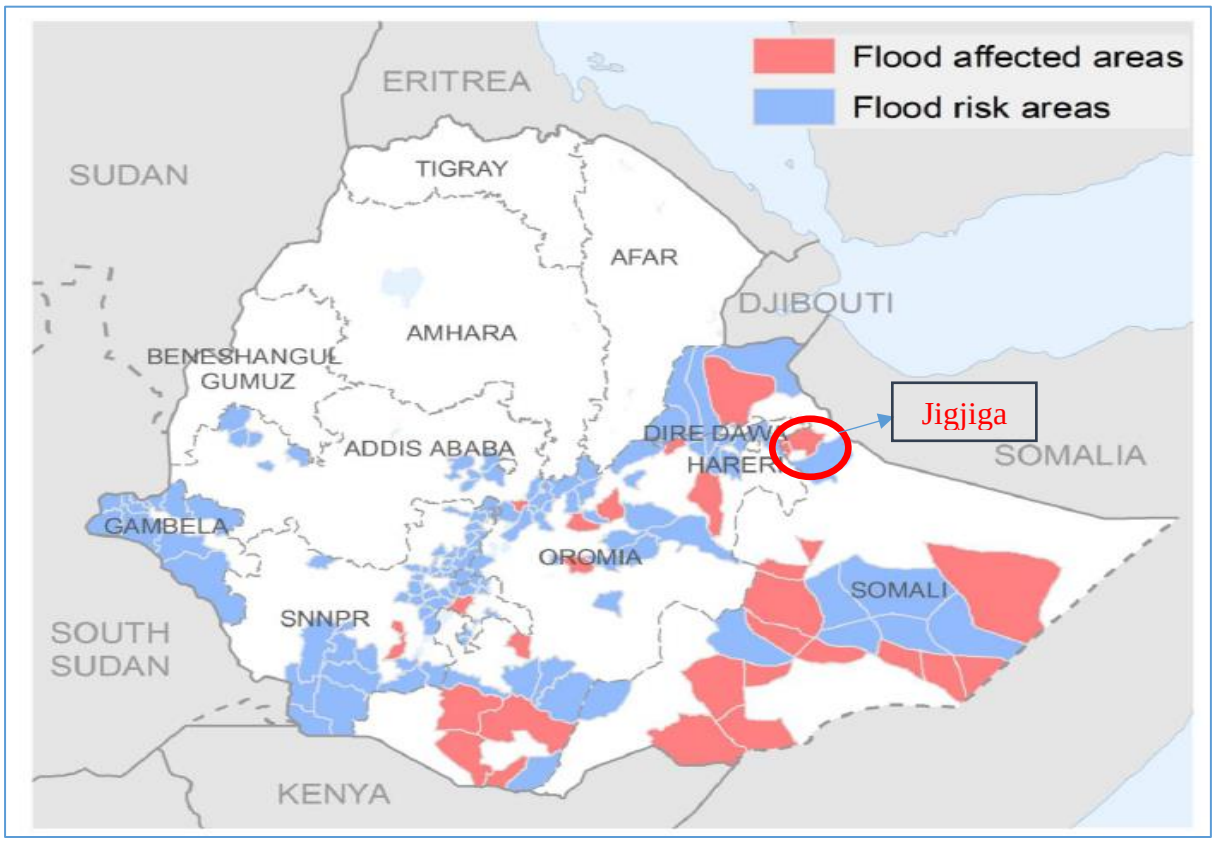


Figure 2-1: Areas affected and likely to be affected by flood in Ethiopia (FDRE NDRMCEWERD, 2018)

2.3 Previous studies of a flash flood in Ethiopia

Flash flooding is one of the most devastating hazards which has occurred and expected to occur in different parts of the country from the small scale to large scales in the past decades as well as in the future. The country has been threatened by quite unprecedented flooding of abnormal magnitude and damage. Particularly, in 2006 main rainy season has caused most rivers to swell and overflow or breach their courses, submerging the surrounding 'flat' fields or floodplains, which are mostly located in the outlying pastoralist regions of the country. As a result of the extended and widespread heavy rainfall as of the beginning of the 2006 main rainy season, many areas have already experienced devastating damage [25]. According to Dire Dawa Administration Environmental Protection Authority, 2006, altogether some 635 people have died (364 in South Omo, 256 in Dire Dawa and 19 in various other parts of the country). Thousands have lost their property and means of livelihood. In this summer, a total of some 524,400 people were vulnerable to flood disasters throughout the country. Out of this population, 199,900 people are actually affected by the flood disaster in various areas.

In Ethiopia, flood disasters and the toll paid in terms of human lives and property damage show an increasing trend (Figure below). If for East Africa, this is due to a lack of the literature on changes in heavy precipitation (IPCC. 2012), the same is patent in Ethiopia for the same reason; though [27], [28] report decreasing trends in heavy precipitation over parts of Ethiopia during the period 1965–2002. The literature on the impact of climate change on river floods (including also flash floods) is scarce, even though the changes in heavy precipitation discussed above may imply flood changes in some regions (IPCC. 2012).

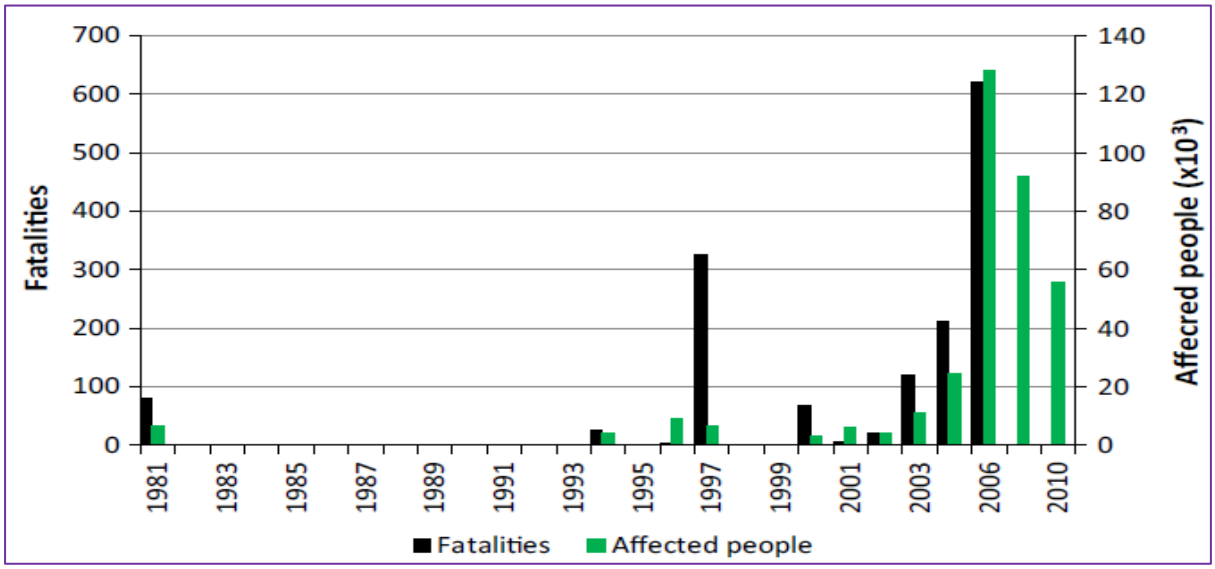


Figure 2-2: Fatalities and people affected by floods in Ethiopia between 1981- 2010 [25]

In fact according to Conway et al. [29], historically floods have never been a major economic hazard in Ethiopia, but in recent years have seen significant socio-economic disruption due to flooding. Recently, according to the reports of Federal Democratic Republic of Ethiopia National Disaster Risk Management Commission, Early Warning and Emergency Response Directorate (2018); flash floods occur in lowland areas when excessive rains fall in adjacent highland areas in Ethiopia; among these areas, Jigjiga in Somali region is most at risk of flooding. The problem of flash floods in the semiarid area of Jigjiga is not new, but their frequency has significantly increased in the past some years. This type of flood is characterized by sudden onset with little lead time for early warning and often resulting in considerable damage to lives, livelihoods, and property. In the last decade, an increasing occurrence of floods is reported for Africa in general, sub-Saharan Africa, and several arid and semiarid areas of Ethiopia. Many studies have been done to asses, forecast, manage, provide mitigation measures, and reduce the devastating damage in Ethiopia in the last some years. For example

Alemu et al. [25], Awash river basin [13], Wodim [30], and Ghibe-Omo basin [2] flood hazards were considered on a large scale while several other small scale flood hazards occurred and studied.

2.4 GIS and remote sensing for flash food

Modeling of a flash flood is a complex exercise where a lot of factors are supposed to be considered. In order to come up with the outstanding assessment and management objectives of flash flood the integrated applications of remote sensing and GIS technologies have been widely used in the world [2], [21], [30]–[32]. Accordingly, these technologies have been used in several studies as the most powerful tools for the evaluation and management of flash flood susceptibility because of RS and GIS are quick and more efficient, which can provide the best opportunity to capture, store, combine, manipulate, retrieve, analyse and display the information for the spatial determination of potential flash food hazard areas. Elkhrachy [21] used remote sensing satellite images and GIS for Flash Flood Hazard Mapping in Najran city, Saudi Arabia. He basically used SPOT and SRTM DEMs data with the causative factors of the study were runoff, soil type, surface slope, surface roughness, drainage density, distance to the main channel and land use which were derived and processed in ARC-GIS. Then he used analytic hierarchy process so that he found the relative significance of each factor to get flood hazard index map (FHI) and finally he developed flood hazard map for the study area by overlying flood hazard index map with the zone boundaries layer. Elmagd [33] presented the assessment of flash flood hazards in the Abu Dabbab drainage basin, Egypt. He was used remotely sensed data to delineate the alluvial active channels, which were integrated with morphometric parameters extracted from digital elevation models (DEM) into geographical information systems (GIS) to construct a hydrological model that provides estimates about the amount of surface runoff at three different outlets (A, B and C) as well as the magnitude of flash floods. As a result, he found that outlet B receives more surface runoff than others do means that this catchment was prone to flash flood than others. According to Samaneta et al [31] remote sensing technique provides a significant contribution in flood mapping and risk assessment they examined the usefulness of integration between remote sensing and GIS in flood susceptibility assessment and mapping. In their study the demarcation of flood susceptibility zones was carried out using ten (10)-folds of geospatial data sets, viz. land use and land cover (LULC), elevation, slope, topographic wetness index (TWI), surface runoff, landform, lithology, distance from the main river, soil texture and soil drainage which were

extracted, derived, processed and retrieved in the combination of RS data and GIS. Finally, they revealed the reclassified flood susceptibility map of the study area based on integrations of RS data, GIS and frequency ratio model [32] were studied Flash flood hazard zonation based on basin morphometry using remote sensing and GIS techniques in case of Wadi Qena basin, Eastern Desert, Egypt. In this study they carried out eleven morphometric parameters including 3-hourly accumulated rainfall TRMM data on 28 January 2013 when the last flash flood occurred of watershed analyzed using remote sensing and GIS techniques were implemented. After all they accomplished their study by classifying flash flood hazard degree as 1 (lowest) and 2 (highest) using weighted sum tool in ArcGIS 10.3 to compute hazard degree for all sub-basins in the watershed. The study of Wondim [30], was done for flood hazard mapping and risk assessment for lower Awash basin, Ethiopia integrating GIS and remote sensing for processing flood causative factors such as slope, elevation, drainage density, soil type and land cover, and finally they combined by weighted overlay in the Arc_GIS environment. The major finding of him was the flood hazard map of Lower Awash sub-basin indicated that 5%, 23%, 39% and 33% of the area considered in Lower Awash Sub-basin were subjected respectively to low, moderate, high and very high flood hazards. Bishaw [34] applied GIS and remote sensing techniques for flood hazard and risk assessment in Dugeda Bora Woreda of Oromiya Regional State, Ethiopia.

2.5 Integration of GIS, remote sensing and logistic regression model for flash flood and hazard susceptibility mapping

Logistic regression is one of the most frequently used multivariate analysis methods for creating spatial flash flood susceptibility prediction maps integrating with GIS and remote sensing [16], [17]. The logistic regression model in GIS processing in this study is frequently used because of its straight forward and understandable concepts. In addition Samaneta et al.[31], [2] Logistic regression can explain the role of factors and it showed a strong prediction ability when compared to other machine-learning techniques [35]. Moreover, this approach is useful for situations in which one wants to be able to predict the presence or absence of a characteristic outcome from a set of predictor variables. The purpose of LR is thus to simulate the relationships between a dependent variable and multiple independent parameters. The merit of LR is that it does not compulsorily require normal distribution data, both continuous and discrete data types can be used as an input for the LR model. In this study flash flood conditioning factors were derived and extracted from remote sensing data

using GIS and its supplementary software so as to derive relationship between flash flood occurrence and causative factors in logistic regression. Then probability of flash flood susceptibility index area map can be generated using the derived logistic coefficients of each factor through the formula that can be expressed as:

$$P = 1/(1 + e^{-Z}) \quad (2\text{-}1)$$

Where; p is the probability of flooding which is acquired between 0 and 1 on an S-shaped curve. Z is a linear combination and it follows that logistic regression involves fitting an equation of the following form of data:

$$Z = \beta 0 + \beta\, 1x1 + \beta\, 2x2 + \ldots + \beta nxn \quad (2\text{-}2)$$

Where β 0 is the intercept of the model,

β i (i = 0, 1, 2, …, n) are the slope coefficients of the logistic regression model,

and the xi (i = 0, 1, 2, …, n) are the flash flood influencing factors.

A positive LR coefficient indicates that the existence of the flash flood conditioning factor in the area increases the probability of the flood creation while the negative logistic coefficient values imply that the occurrence of flooding is negatively related to that specific factor [17]. According to many studies [16], [17], [36], this technique is highly suitable for flash flood hazard evaluation because is a decision-making technique utilized for solving complex problems with many parameters of interrelated objectives or flash flood condition factors. Pradhan et al. [16] progressed flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing within the Kelantan River in Malaysia. He mainly used RADARSAT data for RS, and topographical map, geological map, hydrological map, Global Positioning System (GPS) data, land cover map, geological map, precipitation data, and Digital Elevation Model (DEM) for GIS data. His results showed that delineated flood-prone areas can be performed at 1:25,000 scale which is comparable to some conventional flood hazard map scales qualitatively, the model gave him reasonable results with accuracy was 85%. Pradhan et al. [36] have utilized optical remote sensing (RS) data and radar RS data along with a logistic regression for their study to develop modeling for flood-damaged area delineation in Hoeryeong City in northern North Korea, they considered radar backscattering coefficient difference, elevation map, slope map, land use map, landform map, flow accumulation map and flow direction map were used. After that, study results were

confirmed via comparison with Google Earth images taken after the typhoon. High-resolution web-based satellite imagery was also interpreted to confirm the results of the study. Finally, they got good results of flood-damaged area delineation with the overall accuracy of 88.5%. Tehrany et al. [17] prediction rate of 90.36% flood susceptibility map in china was achieved using the ensemble technique of weight of evidence and logistic regression considering 15 flood conditioning factors included in the spatial database (GIS) were altitude, slope, aspect, geology, distance from river, distancefrom road, distance from fault, soil type, land use/cover, rainfall, normalized difference vegetation index, stream power index, topographic wetness index, sediment transport index and curvature. Pradhan et al. [37] used a Geographical Information System (GIS) and remote sensing to evaluate the hazard of landslides at Penang, Malaysia. In his study Topographical and geological data and satellite images were collected, processed and constructed into a spatial database using GIS and image processing. Then he had chosen some factors which were influence landslide occurrence such as: topographic slope, topographic aspect, topographic curvature and distance from drainage, lithology and distance, land use, vegetation index value. Landslide hazardous areas were analysed and mapped using the landslide occurrence factors by logistic regression model. The results of the analysis were verified using the landslide location data and compared with a probabilistic model.The validation results showed that the logistic regression model is better in predictionthe than probabilistic model. The recent study of Al-Juaidi et al. [38] carried out their study based on logistic regression for evaluation of flood susceptibility mapping using flood conditioning factors which are: digital elevation model (DEM), topographic slope, flow accumulation, rainfall, land use/land cover (LULC), and soil type derived using GIS; where he identified 140 flood locations while 70% were randomly used for data training and the remaining 30% were used for data validation. Quantitatively, they found that the proposed model prediction and success rates were 76 and 81%, respectively which were at reasonable accuracy.

2.6 Flash flood conditioning factors

Selecting flash flood conditioning factors is one of the primary objectives and developing flash flood inventory map is mandatory for flash flood susceptibility prediction and mapping of someone. According to many literatures [31], [39], [1], [40], and there is no still general guidelines and frameworks to select flash flood conditioning factors, but still flash flood inventory map is accomplished based on the past historical flash flood occurred using the

orthorectified aerial imagery and then these surfaces being validated by field surveys in the study area [40]. However, many of this literature consensus that the selection of those factors based on flash-flood characteristics and the availability of geospatial data in the study areas. According to Diakakis et al. [24] GIS is used to determine the significance of the value of potentially flood-influencing parameters, while logistic regression was used to examine the statistical correlation between their values and the respective values at the locations of 1138 flooded buildings. Their results showed that certain factors, including the degree of soil sealing, accumulated rainfall, slope, and others, influence the distribution of flooding to significantly different degrees. In the study, a strong relationship between a combination of factors and flood occurrence with R^2 was reached 0.63 and prediction accuracy of 82.9%. Samaneta et al. [31] have used (10)-folds of geospatial data sets considering to achieve highly accurate results in their study; viz. land use and land cover (LULC), elevation, slope, topographic wetness index (TWI), surface runoff, landform, lithology, distance from the main river, soil texture and soil drainage in order to achieve the demarcation of flash flood susceptibility. Moreover, they selected those factors after consulting local hydrological and natural disaster experts based on their effectiveness in creating a flood. Costache [40] has considered 10 environmental variables: slope angle, land use, lithology, hydrological soil group, convergence index, topographic wetness index (TWI), topographic position index (TPI), aspect, plan curvature and profile curvature; he applied Linear Support Vector Machine (LSVM) method to select flash flood conditioning factors and then using the results of LSVM method, the irrelevant and unnecessary input variables was removed. Information from different data sources was collected and a database was developed in a geospatial environment.

2.7 Some flash flood conditioning factors with flash flood susceptibility

In hydrological assessment, the slope of an area expresses a fundamental role to regulate surface discharge. In such hydrological studies, it is a very important topographic factor [16], [41]. Whenever there is steepness found in the slope of an area, there would be highest surface flow velocity; additionally, gradient partially controls the infiltration process. The surface runoff increases significantly as the gradient increases; consequently, the infiltration decreases in the area. As a result of this, regions with a sudden decrease of the slope, having a higher probability of flood as a massive volume of water causes a severe flood situation [16], [42].

Moore et al. (1991) stated that the SPI controls the potential erosive power of overland flow and catastrophic transformation of the channel because of the high stream power is one of a case in channels. The SPI is a factor that estimates the degree of slope erosion due to flowing water [47]. Therefore, stream power index (SPI) is recommended as a flash flood contributing factor, because it refers to the rate of the erosive power in the drainage network.

Drainage is an important ecosystem controlling the hazards as its densities denote the nature of the soil and its geotechnical properties [12]. This dictates that the higher the density, the higher the catchment area susceptible to erosion, and further resulted in sedimentation at the lower grounds. Moreover, higher probability of flooding is strongly associated with higher drainage density as it indicates greater surface runoff. In the classification process, an area with a higher drainage density is considered to be very highly affected by flood.

Many studies considered the distances of flash flood susceptibility from the drainage network according to the historical susceptible areas. Pradhan [16] considered the regions within 90m distance from the drainage network are more vulnerable to the flood. Samanta et al. [31] considered that the regions located less than 100m distance are highly flood-prone whereas distance more than 2000m has very low flood potential.

2.8 Spatial flash flood susceptibility prediction and mapping

Flash flood is one of the devastating and devastating hydro-meteorological disasters causing considerable damage to highways, settlement, agriculture, and livelihood and many infrastructures in different parts of Ethiopia. It is impossible to avoid its damage or prevent its occurrence totally, however it is possible to work on the reduction of the hazard effects and to reduce the losses which may cause [21]. Therefore, spatial flash flood susceptibility prediction and mapping are among mandatories in order to identify sites in high-risk flash flood zones is one of the powerful objectives for this purpose. Mapping flash floods will be beneficial to urban and infrastructure planners, risk managers and disaster response or emergency services during extreme and intense rainfall events. Moreover, susceptibility is measured by the environmental factors that in combination make places vulnerable to the hazard and prediction represents the probability in the likelihood of a hazardous event to occur in a place. It is of high importance for the city urban planning, watershed management, infrastructure designers and risk managers to map flash flood susceptible areas for future planning.

2.9 Spatial flash flood susceptibility prediction and mapping in scarce gauged rainfall data

Satellite based estimates such as the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA; also TRMM 3B42) product are becoming increasingly attractive as an alternative source of forcing data in data-sparse regions, although their application can be limited by quantitative inaccuracies [43]. TRMM 3B42 provides rainfall values between 1998 and 2014 and presents one of the most valuable 17 years of spatiotemporal rainfall datasets to date.

Many studies revealed that taking gauged rainfall data, as the only alternative for flash flood susceptibility prediction and mapping is not essential unless abandon studying it in scarce rain gauge areas because of the spatial variability/ less distributed and failure of gauges for measuring the rainfall in many study areas. Floods place the problem of un-gauged area prediction under rather extreme conditions [12]. Process understanding is required for flash flood risk management, because the dominant processes of runoff generation may change with the increase of storm severity, and therefore, the understanding based on analysis of moderate floods may be questioned when used for forecasting the response to extreme storms. This means that flood mapping in most developing countries, which are mostly un-gauged, is a complex problem that should integrate data scarcity and uncertainty in the analysis and decision-making process. In such cases, many studies were done to overcome and came up with the outstanding results of flood hazards. Accordingly, the studies of El Maghraby and B. Niyazi [3], Rahmati [9], Mandoza et al [45], used a multi-criteria analysis approach to develop flood susceptibility maps in arid and ungauged catchments. Moreover, others were used satellite TRMM precipitation products data as an alternative in such areas for modeling, predicting, and susceptibility mapping [32], [45], [46] which provide a streamflow forecast for ungauged, complex terrain basins after bias adjustment so as improve application in flood prediction in a given study area.

In many cases, of watershed rainfall is observed at very few locations, and they are being characterized by high spatial and temporal variability. Several remote sensing-products, such as the Tropical Rainfall Monitoring Mission (TRMM) at semi-global coverage with a grid spatial resolution of 0.25^0 x 0.25^0. Still, remote sensing-based rainfall estimates particularly may represent the sole source of precipitation input to any hydrological model for the dryland catchments, particularly when rainfall stations are insufficient or totally absent [47]. In this

study because the climate of observed data is not fluently available and the area is characterized by semi-arid and, no comprehensive studies have been carried out about climate change due to short length of data records and scarcity of climatological stations network. The consequence is that natural resource managers in this watershed must deal not only with low rainfall but also with highly temporal variable rainfall.

Generally, the occasional heavy storms provoke severe flash floods that devastate inhabited areas, hence in this study latest version of satellite TRMM product (3B42V7) precipitation product as an input to develop the spatial mean annual rainfall map thus it was used as one of flash flood conditional factor. These data were then spatially interpolated using Inverse Distance Weighted (IDW) method to obtain the rainfall distribution map [48] and it is the widest used as an interpolation method which can be integrated with GIS is used to estimate the rainfall distribution in a study area. The IDW gets the value by calculating a weighted average of known values within a specific neighborhood. It was defined as everything is related to everything else, but near things are more related than distant things. The IDW method involves the process of assigning values to unknown points by using values from a scattered set of known points [48]. The value at the unknown point is a weighted sum of the values of N known points. Therefore, IDW simply can be used to estimate the unknown spatial rainfall data from the known data of sites that are adjacent to the unknown site.

CHAPTER 3 MATERIAL AND METHODOLOGY

3.1 Description of the study area

3.1.1 Location of Study Area

Jerer watershed is located in eastern Ethiopia particularly in Somali Region which includes Jigjiga city. From administrative point of view, it is located in the Jijiga Zone approximately 80 km east of Harar and 60 km west of the border with Somalia, which is bounded by latitudes and longitudes of 9°21′N 42°E to 9.35°N 42.8°E respectively at an average elevation of 1609 meters above mean sea level. From hydro-meteorological settings, the basin is located in Wabishelle River Basin and it is characterized by frequent semi-annual flash floods. The present study focusses on the upstream part of the Jerer sub-basin covering an area of 1053 km^2.

3.1.2 Climate

The climate of Jijiga watershed is a subtropical highland climate (Köppen climate classification), extremely wet and lush during rainy season, as with the rest of the Ethiopian highlands. Seasonal differences relate only to rainfall, as temperatures year-round are cool to mild in the mornings and uniformly very warm though not hot during the afternoons. There are two rainy seasons: the main *meher* rains occur from July to September, and the short *belg* rains in April and May. The dry season, known as *bega*, is cooler by morning than the wet seasons due to lower cloud cover, but equally hot by the afternoon though less humid. Generally, it has a temperature of 12.27^0c and 27^0c minimum and maximum respectively with minimum and maximum rainfall lying between 400mm and 800mm respectively with an annual mean of 712mm. Relatively the study area is the most drought-prone area in the country due to high inter and intra-annual irregularity in rainfall and high coefficients of variation. The rainy period in most of this area is from March to the end of June followed mostly dry periods with rare precipitation.

3.1.3 Physiography and Drainage

The study area is almost located on flat land with gentle slopes in which some parts of the area are characterized by poor drainage, gully and swampy land features. Gully areas that are found in the western, southern, southeastern, and northern parts of the area require immediate intervention or remedy measures. The topography slopes down from the Karamara Mountain in the northwest spread out in the southeastern border of the town, with a few numbers of

steep-sided valleys and streams. In general, the topography is characterized by gentle morphology and flat land areas. As a result, the stream drains towards southeastern from the Karamara ridge; southeast direction from upper Cinaksan direction and other elevated areas of the eastern outskirts of the city. Wetlands along Biribiris and Toga streams (including Elbahiy and Biyeda streams) and areas south to the southwest of Elbahiy Dam are the major drainage systems in the watershed passing through Jigjig city

3.1.4 Geology and Soils

The surrounding watershed is characterized by the following three categories of geological features: Limestone, sandstone, and basalt. As the Information from Jijiga Woreda's Agriculture Office reveals that the hinterland of the area is dominated by mixed *eutric cambisols*, *chromic vertisols*, black *vertisols*, mixed *Calcic Cambisols*and black *Vertic Cambisols with* clay texture soil types.

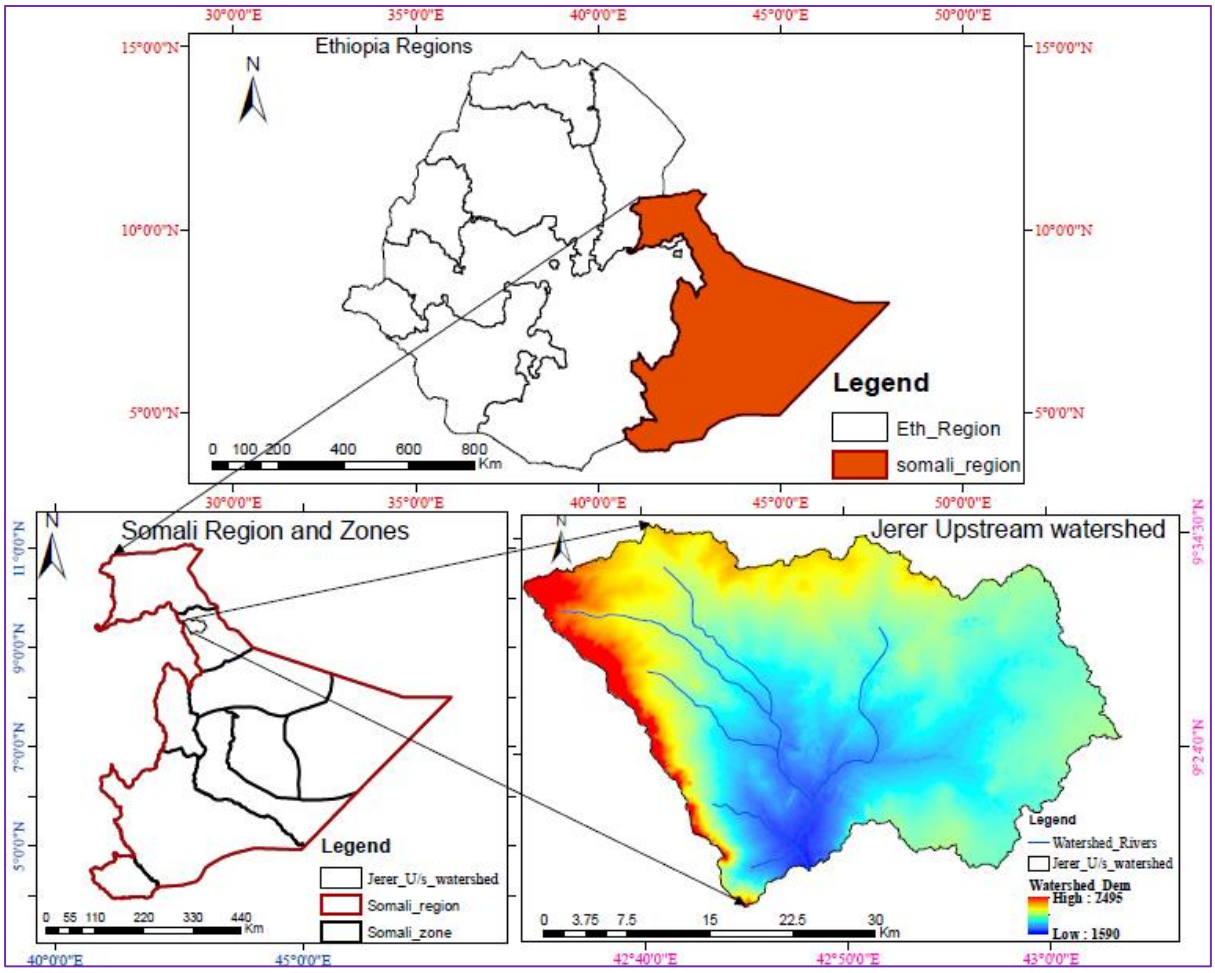


Figure 3-1: Study area location

3.2 Data sources and Tools

In the study, different data were collected from various sources and analyzed using different softwar. Table 3-1 and 3-2 below summarizes important tools, datasets, sources and their purpose in the study.

Table 3-1: Short Summary of tools used and their sources

Name	Purposes	Sources
ArcGIS 10.4.1	• To Delineate study area map • Derive and generate most flash flood conditional factors, and classify them • To combine and flash flood conditional factors according to their weight for developing flash flood susceptibility index • Develops a map of the susceptibility prediction result • Integrate with Google Earth Pro for as supplement for developing flash flood inventory map	Student trial licenses, sources from AAU FTP portal
R studio	• To check and compare the relative importance of flash flood conditional factors (Weights) • To develop a logistic regression model • Evaluate multicollinearity between those factors • To extract the satellite daily TRMM 3B42V7 time series product for a given coordinates	https://www.rstudio.com/products/rstudio/download/
Mendeley Desktop	• To organize all Journals used in this thesis work for inserting citation	https://www.mendeley.com/download-desktop/
ERDAS IMAGINE 14	• Downloaded satellite image stacking and preprocessing • Land use land cover classifications both in Supervised and Unsupervised of the study area • Generate NDVI for the study area • Accuracy assessment of classified images	https://getintopc.com/softwares/3d-designing/intergraph-erdas-suite-2014-free-download/

Table 3-2: Short Summary of data layers used and their sources

Flash flood prediction factors	Source	Gis data type and cell size
Land use land cover	Georeferenced Satellite Landsat 8 images https://earthexplorer.usgs.gov/; Validated with Ground truth values and Google Earth Pro data	Grid (30m)
Lithology	Geology (Geological survey of Ethiopia)	Grid (30m)
Drainage density	Derived from digital elevation model	Grid (30m)
Topographic weightiness index	Derived from the slope (in degree) and flow accumulation pixels	Grid (30m)
Stream power index	Derived from DEM	Grid (30m)
Slope	Derived from DEM	Grid (30m)
Soil types	Arc Swat soil type	Grid (30m)
NDVI	Extracted from Georeferenced Satellite images using ERDAS Imagine 2014	Grid (30m)
SRTMDEM	https://earthexplorer.usgs.gov/	Grid (30m)
Morphometric hazard index	Derived from DEM combining many morphometric characteristics	Grid (30m)
Rainfall	Extracted from Tropical Rainfall Monitoring Mission (TRMM)	Grid (30m)
Distance from main drainage	Derived from DEM	Grid (30m)
Rainfall	From NMA and satellite TRMM 3B42 V7	
Historical flash flood event samples	Derived from field visits survey and satellite images	Point data

3.3 Methodology

In order to achieve the intended objectives, both the aforementioned field and secondary data sets were analyzed in step by step manner. In the subsequent sections below, the detailed methods for flash flood susceptibility analysis and maping procedures were explained. In section 3.3.1, the overall flow of the methods are briefly explained. In section 3.3.2, the derivation procedures for different conditioning factors were grouped in four categories with their detailed methods. Section 3.4 is dedecated for analysis methods of both ground and satellite-based rainfall analysis. The rainfall conditioning factor for flash flood susceptibility analysis was also explained in the same section. In section section 3.5, explains about the landuse land cover data analysis and its use in flash flood susceptibility study. Finally, section 3.6 is dedicated for the overall flash flood susceptibility prediction method by combining the previous section's outputs.

3.3.1 General Methodology

The general methodology in the following figure shows that all the data used and their combination to generate the flash flood susceptiblity prediction map for the study area.

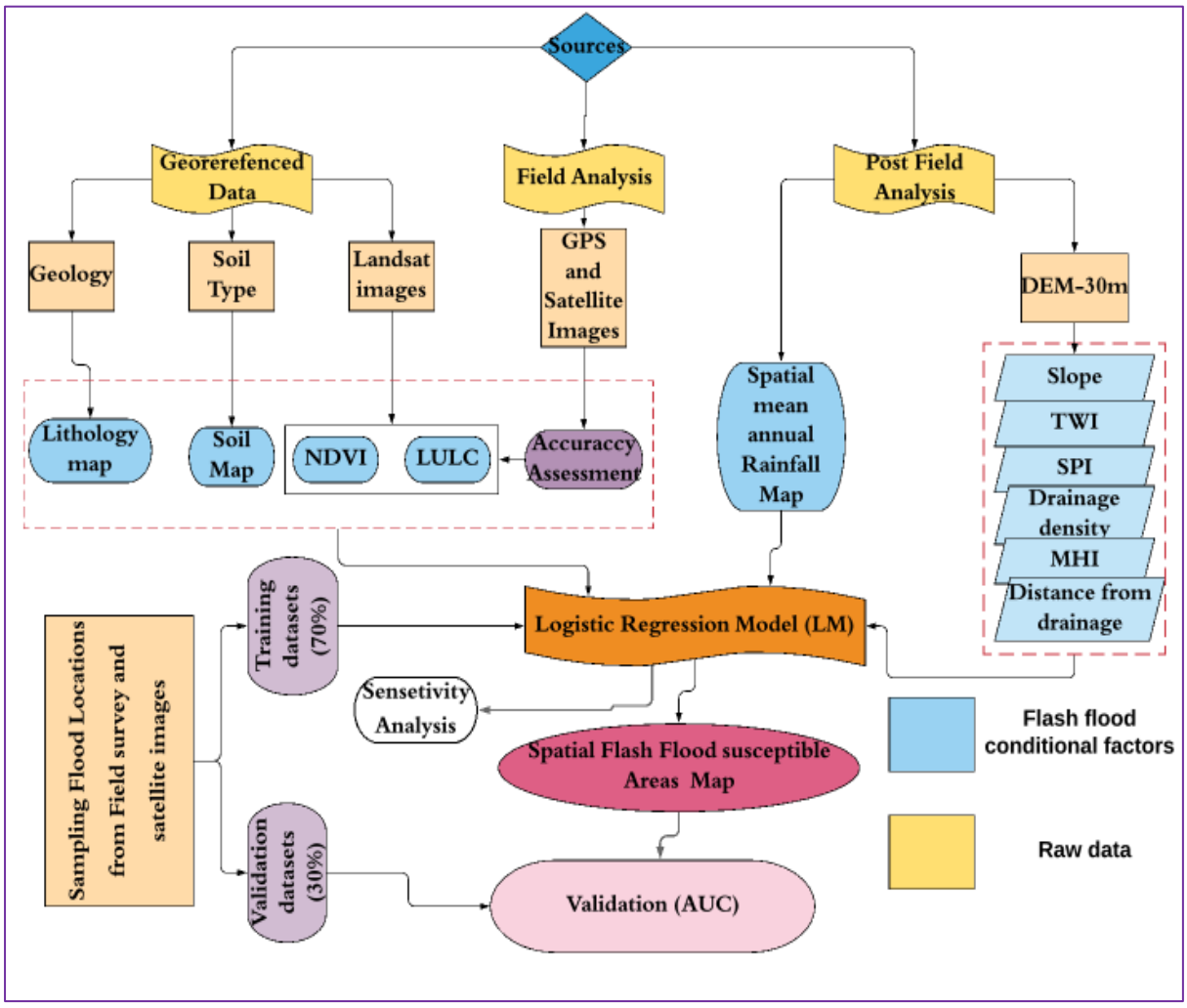


Figure 3-2: Flow chart of the methodology

3.3.2 Flash flood conditioning factors (logistic regression variables) and methods of their processing

As presented in section 3.2 and 3.3.1, about eleven flash flood influence factors (light blue highlighted boxes in the figure) were used in this study. These factors were largely grouped into topographical, hydrological, geological, and land cover factors. Flash flood influence factors were derived and generated using ArcGIS 10.4.1 software and ERDAS Imagine 2014, and then all converted into a raster file with a spatial resolution of 30 × 30 m prior to use as independent factors for logistic regression model so as to produce the Spatial flash flood susceptibility prediction map. In the following section, these factors are explained in detail and their classification was done based on the flash flood susceptibility in the study area. Moreover, the spatial database was produced combining these factors.

3.3.2.1 Topographic factors

Slope is one of the topographic factors considered in the study. The slope degree is a factor that is mainly used to determine groundwater recharge processes, as gentle slope areas have a low surface runoff and high rates of percolation, while the opposite is true for high slope areas. This factor was extracted from the digital elevation model (DEM) using a spatial analyst tool of ArcGIS 10.4.1 software after correcting the data gaps of the elevation model.

The slope is highly related to the flow regulation towards downstream, which can be perceived in-stream power models stated that higher magnitude of the slope may accelerate precipitation related runoff to the downstream. Accordingly, the slope (degree) derived and latter classified as shown in the figure below. In the figure the lowest slope value in the study area is more likely to susceptible for flash flooding than mountainous parts of the topography based on this slope is classified in to five groups where 5 is highest and 1 is lowest spatial flash flood susceptibility zone where its value is between 0 and 58^0.

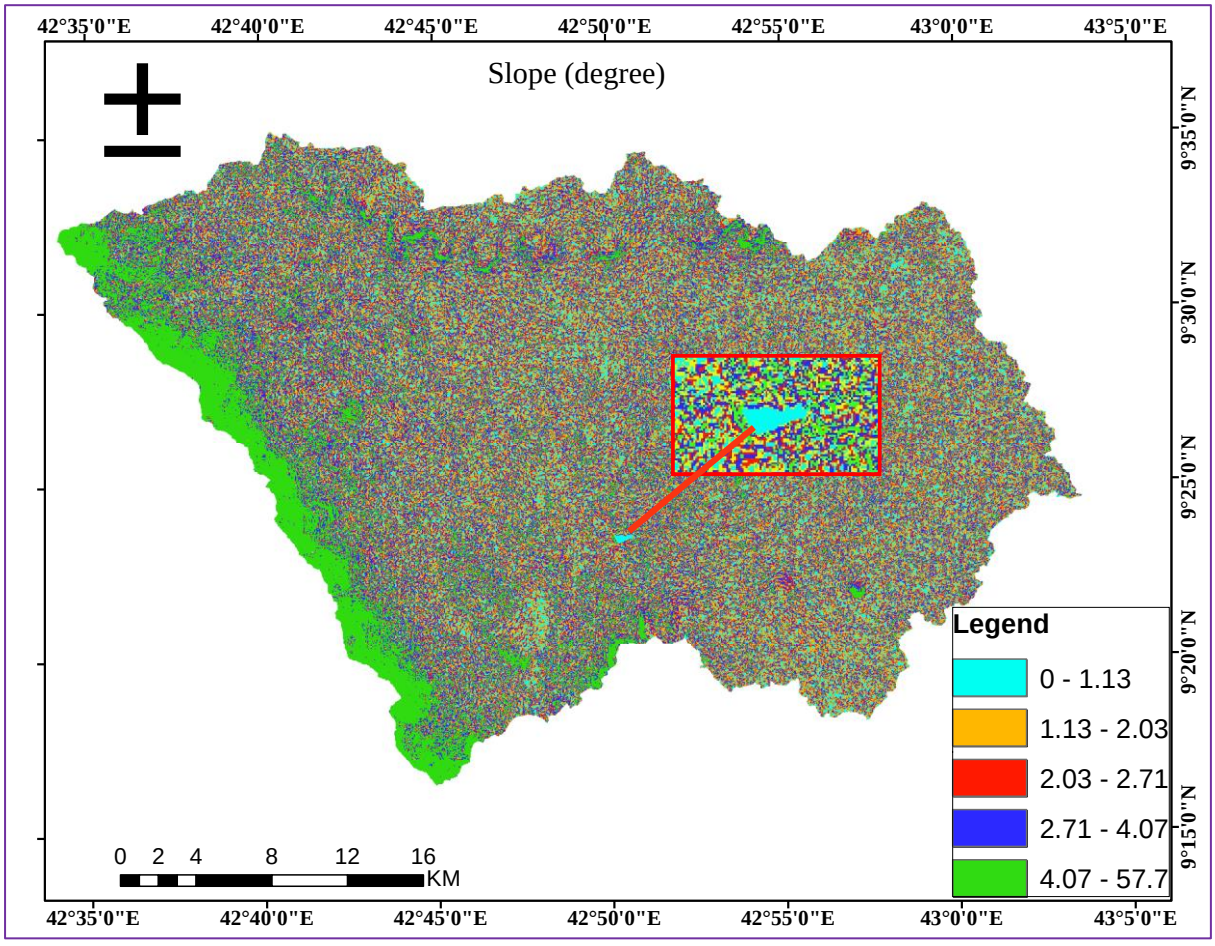


Figure 3-3: Slope in degree

3.3.2.2 Hydrologic related factors

There are different factors that affect the spatial flash flood susceptibility of the area among this factor: topographic wetness index (TWI), stream power index (SPI), distance from drainage, and drainage density and morphometric hazard index were considered.

I) Topographic Wetness Index (TWI): The topographic wetness index has been used to describe spatial moisture patterns and explain the effects of topographic conditions on these patterns. It plays an important role in influencing the movement and accumulation of runoff on the ground surface. The topographic wetness index (TWI) has been used extensively to describe the effect of topography on the location and size of saturated source areas of runoff generation. The index is a function of both the slope and the upstream contributing area per unit width orthogonal to the flow direction. Moore (1991) proposed the equation for the calculation of TWI under the assumption of steady-state conditions and uniform soil properties (i.e. transmissivity is constant throughout the catchment and equal to unity).

Generally, TWI is calculated based on a hydrologically-corrected DEM, implying that all in which sinks (including real depressions) are fully filled before TWI computations. Due to the sink-filling pre-processing, flow directions, flow accumulations, and slopes of all the grids associated with depressions are altered. This alteration also gives rise to artificially high or low TWI values. High TWI values reflect an area that is prone to saturation, whereas low TWI values reflect an area that is not prone to saturation [49]. Areas prone to water accumulation (large contributing drainage areas) and characterized by low slope angle will be linked to high TWI values. On the other hand, well-drained dry areas (steep slopes) are associated with low TWI values. However, which particular TWI values indicate wet soils vary depending on landscape, climate, and scale [50]. Hence, the highest value of TWI is more susceptible to flash flood and the lowest is less. Therefore, in this study, the range of TWI is between -2.46 to 33.96; its derivation and map are shown in the Figure below.

$$TWI = \ln (As/\tan\beta) \quad (3\text{-}1)$$

where As is the flow accumulation, and β is the slope gradient (in degrees) [51].

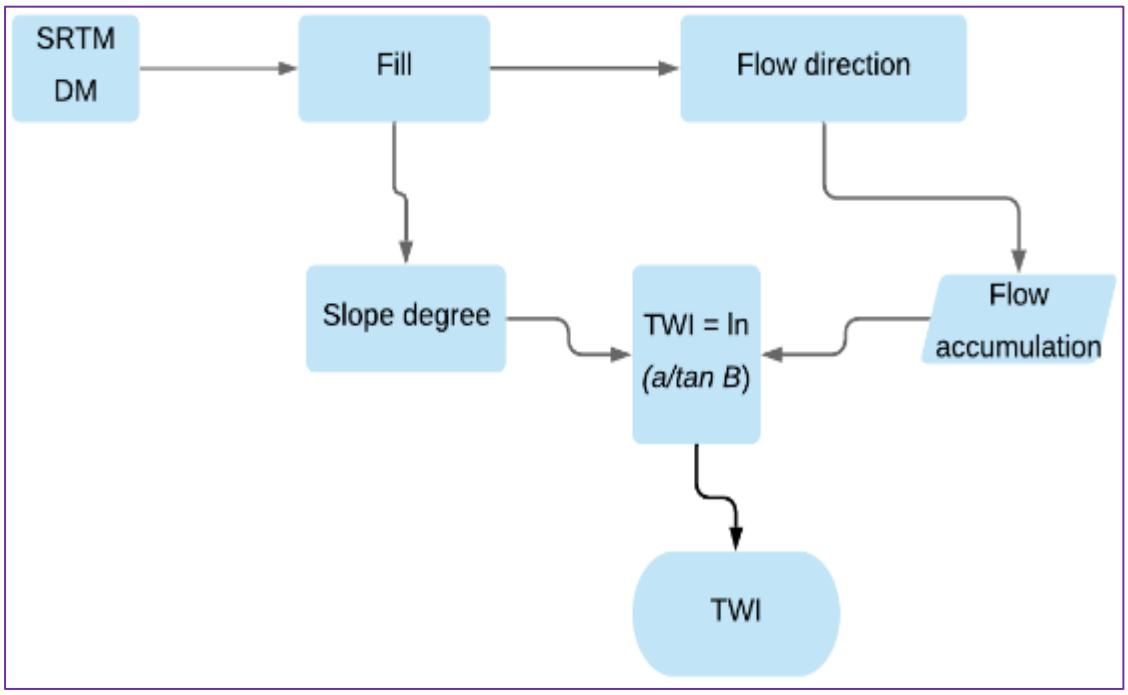


Figure 3-4: General TWI derivation framework procedures

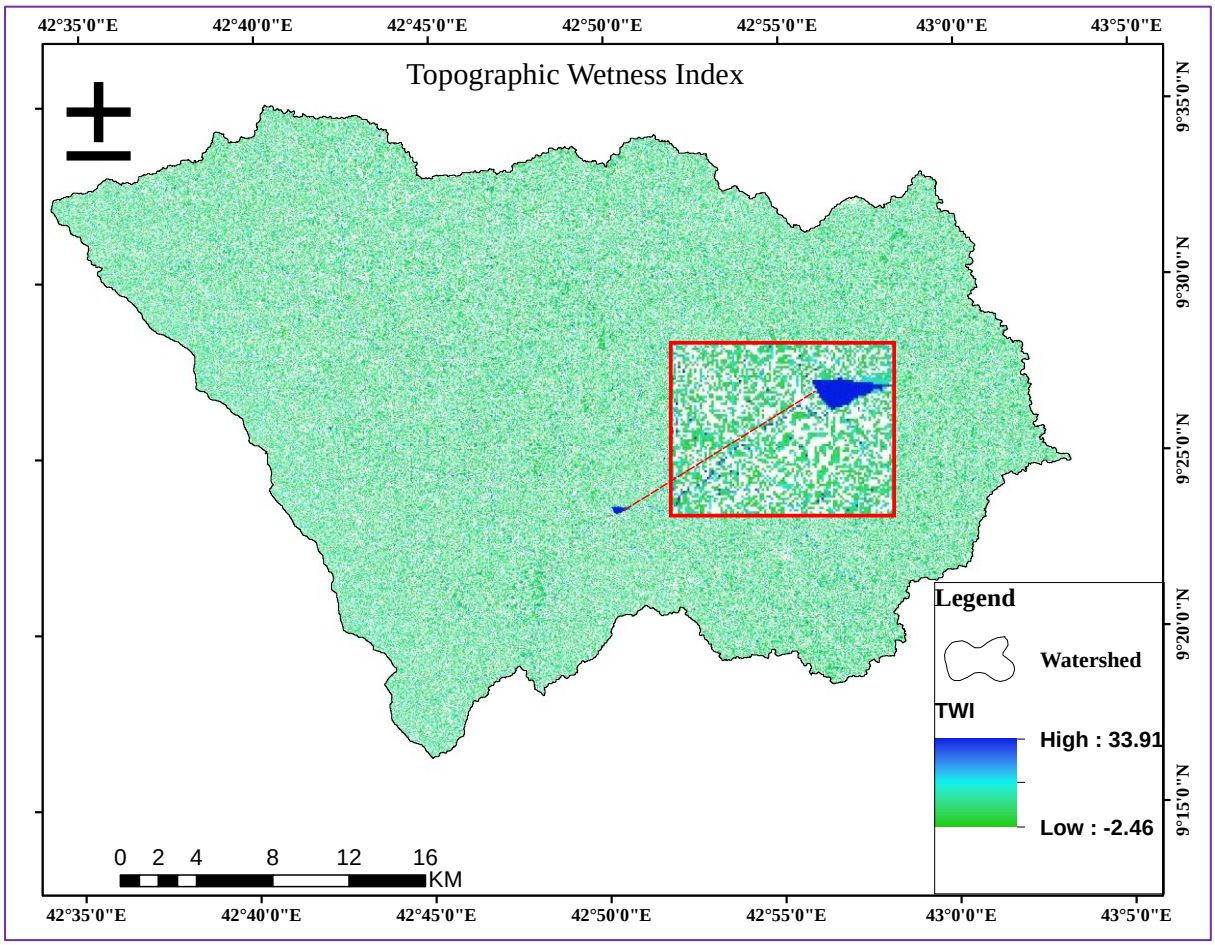


Figure 3-5: Topographic wetness index

II) ***Stream Power Index (SPI):*** is a measure of the erosive power of water flow based on the assumption that discharge (q) is proportional to specific catchment area (As) [84].

$$SPI = As \times \tan\beta \qquad (3\text{-}2)$$

Where As is the specific catchment's area (m^2/m), and β the slope gradient in degrees. As the specific catchment's area and gradient increase, the amount of water contributed by upslope areas and the velocity of water flow increase; hence, the SPI and slope-erosion risk increase [84]. Generally, the SPI map of Jerer upstream watershed was produced using model builder in ArcMap 10.4.1 and classified into five classes with equal intervals where the highest value indicates areas prone to flash flood susceptibility, and the highest values (12.04) were found along the drainage in the study area as shown in Figure below.

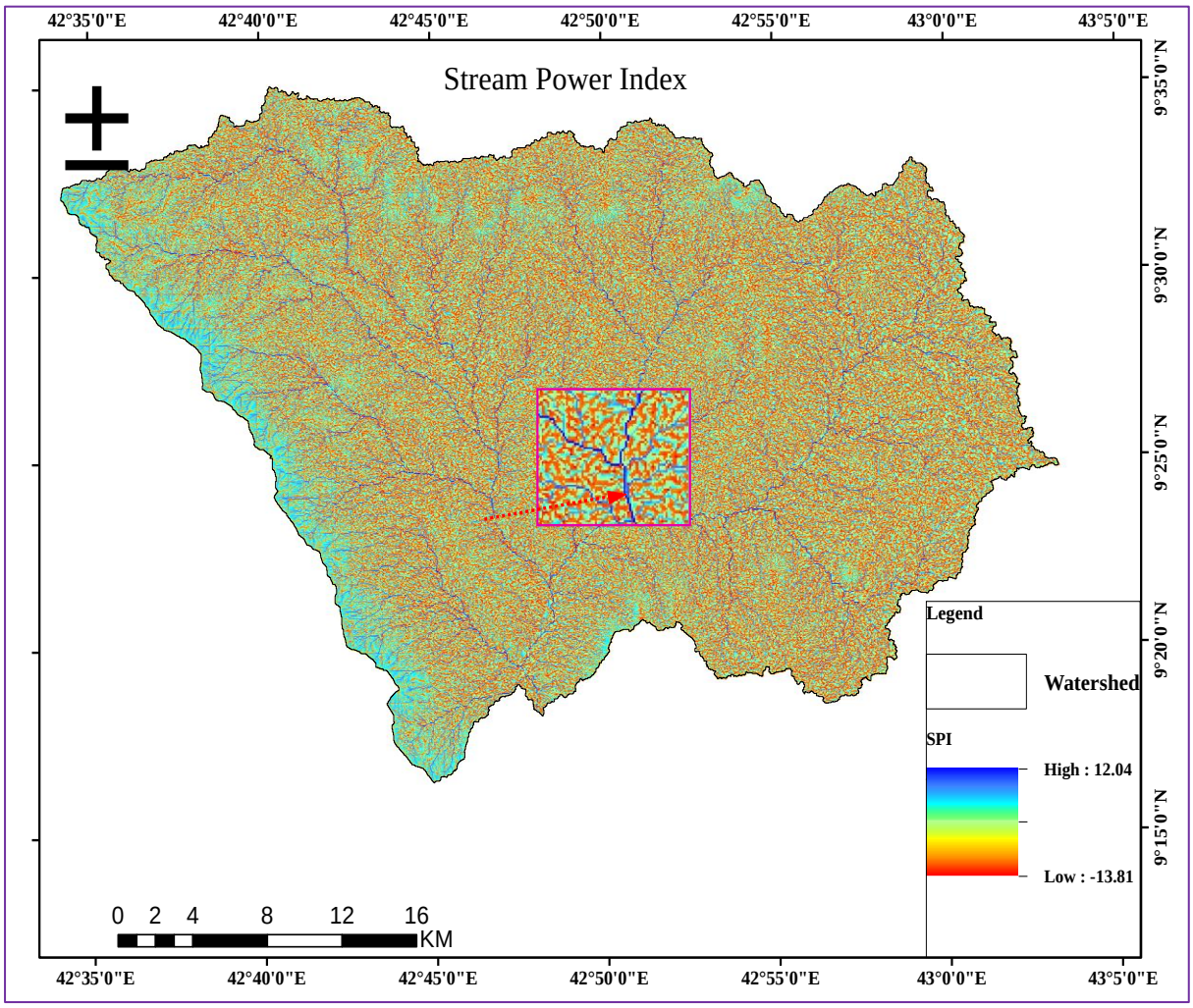


Figure 3-6: Stream Power Index

III) ***Drainage Density (Dd):*** The DEM was used to compute the drainage density (Valleys) using the spatial analyst extension. However, all the valleys do not necessary carry water. The drainage density is the total length of all the streams and rivers in a drainage basin divided by the total area of the drainage basin. The "line density" module in the spatial analyst toolbox of ArcGIS calculates a magnitude per unit area from polyline features that fall within a radius around each cell. The drainage density layer was then classified into five classes.

Since the average distance from watershed divide to channel is approximately 1/ 2Dd, many authors have used Dd as a measure of drainage efficiency and hence as an independent variable for predicting a wide variety of streamflow characteristics in ungauged watershed. Moreover, it is largely a function of infiltration capacity and soil resistance to erosion and

that length of overland flow is one of the most important independent variables affecting both the hydrologic and physiographic development of drainage basin [52]. Horton [52] suggested that the drainage density of the watershed is calculated as:

$$Dd = L/A \quad (3\text{-}3)$$

Where Dd = drainage density of watershed; L = total length of the drainage channel in the watershed (km); A = total area of the watershed (km^2). A watershed with adequate drainage runoff should have a drainage density ≥ 5, while the moderate and the poor ones have drainage density classes 1-5 and <1 respectively. In the study area, streams up to 2^{nd} order were used to calculate the Dd. For the flash flood susceiptibility mapping in the study area, higher weights were assigned to poor drainage density areas and lower weights were assigned to areas with adequate drainage.

The drainage density layer was further reclassified in five sub-groups using the standard classification procedure in ArcGIS (1–5). Areas with very low drainage (high drainage density) were ranked as 1 and those with very high drainage density were ranked with the value of 5 as depicted in the results Figure 3-7.

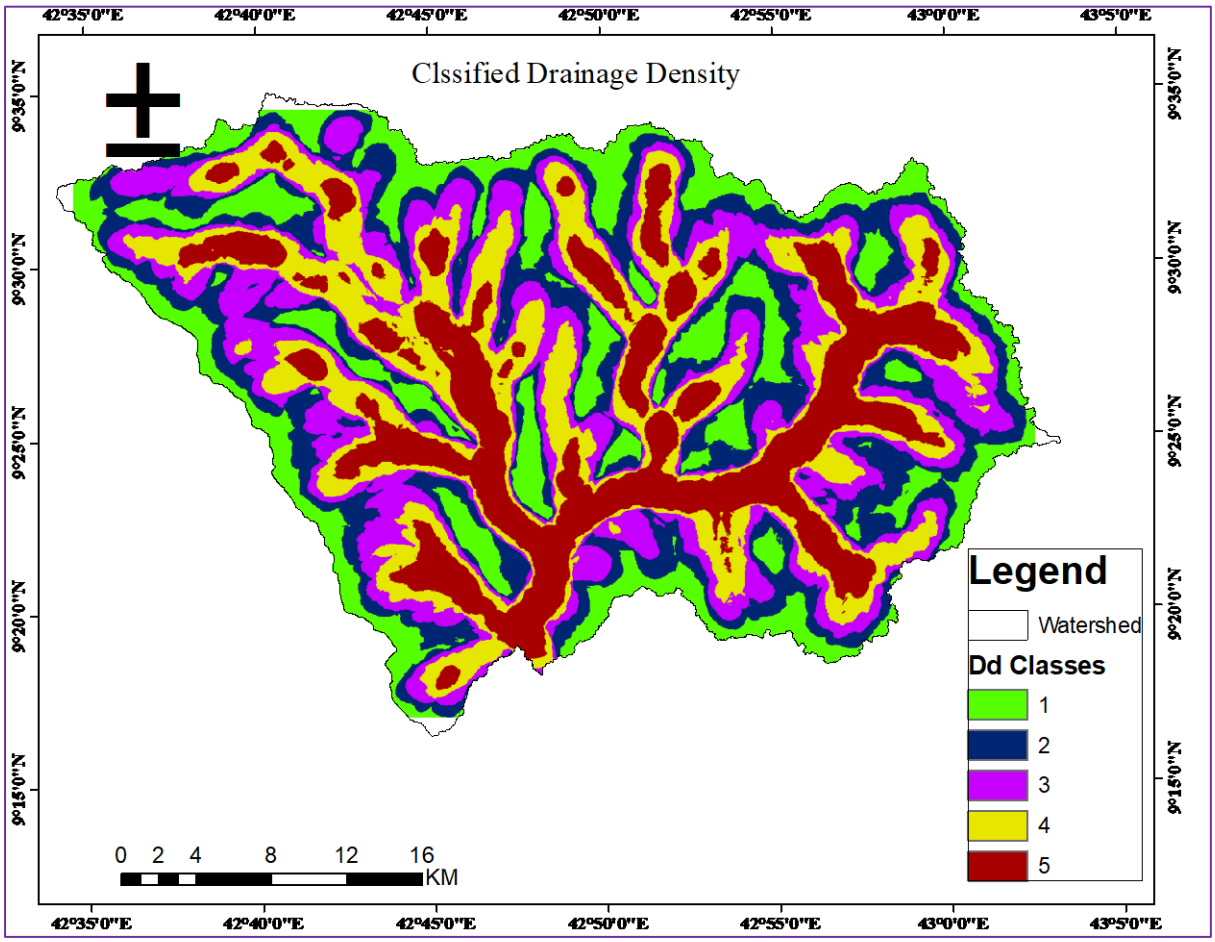


Figure 3-7: Classified Drainage density with a rating of flash flood contribution

IV) ***Distance from the drainage network***: The expansion of a flash flood event depends on the distance of a region from the drainage network [11], and regions where they were located near to the drainage network, generally suffer flash flooding susceptibility higher than areas that are far away as the nearby and, therefore it was one of responsible factor for flash flood susceptibility. Accordingly considering the historical flash flood events happened in the study area within a distance of 150, 450,1000, 1500 and greater than 2000m from the drainage network were classified as very high, high, moderate, low and very low flash flood susceptibility, respectively in this study which is shown in the figure 3-8.

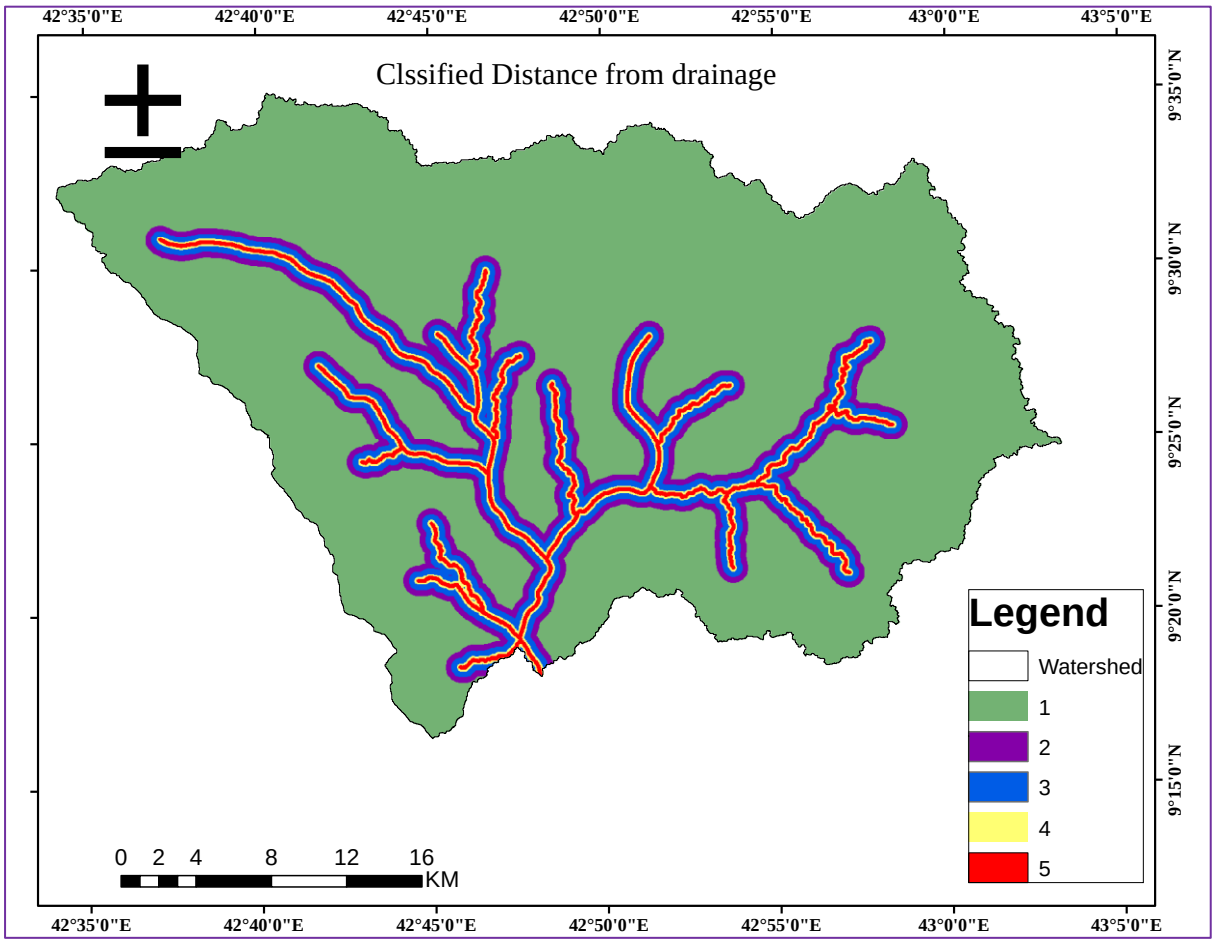


Figure 3-8: Classified distance from drainage

3.3.2.3 Morphometric hazard index

The study of the morphometric analysis of Jerer upstream watershed is mainly based on the tracing of the drainage network using the digital elevation model (DEM) with 30-m resolution. In this study, HEC-GeoHMS and ArcHydro tool were used to extract drainage networks and watersheds from SRTM elevation data. In this way, about 46 number of catchments were generated and the morphometric characteristics of each catchment was calculated as well. The streamlines that generated in this area shows the major known streams which are starting from the Upstream Northwest and Northeast sides and passing through Jigjiga city as well as in the side of the city. After that from the drainage networks and watersheds generated, the morphometric parameters were calculated based on formulas shown in table below. In the process morphometric parameters were calculated according to basin dimensions, basin shape, basin surface, and drainage network where the basin dimensions include area (km^2), perimeter (km), length (km) and width in (km). The basin

shape measured elongation ratio, circulation ratio, basin form factor and shape factor. The basin surface characteristics include relief ratio, relative relief ratio, ruggedness number, hypsometric index, and slope. The drainage network indices include total stream number, total stream length, texture ratio, drainage density, stream frequency, stream order, and bifurcation ratio.

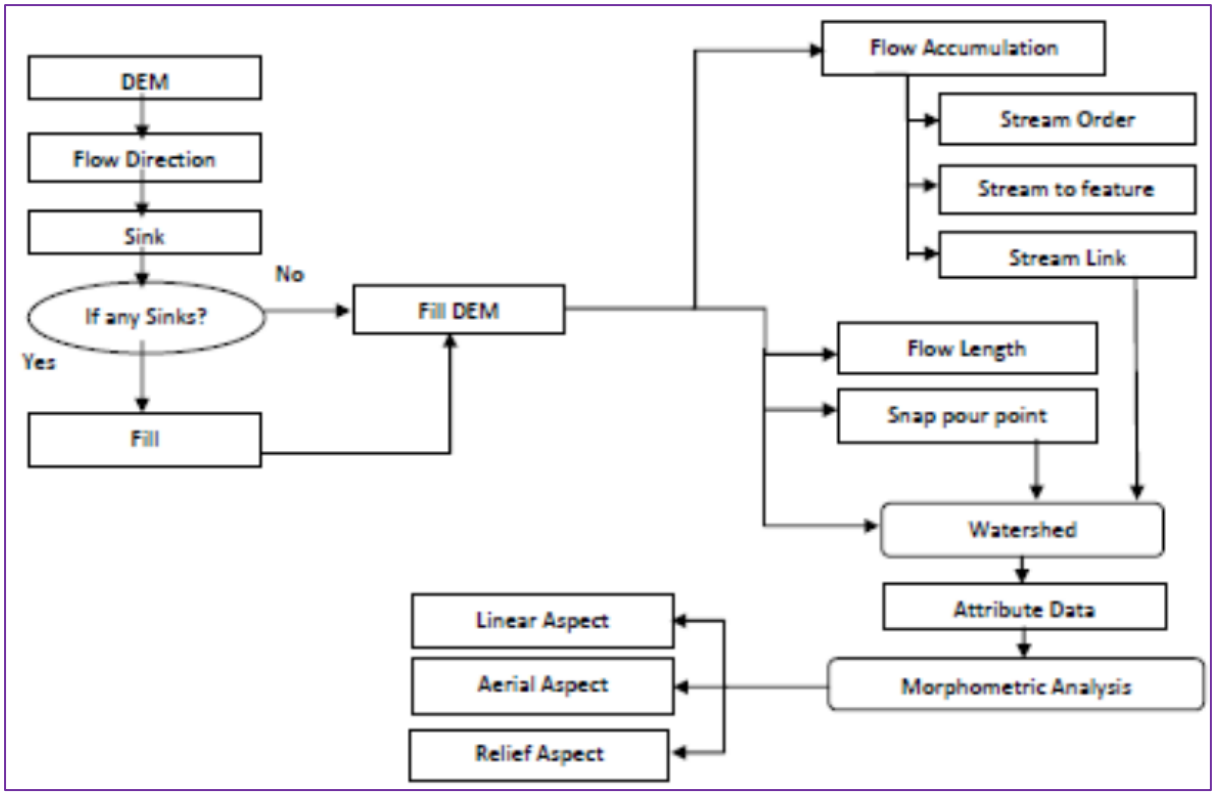


Figure 3-9: The general framework of Morphometric analysis in the study area

Table 3-3: Watershed Morphometric analysis formulas

Sl. No.	Morphometric Parameters	Formula	Reference
A	**Drainage Network**		
1	Stream Order	Hierarchical Rank	Strahler [1]
2	Total Stream order	Sum of Stream order	
3	Stream number (Nu)	Nu = N1+N2+ …….+Nn	Horton [53]
4	Stream length (Lu) (km)	Length of the stream	Strahler [1]
5	Stream length ratio (Lur)	Lur =Lu/(Lu-1)	Strahler [1]
6	Bifurcation ratio (Rb)	Rb= Nu/Nu+1	Strahler [1]

B	**Basin Geometry**		
7	Basin Perimeter (P)	GIS software analysis	Schumm [2]
8	Basin Length (Lb) (km)	GIS software analysis	Schumm [2]
9	Basin Area (km2) (A)	GIS software analysis	Schumm [2]
10	Form factor Ratio (Rf)	$Ff = A / Lb^2$	Horton [53]
11	Elongation Ratio (Re)	$Re= 2\sqrt{(A/\pi)}/L$	Schumm [2]
12	Shape Factor Ratio (Sf)	$Sf=Lb^2/A$	Horton [53]
13	Circularity Ration (Rcn)	Rcn = A / P	Strahler [1]
14	Relative Perimeter (Pr)	Pr = A / P	Schumm [2]
C	**Drainage Texture Analysis**		
15	Drainage Density (Dd)	Dd = Lu / A	Horton [53]
16	Stream Frequency (Fs)	Fs = Nu / A	Horton [53]
18	Length of overland flow (Lo)	Lo= 1/Dd×2	Horton [53]
D	**Relief Characterization**		
19	Maximum Basin Height (Z) (m)	GIS software analysis	
20	Minimum Basin Height (z) (m)	GIS software analysis	
21	Total Basin relief (H) (m)	H = Z - z	Strahler [1]
22	Hypsometric index (HI)	(Emean- Emin)/(Emax-Emin	Pike and Wilson [3]
23	Relief Ratio (Rhl)	Rhl = H / Lb	Schumm [2]
24	Relative Relief Ratio (Rhp)	Rhp = H * 100 / P	Melton [54]
25	Ruggedness Number (Rn)	Rn = Dd * (H / 1000)	Patton & Baker [3]
26	Melton Ruggedness Number (MRn)	MRn = H / A^0.5	Melton [54]

The study of Slater et al. [7] the regions located in the low lying flood plain are more prone towards flooding compared to the structural hilly regions. Accordingly, the study area in Jerer Upstream watershed is defined as well; because of its low land area was prone to flash flood in the last subsequent years.

After the calculation of each morphometric parameters in every 46 catchments in the study area and they were grouped into two classes based on the relations with flash flood. These parameters were considered in many studies [55], [6] grouped direct proportion with the flash flood susceptibility and inversely proportion with the flash flood susceptibility. Accordingly, in this study, these parameters were grouped into two as shown in table below. The ranking score for each morphometric parameter was assigned based on its relationship to the flash flood susceptibility using simple statistically method according to the following steps:

- Determine minimum (Xmin) and maximum (Xmax) value of each morphometric parameter for all sub-basins of the study area
- Normalize all the values of each direct proportion parameters (X) (Group I) based on the following equation.

  (X-Xmin)/ (Xmax-Xmin) (3-4)

- Normalize all the values of each inverse proportion parameters (X) (Group I) based on the following equation.

  (X-Xmax)/ (Xmax-Xmin) (3-5)

Finally, sum the rank for the morphometric parameters to calculate the susceptibility value for each sub-basin. All the results are found in the appendix.

Table 3-4: Groups of morphometric analysis based on flash flood contribution

Morphometric analysis groups	
Group I	**Analysis**
Basin Area (km2) (A)	GIS software analysis
Form factor Ratio (Rf)	$Ff = A / Lb^2$
Circularity Ration (Rcn)	Rcn = A / P
Drainage Density (Dd)	Dd = Lu / A
Stream Frequency (Fs)	Fs = Nu / A
Relief Ratio (Rhl)	Rhl = H / Lb
Relative Relief Ratio (Rhp)	Rhp = H * 100 / P
Ruggedness Number (Rn)	Rn = Dd * (H / 1000)
Group II	
Bifurcation ratio (Rb)	Rb= Nu/Nu+1

Elongation Ratio (Re)	$Re= 2\sqrt{(A/\pi)}/L$
Shape Factor Ratio (Sf)	$Sf=Lb^2/A$
Hypsometric index (HI)	(Emean-Emin)/(Emax-Emin)

As shown in figure below the morphometric hazard index based on flash flood was high hazard group sub-basins (1,6,5,3,8) in the upstream Northwest (Chinaksen town), Northeast (Lefeissa town) and in West (19-21,30,41,43 and 46). This means that there is no chance for recharging the groundwater and infiltration while surface runoff in these sub-basins produced and the flash flood directly comes to downstream of the watershed (Jigjiga City).

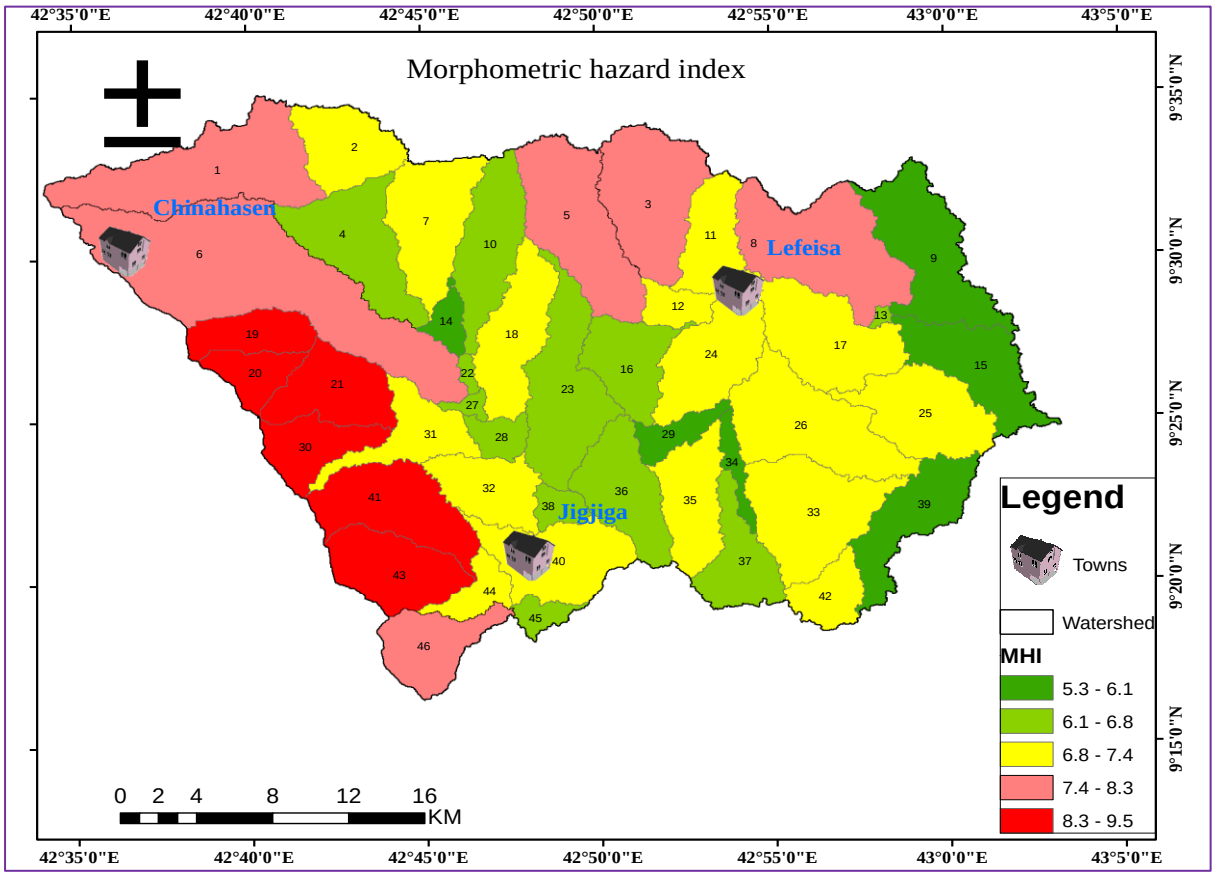


Figure 3-10: Morphometric hazard in the study area

3.3.2.4 Geology and Soil related factors

I) Geology related foctors: Lithology is considered one of the most influential factors in flash flood susceptibility mapping because of its influence on the geo-mechanical characteristics of a terrain [30]. In this study, the geology was obtained from Ethiopian Geological Survey and then the lithology was derived. Moreover, all the discriscribtins and

their symbol expanations were obtained from Federal Democratic Republic of Ethiopia Ministry of Mines and Energy Procedures and Guidelines Forwork Processes (Part I) and Content and Digital Attributes for Products (Part II) documents (2007). Accordingly, the following table shows symbol, discription and the age of each geology in the study area.

Table 3-5: Geological classes and their descriptions

Name	Lithology_Name	Age
Jh2	Evaporite with dolostone & dolomitized limestone	MESOZOIC-Jurassic
Qa	Alluvial and lacustrine deposits	CENOZOIC-Quaternary
Ja	Shale, marl, and limestone	MESOZOIC-Jurassic
Pc2	Sandstone, medium to coarse and contains feldspar	PALEOZOIC-Ordovician
PNa1	Sandstone, shale, conglomerate, and tillite.	PALEOZOIC-Permian

The Figure below shows the geology of the study area divided into five geological classes, which range from MESOZOIC to PALEOZOIC rocks.

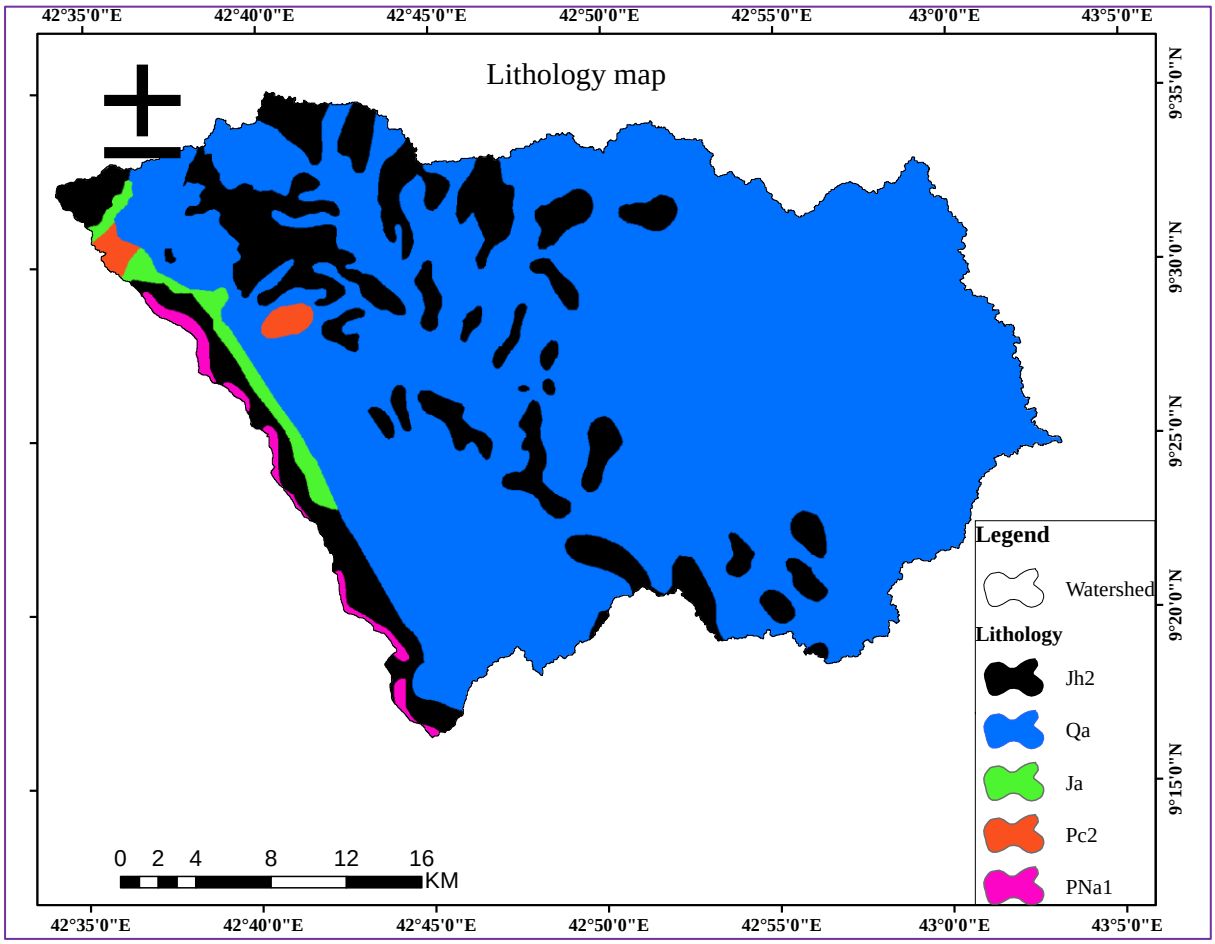


Figure 3-11: Geological classes of the study area

II) ***Soil types:*** Soil type and texture are very important factors in determining the water holding and infiltration characteristics of an area and consequently affect flood susceptibility [56], where infiltration of water mainly depends on soil texture and soil texture is a major factor which has to be fulfilled before the generation of surface runoff. As a general rule, runoff from intense rainfall is likely to be more rapid and greater with clay soils than with sand; additionally in the study area clay has more coverage with a low sand content and therefore of high susceptibility zones.

The study area has four different soil types as shown in the figure below. In the study area about 95% of the area is covered with the clay soil type which has less infiltration and percolation capacity, hence the surface runoff is produced easily when the raindrops. The rest other small coverage Loam, sandy loam, and loamy sand soils cover the Northwest, west, and south of the study area (5%) as shown in the figure below.

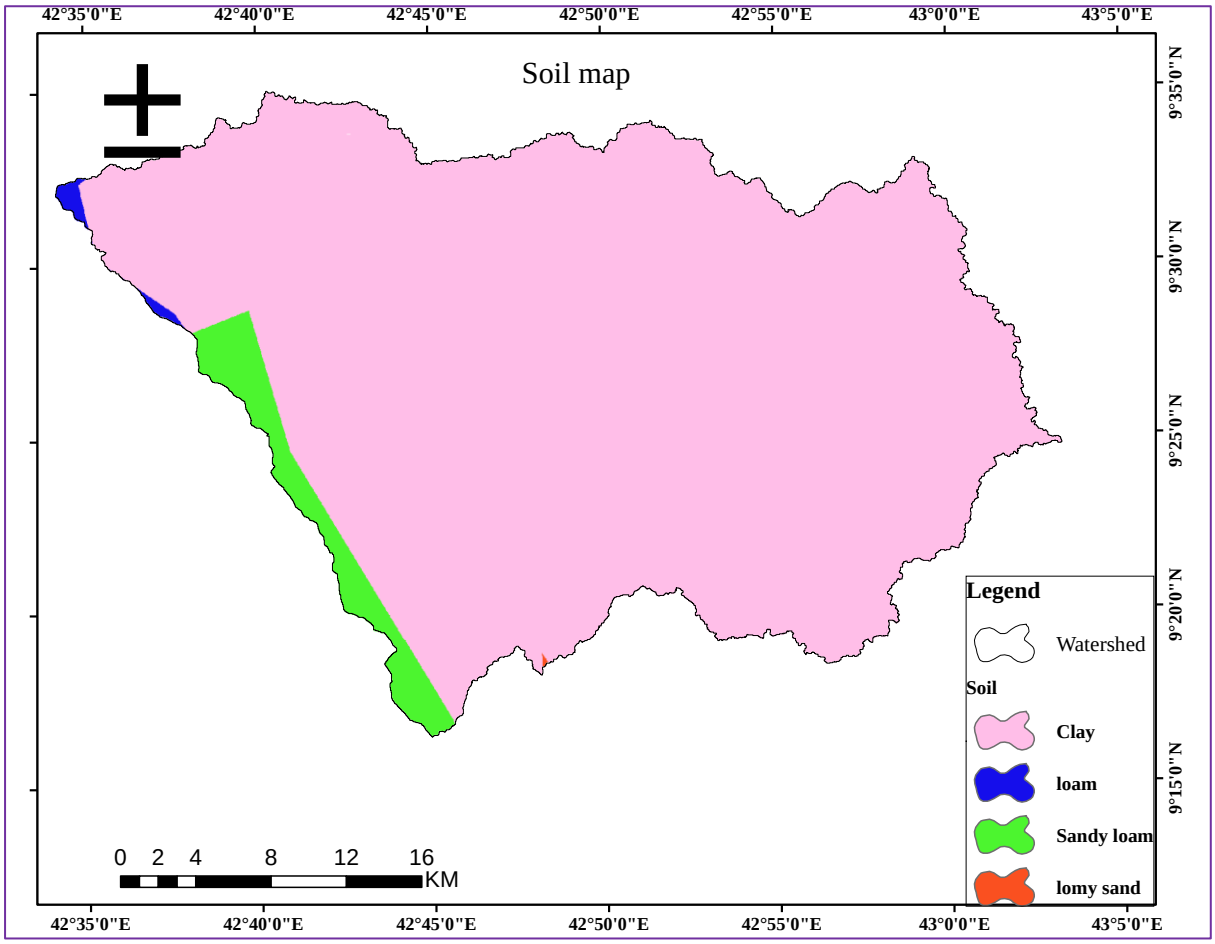


Figure 3-12: Soil types

According to the reports [57], East Jijiga Plains are covered by vertisols, also known as black cotton soils. Vertisols are normally black or dark grey soils with very high clay content. Because of the heavy soil texture and presence of expanding clay minerals the soils' range between moisture stress and water excess is very narrow. The soils are sticky when wet and crack when dry. Due to swelling the infiltration capacity is extremely low, resulting in high runoff rates. Vertisols are typically low in organic matter, have a medium moisture storage capacity, have a poor drainage capacity and are very prone to erosion. Accordingly, these soil types were classified based on the flood susceptibility, with clay classified as 5 value and sandy loam classified as value 1.

3.4 Rainfall data analysis

3.4.1 Rain gauge data

The study area has got limited record of rainfall data. Ground based rain gauge recorded rainfall data were used for validation of the satellite products comprised of 4 stations located within and surrounding the basin. These data were provided by the Ethiopian National Meteorological Service Agency (NMA). The datasets cover daily data for the period from 2008 to 2018. Although the number of stations is relatively good, their distribution over the basin is not uniform. Most of the gauges are located in easily accessible areas and the distribution of gauges in the lowland areas were sparse. Most rainfall stations relatively located in the highland areas where the spatial variability of rainfall is very high. Quality control of rainfall data from each station was done. After data screening using Thessein polygon, 11 stations out of the 4 in the study area were found to be enough for contributing to the area.

3.4.2 Acquisition and analysis of TRMM rainfall data for the study area

In this study the short-term rainfall in the study area is obtained from the Tropical Rainfall Measuring Mission (TRMM) data for the period 2008–2018 and its daily data was prepared using the data obtained from the NASA's Tropical Rainfall Measuring Mission (TRMM). The TRMM rainfall amounts are calibrated with respect to the rain-gauge data recorded at 4 stations across the study area. The data has a temporal resolution of one day and corresponding to an area-averaged over 0.25 x 0.25 degrees latitude-longitude grid. The program written in R studio derived the needed data corresponding to the grid where the each station in the study area lies.

These data were then spatially interpolated using Inverse Distance Weighted (IDW) method to obtain the rainfall distribution map [2]. The inverse distance weighting (IDW) method integrated with GIS is used to estimate the rainfall distribution in the study area [2]. The IDW method involves the process of assigning values to unknown points by using values from a scattered set of nearest known points.

Generally, TRMM data in this study was acquired covering the last 11 years (2008–2018) and downloadedfromhttps://disc2.gesdisc.eosdis.nasa.gov/thredds/catalog/aggregation/catalog.html. Next the acquired data was extracted in Rstudio statistical tool using the selected stations in the study area and analyzed to utilize it as a flood susceptibility factor. Evaluation and

comparison of the Tropical Rainfall Measuring Mission (TRMM) based on observed data was done. The spatial distribution of rainfall data measurements (i.e Rain gauges) are not uniformly distributed through the country Ethiopia; moreover, in the eastern parts, the spatial distributions of these stations are scarce. Therefore, the required regular measured rainfall data in these areas for flash flood study is the most challenging part. Generally, in this study efforts have been made to utilize TRMM rainfall data as one parameter in flash flood susceptibility conditioning factor in semi-arid eastern Ethiopia region which is Upstream Jerer watershed. This work is conducted through calibrating the TRMM rainfall with reference to the observed data, for utilization in application-oriented tasks in the study. In the study, Daily rainfall data collected National meteorological agency (NMA) at 4 stations which have not 100% missing values in the study area and the data for these stations collected between 2008 and 2018. The rain-gauge station names, together with latitude, longitude, altitude (m) are presented in Table below. Moreover, the distributions and rain-gauge networks over the study area are displayed as shown below where the green colors of stations represent the locations of the rain gauges which contribute to the study area were selected based on the thiessen polygon derived in ArcGIS as shown in figure below.

Table 3-6: Observed stations location

No	Stations_Name	Lon	Lat	Altitude
1	Chinaksen	42.561	9.5326	2233.56
2	Life Isa	42.8948	9.4873	1802
3	Hadew	42.6523	9.3563	1697.85
4	Jijiga	42.7885	9.3457	1682.32

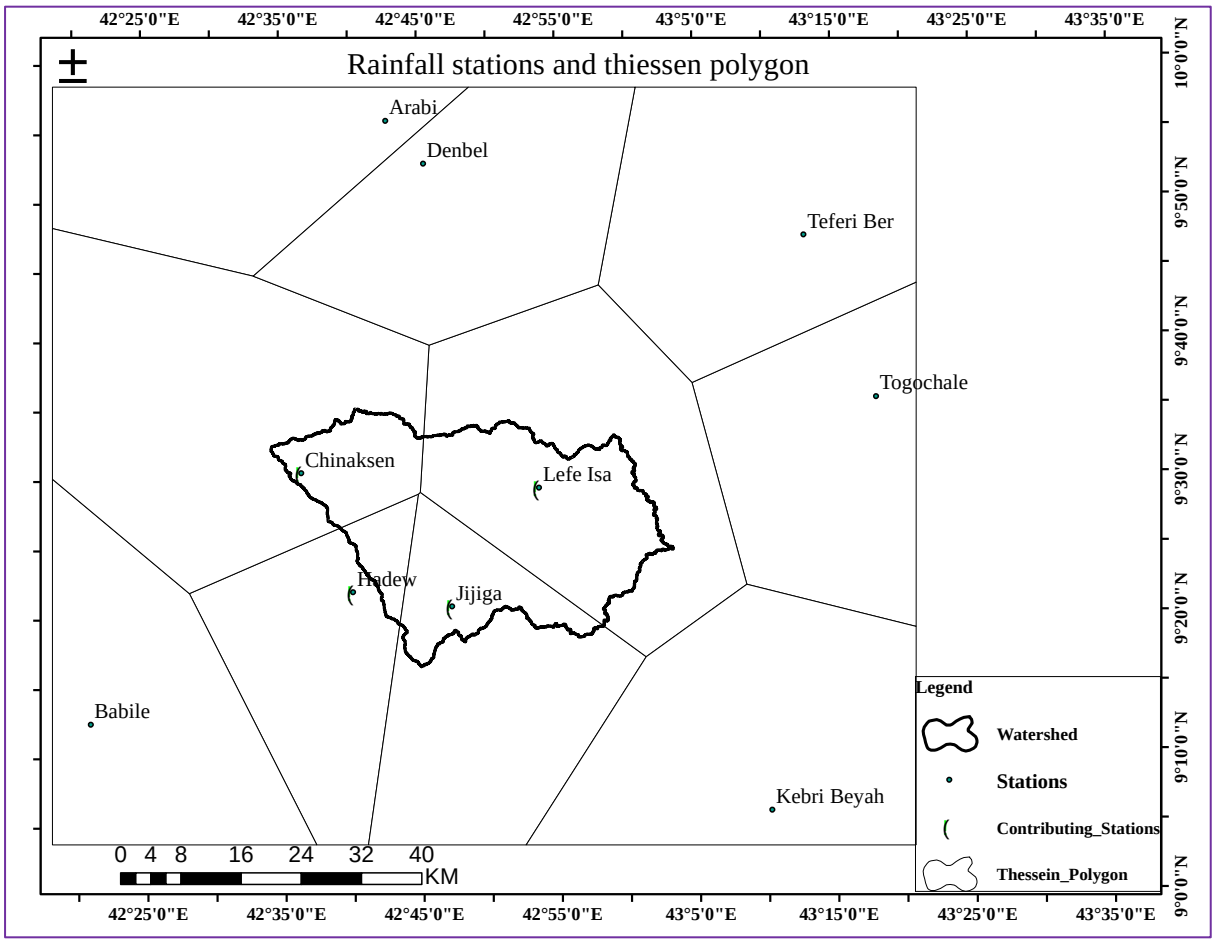


Figure 3-13: Spatial locations and Thiessen polygon of stations

According to Prudhomme and Reed [58], interpolation methods described in the literature, ordinary Kriging and modified residual Kriging are attractive because of their simplicity and ease of use. However, there are only 4 stations over the varied topographic variations of Jerer upstream watershed and so none of them is useful in gridding rain-gauge rainfall. Therefore, rainfall data obtained from the TRMM near the closest point to a station's location in a 25 km grid box were compared with the data from that same observed rain gauge. The rain gauge situated in a grid box represents the observed amount corresponding to the TRMM measured rainfall in that grid.

A comparison of the TRMM day-to-day rainfall with the rain-gauge data for each observation site is then carried out (i.e pixcel to point approach) [59]. Following this, the monthly, seasonal and annual rainfall data obtained were compared. There are three categorical variables that indicate the performances of Seasonal rainfall distribution over the

study area. The rainfall distribution data for the winter, spring, summer, autumn, dry and wet seasons for each station were displayed in Figures 3-17-3-18.

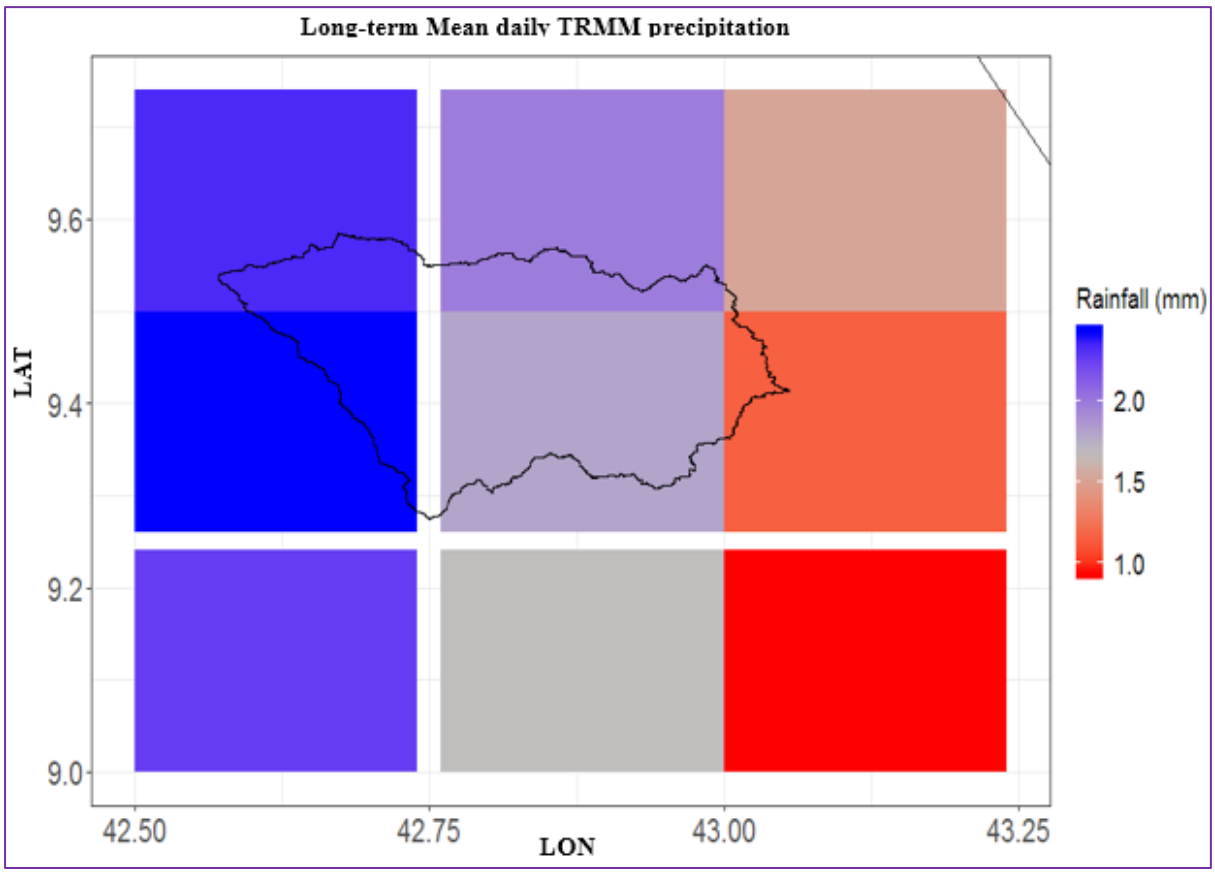


Figure 3-14: Spatial Distribution of TRMM rainfall in the Watershed

In the TRMM data bias correction, there are many bias correction methods like linear scaling, local intensity scaling, (3) power transformation, variance scaling, distribution transfer and the delta-change approach [8].The linear-scaling (LS) approach [5], [6], [7], [8] was based on monthly correction factor, which was the ratio between long-term monthly mean data for ground observation data from stations and TRMM_3B42-V7. By definition, corrected TRMM data will perfectly agree in their monthly mean values with the observations. Accordingly, the following equations were used to correct bias in this study.

$$\mathrm{CFn} = \overline{\mathrm{OBSn}}/\overline{\mathrm{TRMMn}} \quad (3\text{-}6)$$

$$TRMM\ {}_{i,n}^{corrected} = \mathrm{CFn} * \mathrm{TRMM\ i,n} \quad (3\text{-}7)$$

Where CFn is the monthly mean change factor at month m, OBSn and TRMMn represent the mean of ground observation and TRMM data at month n, respectively while $TRMM\ {}_{i,n}^{corrected}$

and TRMM i,n, are the corrected and to be corrected using the correction factor (CFN) respectively. Generally, the observed, TRMM raw, and TRMM corrected are displayed as follows based on the above equations.

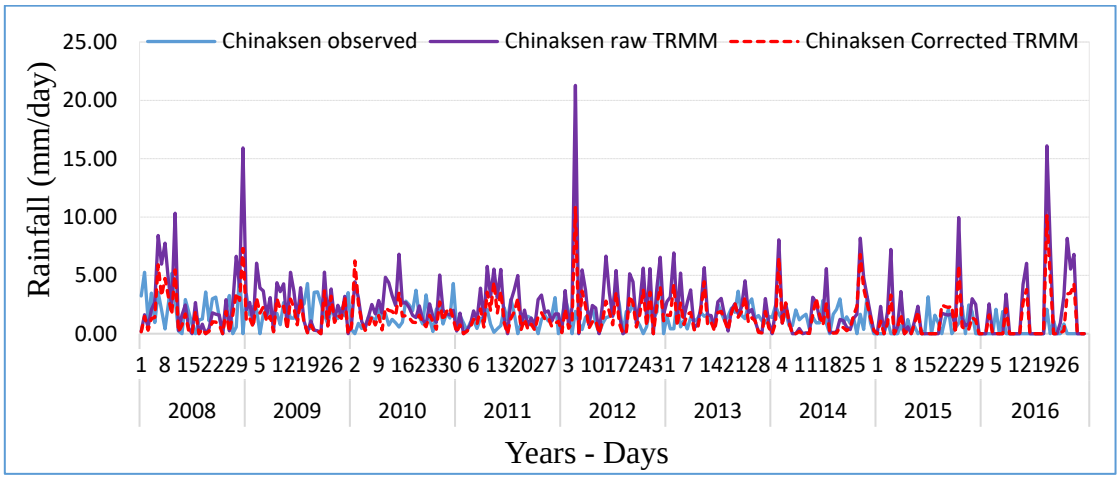


Figure 3-15: Averaged day-to-day comparison of rainfall (mm/day) obtained from the TRMM and observed data during 2008–2016 for Chinksen station

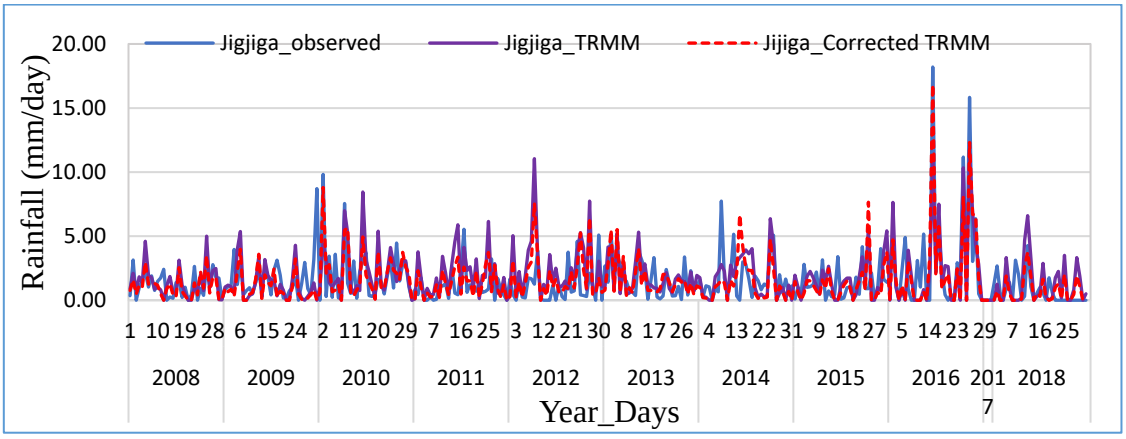


Figure 3-16: Averaged day-to-day comparison of rainfall (mm/day) obtained from the TRMM and observed data during 2008–2018 for Jigjiga station

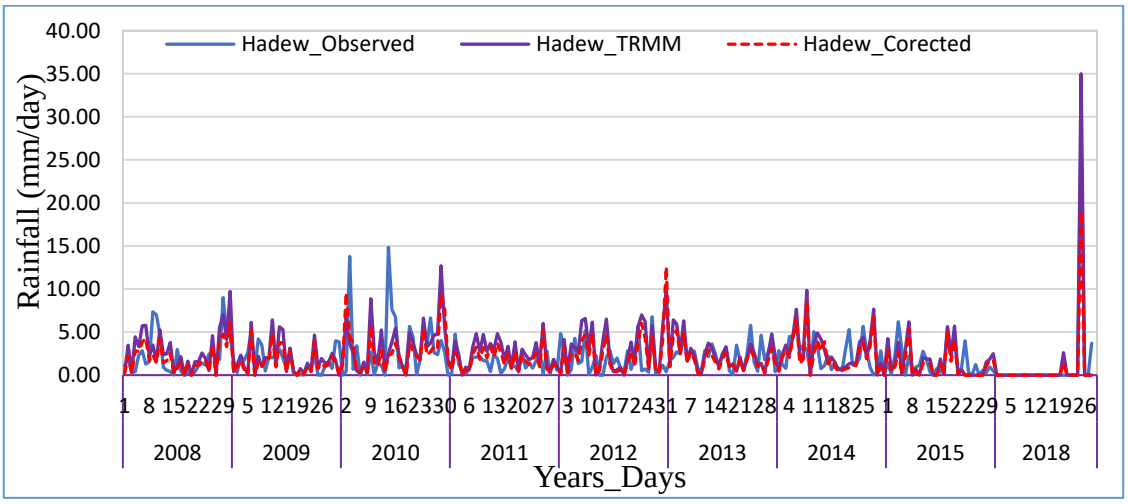


Figure 3-17: Averaged day-to-day comparison of rainfall (mm/day) obtained from the TRMM and observed data during 2008–2018 for Hadew station

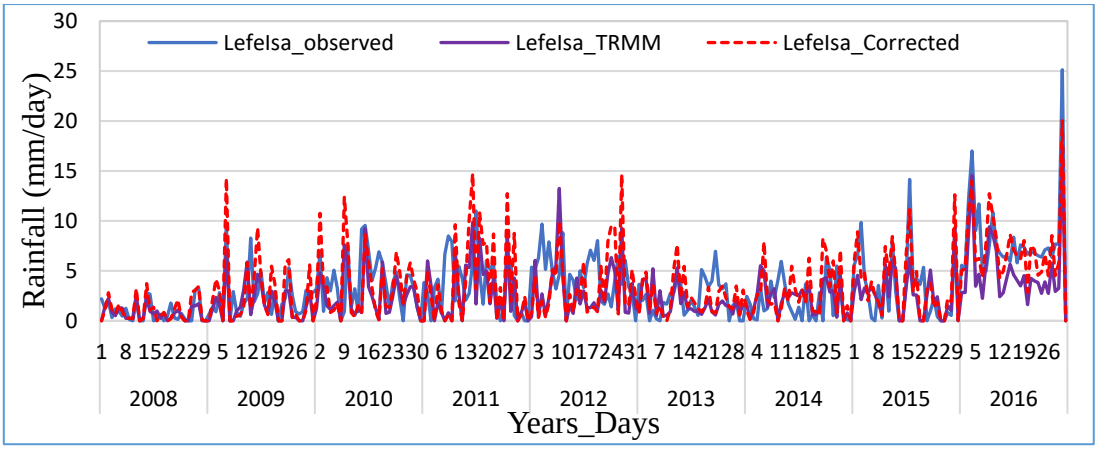


Figure 3-18: Averaged day-to-day comparison of rainfall (mm/day) obtained from the TRMM and observed data during 2008–2018 for Lefeisa station

3.4.3 Performance evaluation and Comparison TRMM data with Gauge Rainfall Data

The satellite TRMM product and observed stations time series were compared to analyze their variation and relationship using various statistical performance measures: Root Mean Squared Error (RMSE), Correlation Coefficient (CC), Nash-Sutcliffe Efficiency (NSE), and Percent Bias (PBIAS). The statistical performance measures for evaluating the TRMM rainfall data are described in the following table [60]–[62].

Table 3-7: Statistical performance measures for evaluating the TRMM rainfall data

Methods	Explanation	Equation
Root Mean Square Error (RMSE	The square root of the average of the difference between the observed value and median of the forecast.	$RMSE = \sqrt{\frac{1}{N}} \sum_{i=1}^{n} (Si - Oi)$
Nash Sutcliffe Efficiency (NSE)	It indicates how well the plot of observed vs forecast values fits the 1:1 line. NSE range from -Inf to 1. The closer to 1, the more accurate the model is	$NSE = \frac{\sum_{i=1}^{n} (Si-Oi)^2}{\sum_{i=1}^{n} \bar{O}}$ *100
Percentage Bias (PBIAS)	A measure of the average tendency of the forecast values to be larger or smaller than their observed ones	$SE = 1 - \frac{\sum_{i=1}^{n} (Si - Oi)^2}{\sum_{i=1}^{n} (Si - \bar{O})^2}$
The correlation coefficient (CC)	A measure of the strength and direction of the linear relationship between two variables. The correlation coefficient may take any value between -1.0 and +1.0	$C = \frac{\sum_{i}^{n} (Si-\bar{S}) * (Oi-\bar{O})}{\sqrt{\sum_{i=1}^{n} (Si-\bar{S}) * \sum_{i=1}^{n} (Oi-\bar{O})^2}}$

Where S_i is the estimated values, O_i is the observed values and N is the number of samples

The CC was used to evaluate the degree of agreement between TRMM and gauge data, with values ranging from -1 to 1. Positive CC values indicate a positive correlation, while negative values showed a negative correlation. The RMSE was used to represent the average error magnitude. Accordingly, the Statistical performance measures are calculated for the raw and corrected satellite TRMM rainfall data in the following tables. In the calculation, the average monthly and annual scale were used respectively.

Table 3-8: The statistical performance measure scores calculated based on an average monthly and annual TRMM using the observed rain gauge measurement.

Station Name	Data	Monthly				Annual			
		CC	RMSE (mm/d)	NSE	PBIAS (%)	CC	RMSE (mm/d)	NSE	PBIAS (%)
Jigjiga	TRMM raw	0.63	1.65	-0.2	26.2	0.77	0.58	0.43	20.5
	TRMM corrected	0.78	0.99	0.57	0.2	0.9	0.34	0.8	-3.1
Chinaksen	TRMM raw	0.63	1.99	-0.54	58.7	0.53	1.08	-2.98	71.6
	TRMM corrected	0.69	1.21	0.43	0.75	0.72	0.42	0.75	7.6
Hadew	TRMM raw	0.72	1.84	0.02	28.2	0.76	0.67	-0.36	25.8
	TRMM corrected	0.76	1.3	0.53	-0.1	0.82	0.37	0.59	-3.4
Lefeissa	TRMM raw	0.7	3.28	0.44	-27.7	0.97	1.24	0.59	-25.7
	TRMM corrected	0.85	2.3	0.72	-4.8	0.97	0.54	0.97	0.2

3.4.4 Seasonal rainfall distribution and flash flood relationship assessment in the study area

The figures shown below indicate that the average seasonal rainfall distributions at stations in the study area in which most of these stations indicate the highest seasonal rainfall distributions between spring and summer. According to the reports of the Federal Democratic Republic of Ethiopia National Disaster Risk Management Commission, Early Warning and Emergency Response Directorate 2016, and 2018 floods events were occurring in the study area especially in the spring to the early beginning of summer. Moreover, the following

average seasonal rainfall results indicate the high seasonal rainfall in that spring and summer seasons.

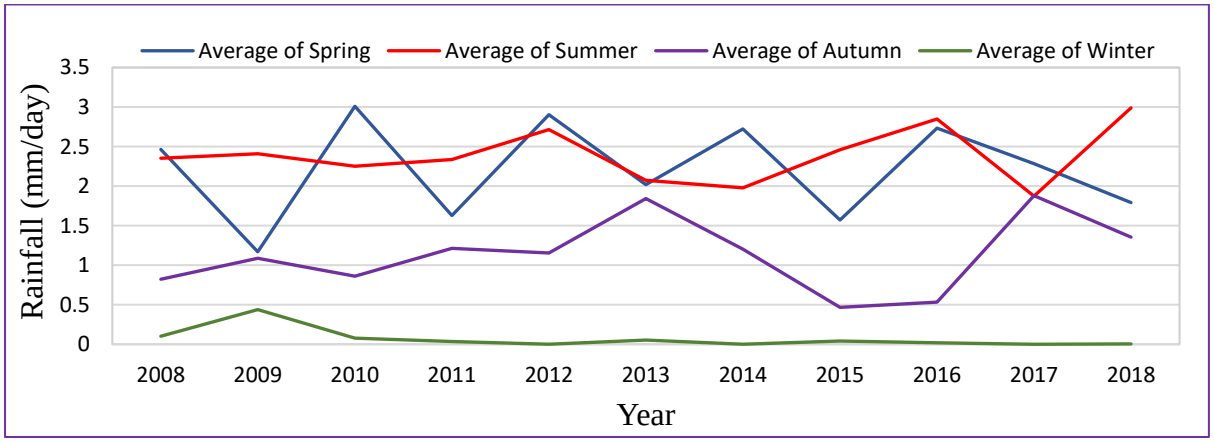


Figure 3-19: Average seasonal rainfall distribution of Chinksen station

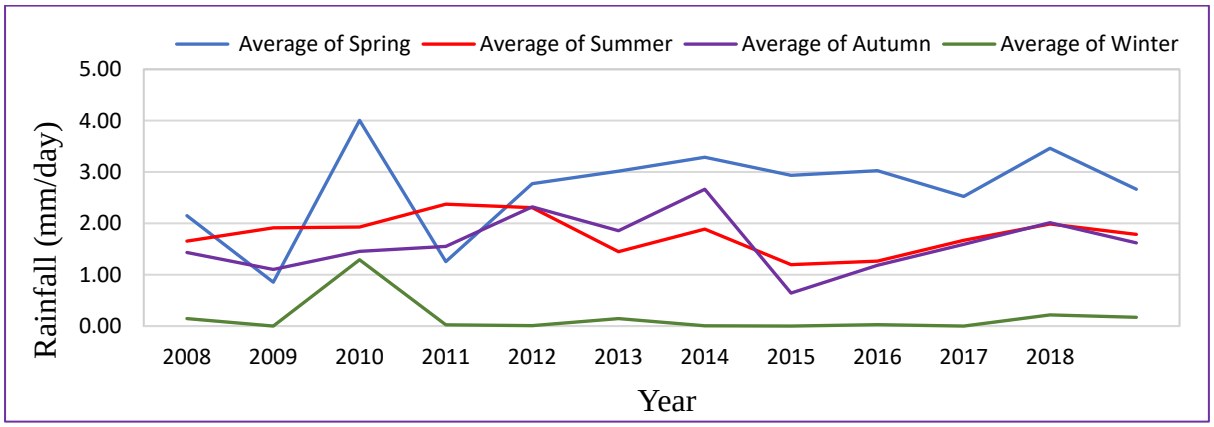


Figure 3-20: Average seasonal rainfall distribution of Jigjiga station

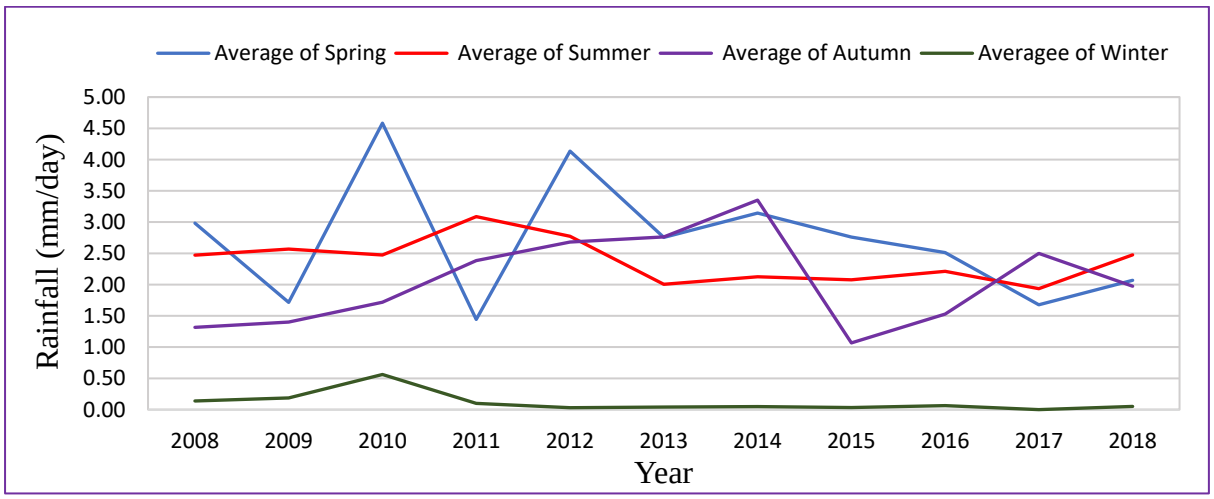


Figure 3-21: Average seasonal rainfall distribution of Hadew station

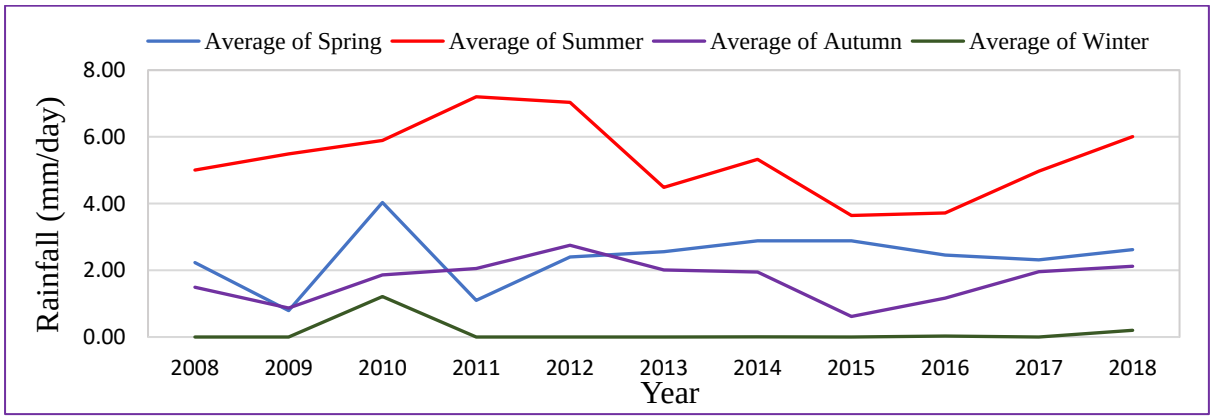


Figure 3-22: Average seasonal rainfall distribution of Lefeisa station

Floods are associated with extremes in rainfall, any water that cannot immediately seep into the ground flows downslope as runoff. The amount of runoff is related to the amount of rain the area experiences although many factors restrict it (i.e LULC, NDVI, Soil and etc.). The level of water in rivers, lakes and drainage channels rises due to heavy rainfalls. When the level of water rises above the river and drainage banks, the water starts overflowing, hence causing flood hazards around.

In the area it was observed that while the local rainfall is relevant for flooding, rainfall amounts on the upstream catchments in Chinaksen and Lefeissa contribute to flash flood susceptibility and risk in the downstream (Jigjiga). Therefore, in this study both the local and upstream stations rainfall (Jigjiga, Lefeissa, Hadew, and Chinaksen) were integrated into the analysis, for Spatial Flash flood Susceptibility prediction. Mean annual rainfall for eleven years (2008–2018) were considered and interpolated using Inverse Distance Weighting (IDW) to create a continuous raster rainfall data within and around the study area..

As shown in the Figure below the long-year mean rainfall pattern indicated that there is relatively high precipitation in the Northeast highlands, while there is low rainfall in the Northwest, southern lowlands of the study area.

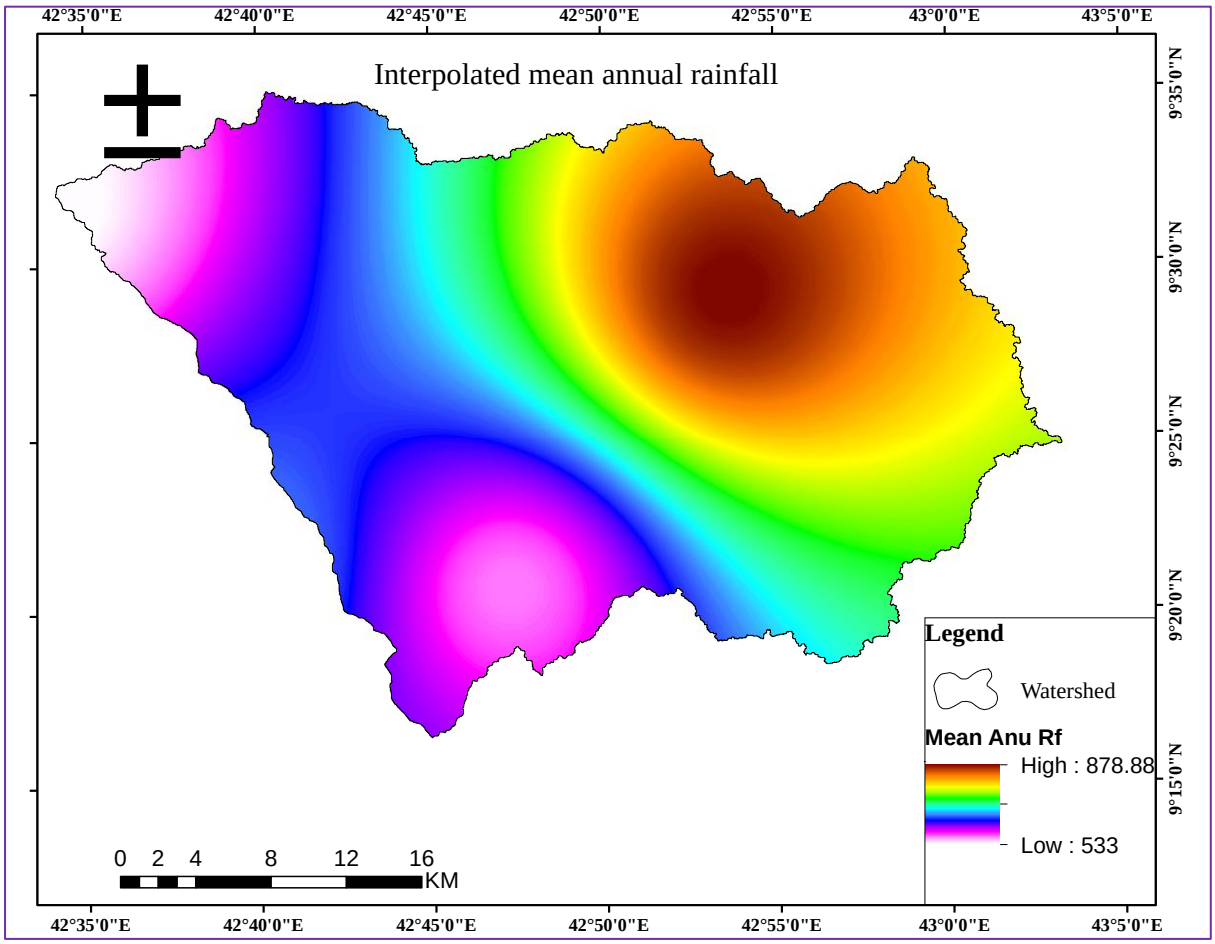


Figure 3-23: Interpolated mean annual rainfall distribution

3.5 Land use land cover data aquisition and analysis

3.5.1 NDVI

The NDVI can be used to reflect the vegetation coverage of the study area. It is used as one of the conditioning factors in flash flood susceptibility study. If the normalized vegetation index is less than zero, it means that the ground is covered with water or snow. If the normalized vegetation index is equal to zero, it means that there is bare land or rock. If the normalized vegetation index is greater than zero, it indicates that there is vegetation cover and the greater the value, the higher vegetation coverage [63].

The Landsat 8 image downloaded from (https://earthexplorer.usgs.gov/) was used to generate the NDVI map in ERDAS 1014. The Normalized Difference Vegetation Index (NDVI) is calculated from the measured radiance on the red and near-infrared part of the electromagnetic spectrum. The index is based on the fact that highly vegetated areas will have high reflectance on the near-infrared and low reflectance on the red, while water and

bare soil will generally have higher reflectance on the red than in the near-infrared part of the electromagnetic spectrum [64]. This opposite behavior of the red and near-infrared (NIR) band leads to maximum separable in classification. Each object has its own unique reflectance characteristics in each band however in these two bands (NIR& Red) there is maximumly separable among objects and hence having many advantages rather using other bands for the purpose of land use mapping and in this context using NDVI is beneficial as it ignores some mathematically undefined results, shadow effects in hilly terrain etc. NDVI ranges from -1 to +1. Generally, NDVI is calculated based on the following formula using Landsat 8 Band 5 and Band 4.

NDVI = (NIR –Red)/ (NIR+Red) = (landsat8 Band 5 – lansat8 Band 4)/ (landsat8 Band 5+lansat8 Band 4) (3-8)

Type of Land Cover	NDVI (scale from -1 to 1)
Thick Vegetation	0.500<=NDVI<=1
Medium Vegetation	0.140<=NDVI<0.500
Scarce Vegetation	0.090<=NDVI<0.140
Bare ground	0.025<=NDVI<0.090
Clouds	0.002<=NDVI<0.025
Ice and snow	-0.046<=NDVI<0.002
Water	-1<=NDVI<-0.046

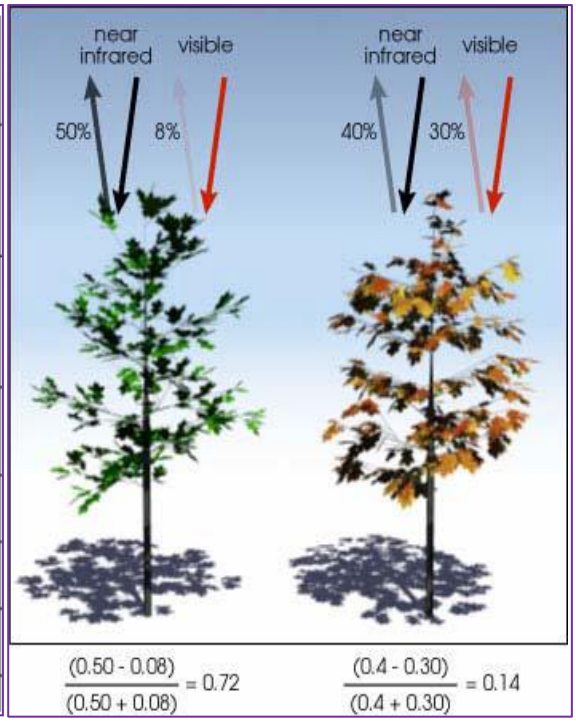


Figure 3-24: General classifications of NDVI

According to the above descriptions, the normalized vegetation index in the study area was generated and has the values between low: -0.19 and high: 0.49 and these values are classified in regarding the flash flood susceptibility. Wherever the high value of NDVI indicates the healthy vegetation coverage in the study area, the spatial flash flood distribution is very low.

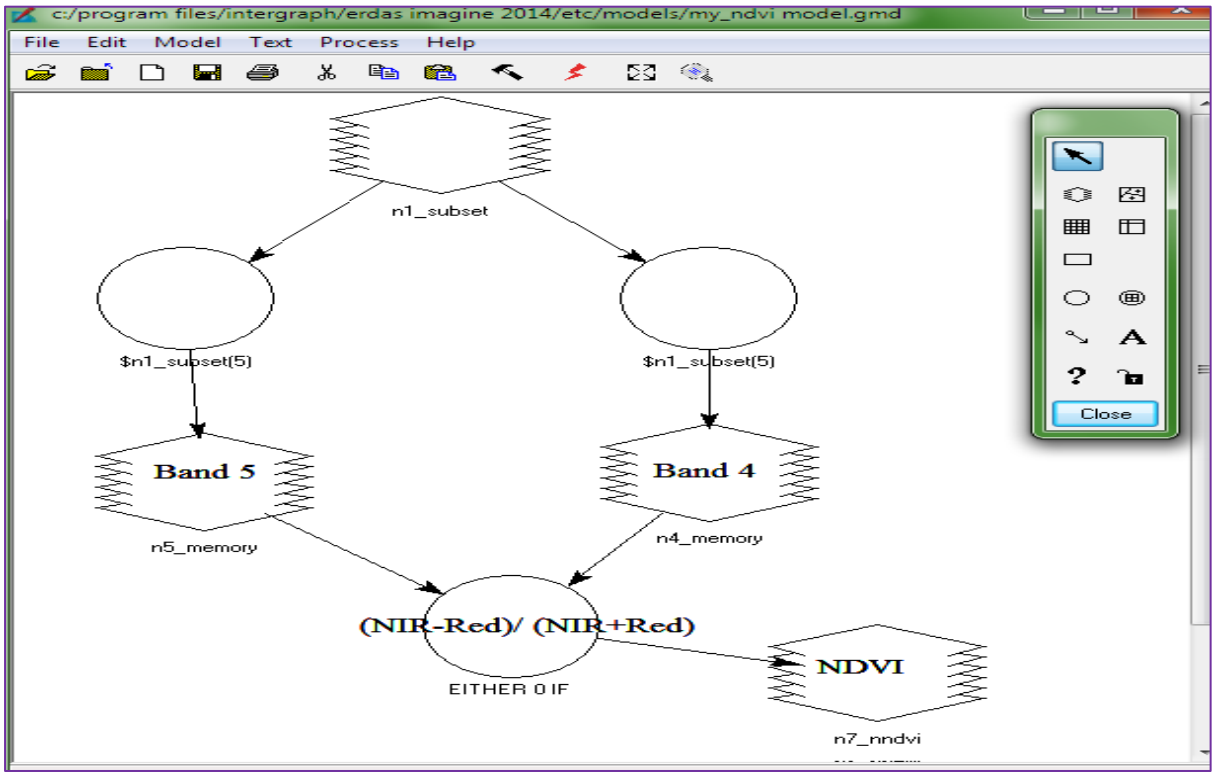


Figure 3-25: General framework of NDVI extraction in ERDAS 2014

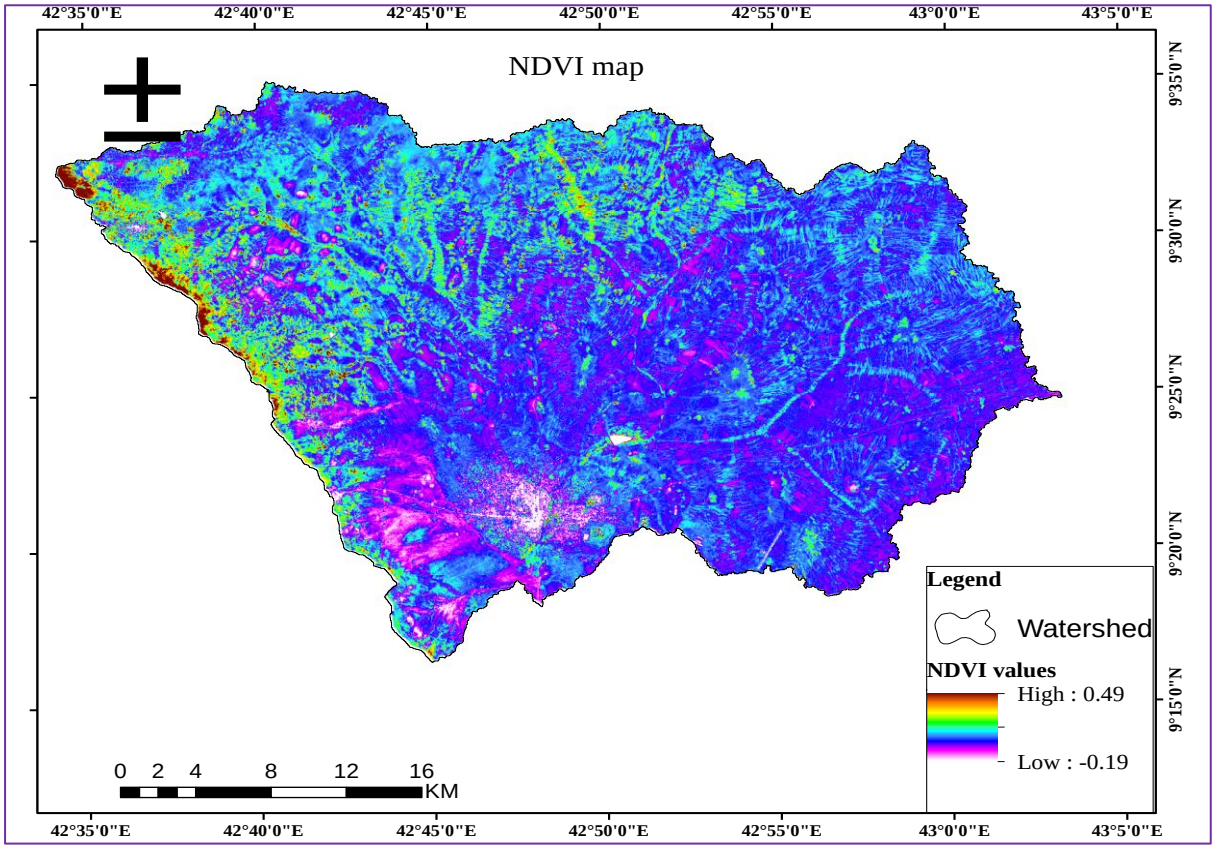


Figure 3-26: Computed NDVI map

3.5.2 Sattelite data download and image preprocessing

Satellite LANDSAT 8 OLI/TIRS images 2016 and 2019 were downloaded from https://earthexplorer.usgs.gov/ which used for image classification. The Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) are instruments on board the Landsat 8 satellite, which was launched in February of 2013. The satellite collects images of the Earth with a 16-day repeat cycle, referenced to the Worldwide Reference System-2. The satellite's acquisitions are in an 8-day offset to Landsat 7 (see Landsat Aquisition).

Table 3-9: Details of Landsat 8 OLI/TIS used for classification

Sattelite	Sensor_ID	Path/Row	Layers	Date of Aquisition	Grid cell size
Landsat 8 OLI/TIS	LC08_L1TP_166054_20160115_T1	166/53	11	15/01/2016	30 m and band 8 is 15m
	LC08_L1TP_166054_20190627_T1	166/53	11	27/06/2019	30 m and band 8 is 15m

3.5.3 Image preprocessing

Satellite Image preprocessing also called image restoration involves the correction of distortion, degradation, and noise introduced during the image process. Therefore, in order to enhance the nature of some distortion in Satellite images like its noise, haze, and stripes, image pre-processing activities were done before processing the data. The process basically produces a corrected image that is as close as possible, both geometrically and radiometrically to the radiant energy characteristics of the original scene. Preprocessing includes importing, layer stacking and subsetting of the image based on the boundary of in the study area (i.e. Jerer upstream watershed-Jigjiga), geometric correction, radiometric correction, and removal of stripes, pan sharpening, and other image enhancement techniques have been done in ERDAS Imagine 2014. Radiometric correction is removal of atmospheric noise to make more representatives of the ground truth conditions based on the sensors. These all previously mentioned activities done were to improve visible interpretability of an image by increasing the apparent distinction between the features in the scene.

In addition, the image was also the georeferenced-using boundary of the Woreda. During the georeferencing and reprojecting process, Adindan UTM Zone 37N coordinate system was

followed for raster and vector data in the study to maintain uniformity. Adindan UTM Zone 37N is a local datum that Ethiopia used.

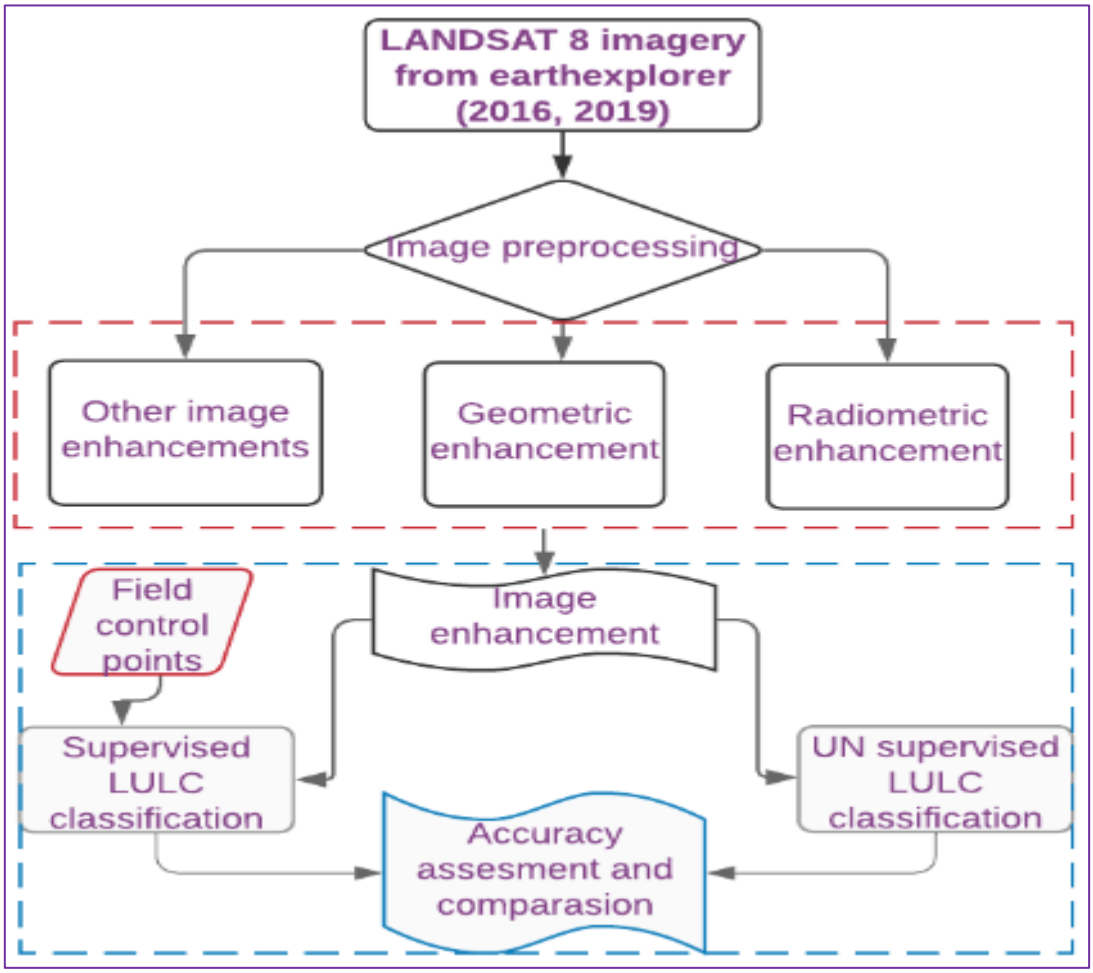


Figure 3-27: General LULC generation framework

3.5.4 Unsupervised LULC classification

Unsupervised classification methods do not require prior knowledge of land cover types before classification and the interpreter is responsible for assigning a class to each cluster of pixels [65]. In this classification, the outcomes (groupings of pixels with common characteristics) are based on the software analysis of an image without the user providing sample classes. The computer uses techniques to determine which pixels are related and groups them into classes. The user can specify which algorism the software will use and the desired number of output classes but otherwise does not aid in the classification process. However, the user must have knowledge of the area being classified when the groupings of pixels with common characteristics produced by the computer have to be related to actual

features on the ground. In this study the unsupervised classification was done to classify the land use land cover of the study area so that the six classes were identified as shown in the figure below, which are Built-up/ Settlement, Agricultural/ Fallow land, Vegetation/ shrubland, Waterbody, open area, bare lands, and open areas.

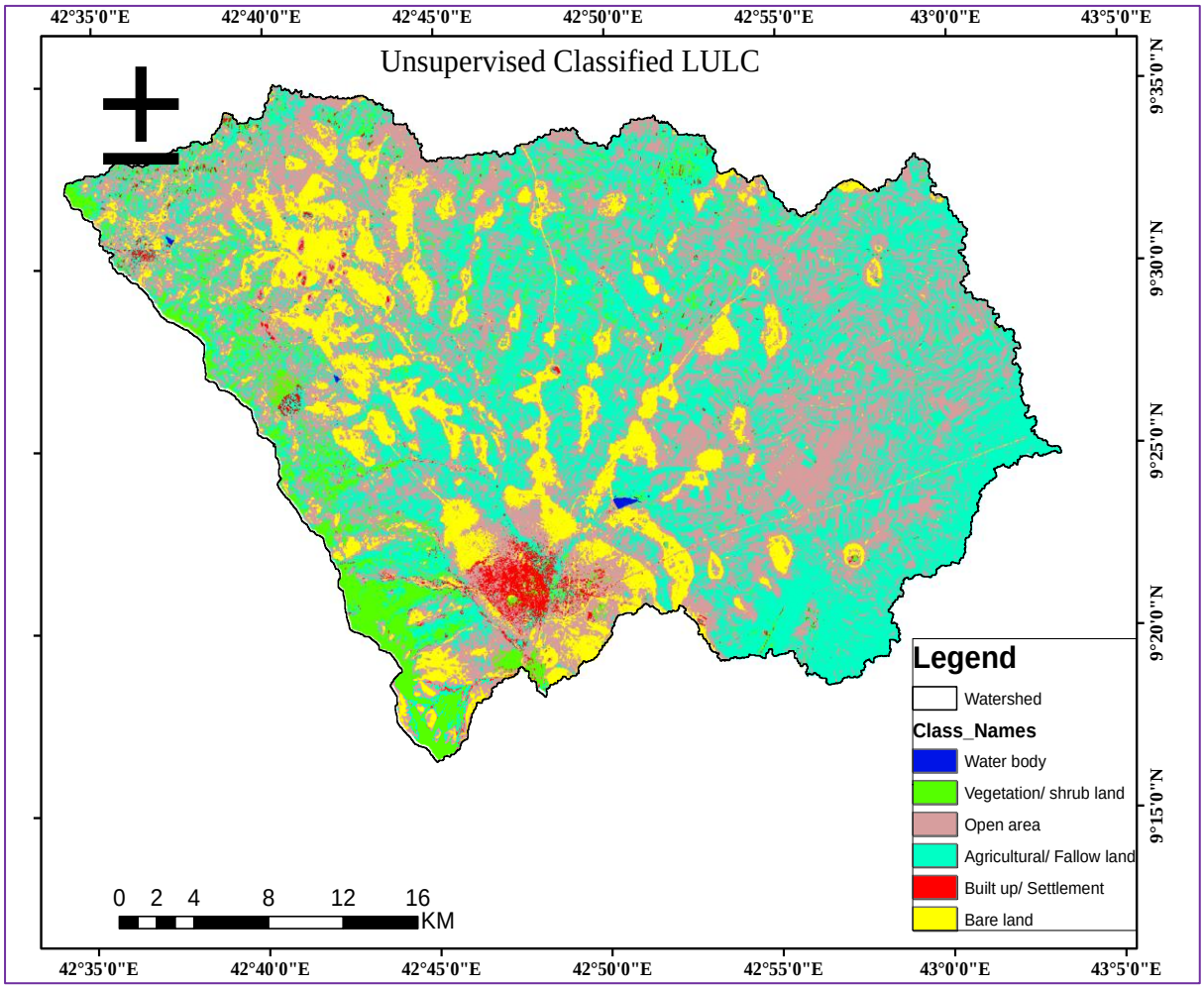


Figure 3-28: Unsupervised classified land use land cover

3.5.5 Supervised classification

This classification is based on the idea that a user can select sample pixels in an image that are representative of specific classes and then direct the image processing software to use these training sites as references for the classification of all other pixels in the image. Training sites (also known as testing sets or input classes) are selected based on the knowledge of the user. The user also sets the bounds for how similar other pixels must be to group them together. These bounds are often set based on the spectral characteristics of the training area, plus or minus a certain increment (often based on "brightness" or strength of

reflection in specific spectral bands). The user also designates the number of classes that the image is classified into.

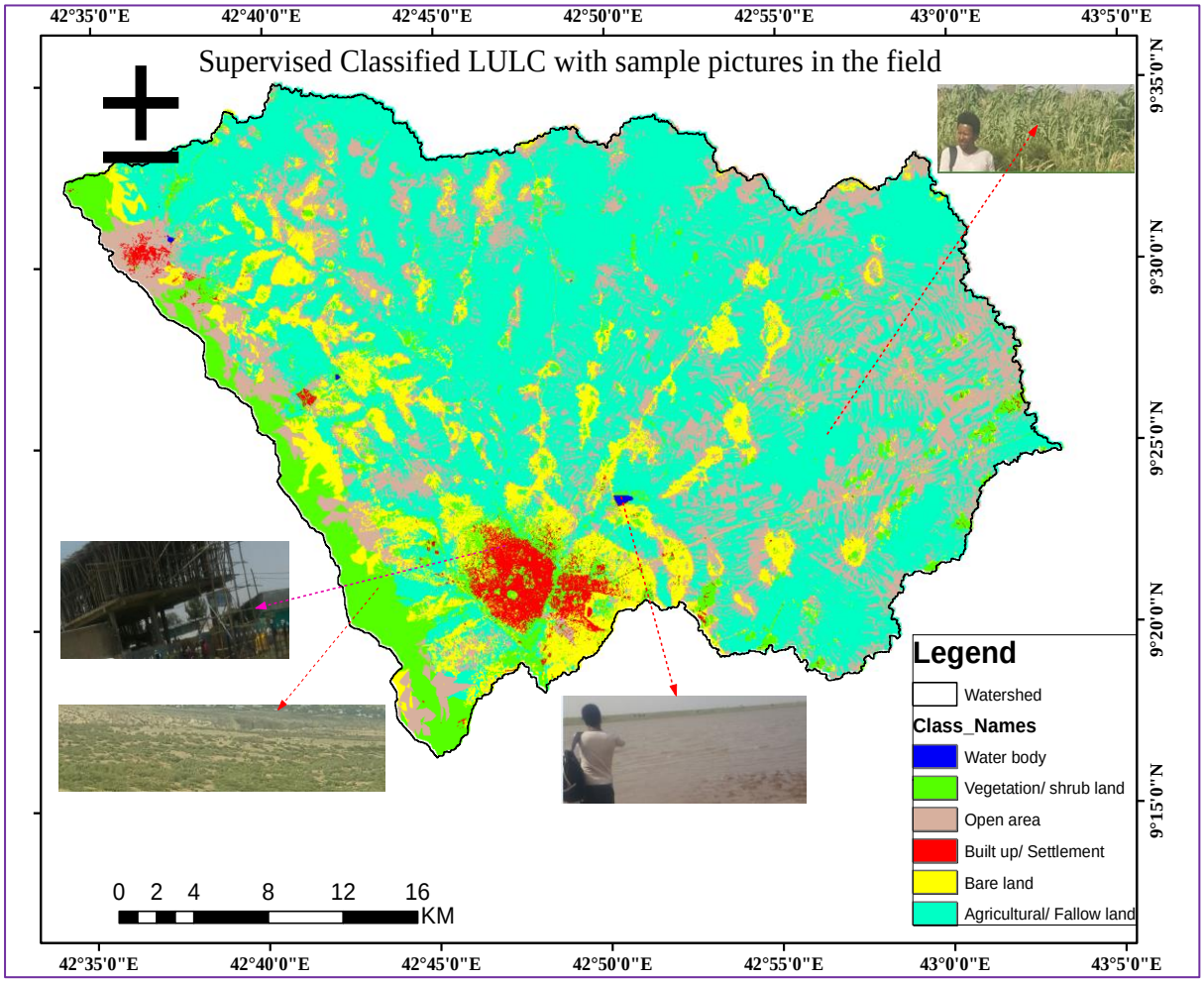


Figure 3-29: Supervised Classified LULC with sample pictures in the field

Generally, the above classification was done taking the Landsat 8 data in June 27-2019 and an intensive filed visit while collecting ground truth data in each class of land use land cover in the study area was done.

3.5.6 Confusion matrix for accuracy assessment in both classifications

The accuracy assessment has been made through a confusion or error matrix. A confusion matrix contains information about actual and predicted classifications done by a classification system [6]. The pixel that has been categorized from the image is compared to the same site in the field. The result of an accuracy assessment typically provides the users with an overall accuracy of the map and the accuracy for each class on the map. The percentage of overall accuracy was calculated using the following formula [66].

Overall accuracy = Total number of correct samples x 100%/Total number of samples

A total of 124 sample sites from Google earth and ground truth values were selected for accuracy assessment. Besides the overall accuracy, the classification accuracy of individual classes is calculated in a similar manner. The two approaches are user's accuracy and the producer's accuracy. The producer's accuracy is derived by dividing the number of correct pixels in one class divided by the total number of pixels as derived from reference data. The producer's accuracy measures how well a certain area has been classified. It includes the error of omission which refers to the proportion of observed features on the ground that is not classified on the map. Meanwhile, the user's accuracy is computed by dividing the number of correctly classified pixels in each category by the total number of pixels that were classified in that category [6]. The user's accuracy measures the commission error and indicates the probability that a pixel classified into a given category actually represents that category on the ground [6].

The producer's and user's accuracy are derived from the following formula [66].

Commission Error =off diagonal row elements/ Total of columns (3-9)

Producer's accuracy (%) = 100% - error of omission (%) (3-10)

User's accuracy (%) = 100% -error of commission (%) (3-11)

Kappa coefficient (Kc) is another measurement used in this study following [67],[66]. Accordingly, it was calculated by multiplying the total number of pixels in all the ground verification classes (N) by the sum of the confusion matrix diagonals (Xii), subtracting the sum of the ground verification pixels in a class time the sum of the classified pixels in that class summed over all classes, and dividing by the total number of pixels squared minus the sum of the ground verification pixels in that class times the sum of the classified pixels in that class summed over all classes. The value of Kappa lies between 0 and 1, where 0 represents agreement due to chance only and 1 represents the complete agreement between the two data sets. It is calculated based on the following formula [66].

$$k = \frac{N\sum_{i=1}^{r} xii - \sum_{i=1}^{r}(xi + *xi + i)}{N^2 - \sum_{i=1}^{r}(xi + * xi + i)}$$

$$k = \frac{(\text{Total Sum of correct}) - \text{Sum of the all the}(\text{row total column total})}{\text{Total squared} - \text{Sum of the all the}(\text{row total column total})}$$

Kappa is always less than or equal to 1. A value of 1 implies perfect agreement and values less than 1 imply less than perfect agreement. In rare situations, Kappa can be negative. This is a sign that the two observers agreed less than would be expected just by chance. It is rare that we get perfect agreement. Different people have different interpretations as to what is a good level of agreement. In the evaluation of Kappa statistics value, Sophia [67] puts that the strength of agreement between images.

Table 3-10: Rating criteria of Kappa statistics

S.No	Kappa statistics	Strength of agreement
1	<0.00	Poor
2	0.00 - 0.20	Slight
3	0.21 - 0.40	Fair
4	0.41 - 0.60	Moderate
5	0.61 - 0.80	Substantial
6	0.81 - 1.00	Almost perfect

Generally, in this study the supervised classification accuracy is 91.80%, kappa statistics is 0.9007 as shown in the table below, unsupervised classification accuracy is 89.39%, and kappa statistics is 0.8638 as shown in table below.

Table 3-11: The overall accuracy assessment results of supervised classified LULC in the study area in ERDAS 2014

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Bare land	82	80	77	93.90%	96.25%
Built up/ Settlement	32	46	32	100.00%	69.57%
Agricultural/ Fallow land	156	150	133	85.26%	88.67%
Open area	170	160	144	84.71%	90.00%

Vegetation/ shrub land	55	59	52	94.55%	88.14%
Water body	42	42	42	100.00%	100.00%
Totals	53	537	480		
Overall Classification Accuracy =		89.39%			
Overall Kappa Statistics = 0.8638					

Table 3-12: The overall accuracy assessment results of supervised classified LULC in the study area in ERDAS 2014

Class Name	Reference Totals	Classified Totals	Number Correct	Producers Accuracy	Users Accuracy
Bare land	18	18	17	94.44%	94.44%
Water body	15	9	9	60.00%	100.00%
Open area	19	19	18	94.74%	94.74%
Built up/ Settlement	21	24	21	100.00%	87.50%
Vegetation/ shrub land	24	22	22	91.67%	100.00%
Agricultural/ Fallow land	25	30	25	100.00%	83.33%
Totals	122	122	112		
Overall Classification Accuracy =		91.80%			
Overall Kappa Statistics = 0.9007					

Table 3-13: Area coverages of unsupervised LULC classes in the study area

Class No	Class_Names	Pixels Count	Area (Km^2)	Area_Percentage (%)
0	Unclassified	6043	5.44	0.52
1	Bare land	153094	137.78	13.10
2	Built up/ Settlement	17177	15.46	1.47
3	Agricultural/ Fallow land	436191	392.57	37.33
4	Open area	486711	438.04	41.65

5	Vegetation/ shrub land	68515	61.66	5.86
6	Water body	813	0.73	0.07
	Total	1168544	1051.69	100.00

Table 3-14: Area coverages of supervised LULC classes in the study area

Class No	Class Names	Pixels Count	Area (Km^2)	Area_Percentage (%)
0	Unclassified	10046	9.04	0.86
1	Bare land	141312	127.18	12.09
2	Built up/ Settlement	24485	22.04	2.10
3	Agricultural/ Fallow land	618212	556.39	52.90
4	Open area	272810	245.53	23.35
5	Vegetation/ shrub land	100780	90.70	8.62
6	Water body	899	0.81	0.08
	Total	1168544	1051.69	100.00

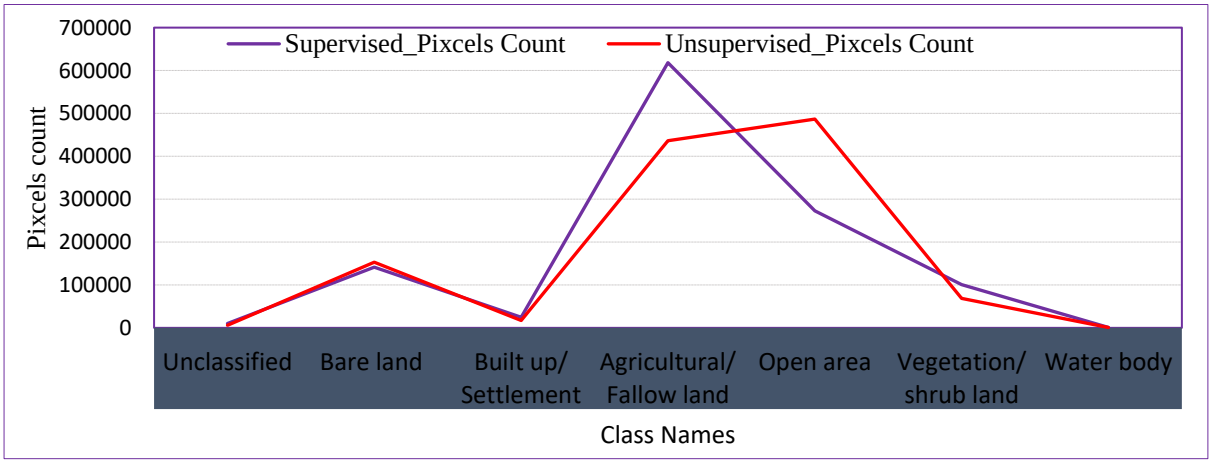


Figure 3-30: Comparison of Supervised and Unsupervised LULC classification based on pixels

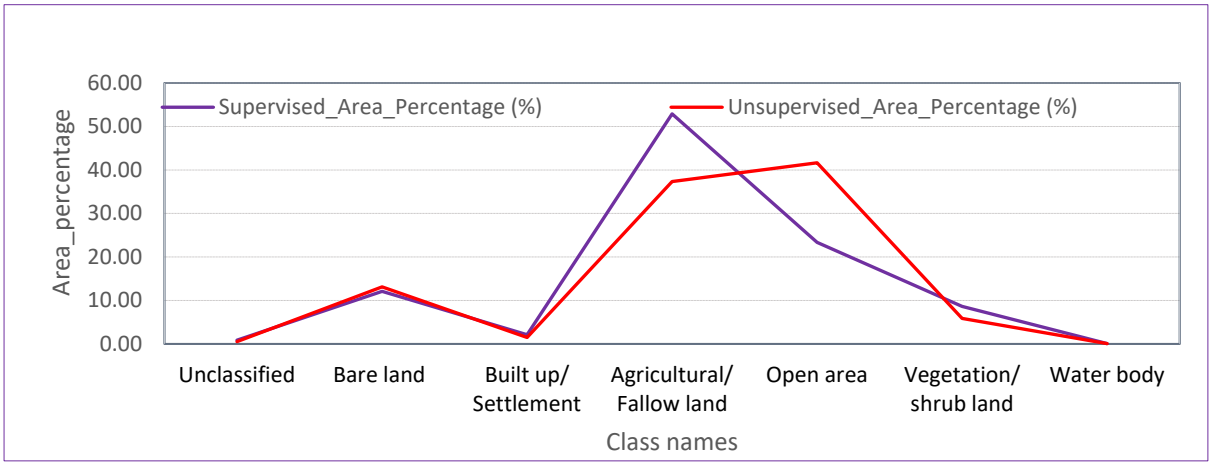


Figure 3-31: Comparison of Supervised and Unsupervised LULC classification based on relative area cover percentage

The above classifications and comparison was done for Unsupervised (January 2016) before rainy seasons and flood events in the study area and supervised classification (June 2019) derived after rainy seasons which occurred in the last three to four months in the study area, and the results indicates that classes Built-up/ Settlement, Agricultural/ Fallow land, Vegetation/ shrubland, Waterbody were increased due to the change in year and rainy seasons influences, on the other hand, open area, bare lands, and open areas were decreased because of the above reasons and more due to agricultural and cropping expansion after the rainy season.

3.5.7 Land use land cover and flash flood Susceptibility

Land use/land cover is an important factor to identify those zones that have shown high susceptibility to flooding Vegetated areas have a low potential for flooding due to the negative relationship between flooding and vegetation density. On the other hand, residential areas and roads, which are mostly made by impervious surfaces, and bare lands increase the storm runoff. In the LULC map, six land use classes were identified: Built-up/ Settlement, Agricultural/ Fallow land, Vegetation/ shrubland, Waterbody, open area, bare lands, and open areas. Logistic Regression for flash flood inventory mapping

3.5.8 Development of flash flood inventory map

Flash flooding locations were identified and extracted from various sources such as extensive fieldwork surveying, the examination of aerial photographs considering the flash flood events happened in different years previously, and then an inventory map containing 560 flooded and nonflood locations were generated over the whole study area. In order to come up with

the good result of logistic regression model result equal proportions of flood and nonflood locations were taken, where flood locations assigned value 1 and nonflood locations were assigned value 0; moreover, 1 representing the existence, and 0 the absence of flooding. Non-flooded areas where randomly selected and generated on a topographic map by choosing sites high on hills where flooding would not be possible [6], [7] and those flood locations collected more near the main drainage lines of the study area, so that the approach used in this study is expected to increase the precision of the results. Many of literatures provide several suggestions regarding the size of samples to be used for training and validation (testing) the logistic regression model [68], [41], [38], [40]. On the basis of those literature review, 70 % of the flooded locations were used randomly to produce the dichotomous data (0 and 1) for the purpose of training in order to assess the spatial distribution of flooding which are about 392 numbers from total [69]. The rest of the locations of 30% (validation data) were used for testing the models which are 168 in numbers. Figure below shows a map of flooded and non-flooded areas in the Jerer up-stream watershed collected.

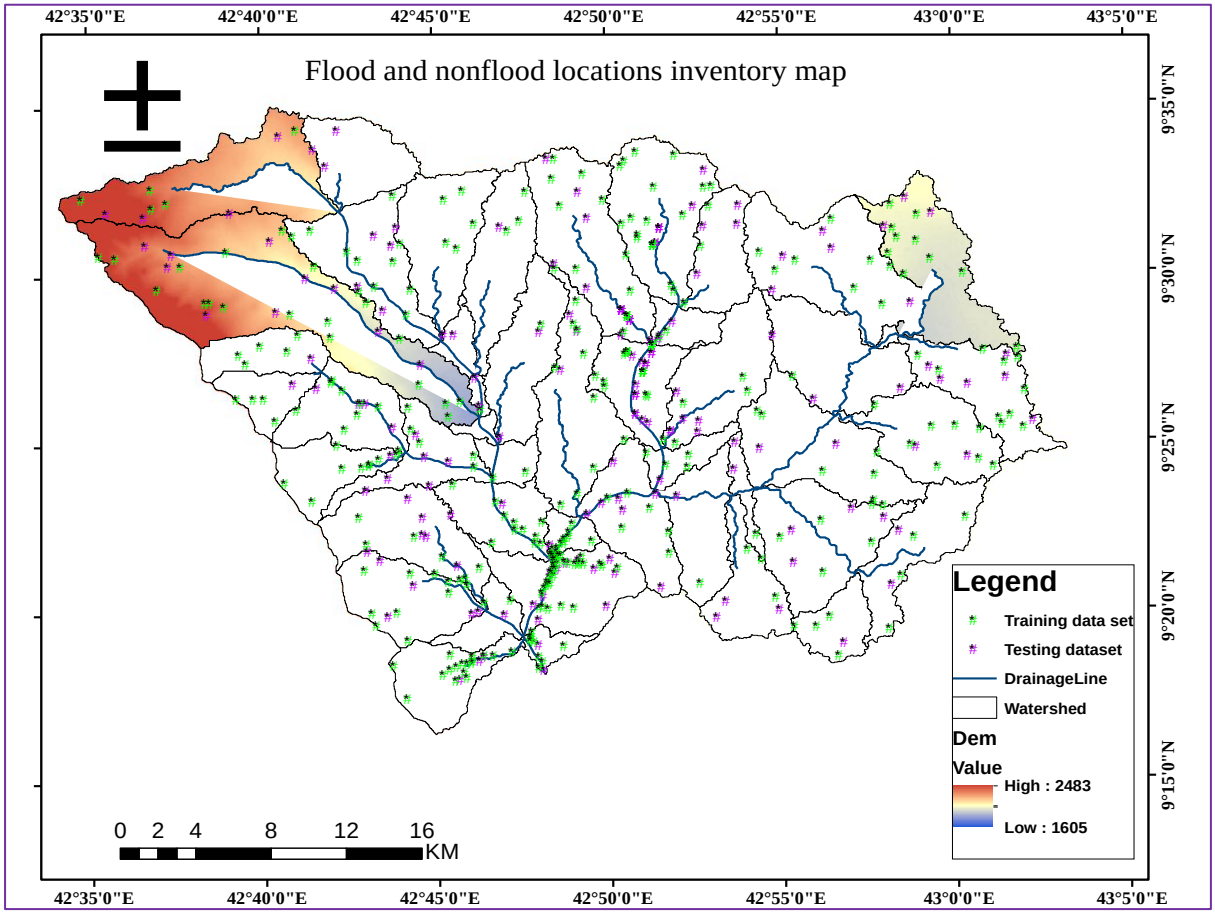


Figure 3-32: Flash flood inventory map of the study area (partitioned training 70% and testing datasets 30%)

In the following pictures and appendix; the Google earth image, field surveys (Ground truth points) and communication with people faced with the hazard after the flash flood happened in the study area were collected in order to develop the flash flood susceptibility inventory map as shown in the figure above.

Figure 3-33: Sample flash flood susceptible locations to develop flash flood inventory map of the study area

3.5.9 Data preparation for Logistic regression method and the model

3.5.9.1 Classification and Normalization of Influencing Factors

The conditioning factors were classified using quintile method in Arc Gis 10.4.1 software based on each conditional factors influence to flash floos suscesptibility. According to Ayalew [70], there are different types of classification methods such that natural breaks, quantiles, equal intervals, and standard deviations. Here it was attempted to choose the one that best suits the objectives for this study. While these classifiers are well established in statistics, they often lead to different results, as they make very different statements about how values should be divided. The classification scheme that relies on natural breaks, for example, identifies breakpoints by looking for patterns inherent in the data. In quantile classification, features are grouped by equal numbers in each class. The equal interval scheme divides the range of values into equal-sized subdivisions. A standard deviation is used to classify data based on the amount a value varies from the mean. Although each classification methods have their own advantages as listed above, the drawbacks in each classification methods are many. The quantile classification has a disadvantage in that it places widely different values into the same class. Using equal intervals was also found to be not practical since it emphasizes one class of susceptibility relative to others. In natural breaks, boundaries are set where there are relatively big jumps in data values. In Standard deviation map doesn't show actual values, only how far values are from the mean and very high or low values can skew the mean.

Many researchers revealed and suggested that quantile classification method was suitable for flash flood susceptibility conditioning factors classification [71], [41], [72], [73] and [68] though Ayalew [70] stated that the quantile classification method has the drawback as it classifies broadly different values into the same category. Finally, all influencing factors were classified as very low (1), low (2), moderate (3), high (4) and very high (5) based on the rates for flash flood susceptibility.

Furthermore, dependent (binary data) and reclassified independent (influencing) variables derived in Arc Gis were converted from raster to ASCII format and normalized to be used in the logistic regression analysis. The normalization of all independent variables is necessary which facilitates final analysis and interpretation [74]. In the Normalization the value of each cell in the reclassified rasters (Xcell) is subtracted from the minimum value (Xmin) present in

the data of the file divided by the result of the maximum value in the raster class (Xmax) minus the minimum value (Xmin) of the raster class as shown in the following equation.

$$H = {(Xcell - Xmin)}\Big/{(Xmax - Xmin)} \quad (3\text{-}12)$$

3.5.9.2 Establishment of Spatial Relationship between flash flood susceptibility and Influencing Factors using frequency ratio

The Frequency Ratio in this study was used to analyze the probabilistic relationship between spatial flash flood susceptibility and those influencing factors. It is defined as the ratio between areas in which Flash flood susceptibility occur to the total study area, and is calculated using the following equation [40],[75]. The frequency ratio was assessed by applying the ratio between the numbers of pixels with flash flood phenomena inside a factor's class and the total number of pixels with flash flood phenomena over the whole study area.

$$FR = \frac{Np\ (Xi)}{\sum Np\ (Xi)} \Big/ \frac{Np\ (Xj)}{\sum Np\ (Xj)} \quad (3\text{-}13)$$

Where $Np\ (Xi)$ is the number of pixels with flash flood for each influencing factor, $\sum Np\ (Xi)$ is the total number of flash flood and non-flash flood numbers in the study area, $Np\ (Xj)$ is the number of pixels in the factor's class, and $\sum Np\ (Xj)$ is the total number of pixels in the study area. FR is considered to show an average relationship if its value is 1, high correlation if larger than 1, and low correlation if lower than 1 [76]. In order to do the frequency ratio in GIS, all the scaled flood conditioning factors were classified, using quintile method among many classification techniques exist based on its popularity [68].

3.5.9.3 Sensitivity and multi-collinearity tests for conditional factors

In this study developing a flash flood susceptibility prediction model involves a set of eleven factors; among these factors, some of them might be strongly related to each other in flash flood susceptibility [77]. Hence, a sensitivity analysis is a process which eliminating one of strongly related flash flood conditional factors from the model and then comparison of each weights (relative impacts) in the model.

After the selection of the flash flood conditioning factors, assessment of multicollinearity problems was done before the logistic regression model implemented. Multicollinearity refers

to a high correlation, or, in other words, interdependent that exists among two or more variables during logistic regression [75]. If multicollinearity exists among the independent variables during regression analyses, the variance of the regression coefficient increases. The error also increases with multicollinearity, reducing the accuracy of the model's prediction.

There are several multicollinearity diagnostics methods to quantify factors including Pearson's correlation coefficient, variance decomposition proportions, conditional index, and Variance Inflation Factors (VIF) and tolerances (TOL). In many logistic regression model applications, the Variance Inflation Factors (VIF) and tolerances (TOL) are common [78], [73]. In the present study, the VIF and tolerance methods were used to test the correlation among the eleven flash flood influencing factors (ground slope, distance from drainage, drainage density, TWI, SPI, rainfall, land use land cover, NDVI, and MHI). The Variance Inflation Factors (VIF) and tolerances (TOL) were computed based on the formula $1/(1-R^2j)$ and the reciprocal of VIF respectively, where R^2 j is the R^2 found when regressing all other predictors onto the predictor *j*.

Multicollinearity can be assessed by various means, and tolerance (TOL) and variance inflation factor (VIF) assessments were used in this study. If the values of TOL and VIF were ≤ 0.1 and ≥ 10 respectively, there would be multicollinearity among independent variables. In this study, the values of TOL and VIF was computed in R statistical Package. The results show that both TOL and VIF were in the required limit; means there were no multicollinearity in the independent variables.

3.5.9.4 Logistic regression model

The goal of logistic regression is to find the best model to describe the relationship between a dependent variable and multiple independent variables [8], [79], [80]. Therefore, logistic regression model allows one to perform a multivariate regression relationship between a dependent variable and several independent variables which are useful for predicting the presence or absence of a characteristic or outcome based on values of a set of predictor variables [68],[63]. The dependent variable of a logistic regression could be binary or categorical and the independent variables of a logistic regression could be a mixture of continuous and categorical or binary variables. The assumption of normality is not needed for logistic regression. Hence, logistic regression is advantageous compared to linear regression and log-linear regression. Logistic regression establishes a functional relationship between

the binary coded flash flood locations (presence or absence of a flood) and different factors. Detailed descriptions of the logistic regression technique can be found in the general form of logistic regression is as follows:

$$Z = \beta 0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \cdots \cdots \cdots \beta n Xn \quad (3\text{-}14)$$

$$P = {}^{1}/_{(1+e^{-z})} \quad (3\text{-}15)$$

where xi (i = 0, 1, 2, …, n) are independent variables (flash flood influencing factors) and Z is a linear combination function of the conditional or explanatory variables representing a linear relationship which is flash flood susceptibility index. The parameters $\beta_1, \beta_2, \ldots, \beta n$ are the regression coefficients to be estimated in the model. The LR produces the weights for each conditioning factor (independent variables), which can be used in GIS to produce the probability map of flood occurrence. Moreover, in this study Z denoted as a binary response variable (0 or 1), where value 1 is the presence of flood, and value 0 shows the absence of flood locations. P refers to the probability of the occurrence of a flash flood, in which P-value equals 1 is there would be flood flood while P-value equals 0, there would not be a flash flood.

3.5.9.5 Accuracy assesment

The statistical evaluation measures of overall accuracy, specificity, sensitivity, positive predictive value (PPV), and negative predictive value (NPV) were applied to measure the comparative performance of the model [35]. PPV and NPV estimate the probability of training and testing dataset samples correctly classified to the flash flooded class and non-flash flooded class, respectively. These were computed based on the following formulas [41].

Overall Accuracy $= {}^{Tp+Tn}/_{Tp+Tn+Fp+Fn}$	PPV $= {}^{Tp}/_{Tp+Tp}$
Sensitivity $= {}^{Tp}/_{Tp+Fn}$, Specifity $= {}^{Tn}/_{Fp+Tn}$	NPV $= {}^{Tn}/_{Fn+Tn}$

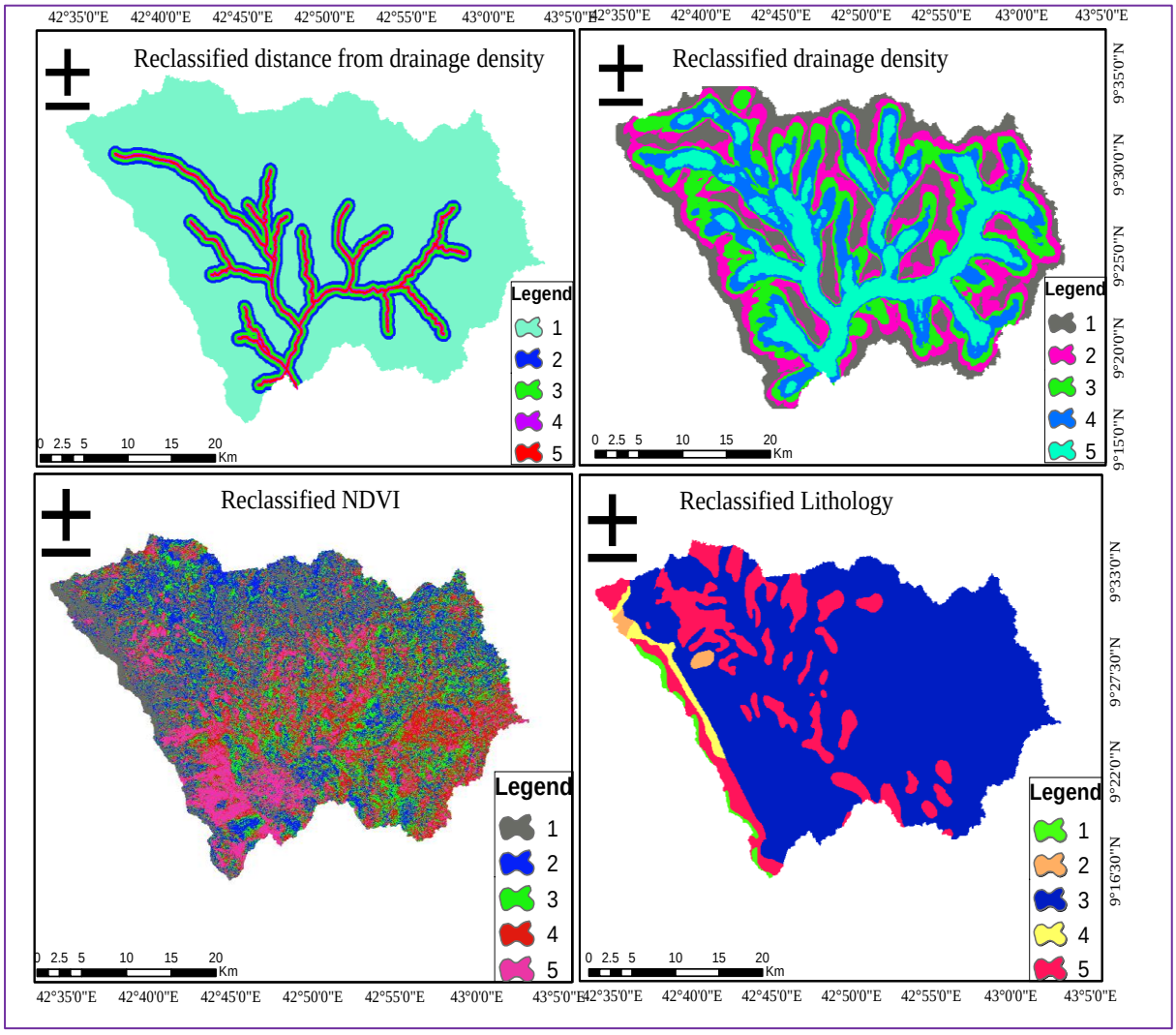
where True Positive (TP) and True Negative (TN) are the number of samples in the training and validation datasets, correctly classified to the flood and non-flood class, respectively.

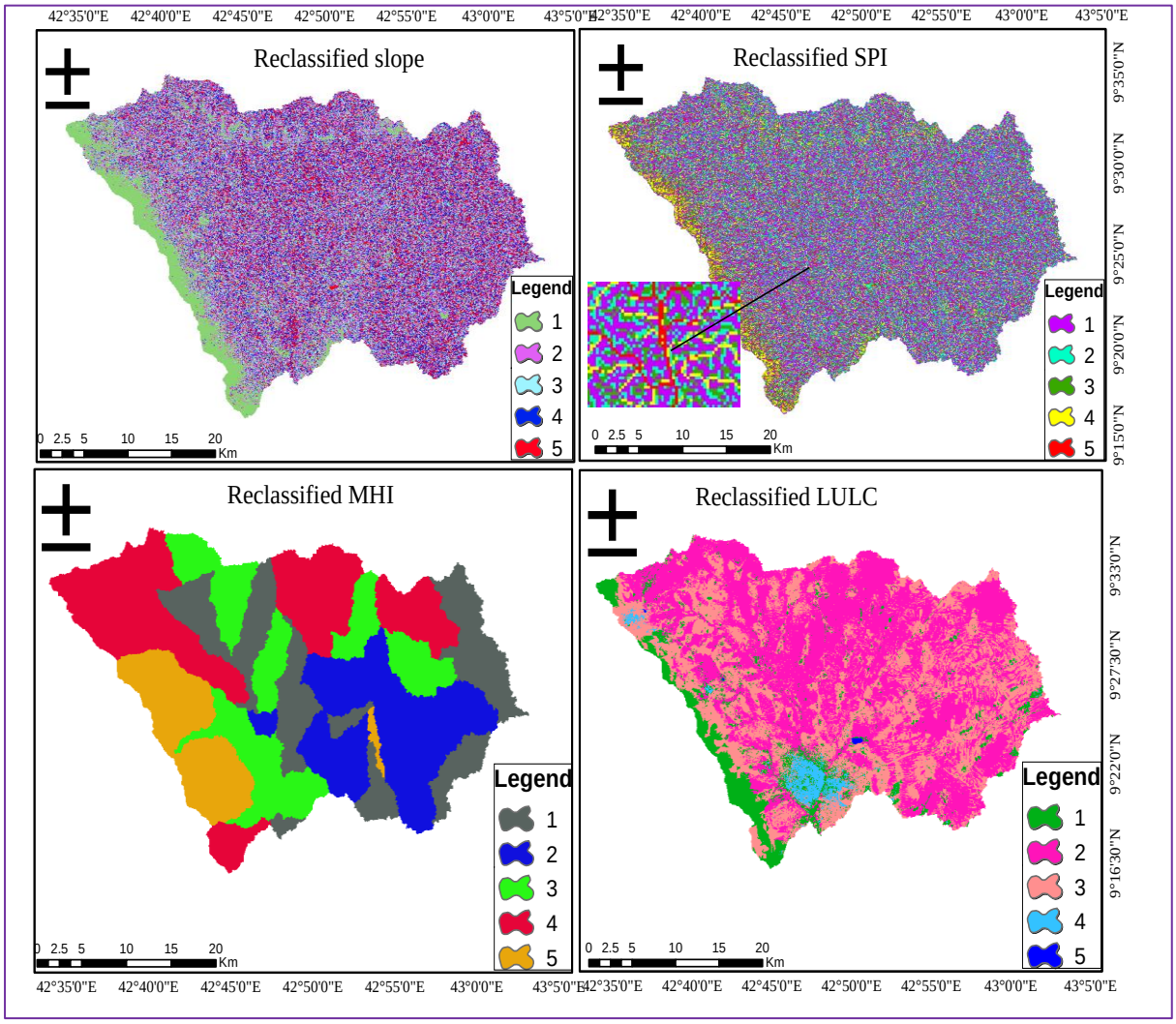
False Positive (FP) and False Negative (FN) are the number of samples in the training and validation datasets that were erroneously classified.

CHAPTER 4 RESULTS AND DISCUSSION

4.1 Classified and Normalized conditional factors for logistic regression

The following figures show all the classified 11 flash flood conditioning factors according to the susceptibility of each classes for flash flood in the study area; where the values for each classes were assigned as 1, 2, 3, 4, and 5, indicating that very low, low, moderate, high and very high susceptibility rates respectively.




Reclassified slope
Reclassified SPI
Reclassified MHI
Reclassified LULC
Legend
1
2
3
4
5
0 2.5 5 10 15 20
Km
42°35'0"E
42°40'0"E
42°45'0"E
42°50'0"E
42°55'0"E
43°0'0"E
43°5'0"E
9°35'0"N
9°30'0"N
9°25'0"N
9°20'0"N
9°15'0"N
9°33'0"N
9°27'30"N
9°22'0"N
9°16'30"N

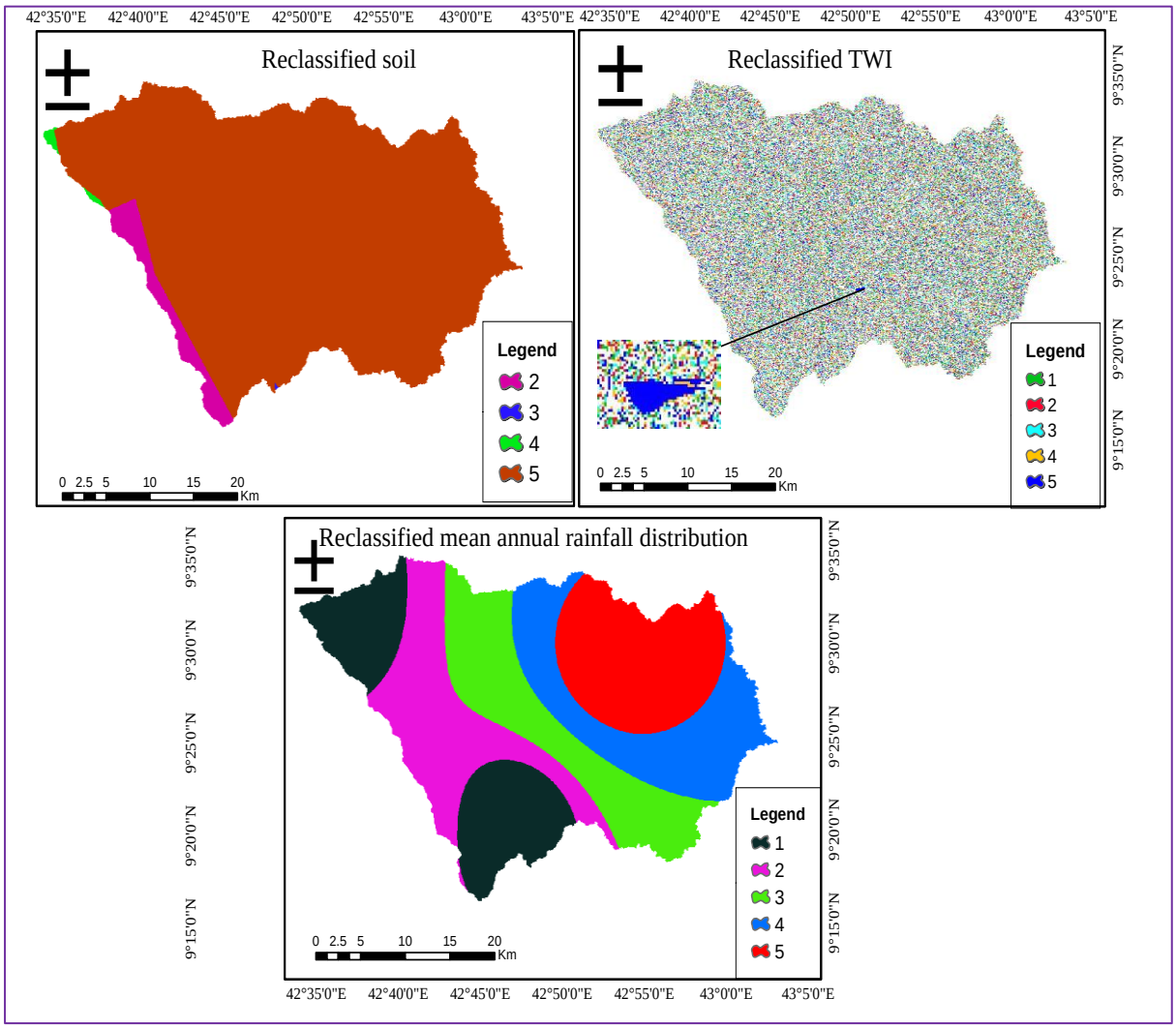
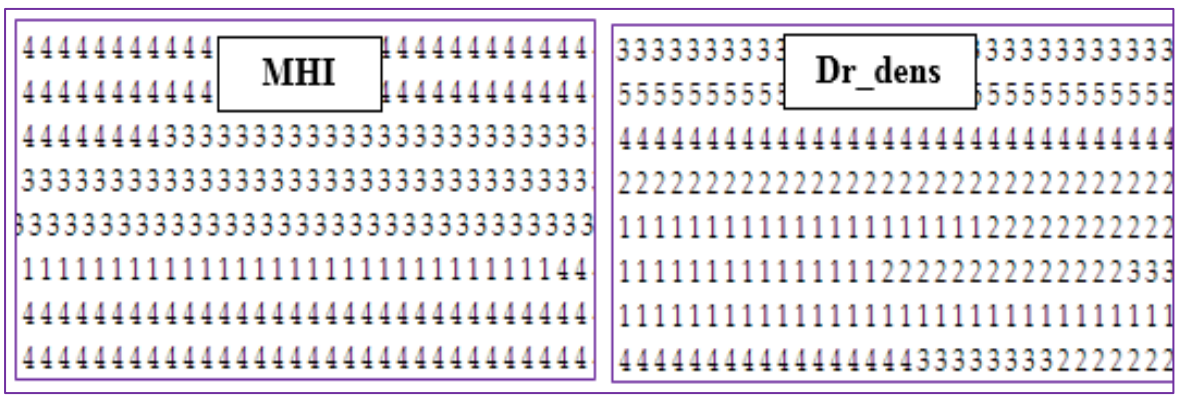


Figure 4-1: All classified flash flood conditioning factors

The results of all classified raster conditional factors were converted to ACSII data format (Acronym for the American Standard Code for Information Interchange) format in ArcGIS as shown sample results below.

Table 4-1: Some reclassified raster to ASCII converted flash flood conditional factors



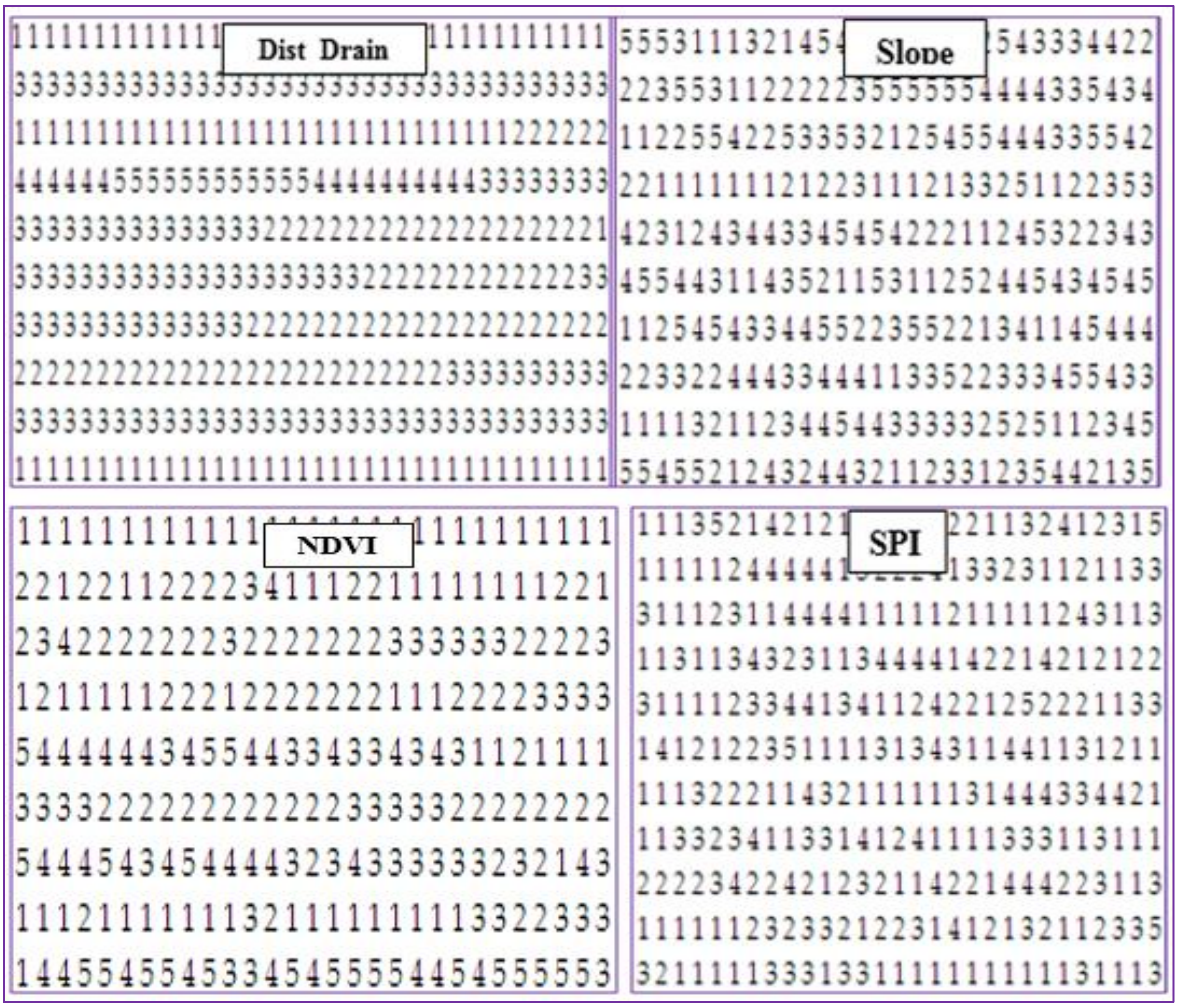


Table 4-2: Sample normalized ASCII files for the logistic regression model

ID	NDVI	SPI	Soil	Slope	LULC	Drain_Dens	MHI	Lith	IDW	TWI
1	0.5	0.5	1.00	0.25	0.5	1	0.25	0.5	1	0.67
1	0.5	0.5	1.00	0.25	0.5	1	0.25	0.5	1	0.67
1	0.5	0.5	1.00	0.25	0.5	1	0.25	0.5	1	0.67
1	0.5	0.5	1.00	0.25	0.5	1	0.25	0.5	1	0.67
1	0.75	0.5	1.00	0.75	0.25	0.5	0.25	0.5	1	0.67
1	0.75	0.5	1.00	0.75	0.25	0.5	0.25	0.5	1	0.67
1	0.75	0.5	1.00	0.75	0.25	0.5	0.25	0.5	1	0.67
1	0.75	0.5	1.00	0.75	0.25	0.5	0.25	0.5	1	0.67
1	0.75	0.5	1.00	0.75	0.25	0.5	0.25	0.5	1	0.67
1	1	0.5	1.00	0.75	0.25	0.75	1	0.5	0.25	0.67
0	0.5	0.5	1.00	0.5	0.25	0.5	0	0.5	0.75	0.67
1	0.5	0	1.00	0	0.25	1	0.25	0.5	1	0.67
1	0.5	0	1.00	0	0.25	1	0.25	0.5	1	0.67
1	0.5	0	1.00	0	0.25	1	0.25	0.5	1	0.67
1	0.5	0	1.00	0	0.25	1	0.25	0.5	1	0.67
1	0.5	0	1.00	0	0.25	1	0.25	0.5	1	0.67
1	0.25	0	1.00	0.75	0.25	1	0.75	0.5	0.5	0.67
0	0.5	0.5	1.00	0.75	0.5	0.25	0	0.5	0.75	0.67
0	1	0.75	1.00	0.25	0	0.75	0	1	0.5	0.67
0	0.25	0.25	0.00	0.25	0.5	0.25	1	0.75	0.25	0.67
0	0.75	0	1.00	1	0.5	0.75	0	0.5	1	0.67
0	0.75	0.25	1.00	1	0.5	0.75	0	0.5	0.5	0.67

4.2 Spatial Relationship between flash flood susceptibility and Influencing Factors

The following table shows the computed frequency ratio that indicates the spatial relationshiops the flash flood susceptibility (dependent variable) and influencing iactors (independent variables).

Table 4-3: Frequency ratios for flash flooding and conditional factors

Conditional Factors	Class intervals	Susceptibility degree	Pixels number in the domain	Pixels number in domain %	Flood susceptibility number	Flash Flood area %	Frequency ratio (FR)
Slope	0 - 1.35	5	247158	21.15	40	29	1.37
	1.35 - 2.05	4	249802	21.38	89	32	1.49
	2.05 - 2.86	3	228443	19.55	58	21	1.06
	2.86 - 3.95	2	223207	19.10	73	15	0.79
	3.95 - 58	1	219934	18.82	20	7	0.38
LULC	Bare land/Open area	3	414122	35.75	155	53	1.50
	Water body	5	899	0.08	2	1	8.89
	Built up/ Settlement	4	24485	2.11	21	7	3.43
	Vegetation/ shrub land	1	100780	8.70	4	1	0.16
	Agricultural/ Fallow land	2	618212	53.36	108	37	0.70
Soil	Clay	5	1120345	95.88	280	100	1.04
	loam	4	3634	0.31	0	0	0.00
	Sandy loam	2	44434	3.80	0	0	0.00
	lomy sand	3	131	0.01	0	0	0.00

Drainage density	0 - 0.73	1	222194	19.11	0	0	0.00
	0.73 - 1.6	2	241010	20.72	1	0.36	0.02
	1.6 - 2.62	3	237247	20.40	17	6.07	0.30
	2.62 - 3.92	4	230837	19.85	44	15.71	0.79
	3.92 - 8.2	5	231683	19.92	218	77.86	3.91
TWI	-2.26 - 5.49	5	121657	20.86	0	0.00	0.00
	5.49 - 6.52	4	132509	22.72	100	35.71	1.57
	6.52 - 7.15	3	120891	20.73	103	36.79	1.77
	7.15 - 8.03	2	125980	21.60	51	18.21	0.84
	8.03 - 13.92	1	82233	14.10	26	9.29	0.66
SPI	-13.82 - -6.95	1	508861	43.55	69	24.64	0.57
	-6.95 - -2	2	263554	22.55	87	31.07	1.38
	-2 - 0.02	3	227965	19.51	72	25.71	1.32
	0.02 - 2.85	4	124413	10.65	33	11.79	1.11
	2.85 - 12.04	5	43751	3.74	19	6.79	1.81
MHI	5.53 - 6.71	1	249411	21.34	38	13.57	0.64
	6.71 - 7.06	2	240186	20.55	89	31.79	1.55
	7.06 - 7.41	3	229815	19.67	64	22.86	1.16
	7.41 - 8.3	4	310750	26.59	56	20.00	0.75
	8.3 - 9.50	5	138382	11.84	33	11.79	1.00
NDVI	-0.19 - 0.11	5	171384	14.67	84	30.00	2.05
	0.11- 0.12	4	251804	21.55	65	23.21	1.08
	0.12- 0.13	3	289768	24.80	69	24.64	0.99

IDW	0.13 - 0.14		2	265691	22.74	42	15.00	0.66
	0.14 – 0.50		1	189897	16.25	20	7.14	0.44
	533.03 - 615.44		1	226732	19.40	110	39.29	2.02
	615.44 - 654.62		2	235537	20.16	29	10.36	0.51
	654.62 - 737.03		3	237633	20.34	22	7.86	0.39
	737.03 - 809.98		4	231118	19.78	36	12.86	0.65
	809.98 - 878.88		5	237524	16.89	83	29.64	1.75
Distance from Drainage	0 - 150		5	81905	7.01	145	51.79	7.39
	150 - 450		4	137384	11.76	86	30.71	2.61
	450 - 1000		3	197627	16.91	34	12.14	0.72
	1000 - 1500		2	217988	18.65	9	3.21	0.17
	>1500		1	533640	45.67	6	2.14	0.05
Lithology	Jh2		5	198930	17.03	16	5.71	0.34
	Qa		3	933025	79.86	264	94.29	1.18
	Ja		4	17761	1.52	0	0.00	0.00
	Pc2		2	7937	0.68	0	0.00	0.00
	PNa1	1	10657	0.91	0	0.00		0.00

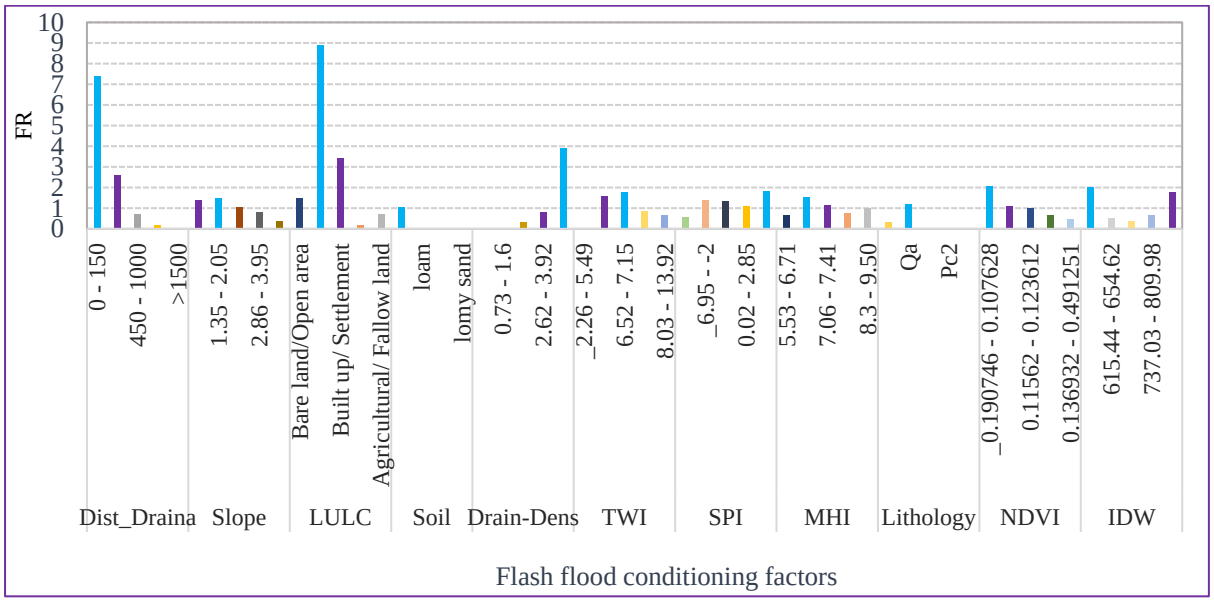


Figure 4-2: Correlations between flash flood susceptibility frequency and the conditional factors

The results of the Frequency ratio computed in the above table were measured for each class of flash flood conditioning factor by dividing the flood occurrence ratio by the area ratio. According to the results in table above and Figure above shows that a high-frequency ratio in the class weight shows a stronger correlation between that class and the flash flood occurrence, and subsequently represents a higher probability of flash flood occurrence in that defined class. For instance, the results in the table showed that the first four classes of slope 0 - 1.35, 1.35 - 2.05 and 2.05 - 2.86 had Fr > 1 which confirms the concept that flooding mostly occurs in the low-elevated lands. The received the highest FR values. The class 0 – 150m in the distance from drainage gained the highest FR value of 7.39 which indicated that the closer the distance to the drainage line, the higher the flash flood susceptibility followed by 1.5 - 4.52 class with FR value 2.61. In LULC classes Bare land/ open area, water body and built up/ settlement classes were highly prone to flooding because of the FR values indicated that 1.50, 8.89, and 3.43 were > 1 respectively. Vegetated areas and agricultural areas showed a relatively low correlation with flash flood susceptibility because they offer a protective cover and infiltration respectively, moreover, a negative relationship exists between a flash flood event and vegetation density. For the soil type classes, the poorly drained soil class clay had an FR value of 1.04 and others loam, sandy loam and loamy sand soils had the moderately to well-drained capacity so that their FR values were 0. In the analysis the frequency ratio was high in classes 533.03 - 615.44 and 809.98 - 878.88 indicates that the effects of rainfall were

a triggering factor in flood generation; in the class 809.98 - 878.88 small rate of flooding happened in the area generated and despite the rainfall amount was lowest in 533.03 - 615.44 class, the FR value was > 1 which means other factors were constrained the amount of infiltration in the study area so that this class revealed the high correlation with flash flood susceptibility. In the NDVI classes of 0.11- 0.12 gave high correlation with the flash flood susceptibility followed by 0.12- 0.13 NDVI classes as the FR values result, which generally indicates the low value of NDVI low vegetation and high flash flood susceptibility.

Generally, based on the results of the FR analysis, the classes of flash flood influence factors used in this study, some of the conditioning factors showed the highest value in each class. For example, the land use land cover had an FR range of 0.38-9.44, and Distance from drainage 0.07-4.34. Moreover, in each flash flood susceptibility factor it was found at least one class with an FR value larger than 1, showing a high correlation with flash flood occurrence in that class while FR <1 indicated that lower correlation. Therefore, all the flash flood susceptibility conditional factors were appropriate and used as input for spatial flash flood susceptibility prediction factors.

4.3 Training and testing data preparation for the logistic regression model

In order to estimate the coefficients of every conditional factor in the logistic model, ArcMap/ArcGIS and R statistical software were used as analytical tools for spatial and statistical analysis of all the dependent and independent (predicting) variables and the partitioned datasets were created containing the values of the independent variables and the status of the dependent variable for the same pixels. A table containing an equal number of flash flood location pixels and non-flash flood location pixels (in highland areas) were prepared. The first column contained the status of the dependent variable with 560 total pixels. The other columns, one per independent variable, contained the values of the independent (predictor) variables.

The training dataset and the testing dataset were generated using the selected influencing factors in conjunction with flash flood inventory binary data as 1 and 0 assigned for present and presence of flash flood for the specified grid cell respectively and these datasets were generated with the selected flash flood-influencing factors. After importing the database of dependent (binary data) and independent (flash flood conditioning factor) variables analyzed well for the missing cell data, the data were randomly partitioned into training and testing

datasets in Rstudio. For this, flash flood inventory map was divided into two parts, 70% (392 locations) used for the training dataset and 30% (168 remaining locations) used for the testing (validation) dataset out of total 560 datasets. The following R code was used import and derive all (560) flood and non flood locations for developing Logistic Regression Model.

```
# Importing the dependent and inependent variables data

Project4= read.table("GPS_Flood_and_Nonflood_locations_classes.csv",sep=",",header = T)

sapply(Project4,function(x) sum(is.na(x))) #check missing values

Project4[Project4=="<Null>"] <- NA # change no data(Null) to Na
sapply(Project4,function(x) length(unique(x))) # no of unique values for each column

# Logistic regression
My_table4 = (subset(Project4,select=c(5,6,7,8,9,10,11,12,13,14,15,16))) # produce new table
#for log-regression
View(My_table4)
```

After, setting all the training and testing data sets, the training (70%) and testing (30%) were randomly partitioned and prepared for logistic regression model, they were introduced for training and validating the logistic regression model using the R studio, where these sample tables were put in the following tables. Thenafter, the output intercept (α) and coefficient (β) values were generated as shown in table 4-9 to calculate the linear logistic model of spatial flash flood susceptibility index (z) in raster calculator (Arc-GIS) that is shown in equation 4.1. Finally, spatial flash flood susceptibility index (z) was produced and then the flash flood susceptibility probability was produced as shwon in figure 4-5.

Table 4-4: Sample training and testing datasets from all datasets

	ID	TWI	Dist_Drainage	SPI	MHI	Soil	Slope	LULC	Drain_Dens	Lith	IDW	NDVI
1	1	0.25	0.50	0.00	0.25	Clay	0.00	Agricultural land	0.75	Qa	0.25	0.00
2	1	0.25	0.00	0.00	0.00	Clay	0.00	Agricultural land	0.75	Qa	0.50	0.00
3	1	0.25	0.50	0.00	0.25	Clay	0.00	Agricultural land	0.75	Qa	0.00	0.00
4	1	0.25	0.50	0.00	0.25	Clay	0.00	Agricultural land	0.75	Qa	0.25	0.50
5	1	0.00	0.75	0.00	0.00	Clay	0.75	Agricultural land	0.75	Qa	0.00	0.25
6	1	0.00	0.75	0.00	0.00	Clay	0.75	Agricultural land	0.75	Qa	0.00	0.50
7	0	0.75	0.00	0.25	0.00	Clay	1.00	Agricultural land	0.75	Qa	0.00	0.50
8	0	1.00	0.25	0.00	0.00	Clay	0.75	Agricultural land	0.75	Qa	0.50	0.25
9	0	1.00	0.25	0.50	0.00	Clay	0.50	Agricultural land	0.75	Qa	0.75	0.75
10	0	1.00	0.25	0.25	0.00	Clay	0.50	Agricultural land	0.75	Qa	0.25	0.25
11	1	0.25	0.25	0.00	0.25	Clay	0.75	Agricultural land	0.50	Qa	0.50	0.25

Showing 1 to 13 of 560 entries, 12 total columns

4.4 The logistic regression model in R studio

All the data were randomly partition in to the training (70%) and testing (30%) datasets, where the training datasets used for training the logistic regression model while the rest testing datasets were used for validating and evaluating of the logistic regression. After importing and analyzing all (560) flood and non flood locations, the following R codes were written and split the data in to training (70%) and testing (30%) datasets randomly.

```
# Split training and testing datasets

# Produce the training datasets for LR model (70%)

training_data <- 0.7 # 70% of all binary data

training<- sample.int(length(My_table4$ID),
                      round(length(My_table4$ID)*training_data))
Train_sample<-My_table4[training,]
View(Train_sample)

# generate the validation datasets

Testinig_data<-My_table4[-training,] # the rest 30% of all the binary data
View(Testinig_data)
```

Having derived the training and validation data sets, the logistic regression model were developed for both the training and testing datasets. Furthermore, application of a binomial distribution of GLM which is a powerful tool for binary response variables, to predict the distribution of spatial flash floods for logistic regression (LR) model in Rstudio [81], [75]. Hence, values obtained from R were applied in the flash flood susceptibility index (the

function Z). These values are coefficients (β) of LR for all the predicting factors and the intercept alpha (α) were generated based on the training datasets. Positive regression coefficients indicate that the dependent variable has a positive correlation with the independent variables, whereas negative regression coefficients indicate that the independent variables have a negative effect on the dependent variable. The following codes written and then generated training and testing logistic regression models respectively as shown below.

```
# Develop multiple logistic regression models usin various variables
#(mhi+lulc+idw+soil+twi+ndvi+spi+lith+drai+slope+disdra)

Model1 <- glm(ID ~ mhi+lulc+idw+soil+twi+ndvi+spi+lith+drai+slope+disdra
,family = binomial,data=Train_sample,maxit=25,na.action=na.omit)
summary(Model1)

# Plot the results of A model
plot(Model)
plot(Model$fitted.values,Train_sample$ID)

# Model using the testing datasets

Model2 <- glm(ID ~.,family = binomial,data=Testinig_data)
summary(Model2)
```

Table 4-5: sample data input for sensitivity analysis

	ID	TWI	Dist_Drainage	SPI	MHI	Soil	Slope	LULC	Drain_Dens	Lith	IDW	NDVI
471	0	1.00	0.00	0.50	0.75	Clay	0.50	Bare land	0.00	Qa	0.25	0.25
289	1	0.00	0.75	0.25	0.50	Clay	0.75	Bare land	0.75	Jh2	0.00	0.00
13	0	1.00	0.50	0.25	0.00	Clay	0.50	Agricultural land	0.50	Qa	0.00	0.25
511	0	0.75	1.00	0.50	0.25	Sandy Loam	1.00	Vegetation	1.00	Jh2	0.25	0.50
385	1	0.00	0.50	0.00	0.50	Clay	0.00	Bare land	0.50	Qa	1.00	0.25
299	0	0.75	0.75	0.25	0.50	Loam	0.00	Bare land	0.75	Jh2	0.50	0.25
377	1	0.75	0.50	0.50	0.25	Clay	1.00	Bare land	0.50	Qa	0.00	0.00
335	1	0.75	1.00	0.00	0.25	Clay	1.00	Bare land	1.00	Qa	0.25	0.50
191	0	1.00	0.50	0.00	0.25	Clay	0.50	Agricultural land	0.50	Jh2	0.00	0.50
129	0	0.75	0.00	0.00	0.00	Clay	0.25	Agricultural land	0.75	Jh2	0.50	0.00
409	0	0.75	0.50	0.50	0.50	Clay	0.00	Bare land	0.50	Qa	0.50	0.75

Showing 1 to 13 of 392 entries, 12 total columns

	ID	TWI	Dist_Drainage	SPI	MHI	Soil	Slope	LULC	Drain_Dens	Lith	IDW	NDVI
6	1	0.00	0.75	0.00	0.00	Clay	0.75	Agricultural land	0.75	Qa	0.00	0.50
10	0	1.00	0.25	0.25	0.00	Clay	0.50	Agricultural land	0.75	Qa	0.25	0.25
14	0	0.00	0.50	0.00	0.00	Clay	0.50	Agricultural land	0.50	Qa	0.00	0.00
18	0	0.00	0.25	0.50	0.00	Clay	0.25	Agricultural land	0.25	Qa	0.25	0.25
23	0	0.25	0.25	0.00	0.00	Clay	1.00	Agricultural land	0.25	Qa	0.25	0.50
26	0	0.75	0.25	0.00	0.00	Clay	0.00	Agricultural land	0.25	Qa	0.75	0.00
28	0	0.25	0.25	0.00	0.25	Clay	0.00	Agricultural land	0.25	Qa	0.00	0.25
29	0	1.00	0.00	0.25	0.00	Clay	0.75	Agricultural land	0.25	Qa	0.00	0.00
31	0	1.00	0.00	0.25	0.00	Clay	0.25	Agricultural land	0.25	Qa	0.00	0.25
36	0	1.00	0.25	0.50	0.00	Clay	1.00	Agricultural land	0.25	Pc2	0.00	0.50
40	1	1.00	1.00	0.75	0.25	Clay	0.50	Agricultural land	0.25	Qa	0.00	0.75

Showing 1 to 13 of 168 entries, 12 total columns

Some of inter-related factors in the model were soil type and LULC, SPI and Drainage density as well as Drainage density and distance from drainage. A sensitivity analysis of these inter-relationships was conducted to check how sensitive the results are to changes in the list of susceptibility factors considered in the logistic regression model. Therefore the following tests were done by eliminating: (A) soil type from the model, (B) LULC from the model, (C) Drainage density from the model, and (D) distance to drainage from the model and (D) SPI from the model. The following sample R codes were written for the purpose of sensitivity analysis of weghts in flash flood conditional factors for logistic regression model eliminating each variable.

```
# Sensitivity analysis in logistic regression model eliminating a variable from the model

Model1 <- glm(ID ~ mhi+lulc+idw+twi+ndvi+spi+lith+drai+slope+disdra # Removing Soil
,family = binomial,data=Train_sample,maxit=25,na.action=na.omit)
summary(Model1)

coeff<-(Model1$coefficients)
write.table(coeff, "coefficientsoill.csv", col.names=TRUE,sep=",")
```

```
# Sensitivity analysis in logistic regression model eliminating a variable from the model

Model1 <- glm(ID ~ mhi+idw+soil+twi+ndvi+spi+lith+drai+slope+disdra # Removing LULC
,family = binomial,data=Train_sample,maxit=25,na.action=na.omit)
summary(Model1)

coeff<-(Model1$coefficients)
write.table(coeff, "coefficientsoill.csv", col.names=TRUE,sep=",")
```

```
# Sensitivity analysis in logistic regression model eliminating a variable from the model

Model1 <- glm(ID ~ mhi+lulc+idw+soil+twi+ndvi+spi+lith+slope+disdra # Removing drainage density
,family = binomial,data=Train_sample,maxit=25,na.action=na.omit)
summary(Model1)

coeff<-(Model1$coefficients)
write.table(coeff, "coefficientsoill.csv", col.names=TRUE,sep=",")
```

Accordingly, the following table shows the relative impacts/ weights and the percentage increase or decrease of the independent variables in the model after eliminating the each of strongly related variables in the model.

Table 4-6: Sensitivity analysis for weights in flash flood conditional factors in the model

			Coefficients/weghts (β) of LR model factors eliminating some parameters									
Conditional factors	All factors	Percentage	A	Percentage	B	Percentage	C	Percentage	D	Percentage	E	Percentage
Intercept	13.86	-23.37	14.09	-24.17	-1.61	-10.49	14.24	-23.73	14.02	-23.50	13.87	-23.27
TWI	0.03	-0.04	0.03	-0.05	0.03	0.17	0.05	-0.08	0.00	0.00	0.05	-0.08
Agricultural land	-14.53	24.49	-14.55	24.96			-14.87	24.78	-14.71	24.65	-14.51	24.34
Bare land	-15.11	25.47	-15.15	25.99			-15.41	25.67	-15.16	25.41	-15.07	25.27
Settelement	-13.88	23.39	-13.88	23.82			-14.02	23.36	-14.05	23.55	-13.85	23.22
Vegetation	-15.45	26.05	-15.51	26.60			-15.98	26.62	-15.39	25.79	-15.35	25.75
IDW	0.47	-0.80	0.47	-0.81	0.41	2.69	0.47	-0.78	0.41	-0.69	0.46	-0.77
Clay	0.18	-0.30			0.75	4.87	0.24	-0.40	0.31	-0.52	0.23	-0.38
Loam	-1.17	1.97			-0.74	-4.80	-1.08	1.81	-0.94	1.57	-1.37	2.30
NDVI	-0.08	0.14	-0.14	0.24	0.03	0.22	-0.12	0.20	0.03	-0.04	0.25	-0.41
SPI	0.75	-1.26	0.80	-1.38	0.68	4.42	0.67	-1.11	0.76	-1.27		
Jh2	-0.66	1.12	-0.69	1.19	-0.28	-1.83	-0.41	0.69	-0.59	0.99	-0.60	1.00
Pc2	-15.51	26.13	-15.50	26.59	14.56	94.61	-15.25	25.39	-15.59	26.13	-15.49	25.98
Qa	-0.30	0.50	-0.31	0.54	-0.26	-1.70	-0.12	0.20	-0.28	0.48	-0.29	0.48
Drain_Dens	0.90	-1.52	0.90	-1.55	1.11	7.24			1.31	-2.20	0.82	-1.38
Slope	-0.54	0.90	-0.55	0.95	-0.59	-3.85	-0.55	0.92	-0.50	0.84	-0.55	0.92
MHI	0.70	-1.17	0.71	-1.21	0.54	3.51	0.69	-1.16	0.72	-1.20	0.75	-1.26
Dist_Drainage	1.00	-1.69	0.99	-1.70	0.76	4.95	1.43	-2.37			1.02	-1.71

As shown in the above table a sensitivity analysis excluding certain conditional factors from the model will result in different weights. For example, eliminating: (A) Soil type from the model the weights of other key factors such as NDVI, increased from 0.14 % to 0.24 %, Agricultural land (24.49 to 24.96%), Bare land (25.47 to 24.99%), Settlement (23.39 to 23.82%) and Vegetation (26.05 to 26.60%) while other factors increased 0.1 to 0.05. If (B) LULC eliminated from the model, weights of other factors were largely vary which means some were largely increased and some were largely decreased, for instance slope, lithology (Qa), lithology (Jh2), and soil (Loam) were decreased from 0.90 % to -3.85%, 0.5% to -1.70 %, 1.12 % to -1.83 %, and 1.97 % to -4.80 % respectively; while other factors were

largely increased as shown in the table. Moreover, the sensitivity analysis for each cases of eliminating shown in the table above, as a result some susceptibility factors have a larger influence than others did on the final weights. Generally, as the analysis revealed in the table, LULC flash susceptibility factor has largest influence in this study which is highest sensitive.

4.5 Multi-collinearity tests result

The following codes and table were produced and the results indicated that there was no problem of multicollinearity among the independent variables used in this study, so logistic regression model was performed with all of the independent variables selected for the spatial flash flood susceptibility prediction.

```
# Check multicollinerity using VIF and TOL
library(car)
vif(Model1)
write.table(vif(Model1), "Vif.csv", col.names=TRUE,sep=",")
tolerance<-1/(vif(Model1))

# Determine how the conditioning factors predict the dependent variable
library(rcompanion)
snell<-nagelkerke(Model1)

Odds_ratio<-exp(coeff(Model1))
View(Odds_ratio)

write.table(Odds_ratio, "Odds_ratio.csv", col.names=TRUE,sep=",")
```

Table 4-7: Multicollinearity checking results of flash flood conditional factors

No	Flash flood conditional factors	Multi-Collinearity Statistics	
		VIF	TOL
1	NDVI	1.28	0.78
2	Slope	1.08	0.93
3	Soil	2.43	0.41
4	SPI	2.15	0.47
5	TWI	1.14	0.88
6	MHI	1.09	0.92
7	LULC	1.4	0.72
8	Lithology	1.29	0.78
9	IDW	1.29	0.77
10	Distance from drainage	1.76	0.57
11	Drainage Density	1.72	0.58

Furthermore, in order to check whether predictor (dependent) variables could explain the response variable the Cox and Snell and Nagelkerke pseudo R^2 were calculated; where the pseudo R^2 value, which can be calculated from 1-(Ln likelihood $_{\text{final step}}$/Ln likelihood $_{\text{initial}}$), indicated that how the logit model fits the dataset [74]. Moreover, a pseudo R^2 equal to 1 indicates a perfect fit, whereas a value of zero shows no relationship [74]. In this study, the results for the final model, Cox and Snell, and Nagelkerke pseudo R^2 values were 0.61 and 0.82 respectively as shown in the table below, which indicated that the model could explain the response variable and the model fits the dataset.

Table 4-8: Pseudo R^2 test of the logistic model with all 11 flash flood conditioning factors

	Pseudo.R.squared
Cox and Snell (ML)	0.61
Nagelkerke (Cragg and Uhler)	0.82

Finally, the figure below the odds ratios for each flash flood conditioning factors were calculated which were essential to assess the effects of each parameter in assessing the susceptibility. The predictor variables that had an odds ratio greater than one were positively related with susceptibility [82] on other hand odds ratio less than one was negatively related with susceptibility, and if the predictor variables that had an odds ratio equal to one, they were considered as a neutral in assessing the Flash flood susceptibility. Among all conditioning factors soil class (Loam), lithology class (Qa and Jh2), NDVI, and slope class were the predictor variables that had an odds ratio of less one and they were had a negative relationship with susceptibility. Other parameters which were drainage density, TWI, IDW, Distance from the drainage, MHI, and SPI had positive relations with susceptibility which had values of greater than one.

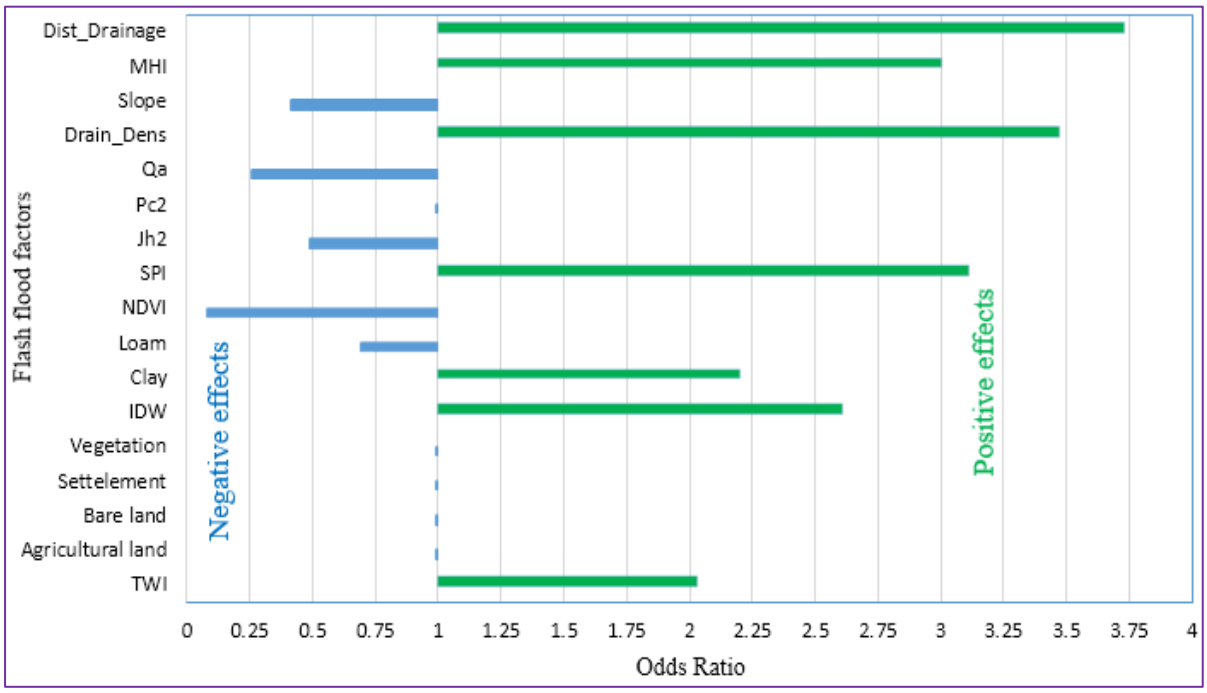


Figure 4-3: Effects and odds ratio for each Flash flood Susceptibility Factor

4.6 Flood susceptibility prediction map

After some iterations of an algorithm in logistic regression in R studio, the relative contributions (β coefficients) of flash flood conditioning parameters were estimated in the logistic regression model. The results of the LR model when it was converged showed that null deviance of 543.06 with 391 degrees of freedom. The residual deviance was 489.67 with 374 degrees of freedom and an Akaike Information Criterion of 525.67. The relationship between independent variables and the results of the analysis using the LR model at different iterations until the model was converged are presented in Table 4-9. Among the independent variables used in this study, factors such as the TWI, SPI, distance from drainage, Drainage density, rainfall (IDW), soil (Clay class), and MHI had positive relationships and effects with the flash flood susceptibility. On the other hand for instance lithology classes (Pc2, and Qa), Land use land cover class, NDVI, Slope degree had negative effects on flash flood susceptibility occurrence. For categorical variables such as the lithology, and land use land cover and soil type classes were multiplied by each class coefficients.

Table 4-9: Determination of flash flood conditioning factors influencing weights

Classes of Flash flood conditioning factors		Coefficients (β) of LR model parameters in each (1-14) iterations						
		2	4	6	8	10	12	14
(Intercept)		1.81	3.86	5.86	7.86	9.86	11.86	13.86
TWI		0.03	0.03	0.03	0.03	0.03	0.03	0.03
Agricultural land		-2.48	-4.53	-6.53	-8.53	-10.53	-12.53	-14.53
Bare land		-3.06	-5.10	-7.11	-9.11	-11.11	-13.11	-15.11
Settelement		-1.83	-3.87	-5.88	-7.88	-9.88	-11.88	-13.88
Vegetation		-3.40	-5.45	-7.45	-9.45	-11.45	-13.45	-15.45
IDW		0.47	0.47	0.47	0.47	0.47	0.47	0.47
Clay	Soil	0.18	0.18	0.18	0.18	0.18	0.18	0.18
Loam		-1.16	-1.17	-1.17	-1.17	-1.17	-1.17	-1.17
NDVI		-0.08	-0.08	-0.08	-0.08	-0.08	-0.08	-0.08
SPI		0.74	0.75	0.75	0.75	0.75	0.75	0.75
Jh2		-0.66	-0.66	-0.66	-0.66	-0.66	-0.66	-0.66
Pc2	Lithology	-3.45	-5.50	-7.50	-9.51	-11.51	-13.51	-15.51
Qa		-0.30	-0.30	-0.30	-0.30	-0.30	-0.30	-0.30
Drain_Dens		0.90	0.90	0.90	0.90	0.90	0.90	0.90
Slope		-0.53	-0.54	-0.54	-0.54	-0.54	-0.54	-0.54
MHI		0.69	0.70	0.70	0.70	0.70	0.70	0.70
Dist_Drainage		1.00	1.00	1.00	1.00	1.00	1.00	1.00

4.7 Flash flood susceptibility index model

Finally, the flash flood susceptibility index was generated summing up all the flash flood conditioning factors which were multiplied by the coefficients obtained in the analysis of logistic regression model as shown in equation (4-1). The result of flood susceptibility index was reclassified into five classes as indicated in figure below.

$$\begin{aligned} Z = \big[&(\text{IDW} * 0.47) + (\text{Distance Drainage} * 1.0) \\ &+ (\text{Drainage density} * 0.90) + (\text{NDVI} * -0.08) \\ &+ (\text{MHI} * 0.70) + (\text{SPI} * 0.75) + (\text{TWI} * 0.03) \\ &+ (\text{Slope} * -0.54) + \text{Lulc}(\text{Vegetation} * -15.45) \\ &+ (\text{Bare land} * -15.11) + (\text{Settlement} * -13.88) \\ &+ (\text{Agricultural land} * -14.53) \\ &+ \text{Lithology}((\text{Pc2} * -15.51) + (\text{Qa} * -0.30) \\ &+ (\text{Jh2} * -0.66) \\ &+ \text{Soil}\big((\text{Loamy sand} * -0.14) + (\text{Loam} * -1.17) \\ &+ (\text{Clay} * 0.18)\big) + 13.86\big] \end{aligned} \quad (4\text{-}1)$$

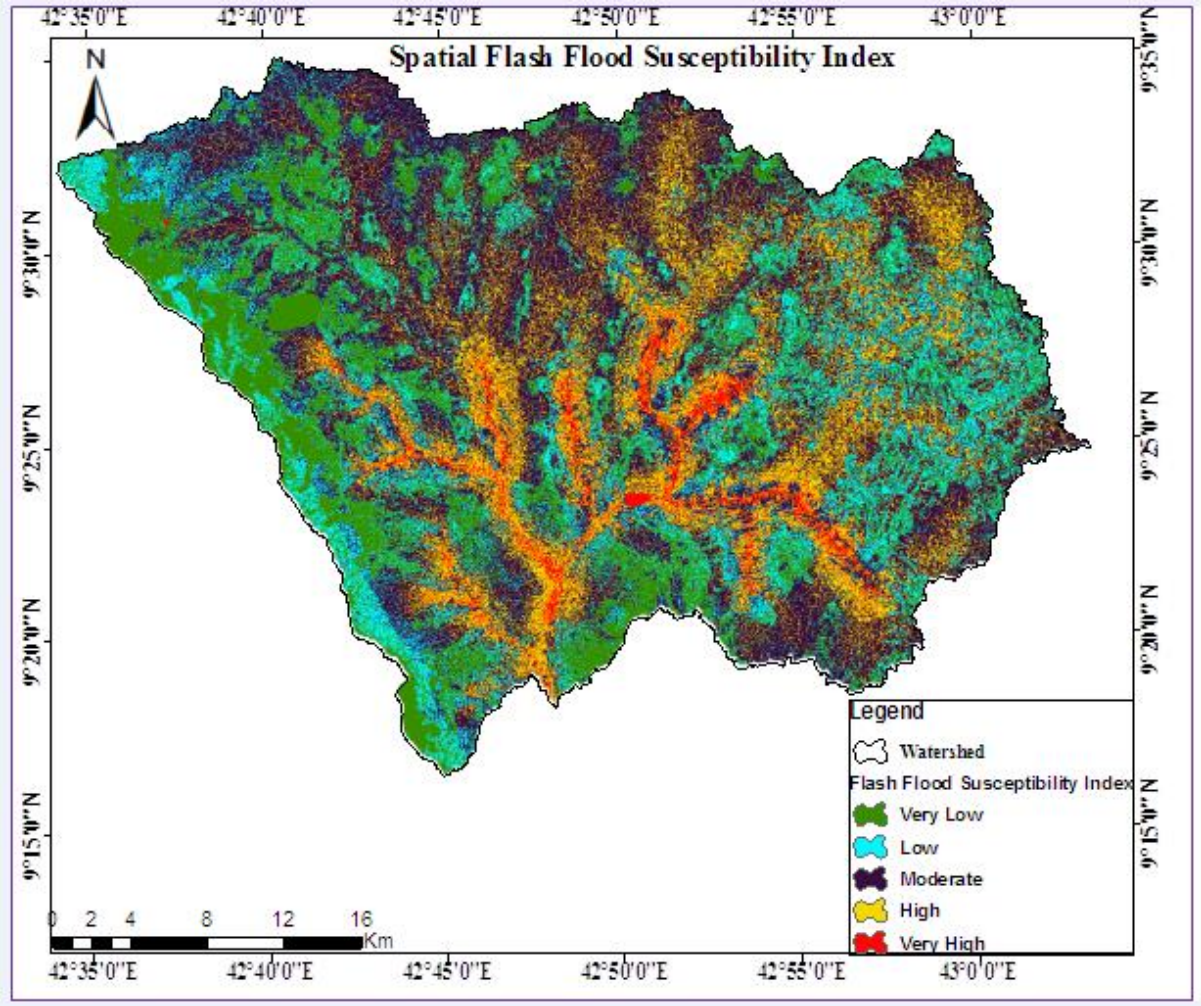


Figure 4-4: Flash flood Susceptibility index

In addition, the spatial flash flood probability map was produced based on the equation (3-15) discussed in section 3 which was between 0 and 1 values. The spatial flash flood susceptibility probability map was generated in Arc-GIS raster calculator showing the nearest locations to the drainage, low laying areas, high stream densities, stream orders (4, 5 and 6), and settlement, waterbody, bareland, waterbody, and settlement classes of LULC, were more susceptible for flash flood in the study area. The generated map is shown in figure 4-5.

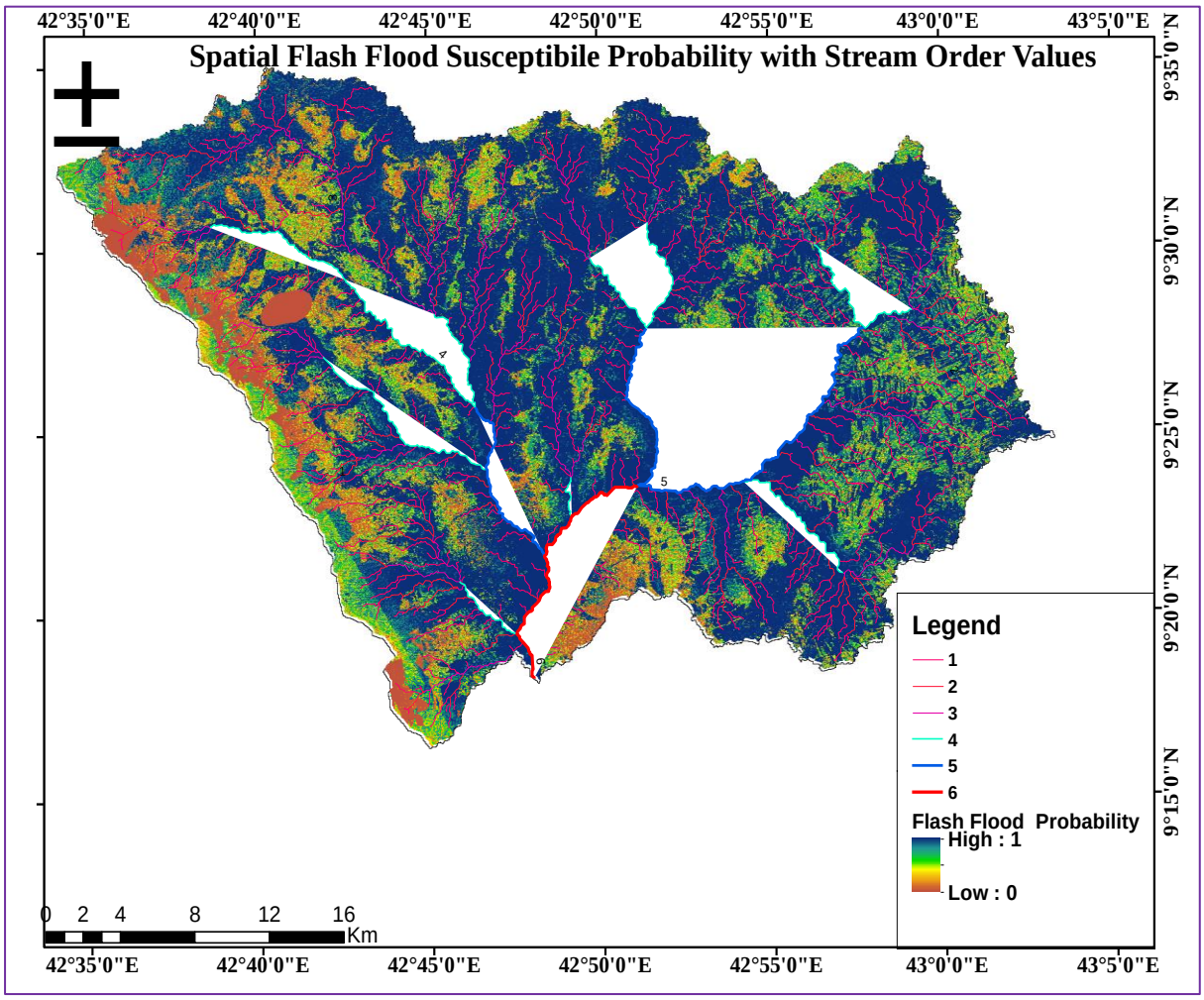


Figure 4-5: Flash flood probability map with stream orders

To produce flood susceptibility map, the values obtained from Rstudio were exported into ArcGIS software to build the model using equation (4-1). After the results of the estimated spatial flash flood probability was produced in a raster map which varied from zero to one, the probabilities were then classified into five categories of Very High, High, Moderate, Low and Very Low susceptibility zones to produce the flood susceptibility maps (Figure 4-7).

The spatial flash flood susceptibility prediction map in the figure 4-5 showed that areas of very high and high susceptibility accounted for 2% and 20% of the whole catchment, based on the respectively. These susceptible classes were located mainly in the low elevation land and near the drainage which passes through and at the side of capital city of Somali regional state of Ethiopia, which is Jigjiga. Moreover, the susceptibility of floods was high in the nearest parts of the drainage line, and as the distance from the drainage increases, the occurrence of floods decreases; this was confirmed during the last consecutive flash flood

occurrences in the area in the field. People who live in these areas should be aware of the potential hazards caused by flash flooding because settlements of people are located in highly susceptible areas.

As Figure below shown, very low (20%) and low susceptibility (29%) classes are mainly located in high elevation and high slope angle areas.

Table 4-10: Classification of spatial flash flood susceptibility prediction map into five classes

No	Count Pixel	Susceptibility Class	Area in Percent
1	228362	Very Low Susceptible	20%
2	333348	Low Susceptible	29%
3	305131	Moderate Susceptible	26%
4	262917	High Susceptible	23%
5	22141	Very High Susceptible	2%

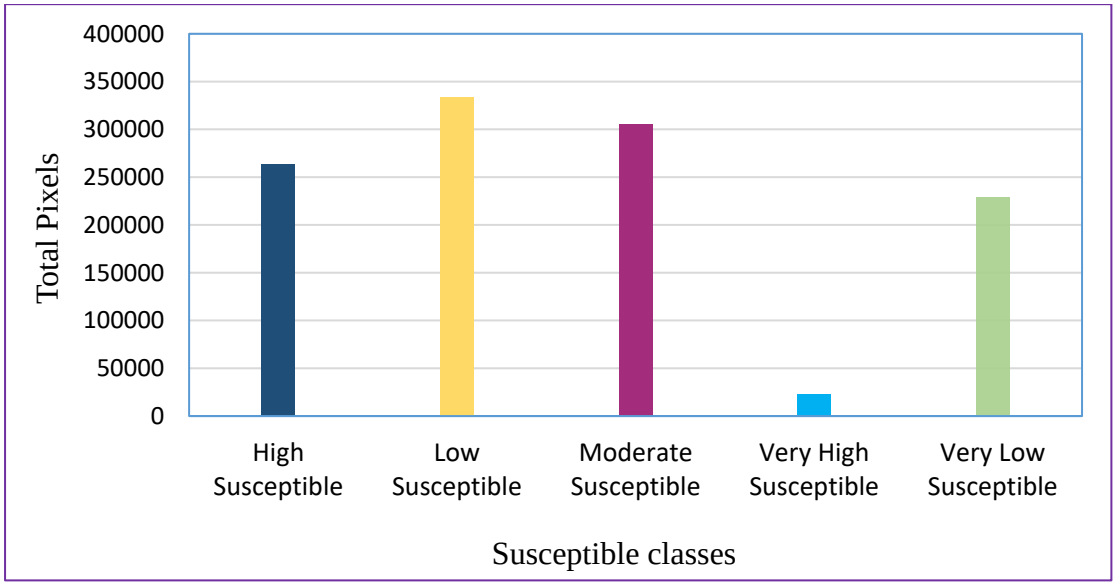


Figure 4-6: Flash flood Susceptibility classes pixel value

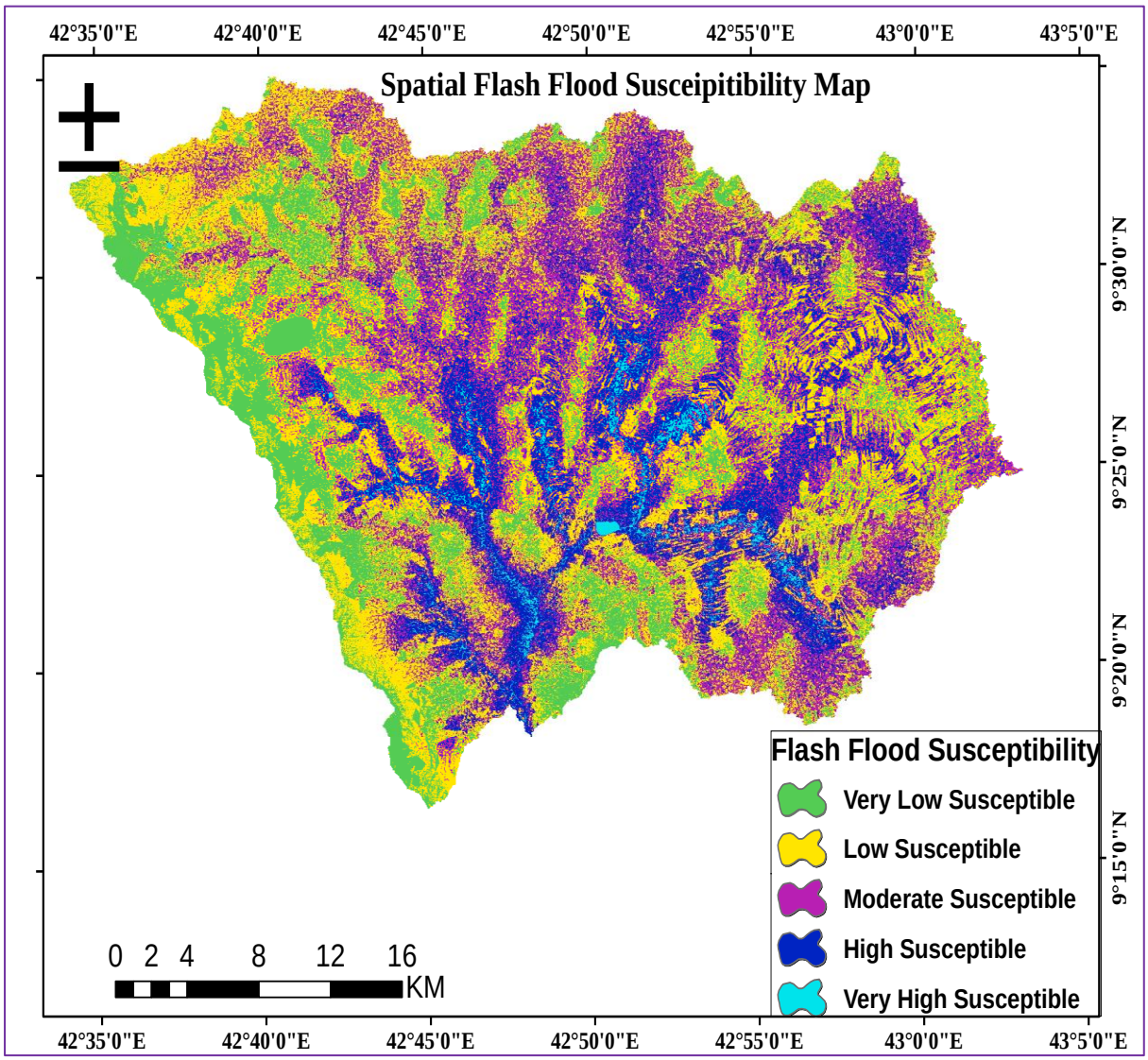


Figure 4-7: Spatial flash flood susceptibility map produced by LR model

4.8 Accuracy assessment and model validation

4.8.1 Accuracy assessment results

Overall accuracy, sensitivity, and specificity measure the proportion of training and testing, flooded, and non-flooded samples, respectively, that are correctly classified.

Generally, the following Rstudio codes were written for analyzing the classification accuracy.

```
# Ealuate the classification accuracy of Training and Testing dataset models

pred<-predict(Model1,Train_sample,type="response")
View(pred)

predd2<-predict(Model2,Testinig_data,type="response")
View(predd2)

# Overall model accuracy

Predictt<-predict.glm(Model,Testinig_data,type="response") # This gives the probabilitie b/n 0 & 1
pr<-round(Predictt) # round values >0.5 (cutoff value) to 1 else 0 for Training dataset
pr1<-round(pred)
pr2<-round(predd2) # round values >0.5 (cutoff value) to 1 else 0  for Validation Datasets

# Comparasion of predicted and observed data with contingency matrix for classification accuracy

contg_matrix<- table(observed = Testinig_data$ID, predicted =pr) # Over all Model accuray code

# Classification accuracy for both Training and testing Datasets

contg_matrix1<-table(observed = Train_sample$ID, predicted =pr1)
contg_matrix2<-table(observed = Testinig_data$ID, predicted =pr2)
write.table(contg_matrix2, "contg_matrix22.csv", col.names=TRUE,sep=",") # produce table for cont_matrix
# %tage of Classification Accuracy

Accuracy <- sum(diag(contg_matrix))/sum(contg_matrix)*100 # 92.26% accurate

# OR
confusionMatrix(contg_matrix) # caret package
```

Table 4-11: Classification accuracy using training dataset summary of observed and predicted Flash flood susceptibility of the logistic regression model (the cut off value is 0.5)

	Observed Flash flood susceptible absent pixel (0)	Predicted Flash flood susceptible present pixel (1)	Total	Percentage correct
Flash flood susceptible abscent pixcel (0)	185	14	199	0.926
Flash flood susceptible present pixcel (1)	12	181	193	0.938
Total	197	195	392	93.2

Table 4-12: Classification accuracy using validation dataset summary of observed and predicted Flash flood susceptibility of the logistic regression model (the cut off value is 0.5)

Observed	Predicted			
	Flash flood susceptible absent pixel (0)	Flash flood susceptible present pixel (1)	Total	Percentage correct
Flash flood susceptible abscent pixcel (0)	72	9	81	0.889
Flash flood susceptible present pixcel (1)	7	80	87	0.909
Total	82	86	168	89.9

In the above tables, the confusion matrices were created for both the training and testing datasets. The confusion matrix represents the ratio of correctly predicted observations and demonstrates the number of true positive, false positive, true-negative and false-negative observations [72]. Consequently, the outcomes of the confusion matrices produced by the training and testing datasets were compared. In the training dataset totally, 93.2% of the 392 pixels were correctly categorized for the training dataset. Furthermore, 198 Flash flood susceptible absent pixels with 90% were without flash flood occurrence and 194 Flash flood susceptible present pixels with 93% were categorized wrongly. The validation results for testing dataset (Table 4-12) represents that 82 of the total 168 pixels were correctly categorized (88.9 %). However, 5 of 86 pixels with Flash flood susceptible present were wrongly categorized while 81 flash flood susceptible present pixels (90.9%) were correctly categorized as flash flood susceptible present. Finally, models were evaluated using statistical parameters such as Correlation Coefficient (CC), Root Mean Squared Error (RMSE), Mean average error (MAE), and Nash-Sutcliffe Efficiency (NSE). The values of these statistical parameters were 0.85, 0, 1 and 0 for training datasets and 0.89 ,0, 1 and 0 for testing datasets respectively as shown below. Thus, the statistical parameters were approximately equal indicating that the model was most appropriate for predicting.

```
# Statistical evaluation of the model

#Check how good is the model on the training set - correlation^2, RME and MAE

train.corr <- round(cor(pred, Train_sample$ID), 2)
train.RMSE <- round(sqrt(mean((pred - Train_sample$ID)^2)))
train.MAE <- round(mean(abs(pred - Train_sample$ID)))
train.NSE<- 1-round(sum(Train_sample$ID-pred )^2)/sum((pred-mean(Train_sample$ID))^2)
c(train.corr,train.MAE,train.NSE,train.RMSE)
# 0.85 0 1 0

# Check how good is the model on the validation set - correlation^2, RME and MAE
valid.corr <- round(cor(predd2, Testinig_data$ID),2)
valid.RMSE <- round(sqrt(mean((predd2 - Testinig_data$ID)^2)))
valid.MAE <- round(mean(abs(predd2 - Testinig_data$ID)))
valid.NSE<- 1-round(sum(Testinig_data$ID-predd2 )^2)/sum((predd2-mean(Testinig_data$ID))^2)
c(valid.corr,valid.MAE,valid.NSE,valid.RMSE)
# 0.9 0 1 0
```

4.8.2 Validation of the model result (spatial flash flood susceptibility prediction map)

The area under the ROC curve (AUC) is the statistical summary of the overall model performance and used to validate. Moreover, it is a popular, comprehensive quantitative method of accuracy assessment that can be used to evaluate the prediction and success rates in the developed model [19], [35]. A ROC curve is a scientific technique used to describe the efficiency of probabilistic and deterministic detection and prediction systems, and it is basically formed by plotting the trade-off between the true positive rate (sensitivity) on the X-axis and the false positive rate (1-specificity) on the Y-axis [75]. The model generated using training flash flood locations coluld not be validated based on this data because it does not represent the model's real efficiency accurately. ROC curves can be divided into success rate curves and prediction rate curves according to the dataset used. A success rate curve is formed using a training dataset (70%) and represents how well the model fits the flash flood susceptible observed. A prediction rate curve is formed using a testing/validation (30%) dataset and evaluated the accuracy of predicting the spatial flash flood susceptibility, based on the application of conditioning factors for the model [19], [35]. The prediction and success rate curves are plotted in Figures 4-8 and 4-9 respectively. The Area Under the ROC curve was computed in equation (4-2) [83].

$$\mathrm{AUC} = {(\Sigma \mathrm{TP} + \Sigma \mathrm{TN})}\Big/{(P + N)} \qquad (4\text{-}3)$$

where P is the total number of floods and N is the total number of non-flood pixels.

Generally, in this study training datasets of 392 points randomly partitioned (70%) of inventory locations of flash flooding were used to obtain the success rate, but when a model

is generated from the training flood layer, it cannot be used for validation, as it is not representative of the model's actual efficiency as well as it cannot be used for calculating predictive capability, which indicates efficiency to predict flash floods in an area [35]. Thus, the 168 points (30% of flash flooded locations) were partitioned for measuring the prediction rate.

Finally, the value of AUC is between 0.5 and 1.0, and a value closer to 1 indicates a model with a better predictive capability. The value of AUC = 1 represented the highest accuracy, showing that the model was completely capable of predicting the spatial flash flood occurrence without any bias [69]. Therefore, AUC values close to 1.0 are considered to show that a model is precise and trustable [84]. The prediction curve is shown below, the AUC value was found to be 0.96 corresponds to a prediction accuracy of 96%; which indicated that the model was completely capable of predicting the spatial flash flood susceptible occurrence and the adopted flash flood susceptibility method can predict the likelihood an area will be flooded quite accurately. Moreover, the classification accuracy of the model using the validation datasets was 92.3% which is shown in table above. Accordingly, the following Rstudio codes were written for generating the ROC and compute AUC.

```
# Develop Receiver operating Characteristic (ROC) curve
# and compute Area under the curve (AUC)

library(ROCR)

# Prediction rate

pred_roc<- prediction(Predictt,Testinig_data$ID)
evaluation<-performance(pred_roc,"acc")

# success rate

pred_roc2<- prediction(pred,Train_sample$ID)
evaluation2<-performance(pred_roc2,"acc")


# Plot the performance of prediction
par(mar=c(5,4,1,1))
plot(evaluation)
plot(evaluation2)
# Identify the best value of cut off with best accuracy
maxx<-which.max(slot(evaluation,"y.values" )[[1]])
acc<-slot(evaluation,"y.values")[[1]][maxx] # where "y.values" are accuracy measures
cut<-slot(evaluation,"x.values")[[1]][maxx] # where "x.values" are cut off measures

print(c(Accuracy=acc, cutoff=cut))
```

```
# Develop receiver operating characterstics curve for prediction rate

pred_rocr<-performance(pred_roc, measure = "auc", x.measure = "cutoff")
pred_rocr@y.values[[1]]<- round(pred_rocr@y.values[[1]],digits = 2)
prf.tpr.fpr<-performance(pred_roc,"tpr", "fpr") # ture and false positive rate

# Develop receiver operating characterstics curve for success rate

pred_rocr2<-performance(pred_roc2, measure = "auc", x.measure = "cutoff")
pred_rocr@y.values[[1]]<- round(pred_rocr2@y.values[[1]],digits = 3)
prf.tpr.fpr2<-performance(pred_roc2,"tpr", "fpr") # ture and false positive rate

# Visualize Area under curve (AUC) for validation (Prediction rate curve)

plot.new()
par(mar=c(5,4,1,1))
AUC<-plot(prf.tpr.fpr,colorize=T, main=paste("Area under the curve:",
                                             pred_rocr@y.values[[1]]),text.adj=c(-0.5,1.5),
     cex.main=0.8,cex.lab=0.8,cex.axis=0.8,mgp=c(2,2,0),lwd=3)
abline(a=0,b=1,lty = 2,col="Blue",lwd=2)

# Visualize Area under curve (AUC) for Training (success rate curve)
plot.new()
par(mar=c(5,4,1,1))
AUC<-plot(prf.tpr.fpr2,colorize=T, main=paste("Area under the curve:",0.98),text.adj=c(-0.5,1.5),
          cex.main=0.8,cex.lab=0.8,cex.axis=0.8,mgp=c(2,2,0),lwd=3)
abline(a=0,b=1,lty = 2,col="Blue",lwd=2)
```

Table 4-13: Overall Classification accuracy using validation dataset summary of observed and predicted flash flood susceptibility of the logistic regression model (the cut off value is 0.5)

Observed	Predicted		
	Flash flood susceptible absent pixel (0)	Flash flood susceptible present pixcel (1)	Percentage correct
Flash flood susceptible abscent pixcel (0)	75	7	91.5
Flash flood susceptible present pixel (1)	6	80	93.02
Overall Accuracy			92.3

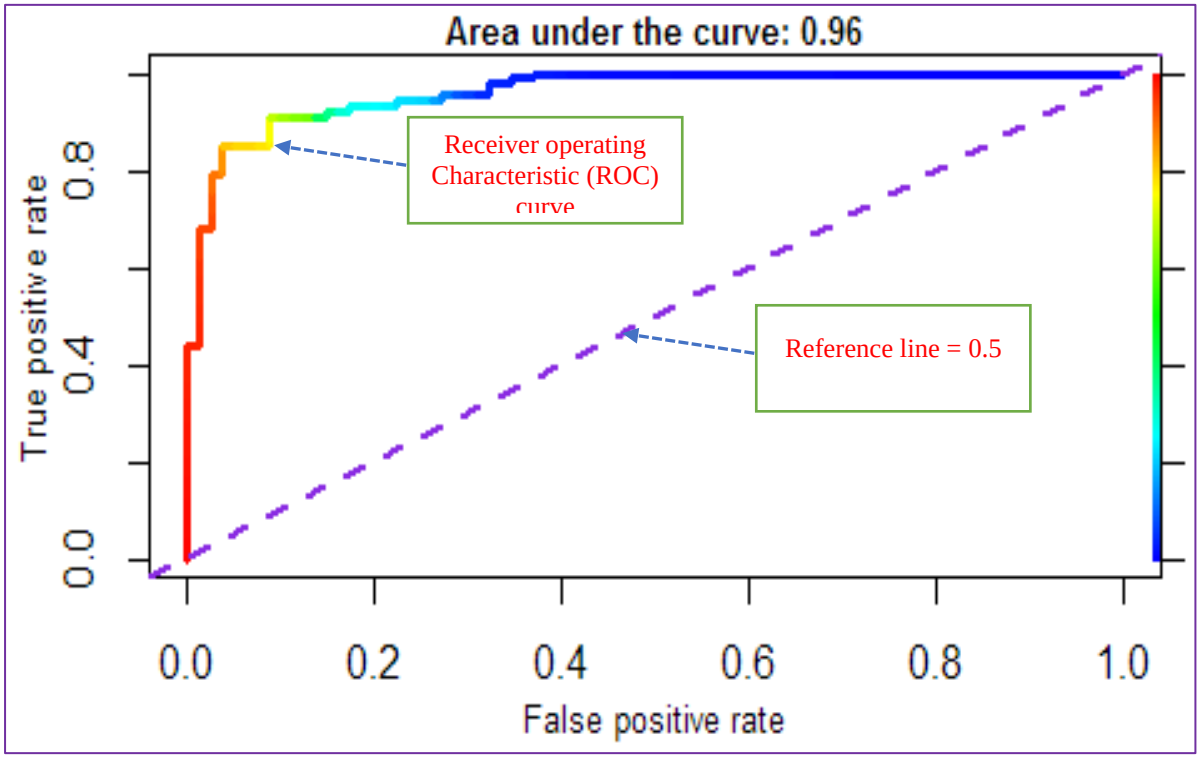


Figure 4-8: Prediction rate (using testing data) curve and its AUC

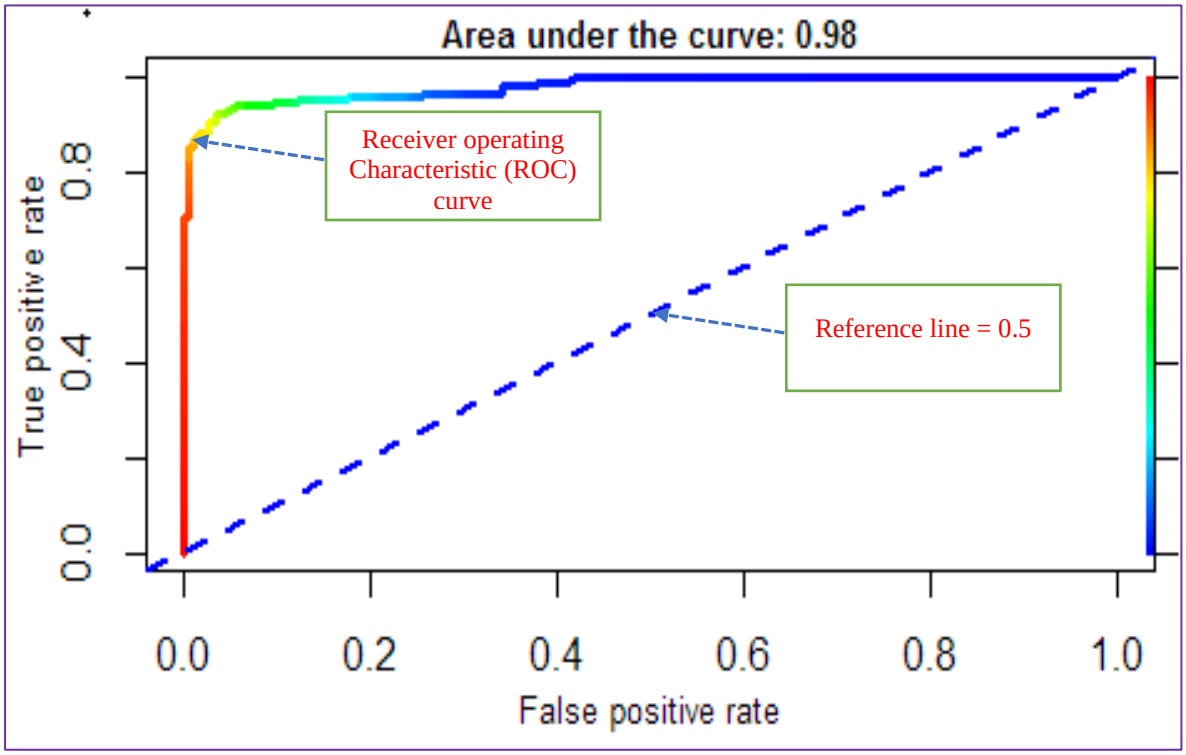


Figure 4-9: Succes rate (Using training data) curve and its AUC

The figures above revealed that the performance measurement Area under the ROC curve were 0.96 and 0.98 with the classification accuracy of 92.3 and 93.2 percent respectively.

4.9 Evaluation of final result with the past flash floods occurred in the study area

In this study, the result of the Spatial Flash Flood Susceptibility Prediction Map from logistic regression was evaluated using overlaying the result on google earth pro image (show the flash flood susceptible areas due flood event occurred in previous area). In addition, when the overlaying was done, the time slider was moved forward and backward between acquisition dates so as to show the historical imagery. Hence, the results of overlaying revealed that the flash flood susceptibility agreement was most completely agree. Some of the results are shown in the following sample pictures in different times of flash flood susceptibility occurred in the area. In the procedure very high and high susceptible pixels used for comparison. Furthermore, the pictures were located along side the drainage networks which basically indicated that the spatial flash flood susceptibility prediction map was adequately predict and showed the sucsceptible areas.

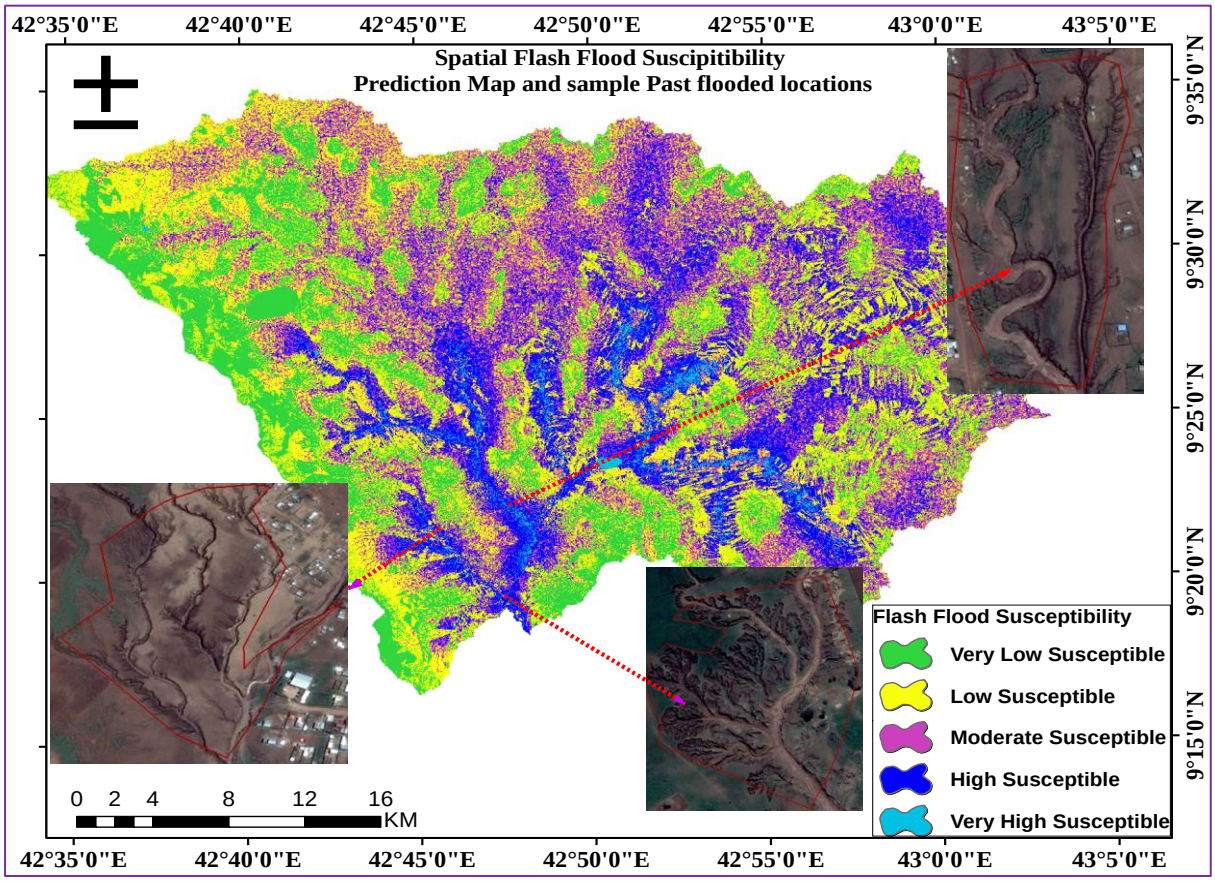


Figure 4-10: Evaluation of the spatial flash flood suceptibility results with past flash flood events in the area

CHAPTER 5 CONCLUSION AND RECOMMENDATION

5.1 Conclusion

Jerer upstream watershed (Jigjiga) is one of the semi-arid watershed in Eastern Ethiopia with average precipitation of 712 mm/year. The present study reavealed or observed that the area was under rapid increase of flash flood susceptibility in the past subsequent years.

Spatial analyst and other tools in ArcGIS, ERDAS 2014 and Rstudio were used for analyzing, combining, correlating, manipulating, deriving, processing and generating all flash flood conditioning factors and historical inventory maps.

In the study, eleven hydrologic, land cover, geologic, topographic flash flood factors were used as independent variables and about 560 flood and non flood locations were used as independent variables in order to produce spatial flash flood susceptible prediction map. Furthermore, 70% (392) of the data were used for training the model. In addition, 168 locations (30%) are used for validation of model. Finally, the logistic regression model developed in R studio was used to determine relative impact weight of flash flood causative factors to get a composite flash flood susceptibility index map. All flash flood conditioning factors data were then integrated to prepare a final flash flood susceptible prediction map for the study area.

The spatial relationshiop and influencing rates of each flash flood factor was evaluated and the weights for each was generated. Accordingly, the final spatial flash flood susceptibility index model and itse map was prepared by summing of weighted multiplications of the reclassified conditioning factor maps. Furthermore, the susceptibility map was produced using spatial flash flood susceptibility index, as a result the probability values between 0 and 1 were generated; followed by its classification based on susceptibility. Finally, the developed spatial flash flood susceptibility prediction map was realy informative, create awareness for public and city planners on probable flash flood-prone areas. Therefore, the logistic regression model can be used as an alternative approach in the spatial flash flood susceptibility prediction unless hydraulic and hydrologic models used due to extensive gauged flow data requirements in arid and semi-arid areas (un-gauged cathments)

Moreover, the reliability of the susceptibility map was validated using the area under the ROC curve using validation set. It showed good accuracy of the model, as the area under the

curve was 0.96, corresponding to a prediction accuracy of 96%; which means that the spatial flash flood susceptibility prediction map was reliable and reasonable. Furthermore, the AUC value with prediction accuracy of 96% which indicated that the model was completely capable of predicting the spatial flash flood susceptible occurrence and the adopted flash flood susceptibility method can predict the likelihood an area will be flooded quite accurately. Consequently, this study could serve as an effective guide for furtherly aid in understanding flood probabilities, defining urban expansion and planning strategies for city planners and watershed management strategies leading to more preparedness for future flooding events.

The Cox and Snell and the Negelkerke pseudo R^2 tests, confusion matrix, statistical parameters, multi-collinearity tests and others were used to evaluate the goodness of the fit of the logistic regression model. Because, all were satisfied, which indicates that the fitting result was good. Finally, the flash flood susceptibility map that produced by the logistic regression model was divided into the following five classes such as very low, low, moderate, high, and very high susceptibile area with percentages of 20%, 29%, 26%, 23 %, and 2% respectively.

5.2 Recommendation

It is recommended that some watershed management in the flash flood susceptible zones especially fourth, fifth and sixth stream order zones as well as near the drainage networks are required for infiltrating, and harvesting runoff water come from upstream sides so as to do, this study provides details on the flash flood susceptible zones. If so, the flash flood factors specially sensitive parameter LULC changed (in the way of reducing flash flood susceptibility), so that it affects the reduction of very high and high flash flood extent obtained by the model.

In order to apply this method in other regions or areas an intensive fieldworks/ surveys and satellite data reviews are mandatory for collecting local historical flash flood inventories. This method is highly dependent on various site-dependent flash flood conditional factors (hydrologic, topographic and environmental flash flood influencing factors) so that selecting, processing, generating and classifying these factors according to the flash flood influence should have to be done properly for precise and accurate results, unless the shortage and uncertainty of flood inventories and conditional factors, might limit the accuracy of model results. If someone properly do the above, the model can easily be applied as an alternative approach for spatial flash flood susceptibility prediction in other interested hydrologic/ hydraulic data-sparse areas.

REFERENCES

[1] Haoyuan H., Wei C., Chong X., Ahmed M. Youssef, Pradhan B. & Dieu T.(2016), "Rainfall-induced landslide susceptibility assessment at the Chongren area (China) using frequency ratio, certainty factor, and index of entropy" pp. 1-15.

[2] Fekadu A. (2018), "Detecting Flash Flood Hazard Areas Using Geo-Spatial – Based Analytic Hierarchy Process in Weidie Watershed South Western Ethiopia," vol. 7, pp. 1–5.

[3] Maghraby M. M. S. El and Niyazi B. (2014)"Assessment of surface runoff in arid, data scarce regions; an approach applied in Wadi Al Hamd, Al Madinah Al Munawarah, Saudi Arabia" pp. 271-289.

[4] Kelsch M. (2001) "Hydrometeorological characteristics of flash floods", pp. 181–193.

[5] Xing C. and Isaacowitz D. M. (2013), "Ocha Weekly humanitarian highlights in Ethiopia", vol. 30.

[6] Abuzied S., Yuan M., Ibrahim S., Kaiser M., and Saleem T. (2016) "Geospatial risk assessment of fl ash fl oods in Nuweiba area , Egypt," vol. 133, pp. 54–72.

[7] Sciences .E. E., N. England, B. Pradhan, U. Putra, M. Neamah, and J. Universitiy (2014) "Flood susceptibility mapping using integrated bivariate and multivariate statistical models," vol.72, pp. 4001-4015.

[8] Lee S. and Resources M. (2016), "The spatial prediction of landslide susceptibility applying artiffcial neural network and logistic regression models , Korea," pp. 117–132.

[9] Rahmati O., Zeinivand H., and Besharat M. (2016), "Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis," vol. 5705, pp. 1000-1017.

[10] United nation office, F. O. R. The, C. Of, and H. Affairs (2018), "United nations office For the coordination of humanitarian affairs annual reports".

[11] Das S. (2019), "Geospatial mapping of flood susceptibility and hydro-geomorphic response to the floods in Ulhas basin , India," *Remote Sens. Appl. Soc. Environ.*, vol. 14, pp. 60–74.

[12] Ouma.Y. O. and Tateishi R. (2014), "Urban flood vulnerability and risk mapping using integrated multi-parametric ahp and gis: methodological overview and case study assessment," pp. 1515–1545.

[13] Gebre G. Y. SL (2015), "Flood hazard assessment and mapping of flood inundation area of the Awash River Basin in Ethiopia using GIS and HEC-GeoRAS/HEC-RAS

Model," vol. 05..

[14] Leuleseged T., Mohamed A., Willems P. (2015), "At-site and regional flood frequency analysis of the upper awash sub – basin in the ethiopian plateau",pp. 1-15.

[15] Yang .J., Townsend R. D., and B. Daneshfar (2005), "A GIS-based approach to river network floodplain delineation", vol. 83, pp. 517–524.

[16] Pradhan .B. (2009), "Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing", vol. 9, pp. 1-18.

[17] Tehrany .M. S., Shaban F.i, Jebur M. N., W. Chen, and X. Xie (2017), "GIS-based spatial prediction of flood prone areas using standalone frequency ratio, logistic regression, weight of evidence and their ensemble techniques", .

[18] Wei Chen, Xiaoshen Xie Jiale Wang, Biswajeet Pradhan, Haoyuan Hong ,DieuTienBui , Zhao Duan , Jianquan M (2017), "A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility" vol.151 , pp. 147–160 ,.

[19] Pradhan .B. and S. Lee (2015), "Utilization of optical remote sensing data and gis tools for regional landslide hazard analysis using an artificial neural network model",vol. 14.

[20] Jones .S. and G. Science (2017), "Evaluating the variations in the flood susceptibility maps accuracies due to the alterations in the type and extent of the flood inventory", vol. XLII, pp. 209–214.

[21] Elkhrachy .I. (2015) , "Flash flood hazard mapping using satellite images and gis tools : A case study of Najran City , Kingdom of Saudi Arabia (KSA)," pp. 261–278.

[22] Komi .K., J. Neal, Trigg M. A., and B. Diekkrüger (2017), "Modelling of flood hazard extent in data sparse areas: A case study of the Oti River basin, West Africa", vol. 10, pp. 122–132.

[23] Prakash P. (2018) "Study the 2013 flood damages and risk assessment in kedarnath, himalaya area using geoinformatic techniques", vol. 08, pp. 10–13.

[24] Diakakis .M., G. Deligiannakis, A. Pallikarakis, and M. Skordoulis (2016), "Factors controlling the spatial distribution of flash flooding in the complex environment of a metropolitan urban area. The case of Athens 2013 flash flood event", vol. 18, pp. 171–180.

[25] Billi .P., Y. T. Alemu, and R. Ciampalini (2015), "Increased frequency of flash floods

in Dire Dawa, Ethiopia", vol. 76, pp. 1373–1394.

[26] Christopher B., Vicente B. Thomas F. Qin D. (2012), "Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation Special"

[27] Easterling D. R., Evans J. L., Groisman P. , Karl T. , Kunkel K., and P. Ambenje *(*2000), "Observed variability and trends in extreme climate events", vol. 81 , pp. 417–426.

[28] Seleshi .Y. and P. Camberlin (2006), "Recent changes in dry spell and extreme rainfall events in Ethiopia" vol. 83, pp. 181–191.

[29] Conway .D. and E. L. F. Schipper (2011), "Adaptation to climate change in Africa", vol. 21, pp. 227–237.

[30] Wondim .Y. K. (2016) , "Flood hazard and risk assessment using gis and remote sensing in lower Awash Sub-basin , Ethiopia," vol. 6, pp. 69–86.

[31] Samanta .S., D. Kumar, and P. Babita, (2018) "Flood susceptibility analysis through remote sensing , GIS and frequency ratio model", vol. 8, pp. 1–14.

[32] Taha M. M. N., Elbarbary S. M., Naguib D. M., and I. Z. El-shamy (2017), "Flash flood hazard zonation based on basin morphometry using remote sensing and GIS techniques : A case study of Wadi Qena basin , Eastern Desert", vol. 8 , pp. 157–167.

[33] El-magd .I. A., Hermas E., and M. El Bastawesy (2010), "GIS-modelling of the spatial variability of flash flood hazard in Abu Dabbab catchment , Red Sea Region , Egypt"., vol. 13, pp. 81–88.

[34] Bishaw .K. (2012), "Application of GIS and Remote Sensing Techniques for Flood Hazard and Risk Assessment: The Case of Dugeda Bora Woreda of Oromiya Regional State, Ethiopia.", pp. 1–17.

[35] Bui .D. T., M. Panahi, H. Shahabi, V. P. Singh, and A. Shirzadi (2018), "Novel Hybrid Evolutionary Algorithms for Spatial Prediction of Floods", pp. 1–14.

[36] Lim .J. (2018), "Flood mapping using multi-source remotely sensed data and logistic regression in the heterogeneous mountainous regions in North Korea", pp. 10–14.

[37] Pradhan .B. and. S. Lee (2010), "Delineation of landslide hazard areas on Penang Island , Malaysia , by using frequency ratio , logistic regression , and artificial neural network models," pp. 1037–1054.

[38] Al-juaidi A. E. M., A. M. Nassar, and Al-juaidi O. E. M. (2018), "Evaluation of flood susceptibility mapping using logistic regression and GIS conditioning factors".

[39] Diakakis .M., G. Deligiannakis, A. Pallikarakis, and M. Skordoulis (2016), "Factors controlling the spatial distribution of fl ash flooding in the complex environment of a metropolitan urban area: The case of Athens 2013 fl ash fl ood event", vol. 18, pp. 171–180.

[40] Costache .R. (2019)," Flash-Flood Potential assessment in the upper and middle sector of Prahova river catchment (Romania). A comparative approach between four hybrid models," vol. 659, pp. 1115–1134.

[41] Tehrany .M. S., L. Kumar, and M. N. Jebur (2019), "Evaluating the application of the statistical index method in flood susceptibility mapping and its comparison with frequency ratio and logistic regression methods", vol. 10, pp. 79–101.

[42] Kamala .M. and Samynathan .M. (2018), "Morphometric analysis of drainage basin using GIS techniques a case study of amaravathi river basin", vol. 9, pp. 28142-28147.

[43] Daniele N., Zed Z., Li-Pen W., Christian O.,Waldo L., and Jean-L. (2015), "A Comparative Analysis of TRMM – Rain Gauge Data Merging Techniques at the Daily Time Scale for Distributed Rainfall – Runoff Modeling Applications", pp. 2153–2168.

[44] González .M. L., M. E. Mendoza, G. Bocco, and B. Solís (2018), "Flood susceptibility in rural settlements in remote zones : The case of a mountainous basin in the Sierra-Costa region of Michoacán , Mexico", vol. 223, pp. 685–693.

[45] Harris .A., S. Rahman, F. Hossain, L. Yarborough, A. C. Bagtzoglou, and G. Easson (2007), "Satellite-based Flood Modeling Using TRMM-based Rainfall Products," pp. 3416–3427.

[46] Furl .C., D. Ghebreyesus, and H. O. Sharif (2018), "Assessment of the Performance of Satellite-Based Precipitation Products for Flood Events across Diverse Spatial Scales Using GSSHA Modeling System", pp. 1-18.

[47] Mohammed E.,Turki H., Khalid B., Ibrahim A. (2013), "Modelling flash floods in arid urbanized areas: Makkah (Saudi Arabia)" vol. 24, pp. 171–181.

[48] Bartier .P. M. and C. P. Keller (1996), "Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (IDW)", vol. 22, pp. 795–799.

[49] Grimm .K., M. T. Nasab, and X. Chu (2018), "TWI computations and topographic analysis of depression-dominated surfaces", vol. 10, pp. 1-12.

[50] Mattivi .P., F. Franci, A. Lambertini, and G. Bitelli (2019), "TWI computation : a comparison of different open source GISs", pp. 1-12.

[51] Hjerdt K. N., McDonnell J. J., J. Seibert, and A. Rodhe (2004), "A new topographic index to quantify downslope controls on local drainage", vol. 40, pp. 1–6.

[52] Dingman .S. L. (1978), "Drainage Density and Streamflow ", vol. 14, pp. 1183-1187.

[53] Horton .B. (1945), "Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology", vol. 56, pp. 275–370.

[54] Melton .M. A. (1957), "Analysis of the relations among elements of climate, surface properties, and geomorphology".

[55] Alamri .N. (2013), "Flash flood hazard mapping based on quantitative hydrology, geomorphology and GIS techniques (case study of Wadi Al Lith, Saudi Arabia)", .

[56] Nyarko .B. K. (2002), "Application of a Rational Model in GIS for Flood Risk Assessment in Accra BT ".

[57] Reinier .V., Mieke .H. and Yitbarek .T. (2016) "An assessment towards building resilience through ecosystem restoration in Somali Regional State, Ethiopia", pp. 1-62.

[58] Prudhomme .C. and D. W. Reed (1999), "Mapping extreme rainfall in a mountainous region using geostatistical techniques : A case study in Scotland," vol. 1356, pp. 1337–1356.

[59] Liechti .T. C., J. P. Matos, J. Boillat, and A. J. Schleiss (2012), "Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River Basin," pp. 489–500,.

[60] Moriasi D. N., Gitau M. W., N. Pai, and P. Daggupati (2015), "Hydrologic and water quality models: performance measures and evaluation criteria" vol. 58, pp. 1763–1785.

[61] Najja .S., M. Zad, Z. Zulkafli, and F. M. Muharram (2018), "Satellite Rainfall (TRMM 3B42-V7) Performance Assessment and Adjustment over Pahang River", pp. 1–24.

[62] Bharti V. and C. Singh (2015), "Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region", pp. 458–473.

[63] Sun, J. Chen, Y. Bao, X. Han, J. Zhan, and W. Peng (2018), "Landslide susceptibility mapping using logistic regression analysis along theJinsha river and its tributaries Close to Derong and Deqin County , Southwestern China", pp. 1–29.

[64] Bharati D. K., V. K. Rai, and P. Kumar (2017), "Land use / land cover analysis using Normalized Difference Vegetation Index (NDVI)", pp. 648–655.

[65] Apollonio C., Balacco G., Novelli A., Tarantino E., P. Bari, and V. E. O. Bari (2016),

"Land use change impact on flooding areas : The Case Study of Cervaro Basin (Italy)".

[66] Pal S. and Ziaul S. (2017)," Detection of land use and land cover change and land surface temperature in English Bazar urban centre", vol. 20, pp. 125–145.

[67] Rwanga S. S. and Ndambuki J. M. (2017), "Accuracy Assessment of Land Use/Land Cover Classification Using Remote Sensing and GIS", vol. 08 , pp. 611–622.

[68] Tehrany M. S., Shabani F., M. N. Jebur, H. Hong, W. Chen, and X. Xie (2017), "GIS-based spatial prediction of flood prone areas using standalone frequency ratio , logistic regression , weight of evidence and their ensemble techniques", vol. 5705,.

[69] Pradhan B. (2010), "Landslide Susceptibility mapping of a catchment area using frequency ratio, fuzzy logic and multivariate logistic regression approaches", pp. 301–320.

[70] Paper O. (2004), "Landslide susceptibility mapping using GIS-based weighted linear combination , the case in Tsugawa area of Agano River , Niigata Prefecture , Japan", pp. 73–81.

[71] Ratan K., Gouri S. Pravat K., Hamid R. (2018), "Flood susceptibility mapping using geospatial frequency ratio technique : a case study of Subarnarekha River Basin , India".

[72] Tehrany M. S., Lee M. J., Pradhan B., M. N. Jebur, and S. Lee (2014), "Flood susceptibility mapping using integrated bivariate and multivariate statistical models", vol. 72, pp. 4001–4015.

[73] Khosravi K., Panahi M., and D. T. Bui (2018), "Spatial prediction of groundwater spring potential mapping based on an adaptive neuro-fuzzy inference system and metaheuristic optimization," pp. 4771–4792.

[74] Ayalew L. and Yamagishi H. (2005), "The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains , Central Japan", vol. 65, pp. 15–31.

[75] Park S., Hamm S., Jeon H., and Kim J. (2017), "Evaluation of logistic regression and multivariate adaptive regression spline models for groundwater potential mapping using Rand GIS".

[76] Ozdemir A. and Altural T. (2013), "A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping:

Sultan mountains, SW Turkey", vol. 64, pp. 180–197.

[77] Mahmoud S. H. and T. Y. Gan (2018), "Multi-criteria approach to develop flood susceptibility maps in arid regions of Middle East", vol. 196, pp. 216–229.

[78] Khabat K.,Binh T., Kamran C., Ataollah S., Himan S., Inge R., Indra P., Dieu T. (2018), "A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran", vol. 627, pp. 744–755.

[79] Ozdemir A. (2011) , "Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey)", vol. 405, pp. 123–136.

[80] Khabat K.,Binh T., Kamran C., Ataollah S., Himan S., Inge R., Indra P., Dieu T. (2018), "A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran", vol. 627, pp. 744–755.

[81] Ghanbarian G. and M. R. Raoufat (2019), "Habitat Suitability Mapping of Artemisia aucheri Boiss Based on the GLM Model in R", pp. 213-228.

[82] Djeddaoui F., Chadli M., and R. Gloaguen (2017), "Desertification Susceptibility Mapping Using Logistic Regression Analysis in the Djelfa Area , Algeria", pp. 1-26.

[83] Richard Mind, Lanhai L., Amobichukwu C., Lamek N., Jean B., Aboubakar G., Mapendo M. (2019), "Flood susceptibility modeling and hazard perception in Rwanda", vol. 38, p. 1-12..

[84] Hamid R., Abbas G., Biswajeet P., Chong X. and Candan G. (2013), "Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran", pp. 349-369.

APPENDIX

Appendices A: Monthly average precipitation data used in the study

	Hadew station											
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	0.40	0.00	0.05	1.59	7.31	0.54	3.67	3.21	1.22	1.74	0.98	0.01
2009	0.56	0.00	0.62	2.26	2.26	2.10	4.24	1.36	1.55	2.19	0.46	0.00
2010	0.00	1.68	6.90	3.57	3.28	3.07	1.85	2.50	4.84	0.16	0.15	0.00
2011	0.16	0.02	0.22	2.02	2.09	1.59	3.78	3.89	6.43	0.21	0.51	0.11
2012	0.02	0.00	2.61	4.92	4.88	1.09	2.50	4.73	7.53	0.33	0.18	0.07
2013	0.12	0.00	1.47	3.90	2.91	0.44	2.47	3.10	2.89	3.41	1.99	0.00
2014	0.00	0.14	2.86	4.66	1.92	0.33	3.87	2.18	4.64	5.42	0.00	0.00
2015	0.07	0.03	1.33	1.30	5.65	1.36	1.84	3.03	2.02	1.06	0.12	0.00
2016	0.06	0.14	1.11	4.27	2.16	1.22	3.49	1.92	2.11	2.30	0.18	0.00
2017	0.00	0.00	0.99	1.52	2.52	1.65	1.85	2.31	7.00	0.50	0.00	0.00
2018	0.00	0.15	0.66	4.93	0.62	1.50	2.32	3.61	1.98	2.74	1.21	0.00

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	0.43	0.00	0.01	1.01	5.43	0.11	2.08	2.77	1.10	2.12	1.08	0.00
2009	0.00	0.00	0.69	1.08	0.80	1.38	3.28	1.08	1.64	1.66	0.00	0.00
2010	0.00	3.87	5.46	2.94	3.61	2.47	1.41	1.91	3.72	0.22	0.43	0.00
2011	0.07	0.00	0.00	1.99	1.78	1.23	2.54	3.35	4.21	0.00	0.45	0.00
2012	0.02	0.00	0.00	4.59	3.73	1.02	2.34	3.55	5.96	0.69	0.31	0.00
2013	0.44	0.00	1.22	5.40	2.43	0.74	1.17	2.44	1.41	2.62	1.53	0.00
2014	0.00	0.01	2.76	5.30	1.81	0.26	3.98	1.42	3.22	4.66	0.11	0.00
2015	0.00	0.00	1.15	2.54	5.12	1.12	1.04	1.42	1.18	0.68	0.07	0.00
2016	0.00	0.08	1.26	6.17	1.65	0.87	1.76	1.16	2.38	1.09	0.08	0.00
2017	0.00	0.00	1.55	3.35	2.68	0.75	2.21	2.04	4.77	0.00	0.00	0.00
2018	0.00	0.65	1.40	8.37	0.61	1.40	2.13	2.44	1.76	2.82	1.46	0.00

	Leffeisa station											
Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	0.00	0.00	0.01	0.68	6.01	0.35	5.46	9.19	1.35	0.74	2.38	0.00
2009	0.00	0.00	0.77	0.72	0.88	4.28	8.61	3.59	2.02	0.58	0.00	0.00
2010	0.00	3.64	6.13	1.97	3.99	7.63	3.70	6.35	4.58	0.08	0.94	0.00
2011	0.00	0.00	0.00	1.33	1.96	3.81	6.67	11.12	5.18	0.00	0.98	0.00
2012	0.00	0.00	0.00	3.07	4.13	3.16	6.15	11.80	7.33	0.24	0.67	0.00
2013	0.00	0.00	1.37	3.62	2.68	2.28	3.07	8.11	1.74	0.92	3.37	0.00
2014	0.00	0.01	3.10	3.55	2.00	0.80	10.46	4.71	3.96	1.64	0.25	0.00
2015	0.00	0.00	1.29	1.70	5.66	3.46	2.74	4.73	1.45	0.24	0.16	0.00
2016	0.00	0.08	1.41	4.13	1.83	2.69	4.63	3.85	2.93	0.38	0.18	0.00
2017	0.00	0.00	1.74	2.24	2.96	2.33	5.81	6.78	5.87	0.00	0.00	0.00
2018	0.00	0.61	1.57	5.61	0.67	4.32	5.58	8.11	2.16	0.99	3.20	0.00

	Chinaksen station											
year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2008	0.30	0.00	0.99	2.39	4.01	0.57	3.22	3.26	1.05	0.28	1.13	0.00
2009	1.18	0.00	1.42	1.25	0.84	2.68	2.54	2.01	2.13	0.93	0.20	0.14
2010	0.00	0.23	5.37	2.41	1.25	1.83	1.48	3.44	2.43	0.15	0.00	0.00
2011	0.03	0.00	0.00	3.46	1.42	0.82	1.77	4.42	2.98	0.00	0.66	0.07
2012	0.00	0.00	2.56	3.34	2.82	0.39	2.56	5.20	3.16	0.20	0.10	0.00
2013	0.16	0.00	1.84	3.25	0.97	0.36	2.12	3.74	2.17	1.70	1.66	0.00
2014	0.00	0.00	2.01	4.74	1.42	0.00	3.50	2.43	2.63	0.92	0.06	0.00
2015	0.00	0.01	1.67	0.80	2.24	1.52	0.31	5.55	1.22	0.00	0.18	0.12
2016	0.00	0.06	1.73	4.73	1.74	1.82	3.49	3.24	1.00	0.33	0.26	0.00
2017	0.00	0.00	4.67	0.71	1.48	0.88	1.54	3.20	5.20	0.44	0.00	0.00
2018	0.00	0.02	0.48	4.51	0.38	2.32	3.14	3.51	1.22	1.22	1.63	0.00

Appendices B: R codes to extract and contour plotting TRMM NetCDf file format in interested gauged stations

```
library(ncdf4)
getwd()
workdir <- "C:/Users/user/Desktop/r/Data/My_Project1/Rainfall" # Netcdf file location
setwd(workdir)
nci <- nc_open("TRMM_3B42_daily.7_Aggregation_2008.ncml.ncml.nc")
dname <- "precipitation"
lon <- ncvar_get(nci, "lon")
lat <- ncvar_get(nci, "lat")
t <- ncvar_get(nci, "time")
nlon <- dim(lon) # Check longitude dimension
nlat<- dim(lat) # Check latitude dimension
nt<- dim(t) # Check time dimension
precipa <- ncvar_get(nci, dname)
climmat=matrix(precipa,nrow = 9,ncol = 366)
mapmat=climmat
lon = seq(  42.125,43.125,length=9) # coordinates of the study area
lat = seq(8.125,10.125,length=366) # is divided equal with time
```

```
# Countour Plotting TRMM netcdf file

par(mar=c(5,4,1,1))

rgb.palette=colorRampPalette(c('lightgrey','blue','red',
                               'yellow','green'),interpolate='spline')
filled.contour(lon,lat,climmat, color.palette=rgb.palette,
               plot.title=title(main="ppt",xlab="Longitude",ylab="Latitude"),
               key.title=title(main="trmmppt",lwd=0.1))
```

```
# Extract the TRMM data by interested guage locations

library(raster)
b <- brick("TRMM_3B42_daily.7_Aggregation_2008.ncml.ncml.nc",
           varname = "precipitation")

pts <- read.csv("book.csv", stringsAsFactors = FALSE) # book.csv is the stations lon and lat
ts <- extract(b, cbind(pts$lat, pts$lon))
ww<-data.matrix(cbind(pts$lat, pts$lon))
aa<-data.frame(ww,ts)

write.csv(aa, "TRMM_2008.csv")
```

Appendices C: Sample pictures showing flood susceptibility in the study area after flood event


2019

Appendices D: The Morphometric Parameters and MHI of the study area

	Basin dimensions				Basin shape				Basin surface				Drainage Network				
Cat_ID	A	P	W	L	Er	Sf	Cr	Fr	Rr	Rv	Ri	Hi	D	Tr	Fq	Br	**MHI**
1	47.86	57.84	3.22	14.85	0.53	4.61	0.18	0.22	0.04	0.01	0.71	0.20	1.11	1.18	1.42	2.13	7.79
2	20.55	28.38	3.00	6.84	0.75	2.28	0.32	0.44	0.02	0.00	0.14	0.57	1.03	1.06	1.46	2.00	7.19
3	35.97	38.34	3.81	9.45	0.72	2.48	0.31	0.40	0.02	0.00	0.22	0.51	1.34	1.23	1.31	2.04	7.68
4	31.24	42.72	2.92	10.69	0.59	3.66	0.22	0.27	0.02	0.00	0.25	0.38	1.24	0.89	1.22	2.00	6.71
5	39.51	45.60	3.29	11.99	0.59	3.64	0.24	0.27	0.02	0.00	0.25	0.53	1.32	1.45	1.67	2.06	7.51
6	92.24	83.58	3.86	23.91	0.45	6.20	0.17	0.16	0.03	0.01	0.93	0.27	1.19	1.51	1.37	2.03	8.30
7	29.11	38.58	2.95	9.88	0.62	3.35	0.25	0.30	0.02	0.01	0.29	0.48	1.32	0.96	1.27	2.06	7.07
8	44.37	50.46	3.97	11.19	0.67	2.82	0.22	0.35	0.02	0.00	0.29	0.27	1.53	1.13	1.28	2.04	7.77
9	35.12	57.72	2.81	12.50	0.53	4.45	0.13	0.22	0.01	0.00	0.12	0.38	1.42	0.73	1.20	2.10	6.10
10	26.04	41.70	2.18	11.95	0.48	5.48	0.19	0.18	0.02	0.00	0.29	0.51	1.43	0.98	1.57	2.05	6.67
11	17.97	29.64	2.42	7.42	0.64	3.07	0.26	0.33	0.02	0.01	0.27	0.37	1.48	0.84	1.39	2.08	7.33
12	7.54	18.66	1.68	4.48	0.69	2.66	0.27	0.38	0.02	0.00	0.11	0.45	1.33	0.70	1.73	1.86	7.09
13	0.95	6.66	0.58	1.64	0.67	2.83	0.27	0.35	0.01	0.00	0.02	0.55	1.74	0.15	1.06	1.00	6.60
14	5.44	17.76	1.16	4.67	0.56	4.02	0.22	0.25	0.01	0.00	0.09	0.45	1.33	0.34	1.10	1.50	5.87
15	25.98	47.46	2.40	10.83	0.53	4.52	0.14	0.22	0.01	0.00	0.11	0.44	1.50	0.59	1.08	2.15	5.87
16	19.69	33.42	2.84	6.94	0.72	2.44	0.22	0.41	0.02	0.00	0.12	0.47	1.14	0.78	1.32	2.00	6.72
17	31.86	41.34	3.32	9.61	0.66	2.90	0.23	0.34	0.01	0.00	0.14	0.37	1.58	0.99	1.29	1.95	7.27
18	25.77	40.98	2.41	10.69	0.54	4.43	0.19	0.23	0.02	0.00	0.25	0.45	1.54	1.12	1.78	2.00	7.22
19	14.26	23.46	2.05	6.97	0.61	3.40	0.33	0.29	0.09	0.03	0.82	0.23	1.24	1.02	1.68	2.18	9.45
20	10.90	21.96	1.87	5.83	0.64	3.12	0.28	0.32	0.08	0.02	0.63	0.35	1.36	0.77	1.56	2.13	8.73
21	22.80	31.20	3.13	7.29	0.74	2.33	0.29	0.43	0.07	0.02	0.61	0.20	1.18	1.06	1.45	1.94	9.01
22	1.62	8.70	0.69	2.36	0.61	3.44	0.27	0.29	0.01	0.00	0.06	0.55	1.77	0.11	0.62	1.00	6.28
23	35.33	53.10	2.55	13.86	0.48	5.44	0.16	0.18	0.01	0.00	0.25	0.42	1.33	1.11	1.67	2.03	6.66
24	28.65	38.94	2.81	10.19	0.59	3.63	0.24	0.28	0.01	0.00	0.18	0.54	1.36	1.08	1.47	2.00	6.92
25	28.77	39.96	3.07	9.36	0.65	3.05	0.23	0.33	0.01	0.00	0.15	0.57	1.74	0.83	1.15	2.06	7.06
26	38.59	48.90	3.45	11.20	0.63	3.25	0.20	0.31	0.01	0.00	0.13	0.47	1.50	1.04	1.32	1.96	7.05
27	2.28	11.46	0.82	2.77	0.62	3.36	0.22	0.30	0.02	0.00	0.05	0.34	0.94	0.44	2.19	1.67	6.35
28	6.53	16.80	1.68	3.88	0.74	2.31	0.29	0.43	0.02	0.00	0.07	0.47	1.19	0.60	1.53	2.00	6.75
29	7.56	22.80	1.28	5.89	0.53	4.59	0.18	0.22	0.02	0.00	0.12	0.46	1.33	0.31	0.93	1.75	5.53
30	16.23	29.10	2.29	7.10	0.64	3.10	0.24	0.32	0.07	0.02	0.90	0.34	1.72	0.72	1.29	2.10	9.11
31	18.09	41.58	1.92	9.44	0.51	4.93	0.13	0.20	0.04	0.01	0.66	0.17	1.65	0.53	1.22	1.83	7.41
32	18.69	31.86	2.54	7.36	0.66	2.90	0.23	0.35	0.01	0.00	0.13	0.43	1.33	1.13	1.93	1.89	7.39
33	36.04	42.36	3.91	9.21	0.74	2.35	0.25	0.43	0.01	0.00	0.17	0.47	1.59	1.02	1.19	3.58	7.01
34	4.97	24.54	0.69	7.20	0.35	10.43	0.10	0.10	0.01	0.00	0.13	0.40	1.67	0.37	1.81	1.80	5.32
35	21.50	35.10	2.35	9.15	0.57	3.89	0.22	0.26	0.01	0.00	0.19	0.43	1.48	0.94	1.53	1.94	6.93
36	24.56	33.42	2.74	8.97	0.62	3.27	0.28	0.31	0.02	0.00	0.20	0.37	1.32	0.84	1.14	2.00	6.79
37	18.50	35.28	2.26	8.19	0.59	3.62	0.19	0.28	0.01	0.00	0.17	0.66	1.39	0.65	1.24	2.09	6.28
38	5.30	14.82	1.57	3.39	0.77	2.16	0.30	0.46	0.02	0.00	0.07	0.44	1.23	0.34	0.94	1.67	6.52
39	24.28	46.02	2.16	11.27	0.49	5.23	0.14	0.19	0.01	0.00	0.07	0.44	1.22	0.74	1.40	2.13	5.59
40	28.54	40.80	2.86	9.97	0.60	3.49	0.22	0.29	0.02	0.00	0.27	0.35	1.43	1.05	1.51	1.87	7.36
41	30.28	41.34	2.96	10.21	0.61	3.45	0.22	0.29	0.05	0.01	0.81	0.24	1.58	1.14	1.55	1.81	8.97
42	11.83	23.04	2.29	5.16	0.75	2.25	0.28	0.44	0.01	0.00	0.12	0.55	1.68	0.56	1.10	2.17	7.06
43	25.11	28.44	3.06	8.22	0.69	2.69	0.39	0.37	0.07	0.02	0.92	0.27	1.58	1.02	1.16	2.64	9.39
44	8.73	23.70	1.61	5.44	0.61	3.39	0.20	0.30	0.04	0.01	0.28	0.25	1.23	0.59	1.60	2.00	7.06
45	4.84	16.98	1.16	4.18	0.59	3.61	0.21	0.28	0.04	0.01	0.16	0.23	1.05	0.35	1.24	1.50	6.31
46	19.72	30.12	2.62	7.54	0.66	2.88	0.27	0.35	0.07	0.02	0.66	3.38	1.26	1.06	1.62	2.13	7.88

Where Cat is Catchment, A is Area (km2), D is Stream density (km-1), Fq is Stream Frequency (km-2), Cr is Circulation ratio, Rv is Relief ratio Relative, Rr is relief ratio, RI is Ruggedness index, Tr is Texture ratio, ratio, TSN is Total stream number Stream order, TSL is total stream length (km), Br is Bifurcation ratio, Er is Elongation ratio, Sf is shape factor and HI is Hypsometric index.