REGIONAL FLOOD FREQUENCY ANALYSIS AND FLOOD INUNDATION MAPPING: THE CASE OF MIDDLE AWASH RIVER BASIN, ETHIOPIA

By

Gadise Tadese Nemera

A Thesis Submitted to the Department of Water Resources Engineering, School of Civil Engineering and Architecture

Presented in Partial Fulfillment of the Requirement for the Degree of Master's of Science in Water Resources Engineering (Specialization in Irrigation)

Office of Graduate Studies
Adama Science and Technology University

Adama, Etiopia
October, 2020

REGIONAL FLOOD FREQUENCY ANALYSIS AND FLOOD INUNDATION MAPPING: THE CASE OF MIDDLE AWASH RIVER BASIN, ETHIOPIA

By

Gadise Tadese Nemera

Advisor: Zelalem Biru (PhD)

Co-advisor: Daniel Alemayehu (PhD)

A Thesis Submitted to the Department of Water Resources Engineering, School of Civil Engineering and Architecture

Presented in Partial Fulfillment of the Requirement for the Degree of Master's of Science in Water Resources Engineering (Specialization in Irrigation Engineering)

Office of Graduate Studies
Adama Science and Technology University

Adama, Ethipia
October, 2020

APPROVAL SHEET

We, the undersigned, members of the Board of Examiners have read and evaluated the MSc Thesis entitled "**Regional Flood Frequency Analysis and Flood Inundation Mapping: The Case of Middle Awash River Basin, Ethiopia**" presented and defended by Mrs. Gadise Tadese Nemera.

This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirement of the Degree of the degree of Master's in Water Resources Engineering (Specialization in Irrigation Engineering)

______________________	____________	____________
Advisor	Signature	Date
______________________	____________	____________
External Examiner	Signature	Date
______________________	____________	____________
Internal Examiner	Signature	Date
______________________	____________	____________
Chair Person	Signature	Date

Final Approval and Acceptance

______________________	____________	____________
Head of Department	Signature	Date
______________________	____________	____________
School Dean	Signature	Date
______________________	____________	____________
Post Graduate Dean	Signature	Date

DECLARATION

I hereby declare that this MSc Thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this have been duly acknowledged.

Name: ____

Signature: ____

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

Name: ____

Signature: ____

Date of submission: October, 2020

ADVISOR’S APPROVAL SHEET

To: **Water Resources Engineering Department**

Subject: **Thesis Submission**

This is to certify that the thesis entitled “Regional Flood Frequency Analysis and Flood Inundation Mapping: the case of Middle Awash River Basin, Ethiopia” submitted in partial fulfillment of the requirements for the degree of Master’s in Irrigation Engineering the Graduate program of Water Resources Engineering and has been carried out by Gadise Tadese Nemera ID. No: APR/16444/10 under our guidance. Therefore, we recommend that the student has fulfilled the requirements and hence hereby she can submit the thesis to the department.

______________________	____________	____________
Major Advisor	Signature	Date
______________________	____________	____________
Co- Advisor	Signature	Date

ACKNOWLEDGEMENTS

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistances in the preparation and completion of this study. In fact, it is almost impossible to list all of them, and therefore the following list is horribly incomplete. However, it would be worse not to make an attempt to acknowledge those who have played a key role in the research work.

First and foremost, I gratefully acknowledge the contribution of my first advisor, Dr. Zelalem Biru, whose patience, dedication and knowledge guided me through the thesis work, it was a great pleasure to work with him. I would like extend my acknowledgement to my second advisor Dr. Daniel Alemayehu for guiding and reviewing my work. I am grateful to the Ministry of Water Irrigation and Electricity of Ethiopia for their collaboration in providing me relevant data

My acknowledgement also goes to Adama Science and Technology University for giving me a full scholarship and funding my research work. Thank you so much ASTU.

Last but not least, I pass my regards and blessings to all my family, friends and the one above all, the Almighty God, for answering my prayers and giving me the strength to finish this thesis work. Thank you so much my dear Lord.

TABLE OF CONTENTS

Contents Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

AD	Anderson-Darling
AM	Annual Maximum
AMF	Annual Maximum Flow
ArcGIS	Geographical Information System
CDF	Cumulative Distribution Function
CV	Coefficient of Variation
DEM	Digital Elevation Model
Di	Discordancy measure
DTEM	Digital Terrain Elevation Model
D/S	Downstream
FFA	Flood Frequency Analysis
GEV	Generalized Extreme Value
GLO	Generalized Logistic
GOF	Goodness of Fit
GPA	Generalized Pareto
GWP	Global Water Partnership
HEC-Geo RAS	Hydrologic Engineering Center's Geographic River Analysis System
HEC-RAS	Hydrologic Engineering Center's River Analysis System
IHR	Identification of Homogeneous Regions
ITCZ	Inter-Tropical Convergence Zone
KS	Kolmogorov-Smirnov
LCk	Linear Coefficient of Kurtosis
LCs	Linear Coefficient of Skewness
LCV	Linear Coefficient of Variation
LMRD	Linear Moments Ratio Diagram
LN	Lognormal
LPIII	Log Pearson Type 3
LS	Least Square
MLM	Maximum Likelihood Method
MML	Method of Maximum Likelihood
MOLM	Method of Linear Moments

MOM	Method of Moment
MoWIE	Ministry of Water, Irrigation and Electricity
PD	Partial Duration Series
P-P	Probability-Probability
PWM	Probability Weighted Method
Q	Extreme Flow (Discharge)
Q_T	Extreme Flow Under T-Year Return Period
Q_H	High Outlier Discharge
Q_L	Low Outlier Discharge
Q-Q	Quantile-Quantile
RFFA	Regional Flood Frequency Analysis
T	Return Period (Recurrence Interval)
TS	Time Series
TIN	Triangulated Irregular Network
USACE	United State Army Corps of Engineering
U/S	Upstream
WMO	World Meteorological Organization
X_T	Estimated Flow Quantiles Value (m3/s)
XS	Cross-Section

ABSTRACT

As flood is a natural disaster and causes loss of life and property destruction, its estimation with a certain frequency and mapping of its inundation area are one of the most important factors for the planning, design and management of flood risk controlling hydraulic structures. The intent of this study is to carry out the regional flood frequency analysis using L-moments and analyze flood inundation mapping using HecGeo-RAS and HEC-RAS models in combination with ArcGIS on middle Awash river basin. The annual maximum flow records of six river gauging stations of middle Awash River Basin from U/S to D/S: Awash below Koka dam, Awash @Wanji, Awash @Nura Hera, Awash @Metehara, Awash @Melka Sadi and Awash @Melka Werer, were examined based on L-moments to investigate and delineate the hydrologically homogeneity of the region using regional heterogeneity measure, identify the discordancy measure of the sites using L-moment based discordancy measure (Di), select best fit regional probability distribution using goodness-of-fit tests by EasyFit software and L-moment ratio diagram, estimate parameters for the best fit distribution, estimate extreme flow quantiles under return period of 2, 5, 10, 20, 25, 50, 100, 200, 500, 1000, 2000, 5000 and 1000-year (Q_T) at each gauging station, develop regional frequency and growth curves, mapping flood inundation and estimating the extent of flood inundation area for the selected Awash river reach using HEC-RAS MApper. The result of the analysis shows that the region is homogeneous, all sites are not discordant, i.e. data from all river gauging stations are suitable for flood frequency analysis, the Generalized Pareto (GPA) was selected as best fit distribution for Awash below Koka, Awash @Wanji and Metehara gauging stations and Generalized Extreme Value (GEV)was selected as best fit distribution for Awash @Nura Hera, Awash Melka Sadi and Awash Melka Werer gauging stations. The estimated extreme flow quantiles for each gauging station from U/S to D/S respectively under a 100-year return period are 6077.64, 991.88, 564.31, 14324, 365.24 and 698.91 m^3/s. The result of flood inundation area for the river reach from U/S to D/S respectively under a 100-year return period are 15.19, 50.97, 61.73, 126.54 and 55.58 Km^2. The mapping result indicated that all parts of Awash below Koka dam up to Wanji, some parts of Awash Nura Hera up to Metehara and some parts of Awash @Metehara up to Melka Werer river reaches were under the impact of 25, 50 100, 200, 500 and 1000-year flood problem.

Keywords: Flood frequency, L-moments, goodness-of-fit, Best fit distribution, Flood inundation mapping

1. INTRODUCTION

1.1 Background

Flood is the most common form of natural disaster that disrupts human activities, causes loss of human lives & destroys properties, particularly in developing countries where there is poor management and protection of the environment due to low economic potential. Among all types of natural hazards threatening the world, flood is probably the most devastating, wide spread and frequent (Vanneuville et al., 2011; Karagiozi E. et al., 2011). Despite the tragic loss of life, an impact of flood includes reduction of the asset base of households, communities and societies through the destruction of agricultural crops, residences, infrastructure and environmental impact. According to reports from the World Meteorological Organization and Global Water Partnership (2009) approximately 70 percent of all disasters occurring on the world are related to hydro-meteorological events.

Ethiopia is one of the eastern countries of Africa, which experienced flooding events in different parts of its basin at different time. Particularly large-scale flooding in the country is common and it occurs in the lowland flat parts of the country. As reported by WMO and GWP (2003) among river basins in the country Awash River Basin, which is located in the Great Rift Valley and with a surface area of 110,000 km^2 is one, where most frequently flooding occurs (Sifan Abera, 2012).

There are just two types of flood in general, namely, the floods that caused by extreme excess of precipitation on the catchment that contributed to the flood in the main and tributaries of the river, and floods owing to unexpected release of an excess amount of water from storage or dam. The flood in Awash River basin is due to the excess precipitation that occurred in the highland of the basin, and finally generates high flood on the downstream part of the basin. The basin originates from the highlands at above 3000 masl receiving 2000 mm of rainfall. The land use of the basin is characterized by extensive land use for agriculture, combined with urbanization and industrialization, is aggravating the generation of runoff.

High flow exceeding danger levels and entering in flood plains is the result of heavy or continuous rainfall exceeding the absorptive capacity of soil, and the flow capacity of the streams. It causes widespread damage to property and life in different parts of the catchment. Despite the fascinating achievements of science and technology in the 21st century, floods and droughts continue to hit every generation of human beings, bringing suffering, death, and material losses.

Knowledge of the magnitude and probable frequency of recurrence of floods is necessary to the proper design and location of structures such as dams, bridges, culverts, levees, highways, waterworks, sewage- disposal plants, and industrial buildings. Structures may be damaged or destroyed by occasional floods of varying magnitude. The frequency with which such damage may occur must be considered in determining the size or strength of the structure, its location, or the feasibility of building it at all. The problem is an economic one, involving computation of the total annual cost of maintaining a structure of a given design compared to the cost for other designs. Either overdesign or under design of structures involves excessive costs on a long-time basis. Also, in case of extreme floods emergency evacuation of people can be carried out well in advance.

Knowledge of flood frequency is necessary also to flood insurance and flood zoning, activities which are now considered on a broad scale. Problems of flood insurance and flood zoning are economically important and they make the development of sound flood-frequency methods imperative. The work of flood-risk management starts from identification and estimation of current and possible areas of higher hazard potential margins (Sitotaw Haile et al., 2018).

In the past, flood frequency analysis techniques were developed to relate the magnitude of floods with their frequency of occurrences (Hosking and Wallis, 1997). Such studies have also been done to estimate flood based on catchment characteristics and statistical analysis. It is understood that a minimum of 30-40 years of records are needed for flood frequency analysis. If the length of records is too short, specifically on inadequate data situation, then regional flood frequency curves together with at-site mean provides consistent estimates of floods.

It is common for a ‘standard of service’ to be defined for a flood defense structure, which is often to provide protection against an event with a specified return period, such as the 100-year flood. Since discharge data for most catchments have been collected for periods of time significantly less than 100 years, the estimation of the ‘design discharge’ (design stage, or water level) necessarily requires a degree of extrapolation, which in turn demands curve-fitting to the existing data.

Thus, this study is focused on assessing regional flood frequency analysis for Middle Awash River Basin and developing its flood inundation mapping.

1.2 Statement of the Problem

As flooding is one of the most prevalent natural hazards that impact negatively upon human livelihood, it requires various responses including construction (downstream flood defenses), forecasting (for warning and evacuation), and land-use management (upstream catchment-scale changes of land use and runoff characteristics). To some degree, all of these require hazard or risk assessment in the form of a flood frequency analysis (FFA) (Kidson R. et al., 2005). Hence, flood forecasting and early warning system is an important tool to give appropriate reliable information of the incoming flood to the vulnerable community.

The frequency of floods with various risks of exceedance is therefore needed in most engineering problems (Vivekanandan, 2015). The estimation must be fairly accurate not only aimed at the preventing of catastrophes, but also at avoiding excessive costs in case of overestimating the flood magnitude, or excessive damage while underestimating flood potential. Thus, engineering works require a reliable estimation of flood quantiles using reliable flood records measured at gauging stations.

However, most of the catchments in developing countries like Ethiopia are poorly gauged or ungauged, which hinders the country's water resources management and flood prediction (Rabba et al., 2017). This is owing to the low density of gauging stations, the operation and maintenance of gauging networks are difficult and the lack of infrastructures required for the acquisition of adequate hydrologic data (Gedefa et al., 2009). These data in both quantity and quality are the primary inputs to the design and successful operation of hydraulic and drainage structures such as dams, spillways, bridges, culverts and flood protection schemes (Betul Saf et al., 2009). Unfortunately, these vigorous inputs are usually inadequate, in most cases incredibly unavailable at points of interest.

Awash River Basin is intensively used for irrigation and other developmental activities. However, the basin is frequently inundated by flood and destroyed irrigation infrastructures and huge areas of farm lands, displaced people from their home, swept way livestock and etc.

There are no significant studies conducted yet in middle Awash River basin on regional flood frequency analysis and mapping of area prone to flooding.

Therefore, determination of regional flood frequency analysis of Middle Awash River Basin and developing its flood inundation mapping will solve these problems

1.3 Objectives of the Study

1.3.1 General Objective

The general objective of this study is to carryout regional flood frequency analysis and mapping the extent of flood inundation of Middle Awash River Basin for flood risk management.

1.3.2 Specific Objectives

The specific objectives of the study were:

i) To identify and delineate hydrologically homogeneous regions of the sub-basin;
ii) To determine the best fit statistical distributions to data of the gauging stations of the middle Awash basin;
iii) To estimate extreme flow quantiles and develop regional flood frequency curves of the selected river gauging stations for different return period
iv) To map flood inundation and estimate its area for different return period.

1.4 Research Questions

i) Are the study regions hydrologically homogeneous?
ii) What is the best-fit probability distribution for prediction of hydrological events of each river gauging station of the basin?
iii) How flood frequency curves of the selected river gauging stations are developed?
iv) What is the extent of flood inundation in the basin?

1.5 Significance of the Study

Floods are recurrent phenomena in Ethiopia from time immemorial. Floods of varying magnitude, affect some or the other parts of the country, almost every year due to different climates and rainfall patterns. With the increase in population and developmental activities in the region, such as, anthropogenic, urbanization, industrialization, and large-scale irrigation expansions have changed the natural environment within Awash River Basin (Awash Basin Authority, 2017).

Therefore, this study is expected to become valuable up to date information for flood risk estimation, forecasting flood for warning and emergency evacuation of people in case of extreme floods, risk minimization, economic evaluation of flood control projects, proper planning, and design of water resources management options on the study area. In addition, the study can be used as a point of reference for policy and decision makers, and any further investigation that will undertake on the middle Awash River Basin.

1.6 Scope of the Study

The study focused on regional flood frequency analysis, mapping the extent flood inundation and estimating the area of flood inundated region under 2, 5, 10, 20, 25, 50, 100, 200, 500, 1000, 2000, 5000 and 1000-year return period. The study was conducted on middle part of Awash river reach from Awash @below Koka dam up to Awash @Melka Werer river gauging stations.

1.7 Organization of the Thesis

The thesis is organized into five chapters from introduction to conclusions and recommendations. The first chapter presents about the introduction part under which the general background, the statement of the problem, objectives of the study, research questions, significance and scope of the study were discussed. The second chapter discusses about the literature review related to regional flood frequency analysis and flood inundation mapping. The third chapter illustrates about the material and methodology to be applied in this research paper, from data collection to data analysis. The fourth chapter discusses the result of the analysis such as quality of data, homogeneity of the study regions, best-fit probability distributions and their estimated parameters, estimates of extreme flow quantile and their frequency and regional growth curves, mapping the extent of flood inundation and the estimating the area of flood inundation. The last chapter of the thesis is about the conclusion and recommendation of the study and this part concludes and recommends on the result of the analysis.

2 LITERATURE REVIEW

2.1 Flood Frequency Analysis

Flood frequency analysis is the most direct method for determining design flood. The primary objective of frequency analysis is to relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions (Chow et al., 1988). It shows the relationship between the magnitude of an event and the frequency with which that event is exceeded. It primarily uses observed annual maximum flood data at a gauging station to estimate flood magnitude. Furthermore, the catchment characteristics, water availability and possible extreme hydrological conditions like floods and droughts at various locations of any river system may be illustrated through the flood frequency analysis (Guru N. and Jha R.). Observed data over an extended period of time in a river system is considered to be independent and identically distributed will be analyzed in frequency analysis

In the planning and design of water resources projects, professionals are interested in the determination of the reasonable estimation of extreme events with defined return periods (Rahmana et al., 2015). These extreme events are necessary in the design of various flow control structures such as levees, culverts, bridges, barrages and dams, reservoir management, economic evaluation of flood protection projects, land use planning and management, flood risk assessment (Rao and Srinivas, 2008; Noto and Loggia, 2009; Bhagat, 2017; Kanti et al., 2017).

A common approach, therefore, is the annual maximum, where for each water year the peak streamflow is determined and a distribution is fitted to this series of data (Schendel and Thongwichian, 2017). These data are used to make frequency distributions for various discharges as a function of their recurrence interval or exceedance probability (Hosking and Wallis, 1997). Data observed over an extended period in a river system are analyzed in frequency analysis. The flood data are considered stochastic and may even be assumed space and time independent (Rao and Hamed, 2000).

2.2 Return Period, Probability, and Plotting Positions

Flood peaks do not occur with any fixed pattern in time or magnitude. Time intervals between floods vary. The definition of return period is the average of these inter-event times between flood events (Cunnane,1989). Large floods naturally have large return periods and vice versa. A given flood q with a return period T may be exceeded once in years. Hence the probability of exceedance is P(QT > q) = 1/T. The cumulative probability of non-exceedance, F(QT), which is the basis for estimating the magnitude of flood QT given its return period T, is given by

$$F(Q_T) = P(Q_T \leq q) = 1 - P(Q_T > q) = 1 - \frac{1}{T} \quad (2.1)$$

Probability plots require an initial estimate of the probability of non-exceedance F = (F (Q_T)), which is called a plotting position. Plotting positions are used to estimate parameters by using the probability weighted moments (PWM) method and it is given by Hosking (1990):

$\boldsymbol{F = \frac{i-0.35}{N}, i = 1, 2, \ldots, N}$ where N is the sample size and I is the rank of the observation

This formula is believed to give acceptable results for some common three-parameter distributions and is used in the PWM method (Hosking, 1990).

2.3 Flood Estimation Techniques

Flood frequency assessment is indispensable for flood management. It addresses the subject of flood risk assessment required in flood zoning and spatial planning, and in the arrangement of flow values for the design of flood alleviation and control works (Murphy et al., 2014; England et al., 2015). The accurate estimation of flood magnitude with corresponding frequency of occurrence is the challenge for hydrologists due to planning, management and design of water resource projects depends on the frequency and magnitude of maximum floods (Chavoshi and Azmin, 2009**;** Saf, 2009; Javelle et al., 2010; Dubey, 2014; Alam et al., 2016) on the sites of interest.

A realistic estimation of flood magnitude for a given return period is also essential in the economic evaluation of flood protection projects; minimizing flood-related costs to government and private enterprises, floodplain management by assessing hazards related to the expansion of floodplains. Over or under-estimation of design floods results in losses like a waste of resources, and infrastructural damage. Investigation in design flood estimation is on the decline and there is a large gap between design flood and practice (Irwin et al., 2014; Arnaud et al., 2017).

The literature identified two comprehensive methods for flood frequency analysis, statistical and derived. Statistical flood frequency analysis is the modern method of determining the frequency of peak stream flows. This method of frequency analysis involves fitting extreme value probability distribution functions to the historical record of annual maximum floods. This method is reliant upon the availability of observed streamflow to fit suitable probability distributions relevant to gauged sites (Kumar and Chatterjee, 2011; Vivekanandan, 2015). The derived techniques of flood frequency analysis involve the quantification of the processes that govern flood behavior which is less dependent upon historical data (Badreldin and Fengo, 2012).

2.4 Flood Frequency Models

Different magnitudes of flooding have a different probability of occurrence. According to Cunnane (1989) and Desalegn *et al.,* (2016), in flood frequency modeling the problems related to the following points have to be pointed out. Choice of model type, choice of distribution to be used and choice of method of parameter and quantile estimation. It should be noted that two separate features are important. These are the descriptive and predictive properties of the method. The descriptive property relates to the requirements that the chosen distribution shape resembles the observed sample distribution of floods and that random samples drawn from the chosen model distribution must be statistically similar to the properties of real flood series, the predictive properties relates to the requirement that quantile estimates are robust with small bias and standard error (Murphy *et al*., 2014).

In FFA, the objective is to determine a Q-T relationship at any required site along a river. At any river site, it is usually assumed that nature affords an exclusive relationship and that Q is a monotonically increasing function of T. In order to estimate this natural relation from a good quality continuous hydrometric record of N year's duration, it is necessary to resort to a statistical or stochastic model of the continuous hydrograph, which retains information in the hydrograph relevant to the relation, and discard the rest (Das and Simonovic, 2012; Desalegn et al., 2016) and the following two models were available for this purpose.

2.4.1 Annual Maximum Series

In the annual maximum flow (AMF) series, only the peak flow in each year of record is considered. Desalegn et al. (2016) discussed that a series of AMF flood is implicit to form a random sample from the stationary population in which is accidental variable with distribution. In the AMF flow series, only the peak flow in each year of record is considered, that may occupy some loss of information (Chow et al.,1988). An AMF is a universally used model by different investigators for the purpose of flood frequency analysis (Badreldin and Fengo, 2012).

2.4.2 Partial Duration Series

In this model, most of the flow hydrograph is disregarded and the hydrograph is viewed as a series of randomly spaced flood peaks of random magnitude. For the case of statistical modeling and identification of the values, which form the series, only the series of peak exceeding an arbitrary threshold are considered. In partial duration series, all peaks above a certain base value are considered. The base is usually selected low enough to include at least one event each year (Rao and Hammed, 2000). Therefore, to avoid the problem of data dependency, the annual maximum flow series model was selected. In addition to this, AMF series is widely and universally used model by different researchers for the purpose of flood frequency analysis (Desalegn *et al*., 2016). As a result, to keep away from the concern of requirement on data, AMF series model was chosen.

2.5 Regionalization

The availability of data is an important aspect in frequency analysis. The estimation of probability of occurrence of extreme floods is an extrapolation based on limited data. Thus, the larger the database, the more accurate the estimates will be. From a statistical point of view, estimation from small samples may give unreasonable or physically unrealistic parameter estimates, especially for distributions with a large number of parameters (three or more). Large variations associated with small sample sizes cause the estimates to be unrealistic (Raghunath H. M., 2006).

In practice, however, data may be limited or in some cases may not be available for a site. In such cases, regional analysis is most useful. Regional analysis is based on the concept of regional homogeneity which assumes that annual maximum flow populations at several sites in a region are similar in statistical characteristics and are not dependent on catchment size (Cunnane, 1989). Although this assumption may not be strictly valid, it is convenient and effective.

In flood frequency analysis regionalization reflects about the identification of homogeneous regions. This suggests grouping of the into homogeneous regions, which contain stations of identical flood producing features through the homogeneity test and selection of appropriate distribution for the identified regions (Sine and Ayalew, 2004). Delineation of hydrologically homogeneous regions is a common major step of any RFFA. Regionalization is performed to transfer the hydrologic characteristics from gauged basins to ungauged basins. But, due to the complexity of factors that affect the generation of floods, there is no universally accepted unbiased method of regionalization (Kachroo *et al*., 2000; Mishra *et al*., 2010).

Regionalization serves two purposes. For sites where data are not available, the analysis is based on regional data (Cunnane, 1989). For sites with available data, the joint use of data measured at a site, called at-site data, and regional data from a number of stations in a region provides sufficient information to enable a probability distribution to be used with greater reliability. This type of analysis represents a substitution of space for time where data from different locations in a region are used to compensate for short records at a single site (National Research Council, 1988; Stedinger et al., 1993).

In regionalization, expectations are about the statistical similarity of the sites in a region. For this analysis the values of coefficient of variation and the site-to-site the coefficient of variation has to be used. For homogeneity test of each site in a region the mean, standard deviation and coefficient of variance have to be calculated (Nobert et al., 2014).

Hydrologic information may be used in a number of ways, depending on the availability of gauging stations within the region. Several attempts have been made by different authors to identify hydrologically homogeneous regions based on either geographical considerations or flood data characteristics, or a combination of both (Kachroo et al., 2000). The set of defined homogenous catchments can then be pooled together and described via statistical properties. This has resulted in regional flood methods being widely used and provide a viable way to estimate discharge in data-poor regions (Zaman et al., 2012).

Many types of regionalization procedures are available (Cunnane, 1988 and 1989). One of the simplest procedures which has been used for a long time is the *index flood* method. The key assumption in the index flood method is that the distribution of floods at different sites in a region is the same except for a scale or index flood parameter, which reflects rainfall and runoff characteristics of each region. The index flood may be the mean flood, although any location parameter of the frequency distribution may be used (Hosking and Wallis, 1991). In this case, regional quantile estimates Q at a given site for a given return period T can be obtained as in equation below, where q_T is the quantile estimate from the regional distribution for the given return period, and μ is the mean flow at the site.

$$Q = \mu_i q_T \tag{2.2}$$

The regional distribution parameters are obtained by using the regional weighted average of dimensionless moments obtained by using the dimensionless rescaled data $\boldsymbol{qij} = \boldsymbol{Q}/\mu_i$. Another method of obtaining the regional distribution parameters is the *station year* approach (Cunnane, 1989) where all the data are pooled, after dividing them by the mean at each site, and are treated as a single sample. The joint use of at-site and regional data is advisable, provided that a reasonably homogeneous flood region can be identified. The data at a site may be used when the record at a station is exceptionally long, or when regional data are not available, or when a region is heterogeneous.

2.6 Identification and Delineation of Homogeneous Regions

The identification of flood-producing natures in data-poor regions has got a considerable attention in recent years (Smith *et al*., 2015). This delineates an area consisting of sites with the same standardized frequency distribution and parameters (Ahmad *et al*., 2016). Such regions are expected to be geographically continuous and basically used for carrying out regional frequency analysis for estimation of flood magnitude of water resources projects (Sine and Ayalew, 2004).

Hosking and Wallis (1997) mentioned all the stages in RFFA involving many sites. The authors discussed that identification of homogeneous regions (IHRs) is usually most difficult and requires the greatest amount of subjective judgment. The grouping into homogeneous regions can be done by the identification of geographically contiguous regions. Geographical proximity does, however, not guarantee hydrological similarity (Patil and Stieglitz, 2012).

The catchments of a given homogeneous region may not be geographically contiguous, but may similar in terms of their flood generation processes. So, it is powerfully preferred to cluster regions on site characteristics and at site statistics in the consequent testing of homogeneity of a suggested set of regions (Hosking and Wallis, 1997). This is to group sites that almost satisfy homogeneity condition of sites. Several authors (Capesius and Stephens, 2009; Noto and Loggia, 2009) have proposed methods for forming groups of similar sites for use in RFFA. This include discordancy measure (D), similarity in high flow regime, similarity in high-flow seasonality, homogeneity in hourly runoff response, comparing at-site versus regional parameters and similarity in L-moment ratios (LMR).

2.7 Tests on Hydrologic Data

Two basic assumptions in statistical flood frequency analysis are the independence and stationarity of the data series. In addition, the assumption that the data come from the same distribution (homogeneity) is made. The following tests, which are also discussed by Bobeé and Ashkar (1991), are commonly used to test for stationarity, homogeneity and independence of data. Other (similar) tests are found in Kite (1977).

2.7.1 Test for Independence and Stationarity

Given a sample of size N, the Wald-Wolfowitz (1943) (W-W) test is used to test for the independence of a dataset and to test for the existence of trends in it. For a data set x_1, x_2, …, x_N the statistic R is calculated from Equation below.

$$R = \sum_{i=1}^{N-1} X_i X_{i+1} + X_1 X_N \qquad (2.3)$$

When the elements of the sample are independent, *R* follows a normal distribution with mean and variance given by the following equations.

$$R_{mean} = \frac{(S_1^2 - S_2)}{N-1} \qquad (2.4)$$

$$Var(R) = \frac{S_2^2 - S_4}{N-1} - R_{mean}^2 + \left(\frac{S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4}{(N-1)(N-2)}\right) \qquad (2.5)$$

Where $s_r = Nm_r'$ and m_r' is the r^{th} moment of the sample about the origin.

The statistics $u = (R - R_{mean})/(var(R))^{1/2}$ is approximately normally distributed with mean zero and variance unity and is used to test the hypothesis of independence at significance level α, by comparing the statistic *u* with the standard normal variate $u_{\alpha/2}$ corresponding to a probability of exceedance $\alpha/2$.

2.7.2 Tests for Homogeneity and Stationarity

The L-moment based homogeneity tests form the basis to check the regions for homogeneity. The regions formed by any regionalization method are heterogeneous in general and need adjustments to make homogeneous (Rao and Srinivas, 2008). To investigate whether those have been meeting or not; many researchers have used the values of the mean of CV and the site-to-site CV of both convention and L-moment of the proposed region. Homogeneity tests based on Cv and LCV are applied to verify if the preliminary identified and delineated region is homogeneous. In this case, the hydrological data have to be used and the region is confirmed to be homogeneous if it satisfies both criteria of homogeneity tests (Nobert *et al*., 2014).

The discordance measure estimates how far a given site from the center of the group. It is also helpful to screen out the data from unusual sites to look for the appropriate datasets for regionalization (Hosking and Wallis,1997; Parida *et al*.,1998; Rao and Hamed, 2000; Noto and Loggia, 2009; Kanti *et al*., 2017). These sites were due to the presence of inaccuracies in data or some other local conditions (Guru and Jha, 2016). It is a useful measure in assessing whether any of the regions obtained from the cluster analysis contain potential outliers and should, therefore, be adjusted accordingly (Smith *et al*., 2015).

In RFFA, the homogeneity of a group of stations is an essential assumption. A standard procedure in hydrology to evaluate this condition is the homogeneity measures, which relates to L-moments (Hosking and Wallis, 1997; Lilienthal *et al*., 2018). In this method stations data in a region can be tested for homogeneity and can be taken as a base for many criteria for any other investigation (Malekinezhad *et al*., 2011). It is unreasonable to expect that a region can be chosen in which the flood frequency distribution at all site are identical. The delineation of the homogeneous region is important for site characteristics to be truly representative of the observed discharge data used to estimate hydrologic design values (Irwin *et al*., 2014).

2.7.3 Test for Outliers

An outlier is an observation that deviates significantly from the bulk of the data, which may be due to errors in data collection, or recording, or due to natural causes. The presence of outliers in the data causes difficulties when fitting a distribution to the data. Low and high out-liers are both possible and have different effects on the analysis. The Grubbs and Beck (1972) test (G-B) may be used to detect outliers. In this test the quantities X_H and X_L are calculated by using the following equations.

$$X_H = \exp(X_{mean} + K_N s) \qquad (2.6)$$

$$X_L = \exp(X_{mean} - K_N s) \qquad (2.7)$$

where X_{mean} and s are the mean and standard deviation of the natural logarithms of the sample, respectively, and K_N is the G-B statistic tabulated for various sample sizes and significance levels by Grubbs and Beck (1972). At the 10% significance level, the following approximation proposed by Pilon et al. (1985) is used, where N Is the sample size.

$$K_N = -3.6220 + 6.28446N^{1/4} - 2.49835N^{1/2} + 0.491436N^{3/4} - 0.037911N \quad (2.8)$$

2.8 Statistical Distributions for Flood Frequency Analysis

The main aim of regional flood frequency analysis is to find a distribution that will yield as accurate as possible quantile estimates for each site. The optimal of distribution is the one which is capable of giving good quantile estimates when several distributions fit the data adequately (Hosking and Wallis, 1997). The choice of distribution for a given application is generally made arbitrarily as there is no sound physical basis to justify the selection (Rahman *et al*., 2013).

The choices of the distribution models were based on the previous studies where most of these have been used and recommended in various countries. This is also influenced by many factors, such as methods of discrimination between distributions, methods of estimation parameters and the availability of data (Kumar and Chatterjee, 2011).

2.8.1 Best Fit Probability Distribution

Probability distribution fitting is judging a suitable probability distribution to a given dataset. In flood frequency analysis accurate estimation of maximum flood are obtained by fitting probability distribution for a specified return period (Vivekanandan, 2015). The objective is to predict the frequency of occurrence of the magnitude of phenomenon in a certain interval. This can lead to a good prediction of flood. The probability distributions most closely fitted to the observed data depends on the nature of the occurrence and the distribution (Athulya and James, 2017).

Thus, choosing the best statistical distribution is the most important factor in frequency analysis. Therefore, different distributions must use and then, the most appropriate distribution of data should be selected (Amirataee *et al*., 2014). In flood frequency analysis, an assumed probability distribution is fitted to the available data to estimate the flood magnitude for a specified return period. Details of commonly used distributions in flood data are found in Rao and Hammed (2000).

The first of error, which is associated with the wrong assumption of a particular distribution for the given data checked to a certain extent by using goodness-of-fit tests (Millington *et* *al*., 2011). A couple of goodness-of-fit tests have been conducted such as Kolmogorov-Smirnov test, Anderson-Darling test along with the chi-square test at significance level (a=0.05) to assess the reasonability and check the adequacy of best-fitting probability distributions to the recorded data. These are statistical tests, which provide a probabilistic framework to evaluate the adequacy of a distribution. The selection of a distribution for flood frequency analysis goes with the selection of the method of parameter estimation (Das and Simonovic, 2012).

2.8.2 Goodness of Fit Tests

The choice of distribution to be used in flood frequency analysis has been a subject of interest for a long time (Rao and Hamed, 2000). These test statistics are used for checking the validity of a specified or assumed probability distribution model. The method of moments, maximum likelihood and L-moments are used for parameter estimations. These parameters are used to calculate the quantiles corresponding to return periods (Ashraful *et al*., 2018). The results of the goodness of fit tests are used to select a distribution for observed flow at the site of interest (Ghosh *et al*., 2016).

2.8.3 Method of L-Moment Ratio Diagram

The suitability of probability distribution can be assessed with the help of L-moment ratio diagrams. This method effectively used in regional frequency analysis to select a distribution for a region (Das and Simonovic, 2012). The LMRDs are considered as a reliable diagnostic tool for identifying a probability distribution. This provides a visual comparison of the sample estimates with the population values of L-moments and is always preferable to product moment ratio diagram for a goodness-of-fit test (Hosking and Wallis, 1997; Amalina *et al*., 2016). An advantage of LMRD is that one can compare the fit of several distributions using a single graphical instrument (Chavoshi and Azmin, 2009).

This is a graph between L-kurtosis and L-skewness. Usually, a two-parameter distribution with a location and a scale parameter plots as a single point on such a diagram while a three-parameter distribution with the location, scale and Shape plots as a line or curve on the diagram. The Clusters of stations are categorized into different regions based on the proximity of stations in the LMRD (Hosking And Wallis, 1997; Badreldin And Fengo, 2012).

2.8.4 Parameter Estimation

The data analysis often requires estimation of parameters for a few probability distributions. Before the analysis can be done, the parameter for each selected distribution needs to be estimated first (Ahmad *et al*., 2011). Since the parameters are estimated from the sample data, the estimates are subject to sampling errors. A method of fitting must be chosen to minimize these errors. A method suitable to estimate the parameters of one distribution might not necessarily be as efficient for another distribution. Hosking and Wallis (1997) noted that even if an acceptable distribution is selected, proper estimation of parameters is important. Some of the parameter estimation methods are:

i. Probability-weighted moments

Probability-weighted moments (PWM) are useful in the deriving expression for the parameters of distributions can be explicitly defined. Methods of parameter estimation obtained in this method are by equating moment of the distribution with the corresponding sample moment of observed data. For a distribution with a parameter, the first sample moments are set equal to the corresponding population moments. The resulting equation then solved simultaneously for the unknown parameters. Parameter estimation by PWM, which is relatively new is as easy to apply as ordinary moments is usually unbiased and is almost as efficient as MML. Indeed, in small samples, PWM may be as efficient as MML; with a suitable choice of distribution PWM, contributes to assessing robustness point of view (Cunnane, 1989).

ii. Method of moment

Method of moment (MOM) is relatively easy and is more commonly used methods of estimating parameters of a probability distribution. It can also be used to obtain starting values for numerical procedures involved in MML estimation. However, MOM estimates are generally not as efficient as the MML estimates. Especially for distributions with a large number of parameters, as higher order moments are more likely be used to obtain starting values for numerical procedure involved in MML estimation and to be highly biased for relatively small samples. The most popularized method to frequency analysis in recent time is that L-moment approach introduced by Hosking and Wallis (1997).

iii. Method of maximum likelihood

The method of maximum likelihood (MML) is considered to be the most accurate method, especially for large datasets since it leads to efficient parameter estimators with Gaussian asymptotic distributions. It provides the smallest variance of the estimated parameters, and hence of the estimated quintiles, compared to other methods. However, with small samples, the results may not converge. This method involves the choice of parameter estimates that produce a maximum probability of occurrence of the observations (Cunnane, 1989). In general, the PWM and MOM are better for estimating the parameters for three and two parameter distributions respectively of the underlying distribution from which the data are sampled. They are less sensitive than others are to sampling variability (outliers), and therefore, they yield more accurate and robust estimates of the characteristics or parameters of the underlying probability distribution (Rao and Srinivas, 2008).

iv. L-Moment method

L-Moments (LMM) are analogous to the method of moments and linear functions of the expectations of order statistics and they are viewed as an alternative system of describing shapes of probability distributions. It is a powerful and efficient method to compute statistical parameters, because such methods can give an unbiased estimate of sample parameters, and cannot easily influence with the presence of outliers (Ghosh et al., 2016; Rao and Hamed, 2000). The L-moments technique depicts accurate predictions of all kinds of statistical analysis and as such, the method can be suggested for policies and decision-making pertaining to hydrological catchment design (Kanti et al., 2017).

Compared to the method of moments and maximum likelihood, L-moments can characterize a wide range of distributions. Sample estimates of L-moments are so forceful, may not be affected by the presence of an outlier in the dataset and less subjected to bias in estimation. L-moments can yield accurate estimates of the parameters of a fitted distribution (Cunnane, 1989).

Quantile estimation

Quantile estimation is the main focus of hydrologic frequency analysis and estimated by applying a distribution function. The selected quantile of under or over design criterion concerning with hydraulic structures is exposed to risk as the return period is determined according to cost and economic-strategic significance of the structure. Selecting a reliable design quantile, are necessary for the delineation of floodplains, the development of floodplain management and flood warning systems, which effects on design, operation, and management of a hydraulic structure, considerably depends on statistical methods used in parameter estimation belonging to the probability distribution (Amalina et al., 2016).

The parameter estimates that maximize the likelihood function are computed by partial differentiation with respect to each parameter and setting these partial derivatives equal to zero and finally solve the resulting set of equations simultaneously. The equations are usually complex as a result of this difficulty; the solution set may not properly found (Cunnane, 1989). Although the use of these parameters yields less biased estimates compared to the two parameter ones, as there is no general agreement in the choice (Parida *et al*.,1998).

When quantiles have to be estimated for sites where no observations have been recorded or observation recorded only for a very small period, and then the estimates using frequency analysis is neither possible nor reliable. RFFA is one of the means to overcome such problems while reasonably quantifying the flood estimates at desired frequencies for series within a more or less hydrological homogeneous region (Dubey, 2014).

After the parameters of a distribution are estimated, quantile estimates (X_T) which correspond to different return periods T may be computed. According to Rao and Srinivas (2008), the return period is related to the probability of non-exceedance (F) by the relation;F = (1 - 1/T) where; F= F (X_T), is the probability of having a flood of magnitude X_T or smaller. The problem then reduces to evaluating X_T for a given value of F. In practice, two types of distribution functions are encountered. The first type is that which can be expressed in the inverse form $X_T = \varphi(F)$. In this case, X_T is evaluated by replacing ϕ (F). In the second type, the distribution cannot be expressed directly in the inverse form $X_T = \varphi(F)$.

2.9 Derivation of Flood Frequency Curves

Flood frequency curves (FFC) describe the relationship between the magnitude of river peak flows and the recurrence interval or return period. FFC, the estimation of flood for various return periods is needed when analyzing flood risk (Das and Simonovic, 2012). Developing FFC for different return period helps to estimate flood quantiles. The curves can be derived from data at flow monitoring stations. Then it regionalized for use at any location along the basin’s river network by relating the spatial differences to geographical regions and to variations in upstream sub basin characteristics inside each region (Willems *et al*., 2009).

Regional flood frequency curves have the ability to considering the spatial pattern of variation of hydrologic phenomena across many gauging sites, can be used for estimating flood quantiles at any ungauged site within the region (Parida *et al*.,1998). In every RFFA, the main goal of the analysis is to develop a regional curve that can represent the averagely weighted distribution of the homogeneous regions. It is the final process of FFA to estimate the normalized regional quantile floods (X_T); FFC for a give return period (Tadesse *et al*., 2011).

For a given region, the model parameters derived from the best-fitted distribution to the observed data are used. This helps to compute standardized quantile estimates and then used to construct a regional flood frequency curve for the homogeneous region. These curves are plots of quantiles representing for all sites of a homogeneous region (Hosking and Wallis,1997; Hailegeorgis and Alfredsen, 2017).

2.10 Previous Studies on RFFA in Ethiopian River Basins

Investigation of regional flood frequency analysis based on monthly rainfall pattern and geographical proximity was conducted by Gebeyehu (1989) for the Blue Nile River Basin. The study had some limitation about the way that it does not delineate homogeneous regions accurately because the responses of the statistical approach in similar rainfall regions are different consequences of changes in basin topography. In his conclusion, Gebeyehu (1989) point out the following information. The regionalization approach provides useful information about the flood frequency of gauged and ungauged catchments, a small amount of site data greatly improves the estimate of the mean annual flood that can be used with a regionally based estimate of X_T relationship and the results of RFFA should always be updated as more relevant information becomes available.

Blue Nile River Basin has also been regionalized into similar flood producing characteristics based on statistics of at site data (Sine and Ayalew, 2004). The author defined a homogeneous region found have to be with geographical proximity and it performs mainly for carrying out regional frequency analysis for estimation of flood magnitude for water resources project planning and design. Identification and delineation of homogeneous regions for all stations of the respective regions satisfy homogeneity criteria. The types of distribution most likely to fit data of each region were identified from the regional average statistical value of L-Moment ratio. The study recommended that selection of best-fit single distribution and dynamic parameter estimation method require further investigation.

Demissie and Michael (2008), Mekoya and Seleshi (2010) established RFFA for upper Awash sub-basin using the application of index flood method. The former recommended that additional testing of stations for homogeneity should be done considering geographical factors are a good method in RFFA of the basin and the later to extend the method of RFFA for the other Ethiopian river basins.

Gedefa and Seleshi (2009) investigated Upper Omo-Gibe sub-basin using index flood estimation based on the observed AMF. L-moment based statistical homogeneity tests were used to identify homogeneous regions. The study concluded that regionalization provides valuable information even in possibly heterogeneous regions, and regional analysis is more accurate and flexible than single-site analysis.

According to Hussein and Wagesho (2016), regionalization of Abaya-Chamo sub-basin was performed based on site characteristics such as elevation, soil type, soil texture, slope, land use land cover and mean annual rainfall. Site statistics were used for testing of homogeneity of the proposed region. The authors concluded that to get reliable quantile estimate more gauging stations should be installed in the basin to infer something for ungagged sites.

Ketsela *et al*. (2017) performed FFA on Awash River Basin using statistical distribution technique. The Easy Fit Software was employed for selection of best-fit distributions and estimation of parameters for stations. Kolmogorov–Smirnov test was used for the choice of a suitable distribution for estimation of maximum flood discharge. According to this study, Awash River basin was delineated into five satisfactory homogeneous regions and recommended software-based techniques like Easy Fit and other alternative statistical software packages to get accurate and reliable flood estimation results.

2.11 Parameter Estimation Model

Data fitting process involves using certain statistical techniques, which allow estimating fitness parameters in accordance to data sample. One advantage of using software to fit the data and interpreting probability data is that they are able to automatically fit data with a variety of known distribution patterns simultaneously. Easy Fit Software is a data analyzer and simulation Software which is capable to fit and simulate statistical distributions with sample data, choose the best model, and then use the obtained result of analysis to provide better decisions. For many distributions, Easy Fit uses the maximum likelihood method regarding the maximization of the log-likelihood function (Mehrannia and Pakgohar, 2014).

Easy Fit Software is an interactive software system to identify parameters, allows the most flexible input of the underlying model in form of Fortran code, and are executable independently from the interface. It consists of a database containing models, data and results, nd of underlying numerical algorithms for solving the parameter estimation problem depending on the mathematical structure (Schittkowski, 2002).

2.12 EasyFit Software

EasyFit software is a data analysis and simulation software which enables us to fit and simulate statistical distributions with sample data, choose the best model, and use the obtained result of analysis to take better decisions. This software can function as a stand-alone windows application or as an add-on for Excel spread sheet (Hossein Mehrannia *et al*. ,2014). Prominent features of this program are: supports more than 50 discrete and continuous distributions, automatic and manual settings, ability to test performed operations, bi-modal graphs, random number generation and integrated help system.

2.13 Flood Inundation Mapping

Flood inundation maps, in particular, can be defined as maps showing inundated area or different parameters such as flood depth and flood velocity and water level. Several important parameters are required for performing hydraulic flood modeling such as topographic data, discharge data to provide model inflow and outflow as boundary conditions, estimation of the manning's roughness coefficient and validation data (Bates 2004). Merz *et al*. (2007) noted that flood mapping is an effective tool for assisting flood hazard management and action to be taken.

The Hydrologic Engineering Center's River Analysis System (HEC-RAS) has the ability to model flood events and produce water surface profiles over the length of the modeled stream. With the companion GIS utility, HEC-Geo RAS, those water surface profiles can easily be converted to flood inundation maps.

3 MATERIALS AND METHODS

3.1 Description of the Study Area

3.1.1 Geographical Location

Awash River Basin is the most intensively utilized and the longest river basin of the Great Rift Valley in Ethiopia and located from 7^0N to 10^0N latitude and 38^0E to 41^0E longitude. The river basin covers a total area of 110,000km^2 with a total length of 1,200 km along its course. Awash river is originated from central Ethiopian highlands, about 150 km west of Addis Ababa, at an elevation of about 3,000 m above mean sea level and flows north-eastwards, where it finally drains into Lake Abe, on the border with Djibouti, some 100 km from the head of the Gulf of Tadjoura. It has been divided into three distinct zones: Upper Basin, Middle Basin and Lower Awash Basin on the basis of various inter-related factors such as location, altitude, climate, topography etc. (MoWE, 2012). For this research part of middle Awash river basin, which is between Awash below Koka dam and Awash Melka Werer stations was considered. Its area covers the total catchment of 106,201.8 km^2.

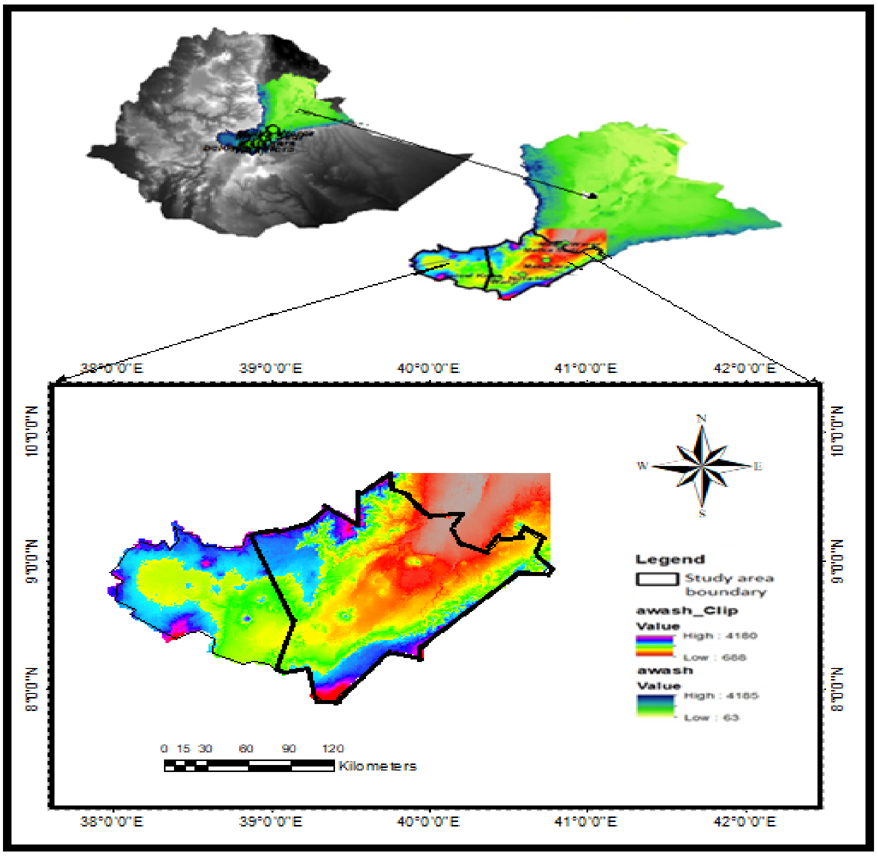


Figure 3. 1 Location map of the study area

3.1.2 Geomorphology

Awash River Basin has two main Physiographic components; the Ethiopian plateau, and the Rift Valley, which covers most of the basin. The topography of the Ethiopian plateau is generally flat with elevations ranging from 2,500m down to 2,000m. the lowest elevation of the plateau, or uplands, is commonly considered to be at 1,500m. Unlike many other Ethiopian rivers, the alluvial plains adjacent to the Awash River are relatively wide in parts extending over 25km.

3.1.3 Climate

The climate of the Awash Basin comes under the influence of the Inter-Tropical Convergence Zone (ITCZ). The seasonal rainfall distribution with in the Basin results from the annual migration of the ITCZ. The mean annual rainfall varies from about 1,600 mm at Ankober, in the highlands North East of Addis Ababa to 160 mm at Asayita on the Northern limit of the basin. The mean annual rainfall over the entire western catchment is 850 mm and the headwaters of the Awash, as gauged at Melka Hombole it is 1,2016 mm. the mean annual temperatures at these stations occurring in June, at 23.8^0C and 33.6^oC respectively. Mean annual wind speed at Koka averages 1.2m/s, the windiest months being June and July with mean monthly values of 1.9 and 1.6m/s respectively (Sifan Abera, cited in MoWE,2011).

3.1.4 Soil and Land Cover

Soils, land cover and rainfall are the major physical catchment characteristics that govern runoff generation. Mostly runoff generation depends on topography, infiltration rate, soil water holding capacity, etc. which are highly related to soil types and land cover. Awash River Basin comprises different soil types. The dominant soils in the study area are leptosoil, chromic, eutric, dystricts, vetric etc. The land covers of the area are the natural vegetation (short grass, savannah, tree/shrubs and marshes, wasteland (desert and sand dunes), agricultural lands and lakes (Haile Arefayne Shishaye and Asmlash Tilahun, 2020).

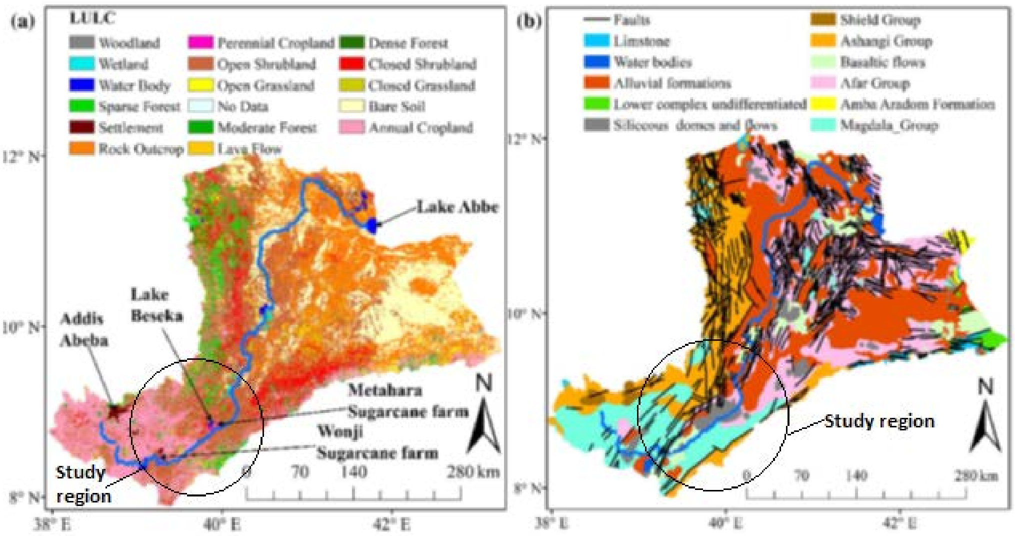


Figure 3. 2 (a) Land use land cover (b) surface geological maps of the Awash River basin (Sourec: Haile Arefayne Shishaye and Asmlash Tilahun, 2020)

3.1.5 Economic Value

Awash River Basin is the most intensively utilized river basin in Ethiopia. According to Likasa, 2013, cited in Behailu, 2004, large irrigation schemes such as Metehara Sugar Factory, Wonji Sugar Factory, Fentale Irrigation Project, Kessem Irrigation Project, horticulture production sites and many more small farm irrigation projects have been functional for many years following the construction of Koka Dam in 1960. has major economic value for the country. There are sugarcane factories (Wanji, Matahara), which are contributing in the nation’s development. Besides these, there are many of large and small-scale irrigation projects implemented in the basin, which has also a great contribution toward the country development.

3.2 Materials Used

Digital elevation model (DEM) and Digital terrain elevation model (DTEM) of the study area were used for mapping of the study area and creating Triangulated irregular network (TIN) which was used to generate river geometries. DEM also employed as basic input for delineation and specifying the location of the gauging stations in the basin. ArcGIS software was used to generate the study area map representing geographical location of gauging stations and delineate hydrologically homogeneous regions. It was also used to create river geometries to map flood inundation area of the selected river reach in combination with Hec-GeoRAS.

Statistical Software (Easy Fit) was used to select the best fit probability distribution with its method of parameter estimations, a goodness of fit tests and to check the estimation accuracy of each of data of stations. Easy Fit allows to automatically or manually fit a large number of distributions to the data and select the best model. It is used as a stand-alone application or with Microsoft Excel, enabling us to solve a wide range of engineering problems with only a basic knowledge of statistics. Microsoft Excel was used for data arrangement and calculate the statistical parameters of hydrological data that was used in the flood frequency analysis.

The Hydrologic Engineering Center's River Analysis System (HEC-RAS) and the Hydrologic Engineering Center's geographic river analysis system (HEC-Geo RAS) softwares were used to develop hydraulic flood modeling to map flood inundation area.

3.3 Data collection and Analysis

The study was begun with collection, analysis and review of available data and information concerning the study area. Hydrological and DEM (digital elevation model) data of Middle Awash River Basin was collected from Ministry of Water, Irrigation and Electricity (MoWIE).

Flood frequency analysis primarily uses observed annual maximum flood data at gauging stations to estimate flood magnitude. Discharge (stream flow) data of various stations in the basin and the site characteristics of stations for this study includes the code of the stations, the name of the river and their gauging sites, the locations (latitude and longitude) and catchment area were collected from MoWIE.

Table 3. 1 The site characteristics of study area river gauging stations

Location of gauging station	Coordinate		Area (Km^2)	Record period	Record length
	Latitude	Longitude			
Awash below Koka	8:28: 0 N	39:10: 0 E	11219	1999-2014	16
Awash@ Wanji	8:27: 0 N	39:14: 0 E	11690	1999-2014	16
Awash @Nura Hera	8:32: 0 N	39:35: 0 E	14173	1999-2014	16
Metehara	8:51: 0 N	39:51: 0 E	16416.8	1999-2014	16
Awash Melka Sadi	9:11:60 N	40:7: 0.012 E	21520	1999-2014	16
Awash Melka Werer	9:19:12N	40:10:12 E	31813	1999-2014	16

In the study area, six gauging stations with length of the at-site AMF records of 16 years were selected for RFFA. All gauging stations satisfied the minimum recording length of 10 years.

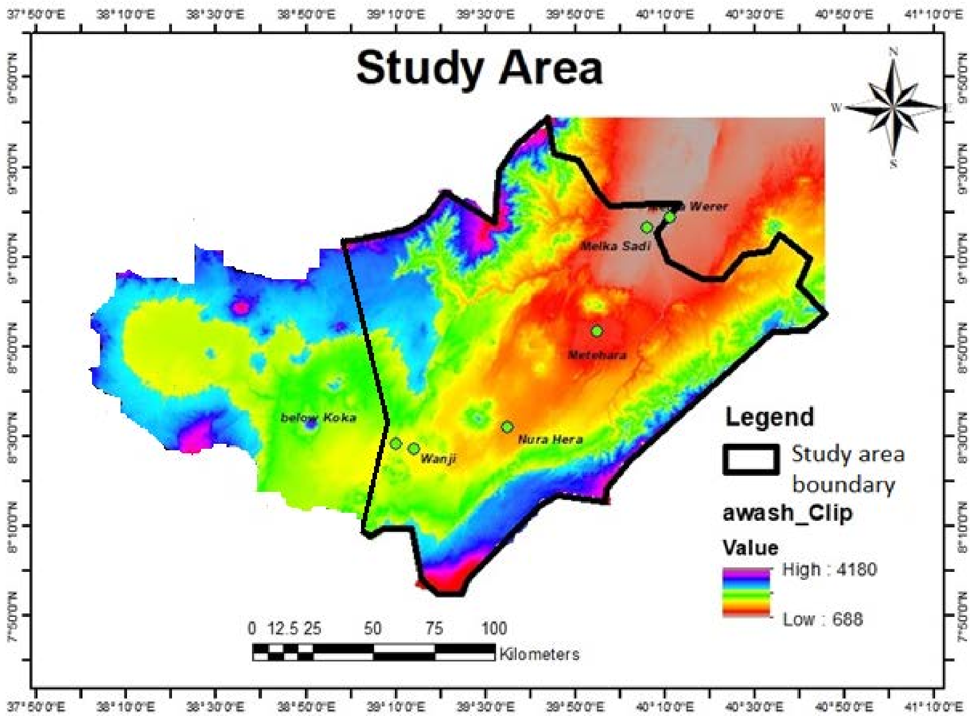


Figure 3. 3 The spatial distribution of gauging stations of the study area

3.3.1 Maximum Flow Time Series

In flood frequency analysis, a unique relationship between a flood magnitude and the corresponding recurrence interval T is very important to extract information from a flow record to estimate the relationship between Q and T. three different models: the annual maximum series (AM), the partial duration series (PD) and the timeseries (TS) model, can be considered for this purpose. The AM model is statistically the most efficient method for flood frequency analysis (Cunnane, 1989). Hence, in this study annual maximum flow (AMF) series was used for regional flood frequency analysis.

The procedures of data analysis in this study includes from the preliminary screening of data to develop a regional flood frequency curve depending on annual maximum flow (AMF) series data. The AMF and monthly maximum flow time series were shown on **Appendix-L.** Fig 3.4 below displays the monthly maximum flow time series data collected from the selected gauging stations for the duration 1999 - 2014. They were analyzed and used for estimation of extreme flow quantiles using flood frequency analysis.

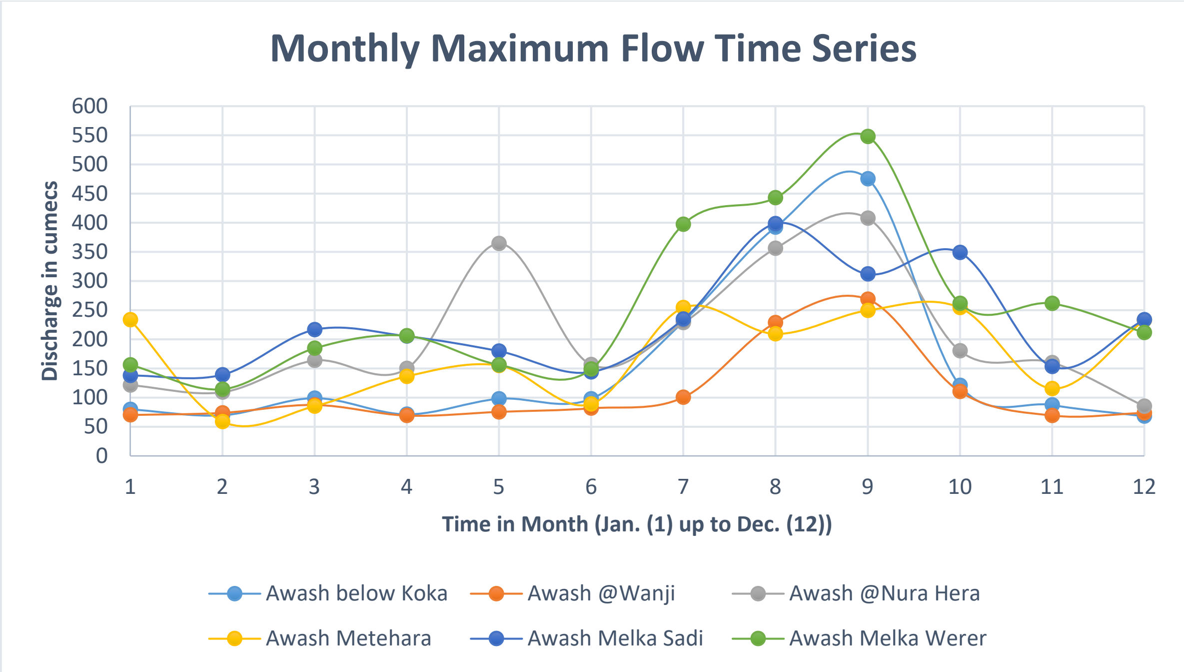


Figure 3. 4 Monthly maximum flow time series of six gauging stations

3.3.2 Checking Data Quality

Basic assumptions in statistical flood frequency analysis are the independence, stationarity and homogeneity of the data series. Data Screening was carried out to check for its continuity, quality and gross errors as careful examination, or validation, of data is essential before they are applied in analysis. Hence, tests such as check for: independency and stationarity, homogeneity and stationarity, outliers and continuity with previous batch of data were conducted to check the quality of data used. The following tests, are commonly used to test for stationarity, homogeneity and independence of data (Rao and Hamed, 2000).

3.3.2.1 Test for Independence and Stationarity

Given a sample of size N, the Wald-Wolfowitz (1943) (W-W) test was used to test for the independence and stationarity of a dataset and to test for the existence of trends in it. For a data set X_1, X_2, …, X_N the statistic R is calculated from Equation below.

$$R = \sum_{i=1}^{N-1} X_i X_{i+1} + X_1 X_N \qquad (3.1)$$

When the elements of the sample are independent, *R* follows a normal distribution with mean and variance given by the following equations.

$$\bar{R} = \frac{(S_1^2 - S_2)}{N-1} \qquad (3.2)$$

$$Var(R) = \frac{S_2^2 - S_4}{N-1} - \bar{R}^2 + \left(\frac{S_1^4 - 4S_1^2 S_2 + 4S_1 S_3 + S_2^2 - 2S_4}{(N-1)(N-2)}\right) \qquad (3.3)$$

Where $S_r = Nm_r'$ and m_r' is the r^{th} moment of the sample about the origin.

The statistics $u = (R - \bar{R})/(var(R))^{1/2}$ is approximately normally distributed with mean zero and variance unity and is used to test the hypothesis of independence at significance level α, by comparing the statistic u with the standard normal variate $u_{\alpha/2}$ corresponding to a probability of exceedance $\alpha/2$. Typical significance levels are: 0.10, 0.05, and 0.01. In this study, the stations' data were concluded to be independent and stationarity at the 5% significance level or 95% confidence level

Table 3. 2 Showing confidence level and critical values (z-value)

Confidence Level	Z-Value
80%	1.28
85%	1.44
90%	1.64
95%	1.96
98%	2.33
99%	2.58

3.3.2.2 Test for Homogeneity and Stationarity

Given a sample of size N, the Mann-Whitney (1947) (M-W) test was used to test for the homogeneity and stationarity of a dataset and to test for the existence of trends in it. In this test two samples of size p and q with are compared. The combined data set of size N = p + q is ranked in increasing order. The Mann-Whitney (1947) (M-W) test considers the quantities V and W in Eqs. 3.4 and 3.5 below.

$$V = R - \frac{\left(p(p+1)\right)}{2} \qquad (3.4)$$

$$W = pq - V \qquad (3.5)$$

R is the sum of the ranks of the elements of the first sample (size p) in the combined series (size N). V represents the number of times an item in sample 1 follows an item in sample 2 in the ranking. Similarly, W can be computed for sample 2 following sample 1. The M-W statistic U is defined by the smaller of V and W. When N > 20 and p, q > 3, and under the null hypothesis that the two samples came from the same population, U is approximately normally distributed with mean and variance from the following formula.

$$U_{mean} = \frac{pq}{2} \tag{3.6}$$

$$var = \left[\frac{pq}{N(N-1)}\right]\left[\frac{N^3 - N}{12} - \sum T\right] \tag{3.7}$$

$$T = \frac{J^3 - J}{12} \tag{3.8}$$

J is the number of observations tied at a given rank. T is summed over all groups of tied observations in both samples of size p and q. The statistic was used to test the hypothesis of homogeneity at significance level α by comparing it with the standard normal variate for that significance level.

3.3.2.3 Test for Outliers

An outlier is an observation that deviates significantly from the bulk of the data, which may be due to errors in data collection, or recording, or due to natural causes. The presence of outliers in the data causes difficulties when fitting a distribution to the data. Low and high outliers are both possible and have different effects on the analysis. In this study, the Grubbs and Beck (1972) test (G-B) was used to detect outliers and the quantities Q_H and Q_L are calculated by using the following formula.

$$Q_H = \exp(\bar{x} + k_N s) \tag{3.9}$$

$$Q_H = \exp(\bar{x} - k_N s) \tag{3.10}$$

where $\bar{x}$ and s were the mean and standard deviation of the natural logarithms of the sample respectively, and k_N is the G-B statistic tabulated for various sample sizes and significance levels by Grubbs and Beck (1972). At the 10% significance level, the following approximation proposed by Pilon et al. (1985) is used, where N is the sample size.

$$K_N = -3.62201 + 6.28446N^{1/4} - 2.49835N^{1/2} + 0.491436N^{3/4} - 0.037911N \tag{3.11}$$

For this study, k_N for N = 16 is 2.279

3.4 Identifying and Delineating Homogeneous Regions

3.4.1 Discordancy Measure of Regions

This is sites screening test to identify discordant sites with the group as a whole based on L-moment ratios of a site from the regional average L-moment ratios f a pooling group (Hosking andWallis,1997). The objective of the discordancy measure (Di) test (Hosking and Wallis 1997) is to identify those sites from a group of given sites that are grossly discordant with the group as a whole and to screen out the annual maximum flow data to check that the data are appropriate for performing the regional flood frequency analysis.

In this study, screening of the data was performed using the L-moment-based discordancy measure (Di). Discordancy is measured in terms of the L-moments of the sites' data, and the aim is to identify those sites that are grossly discordant with the group as a whole. The sample L-moment ratios (τ_2, τ_3, and τ_4) of a site are considered as a point in a three-dimensional space. A group of sites form a cluster of such points in the three-dimensional space. A site is considered discordant if it is far from the center of the cluster. Hosking and Wallis (1997) defined the discordancy measure (Di) for a site i in a group of N sites. Let Ui = $[\tau_2{}^{(i)}\ \tau_3{}^{(i)}\ \tau_4{}^{(i)}]^T$ be a vector containing the sample L-moment ratios τ_2, τ_3, and τ_4 values for site i, analogous to their regional values termed as τ2, τ3, and τ4, expressed in Eq. (3.32), T denotes transposition of a vector or matrix then the discordancy measure may be defined as:

$$Di = \frac{1}{3}(Ui - \overline{Ui})S^{-1}(Ui - \overline{Ui})^T \tag{3.12}$$

$$S = \frac{1}{(N-1)} * \sum_{i=1}^{N}(Ui - \overline{Ui})(Ui - \overline{Ui})^T \tag{3.13}$$

$$\overline{Ui} = \frac{1}{N} * \sum_{i=1}^{N} Ui \tag{3.14}$$

To estimate discordancy values for sites in a region, the sites are considered as points in three- dimensional space of sample L-moment ratios (LCV, LCs, and LCk). If a vector, Ui = $(\tau^i{}_2, \tau^i{}_3, \tau^i{}_4)^T$, which controlled the L-moment ratios for site i, T is the transpose of the vector Ui (Hosking and Wallis, 1997), then

Where N-is the total number of sites

Di-discordancy measure

Ui-is defined as a vector containing the L-moment ratios for site i,

$\overline{Ui}$-is the group averages Ui,

S-sample covariance matrix of Ui

The site i is declared to be discordant, if Di is greater than the critical value of the discordancy statistic Di, given in a tabular form by Hosking and Wallis (1997).

Table 3. 3 Critical values for the discordancy statistic Di (from Hosking and Wallis, 1997)

Number of sites in a region	Critical value	Number of sites in a region	Critical value
5	1.333	6	1.648
7	1.917	8	2.140
9	2.329	10	2.491
11	2.632	12	2.757
13	2.869	14	2.971
>15	3		

A large value of D_i indicates the discordancy of site i with other sites and the definition of large depends on the number of sites in the study region.

Thus, a site is regarded as discordant if D_i exceeds the critical value given in tabular form (see Table 3.6) (Hosking and Wallis 1997).

3.4.2 Identifying Homogeneous Regions

The purpose of this step is to form groups of stations that satisfy the homogeneity condition, that is, stations with frequency distributions that are identical apart from a station-specific scale factor. For this research, a heterogeneity measure called the H-statistic was used to compare site variation in sample L-moments for a group of sites with what would be expected for a homogeneous region (Hosking and Wallis 1997).

Hosking and Wallis (1997) constructed the statistic H based on L-C_v(t)

$$H = \frac{(V_1 - \mu_v)}{\sigma_v} \qquad (3.15)$$

$$V_1 = \sum_{i=1}^{Ns} N_i\left(t^{(i)} - \bar{t}\right)^2 / \sum_{i=1}^{Ns} N_i \qquad (3.16)$$

$$\bar{t} = \left(\sum_{i=1}^{Ns} N_i t^{(i)}\right) / \left(\sum_{i=1}^{Ns} N_i\right) \qquad (3.17)$$

Where, V1 is the weighted standard deviation of (t), Ns is the number of sites, N_i is the record length at each site and $\bar{t}$ is the average value of $t^{(i)}$ and μ_v and σ_v are mean and standard deviation respectively of N_{sim} simulated values of V_1. For this study, in order to obtain reliable values of μ_v and σ_v five hundred simulations were carried out using four parameter Kappa distribution for computing the measures of heterogeneity. The H criteria established by Hosking and Wallis (1993) indicate that the region is acceptably homogeneous if $H < 1$, possibly heterogeneous if $1 \leq H < 2$ and definitely heterogeneous if $H \geq 2$.

3.4.3 Delineation of Homogeneous Regions

The performance of any regional estimation method highly depends on the grouping of sites into homogeneous regions. In this study, the geographical proximity and the method of L-moment ratio diagram (LMRD) were used in order to cluster preliminary regions which then tested for hydrologic similarity. The delineation of homogeneous regions is closely related to the identification of the common regional distributions that apply within each region. A region can only be considered homogeneous if sufficient evidence can be established that at different sites in the region are drawn from the same parent distribution. Under this method the following procedures were followed.

- DEM of Middle Awash River basin was used and the delineation of homogeneous regions was performed by taking in to account the drainage boundaries of the sub-basin with ArcGIS 10.4.1 environment.
- The preliminarily identified regions were checked by various homogeneity tests. All sample stations are located on a digitized map by latitude and longitude

For delineation of the boundary:

- ✓ The values of linear coefficient of skewness (LCs) and linear coefficient of kurtosis (LCk) of each station were computed.
- ✓ The location of stations along the distributions of LMRD for the defined regions was identified using statistical comparison of observed flood data
- ✓ The group was identified based on the identified location of stations
- ✓ Statistical homogeneity was checked using the proposed tests.

3.5 Selection of Best Fit Regional Probability Distribution

3.5.1 Regional Probability Distributions

After confirming the homogeneity of the study region, an appropriate distribution needs to be selected for the regional frequency analysis. In this study, the best fit distribution for the region was identified based on the results of a L-moment ratio diagram and goodness-of-fit test derived from L-moment ratios as described by Hosking and Wallis (1997). When a certain distribution is chosen as data distribution, it is expected to fit suitably with data, so we are ready to practically fit those data with the distributions.

In this study, five of the most frequently used distributions in the analysis of hydrologic extreme variables, viz. Generalized extreme value (GEV), Generalized logistic (GLO), Generalized Pareto (GPA), Log Pearson type 3 (LPIII) and Lognormal (LN) distributions were applied to find out the best-fit probability distribution. Many flood frequency distributions have been practiced for flood modeling, but none has been accepted as universal. Hence, these distributions were considered for the evaluation of the possible distributions that can represent the average frequency distribution of the regional data of the basin. The Mathematical relations for each model have been shown as follows.

Mathematical relations for selected distributions

1. Generalized extreme value (GEV) Distribution

$$\boldsymbol{Parameters}: k = shape\ parameter, \sigma = scale\ parameter\ (\sigma > 0) and$$
$$\mu\ = location\ parameter$$

$$\boldsymbol{Domain}: 1 + k\frac{(x-\mu)}{\sigma} > 0\ for\ k \neq 0\ and - \infty < x < +\infty\ for\ k = 0$$

Probability density function (PDF)

$$f(x) = \frac{1}{\sigma} exp\left(-(1+kz)^{-\frac{1}{k}}\right)(1+kz)^{-1-\frac{1}{k}} \quad for\ k \neq 0 \qquad (3.18a)$$

$$f(x) = \frac{1}{\sigma} exp\big(-z - exp(-z)\big) \quad for\ k = 0 \qquad (3.18b)$$

Cumulative distribution function (CDF)

$$F(x) = exp\big(-(1+kz)^{-1/k}\big) \quad for\ k \neq 0 \qquad (3.18c)$$

$$F(x) = exp\big(-exp(-z)\big) \quad for\ k = 0 \qquad (3.18d)$$

$$z \equiv \frac{x-\mu}{\sigma} \qquad (3.18e)$$

For k >0, the distribution has a finite upper bound at $\mu + \sigma/k$ and corresponds to the EV3 distribution. For $k < 0$, the distribution has a thicker right-hand tail and corresponding to the EV2 distribution. The EV1 distribution is obtained when k=0

2. Generalized Logistic (GLO) Distribution

$\boldsymbol{Parameters}$: $k = shape\ parameter, \sigma = scale\ parameter\ (\sigma > 0) and$

$\mu\ = location\ parameter$

$\boldsymbol{Domain}$: $1 + k\frac{(x-\mu)}{\sigma} > 0\ for\ k \neq 0\ and -\infty < x < +\infty\ for\ k = 0$

Probability density function (PDF)

$$f(x) = \frac{(1+kz)^{-1-1/k}}{\sigma(1+(1+kz)^{-1/k})^2} \quad for\ k \neq 0 \qquad (3.19a)$$

$$f(x) = \frac{\exp(-z)}{\sigma(1+\exp(-z))^2} \quad for\ k = 0 \qquad (3.19b)$$

Cumulative distribution function (CDF)

$$F(x) = \frac{1}{1+(1+kz)^{-1/k}} \quad for\ k \neq 0 \qquad (3.19c)$$

$$F(x) = \frac{1}{1+\exp(-z)} \quad for\ k = 0 \qquad (3.19d)$$

$$z \equiv \frac{x-\mu}{\sigma} \qquad (3.19e)$$

3. Generalized Pareto (GPA) Distribution

$\boldsymbol{Parameters}$: $k = shape\ parameter, \sigma = scale\ parameter\ (\sigma > 0) and$

$\mu\ = location\ parameter$

$\boldsymbol{Domain}$: $\mu \leq x + \infty\ \ for\ k \geq 0\ and\ \mu \leq x \leq \mu - \frac{\sigma}{k}\ for\ k < 0$

Probability density function (PDF)

$$f(x) = \frac{1}{\sigma}\left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-1-1/k} \quad for\ k \neq 0 \qquad (3.20a)$$

$$f(x) = \frac{1}{\sigma}exp\left(-\frac{(x-\mu)}{\sigma}\right) \quad for\ k = 0 \qquad (3.20b)$$

Cumulative distribution function (CDF)

$$F(x) = 1 - \left(1 + k\frac{(x-\mu)}{\sigma}\right)^{-\frac{1}{k}} \quad for\ k \neq 0 \qquad (3.20c)$$

$$F(x) = 1 - exp\left(-\frac{(x-\mu}{\sigma}\right) \quad for\ k = 0 \qquad (3.20d)$$

4. Lognormal (LN) Distribution

$\boldsymbol{Parameters}$: $\sigma = shape\ parameter(\sigma > 0), \mu = scale\ parameter\ and$
$\gamma = location\ parameter(\gamma \equiv 0\ yields\ the\ two\ parameter\ Lognormal\ distribution)$
$\boldsymbol{Domain}$: $\gamma < x + \infty$

Probability density function (PDF)

$$f(x) = \frac{exp\left(-\frac{1}{2}\left(\frac{ln(x-\gamma)-\mu}{\sigma}\right)^2\right)}{(x-\gamma)\sigma\sqrt{2\pi}} \qquad (3.21a)$$

Cumulative distribution function (CDF)

$$F(x) = \Phi\left(\frac{ln(x-\gamma)-\mu}{\sigma}\right) \qquad (3.21b)$$

5. Log Pearson Type 3 (LPIII) Distribution

$\boldsymbol{Parameters}$: $\alpha = shape\ parameter(\alpha > 0), \beta = scale\ parameter\ (\beta \neq 0) and$
$\gamma\ = location\ parameter$

$\boldsymbol{Domain}$: $0 < x \leq e^{\gamma}\ \ for\ \beta < 0\ and\ e^{\gamma} \leq x + \infty\ \ for\ \beta > 0$

Probability density function (PDF)

$$f(x) = \frac{1}{x|\beta|\,\Gamma(\alpha)}\left(\frac{\ln(x)-\gamma}{\beta}\right)^{\alpha-1} exp\left(-\frac{\ln(x)-\gamma}{\beta}\right) \qquad (3.22a)$$

Cumulative distribution function (CDF)

$$F(x) = \frac{\Gamma\,(\ln(x)-\gamma)/\beta(\alpha)}{\Gamma\,(\alpha)} \qquad (3.22b)$$

3.5.2 Selection of Regional Probability Distribution

In this study, the selection was carried out based on the results of L-moment ratio diagram and goodness-of-fit test computed by EasyFit software.

3.5.2.1 Using L-Moment Ratio Diagram

L-Moments

L-moments are an alternative system of describing the shapes of probability distributions. Recently a number of regional flood frequency analysis studies have been carried out based on the L-moment approach. The L-moment methods are demonstrably superior to those that have been used previously and are now being adopted by many organizations worldwide (Kumar R. and Chatterjee C., 2011). The L-moments offer significant advantages over ordinary product moments, especially for environmental data sets, because of the following points (Kumar R. and Chatterjee C., 2011).

i) L-moment ratio estimators of location, scale, and shape are nearly unbiased, regardless of the probability distribution from which the observations arise.
ii) L-moment ratio estimators such as L-coefficient of variation, L-skewness, and L-kurtosis can exhibit lower bias than conventional product moment ratios, especially for highly skewed samples.
iii) The L-moment ratio estimators of L-coefficient of variation and L-skewness do not have bounds which depend on sample size as do the ordinary product moment ratio estimators of coefficient of variation and skewness.
iv) L-moment estimators are linear combinations of the observations and thus are less sensitive to the largest observations in a sample than product moment estimators, which square or cube the observations.
v) L-moment ratio diagrams are particularly good at identifying the distributional properties of highly skewed data, whereas ordinary product moment diagrams are almost useless for this task.

According to the Central Water Commission (2010), L-moments has the following advantages:

a) Characterize most of probability distributions than conventional moments,
b) Less sensitive to outliers in the data,
c) Approximate their asymptotic normal distribution more closely,
d) Nearly unbiased for all combinations of sample sizes and populations.

L-moments arose as modifications of probability weighted moments (PWMs) (R. Kumar and C. Chatterjee, 2011). Probability weighted moment (PWM) is defined as:

$$M_{p,r,s} = E(x^p\{F\}^r\{1-F\}^s) = \int_0^1 \{x(F)\}^p F^r \{1-F\}^s dF \qquad (3.23)$$

Where F = F(x) is the cumulative distribution function (CDF) for x, x(F) is the inverse CDF of x evaluated at the probability F, and p, r and s are real numbers. If p is a nonnegative integer, $M_{p,0,0}$ represents the conventional moment of order p about the origin. If p=1 and s =0,

$$M_{1,r,0} = \beta_r = \int_0^1 x(F)F^r dF \qquad (3.24)$$

For an ordered sample $X_1 \leq X_2, \ldots, \leq x_N$, $N > r$, the unbiased sample PWMs are given by

$$\hat{\beta}_r = \frac{1}{N} \frac{\sum_{i=1}^{N} \binom{i-1}{r} x_i}{\binom{N-1}{r}} \qquad (3.25)$$

For any distribution the *rth* *L*-moment λ_r is related to the *rth* PWM, through

$$\lambda_r = \sum_{k=0}^{r} \beta_k (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} \qquad (3.26)$$

These *L*-moments are linear functions of PWMs. The first four *L*-moments are related to the PWMs using

$$\lambda_1 = \beta_0,$$

$$\lambda_2 = 2\beta_1 - \beta_0,$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0,$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 3.15 \qquad (3.27)$$

Where;

$$\beta_0 = n^{-1} \sum_{j=1}^{n} x_j \qquad (3.28)$$

$$\beta_1 = n^{-1} \sum_{j=2}^{n} x_j \, [(j-1)/(n-1)] \qquad (3.29)$$

$$\beta_2 = n^{-1} \sum_{j=3}^{n} x_j \, [(j-1)(j-2)]/[(n-1)(n-2)] \qquad (3.30)$$

$$\beta_3 = n^{-1} \sum_{j=4}^{n} x_j \, [(j-1)(j-2)(j-3)]/[(n-1)(n-2)(n-3)] \qquad (3.31)$$

Where n is length of the recorded annual maximum flow data of each station and j is the rank of the data after arranged in ascending order ($Q_{l:n}$)

L-Moment Ratios

L-Moments are analogous to their conventional counterparts as they can be directly interpreted as measure of scale and shape of probability distributions and hence are more convenient than the PWMs. The L-moment measure of location, and L-moment ratio measures of scale, skewness and kurtosis were computed by Hosking and Wallis (1997) as:

$$L - coefficient\ of\ variation: L - Cv\ (\tau_2) = \lambda_2 / \lambda_1$$

$$L - coefficient\ of\ skewness: L - Cs\ (\tau_3) = \lambda_3 / \lambda_2$$

$$L - coefficient\ of\ kurtosis: L - Ck\ (\tau_4) = \lambda_4 / \lambda_2 \qquad (3.32)$$

Where; λ_1(mean) is a measure of location, τ_2 $(L - Cv)$ is a measure of scale, L-skewness (τ_3) is: Measure of skewness and L-kurtosis (τ_4) is measure of kurtosis

Analogous to the conventional moment ratios, λ_1 is a measure of location, τ_2 is a measure of scale and dispersion, τ_3 is a measure of skewness, and τ_4 is a measure of kurtosis. Hosking (1990) showed that for $x \geq 0$, the value of τ_2 lies between 0 and 1, while the absolute values of τ_3 and τ_4 lie between 0 and 1. This restriction in the values of the *L*-coefficients works out to be an advantage in their interpretation as opposed to the conventional moments which do not have any bounds (R. Kumar and C. Chatterjee, 2011).

Method of L-moment ratio diagram is used as a tool to give priority for identification of homogeneous regions and distributions based on the statistical principles. It can be used to compare the L-skewness versus L-kurtosis relations of different distributions and data samples. This gives a visual indication of which distribution may be expected to give a good fit to a data points by plotting them on a graph of L-skewness versus L-kurtosis. The main hypothesis of the study is that if the annual maximum flows of different stations come from a single distribution model, then these stations belong to the same group and form a homogeneous region.

L-moment ratio diagram is a useful way of representing the moments of different distributions depending on the statistical nature of data. L-moment statistics are used to group stations comparing with geographical proximity and continuity of gauging stations.

Using L-moment ratio diagram, a model that minimize uncertainties was chosen by applying different distributions. The test can be viewed as plotting the regional L-Skewness and L-Kurtosis values obtained from the sites for the proposed region and identifying the probability distribution(s) whose L-Kurtosis value most closely matches the regional L-Kurtosis value. The following equations were used for each selected probability distributions to generate its L-moment ratio diagram (Mishra et al., 2009).

- Generalized extreme value (GEV),

$$C_k = 2.695079 + 0.185768C_s + 1.753401{C_s}^2 + 0.110735{C_s}^3 + 0.037691{C_s}^4 + 0.0036{C_s}^5 + 0.00219{C_s}^6 + 0.000663{C_s}^7 + 0.000056{C_s}^8$$

- Generalized logistic (GLO),

$$C_k = 4.2 + 2.400505{C_s}^2 + 0.244133{C_s}^4 - 0.00933{C_s}^6 + 0.002322{C_s}^8$$

- Generalized pareto (GPA),

$$C_k = 1.8 + 0.292003C_s + 1.34141{C_s}^2 + 0.090727{C_s}^3 + 0.022421{C_s}^4 + 0.004{C_s}^5 + 0.000681{C_s}^6 + 0.000089{C_s}^7 + 0.000005{C_s}^8$$

- Lognormal (LN),

$$C_s = 3 + 0.025653C_s + 1.720551{C_s}^2 + 0.041755{C_s}^3 + 0.046052{C_s}^4 - 0.00478{C_s}^5 + 0.00196{C_s}^6$$

- Log Pearson type 3 (LP-III)

$$C_k = 3 + 1.5{C_s}^2 \qquad (3.33)$$

3.5.2.2 Using Goodness of Fit

Goodness-of-Fit tests (GOF), as suggested by their very name, can be used to determine whether a certain distribution is fitted properly to the data or not. GOF measure the compatibility of a random sample with a theoretical probability distribution function. In other words, these tests show how well the distribution you selected fits to your data. Calculating statistics of Goodness- of-Fit also helps to rank the fitted distributions according to quality of fit over the raw data. The particular characteristic feature of the EasyFit software is very useful for comparing fitted models to each other.

In this study, to test the statistical hypothesis whether a particular distribution provides an adequate fit to the observed AMF series data the following three most used goodness of fit tests were applied.

- Kolmogorov-Smirnov
- Anderson-Darling
- Chi-Squared

These tests were used at 5% significance level for the selection of the best-fit distribution using EasyFit software. The reason for selecting three different tests is that there is no single test that can give conclusive results and a particular test emphasizes a particular aspect of the goodness- of-fit. The best-fitted distribution is selected based on the minimum error produced, which is evaluated by the following techniques:

When the distributions are fitted, EasyFit will generate a report of goodness-of-fit values in the form interactive tables which includes calculated test statistics and critical values for various significance levels. This value can help to decide which model describes the data in the best way.

i. Kolmogorov-Smirnov Test (KS)

The Kolmogorov-Smirnov (KS) test is a goodness-of-fit statistic that compares an empirical distribution function Fn(x), with a specified distribution function Fo(x) and it measures the distance between the empirical distribution function F.(x) and the specified distribution F(x). the large difference indicates an inconsistency between the observed data and the statistical model (Vikram Kumar et al. 2017). The Kolmogorov-Smirnov statistic (DN) can be computed as:

$$DN = max|F_n(X_i) - F_o(X_i)| \qquad (3.34)$$

1<i<n

The values of FN (x) are predictable as Nj/N, where Nj is the cumulative number of sample events in class i. The value of DN must be less than a tabulated value of DN at the specified confidence level for the distribution to be received (Rao et al.,2000). In this method, the hypotheses take dependability of a specified distributions data of stations.

The hypothesis regarding the distributional form is rejected at the chosen significance level (a) if the test statistic, D, is greater than the critical value obtained from a table. The fixed values of a (0.01, 0.05) are generally used to evaluate the at various significance levels. A value of 0.05 is typically used for most applications.

The hypothesis regarding the distributional form is rejected at the chosen significance level (α) if the test statistic, D, is greater than the critical value obtained from a table. The fixed values of α (0.01, 0.05 etc.) are generally used to evaluate the null hypothesis (Ho) at various significance levels. A value of 0.05 is typically used for most applications, however, in some critical industries, α lower a value may be applied.

The standard tables of critical values used for this test are only valid when testing whether a data set is from a completely specified distribution. If one or more distribution parameters are estimated, the results will be conservative: the actual significance level will be smaller than that given by the standard tables, and the probability that the fit will be rejected in error will be lower.

EasyFit displays the P-values based on the Kolmogorov-Smirnov test statistics (D) calculated for each fitted distribution.

ii. Chi-Squared Test(X^2)

The Chi-Squared goodness of fit test is a non-parametric test used to determine if a sample comes from a population with a specific distribution. It is a technique that checks if a specific distribution of a certain observed event's frequency in a sample is suitable for that sample or not. This test is applied to an arbitrary number of classes (k) to which the observed data are divided, so the value of the test statistic depends on how the data is sorted. (Hossein Mehrannia et al,2016).

In X^2 goodness of fit test, sample data is separated into intervals. Then the numbers of points that drop into the interval are compared, with the predictable numbers of points in every interval. The null hypothesis assumes that there is no notable variation between the observed and the expected value. The degree of freedom depends on the distribution of the data sample (Ghosh et al., 2016).

In this goodness of fit test, the alternative hypothesis assumes that there is an essential variation between the observed and the expected value. Using O, to define the "observed frequency" for class i and "E, to define the "expected frequency" for class i", the Chi-square test statistic (X^2) is calculated by:

$$X^2 = \frac{(O-E)^2}{E} \tag{3.35}$$

Where: X^2= Chi-Square goodness of fit test
O = observed value
E = expected value

If the considered value of X^2 goodness of fit test is less than the table value, will admit the null hypothesis and conclude that there is no important differentiation between the observed and expected value.

The hypothesis regarding the distributional form is rejected at the chosen significance level (α) if the test statistic is greater than the critical value defined as $\mathbf{X^2_{1-\alpha, k-1}}$ meaning the Chi-Squared inverse CDF with k-1 degrees of freedom and a significance level of α. Though the number of degrees of freedom can be calculated as k-c-1 (where c is the number of estimated parameters), EasyFit calculates it as k-1 since this kind of test is least likely to reject the fit in error.

The fixed values of α (0.01, 0.05 etc.) are generally used to evaluate the null hypothesis (Ho) at various significance levels. A value of 0.05 is typically used for most applications, however, in some critical industries; a lower α value may be applied.

Easy Fit displays the P-values based on the Chi-Squared test statistics *(i:)* calculated for each fitted distribution.

iii. Anderson-Darling test (AD)

The AD test is used to test if a sample of data came from a population with a definite distribution. It is a revision of the KS test and gives further influence to the tails than does the KS test. The KS test is distribution free in the logic that the critical values do not depend on the definite distribution being tested. This test makes utilize of the definite distribution in manipulative critical values. This has the benefit of allowing an additional perceptive test and the drawback that critical values should be intended for each distribution. The critical values for the AD test are dependent on the specific distribution that is being tested (Ghosh et al., 2016). The Anderson-Darling statistic (A^2) is defined as

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln\ F(X_i) + \ln\ (1 - F(X_{n-i+1}))\right] \tag{3.36}$$

Where $\{x_{(1)} < \cdots < x_{(n)}\}$ is the ordered (from smallest to the largest element) sample of size n, and F(x) is the underlying theoretical cumulative distribution to which the sample is compared. The null hypothesis that $\{x_{(1)} < \cdots < x_{(n)}\}$ comes from the underlying distribution F(x) is rejected if AD is larger than the critical value AD_{α} at a given significance level (α).

The hypothesis regarding the distributional form is rejected at the chosen significance level (a) if the test statistic, A^2, is greater than the critical value obtained from a table. The fixed values of a (0.01, 0.05 etc.) are generally used to evaluate the null hypothesis (H_O) at various significance levels. A value of 0.05 is typically used for most applications, however, in some critical industries, a lower a value may be applied. In general, critical values of the Anderson-Darling test statistic depend on the specific distribution being tested. However, tables of critical values for many distributions (except several the most widely used ones) are not easy to find. The Anderson-Darling test implemented in EasyFit uses the same critical values for all distributions. These values are calculated using the approximation formula, and depend on the sample size only. This kind of test (compared to the "original" A-D test) is less likely to reject the good fit, and can be successfully used to compare the goodness of fit of several fitted distributions.

Hypothesis Testing

The null and the alternative hypotheses are:

- H**o**: the data follow the specified distribution;
- H_A: the data do not follow the specified distribution.

P-Value

The P-value, in contrast to fixed α value, is calculated based on the test statistic, and denotes the threshold value of the significance level in the sense that the null hypothesis (H_O) will be accepted for all values of α less than the P-value. For example, if P=0.025, the null hypothesis will be accepted at all significance levels less than P (i.e. 0.01 and 0.02), and rejected at higher levels, including 0.05 and 0.1.

The P-value can be useful, in particular, when the null hypothesis is rejected at all predefined significance levels, and you need to know at which level it could be accepted.

3.5.2.3 Using Probability Distributions Performance Evaluations

Although the estimates obtained by L-moment ratio diagram and goodness-of-fit are reliable, still we need certain measures for the assessment of the accuracy of these estimates. The probability-probability (P-P) and quantile-quantile (Q-Q) plots are powerful visualization methods for evaluating the fit of the selected probability distributions through the visual assessment of the linearity of the pattern of points tend to lie reasonably along and close to a straight line and this provide a validation of regional flood frequency curves to model the observed annual maximum flood data. The performance of the best distribution model identified for the respective regions was evaluated by comparing observed with simulated values by employing the P-P and Q-Q plot techniques with Easy Fit Software.

Probability-probability (P-P) plots

The probability-probability (P-P) plot is a graph of the empirical Cumulative Distribution Function (CDF) values plotted against the theoretical CDF values. It is used to determine how well a specific distribution fits to the observed data. This plot will be approximately linear if the specified theoretical distribution is the correct model. EasyFit displays the reference diagonal line along which the graph points should fall. This graph can also be used to determine where the data do and don't follow the theoretical distribution.

The probability-probability plot of analysis and observed annual maximum (AM) discharge with the probability distribution were developed as the following steps:

Step 1: Sort the n values from the smallest to the largest

Step 2: For each of the N data points, compute an empirical (observed) cumulative probability as: $F_n(x_i) = i/n$, for i =1, 2, …, n. This simple formula works well in Easy-Fit software cases, for computing the empirical cumulative probability for use in probability plots.

Step 3: Then compute the theoretical (fitted) cumulative distribution function

Step 4: Finally plot the graph of empirical cumulative probability on abscissa and cumulative distribution function on ordinates.

Quantile-quantile (Q-Q) plots

The quantile-quantile (Q-Q) plot is a graph of the input (observed) data values plotted against the theoretical (fitted) distribution quantiles. Both axes of this graph are in units of the input data set. The quantile-quantile graphs are produced by plotting the observed data values x_i (i = 1, ..., n) against the X-axis, and the following values against the Y-axis:

$$F^{-1}(F_n(x_i) - \frac{0.5}{n}) \quad (3.37)$$

Where; $F^{-1}(x)$ is inverse cumulative distribution function (ICDF), $F_n(x)$ is empirical CDF and n is sample size. The Q-Q plot will be approximately linear if the specified theoretical distribution is the correct model. EasyFit displays the reference diagonal line along which the graph points should fall.

The type of plot utilized for distribution fit and assessment is the quantile–quantile plot, where each ordered data point x(i) is plotted against the inverse of the chosen probability distribution for the empirical plotting position corresponding to that data point. On the quantile-quantile plot, perfect matching of the data along a 45° line on this type plot would suggest a "perfect" fit. The quantile-quantile plot of analysis and observed annual maximum (AM) discharge with the distribution were developed as the following steps:

Step 1: Sort the n values from the smallest to the largest, so that $x_1 \leq x_2 \leq .... < x_n$.
Step 2: Then compute the inverse cumulative distribution function using $F^{-1} (F_n(x_i) – 0.5/n)$
Step 3: Finally plot the graph of the input (observed) data values plotted against the theoretical (fitted) distribution quantiles.

3.5.3 Identification of Best-Fit Probability Distribution

To select the best-fit distribution to the annual maximum flow series, L-moment ratio diagram, goodness-of-fit tests such as Kolmogorov-Smirnov, Anderson-Darling and Chi-squared and probability distribution performance evaluations: P-P and Q-Q plots were applied.

The goodness-of-fit test was fitted to the annual maximum flow data of the study area. The test statistic was computed and tested at (**a** = 0.05) level of significance. Accordingly, the ranking of different probability distributions was marked based on minimum test statistic value.

Using L-moment ratio diagram, P-P and Q-Q plots, the assessment of the best fit probability distribution was based on a visual indication of which distribution may be expected to give a good fit to a data points by plotting their respective graphs.

3.5.4 Parameter Estimation

After the best fit regional probability distribution was selected, the distribution parameters for the homogeneous region were determined by using the appropriate method of parameter estimation. There are a number of well-known methods which can be used to estimate distribution parameters based on available sample data. These include the method of moments (MOM), method of L-moments, the maximum likelihood method (MLM), the probability weighted moments method (PWM), the least squares method (LS), maximum entropy (ENT), mixed moments (MIX), the generalized method of moments (GMM), and incomplete means method (ICM) (Rao and Hamed, 2000).

L-moments are a recent development within statistics (Kumar R. and Chatterjee C., 2011). In a wide range of hydrologic applications, L-moments provide simple and reasonably efficient estimators of characteristics of hydrologic data and of a distribution's parameters (Kumar R. and Chatterjee C., 2011). Like the ordinary product moments, L-moments summarize the characteristics or shapes of theoretical probability distributions and observed samples. Both moment types offer measures of distributional location (mean), scale (variance), skewness (shape), and kurtosis (peakedness).

For this study, EasyFit software was used to estimate parameters of the selected distribution. For every supported distribution, EasyFit implements one of the following parameter estimation methods (Hossein Mehrannia *et al.*, 2014):

- Method of moments (MOM);
- Maximum likelihood method;
- Least squares estimates (LSE);
- Method of L-moments (MOLM)

Table 3.4 below shows the list of supported distributions and methods used.

Table 3. 4 List of distribution and appropriate method of parameter estimation (source: EasyFit software manual)

Distributions	Parameter Estimation Method
Generalized extreme Value (GEV)	Method of L-moments
Generalized Logistic (GLO)	Method of L-moments
Generalized Pareto (GPA)	Method of L-moments
Lognormal (3P) (LN)	Maximum likelihood method
Log Pearson Type 3 (LP-III)	Method of moments

From table above, it was found that method of L-moment was selected for estimation of parameter for the selected best fit distributions: Generalized extreme value (GEV) and Generalized pareto (GPA) distributions.

3.5.5 Estimating Extreme Flow Quantiles and Developing Regional Flood Frequency Curves

3.5.5.1 Estimating Extreme Flow Quantiles

After the parameters of a distribution was estimated, flood quantile estimates (Q_T) which correspond to different return periods were computed. Based on the selected distributions for each station, the quantile can be calculated according to the formula of the selected distributions. For stations with a computed value of scale, location and shape parameter, then it is possible to determine the quantile with different return periods using different equations for different distributions.

For Generalized extreme value (GEV) distribution the flow quantile can be estimated as;

$$Q_T = \mu + \frac{\sigma}{\kappa}(1 - \left(-\ln\left(1 - \frac{1}{T}\right)^{\kappa}\right), for\ \kappa \neq 0 \quad (3.38)$$

$$Q_T = \mu + \sigma(\ln\left(-\ln\left(1 - \frac{1}{T}\right)\right), for\ \kappa = 0 \quad (3.39)$$

For Generalized pareto (GPA) distribution the flow quantile can be estimated as;

$$Q_T = \mu + \sigma\left(\ln\left(\frac{1}{T}\right)\right), for\ k = 0 \quad (3.40)$$

$$Q_T = \mu + \frac{\sigma}{\kappa}\left(1 - \left(\frac{1}{T}\right)^{\kappa}\right), for\ \kappa \neq 0 \tag{3.41}$$

The return period is related to the probability of non-exceedance (F) by the relation,

$$F = 1 - \frac{1}{T} \tag{3.42}$$

Where; σ is scale parameter, μ is location parameter, k is parameter and T is return period

3.5.5.1 Developing Regional Flood Frequency Curves

After estimating the flood discharge for different return period, regional flood frequency curves were developed by drawing return period T versus the estimated flood quantile QT

3.5.5.2 Developing of Regional Growth Curves

The average of the regional growth curves was determined to represent the frequency curves of regions. Index flood method employs data of the gauged catchments to evaluate a regional from the flood magnitudes of various return periods for ungauged catchments to be evaluated (Modi and Mitra, 2017). In this study, the index flood method was used to determine the magnitude and frequency of flood quantiles for sites located within a homogeneous region.

Estimation of index-flood

Derivation of the mean annual flood ($\bar{Q}$) for each station was obtained by relating the annual flood data from each station (Qi) and dividing it by the number of record years. The main assumption of this is that data at different sites in a region follow the same distribution consisting of identified homogeneous regions, determination of best-fit distribution and derivation of the regional flood frequency curves. In this study, the index flood L-moment approach of regionalization was applied depending on the homogeneity of the stations.

Flood quantiles estimation in flood frequency analysis were corresponding to the required return periods. The model parameters for the distributions estimated for each station were used to compute standardized flow estimates conforming to the return periods 2, 5, 10, 20, 25, 50, 100, 200, 500, 1000, 2000, 5000 and 10000 years. Plots of Q/Qave against the Gumble reduced variate (-ln(-ln(1-1/T)) known as growth curves, were generated for each station and used in the derivation of the regional growth curves.

After the regional frequency distribution is determined, the flood quantiles having a return period of T year within a homogeneous region can be estimated based on the equation (3.42) proposed by Hosking and Wallis (1997). The common practice is to get the dimensionless data by dividing the values by an estimate of the at-site mean.

$$X_T = \frac{Q_T}{Q_{ave}} \tag{3.43}$$

Where; Q_{ave} - is the mean annual flood(m^3/s) is the index flood

Q_T - is the quantile (m^3/s) function of fitted distribution at site i

X_T - regional growth curve; this defines the frequency distribution common to all the sites in a homogenous region.

3.6 Mapping Flood Risk Area for Selected Return Period

Flood hazard maps, in particular, can be defined as maps showing inundated area of different parameters such as flood depth and flood velocity and water level. Flood inundation mapping is an important tool for engineers, planners and government agencies used for understanding the extent of flooding and floodwater inundation, to best allocate resources to prepare for emergencies and to generally improve the quality of life (Ali El-Naqa and Mohammed Jaber, 2018).

In this study after computing extreme discharge for return period of interest, flood hazard mapping was computed for the selected study area using HEC-Geo RAS and HEC-RAS models. Besides floodplain mapping, the results of a Geo RAS analysis can be used for flood damage computations, ecosystem restoration, and flood warning response and preparedness (USACE HEC-RAS, 2016). Several important parameters were required for performing hydraulic flood modeling such as topographic data, discharge data to provide model inflow and outflow as boundary conditions, estimation of the manning's roughness coefficient and validation data (Sitotaw Haile et-al., 2018). Besides floodplain mapping, the results of a Geo RAS analysis can be used for flood damage computations, ecosystem restoration, and flood warning response and preparedness (USACE HEC-RAS, 2016).

3.6.1 HEC-Geo RAS

HEC-Geo RAS is a geographic river analysis system developed using ArcGIS Desktop and ArcGIS Spatial Analyst and 3D Analyst extensions. It is used to preparing a geometric data for importing data in HEC-RAS and processing the simulation results exported from HEC-RAS. For the preparation of geometric data, the DEM at 30m resolution of the study area, Awash river below Koka dam upto Awash Melka Werer station, was downloaded from STRM web site (https://earthexplorer.usgs.gov/). This DEM was clipped in such a way that it covers the study area. Then a Triangular irregular network (TIN) was generated by ArcGIS using Raster to TIN tool. This output TIN, as shown in figure 3.5), was used as an input for HC-GeoRAS environment.

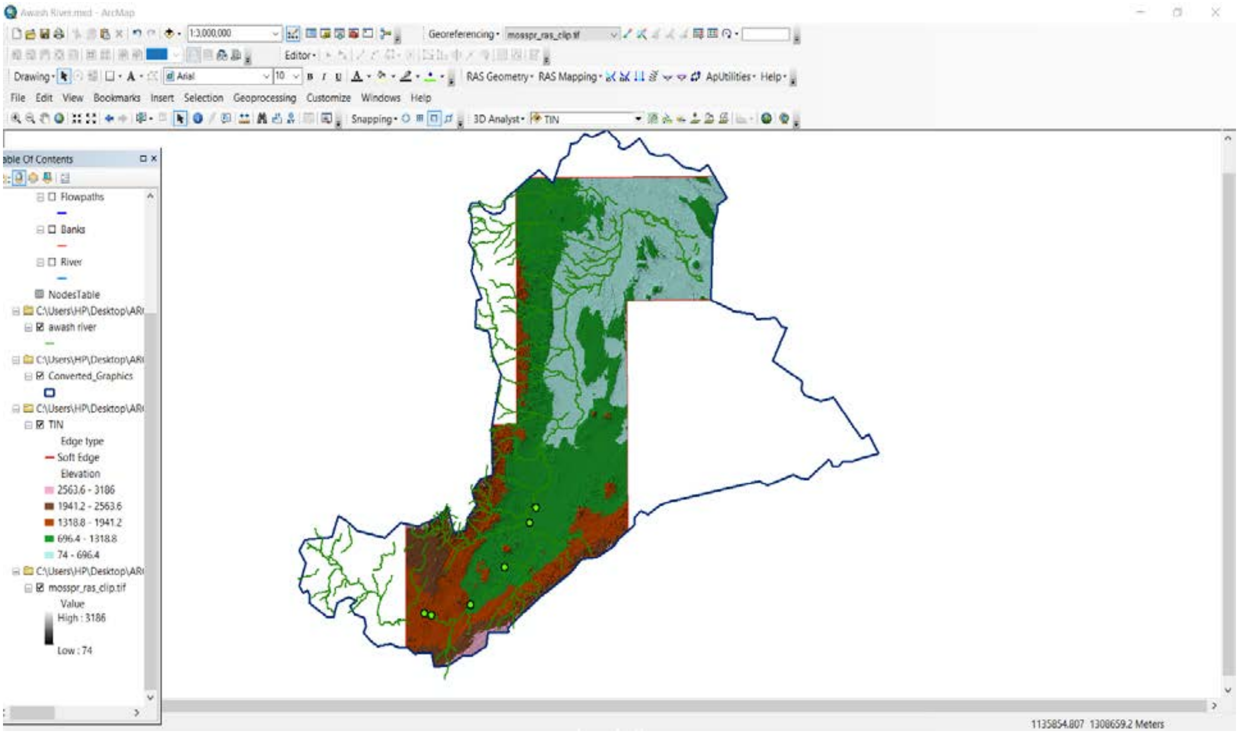


Figure 3. 5 Generated TIN of the study area

The HEC-GeoRAS extension was used to set up the necessary features that would be needed for the HEC-RAS model (i.e., stream centerline, bank lines, cross sections, flow paths, etc.). HECRAS uses these features to obtain an accurate layout of the river, and to establish the river reach network and the cross-sectional elevations of the potential floodplains. The cross sections must extend far enough to ensure that all water from the flood is contained within the cross-sectional area. Methods for setting up the model using HEC-GeoRAS were taken from the HEC-GeoRAS user manual (USACE, 201615). The river geometry was digitized using the ArcGIS editing features. Figure 3.6 and 3.7 respectively show the created RAS geometry and their digitization on top of the TIN. HEC-GeoRAS uses the line features in conjunction with the TIN to extract elevations for the cross sections and flow profile.

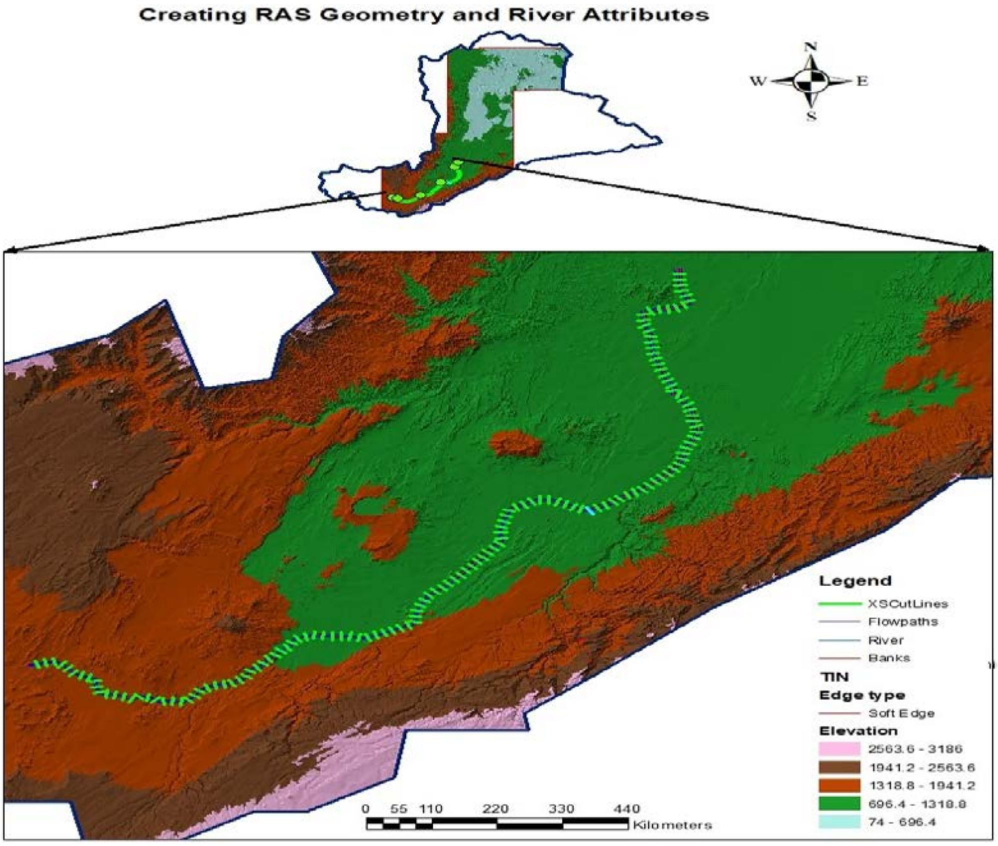


Figure 3. 6 Creating RAS Geometry

As shown in Figure 3.7 below, the length of the rive reach, from Awash below Koka dam up to Awash Melka Werer, is about 281.2Km along the center of its course

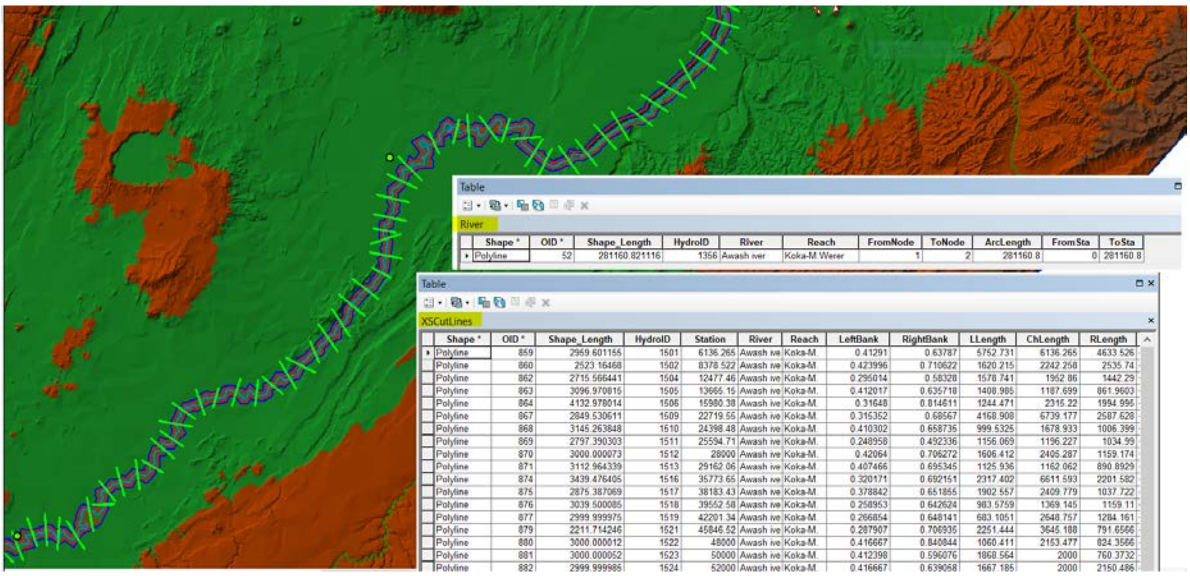


Figure 3. 7 Digitizing RAS Geometry

3.6.2 HEC-RAS

HEC-RAS is one of the most widely used models used by hydraulic engineers in the channel flow analysis and floodplain delineation (USACE HEC-RAS,2016). It is a one-dimensional steady flow hydraulic model developed by the US Army Corps of Engineers. It comprises a set of tools, utilities, and procedures for working with geospatial data. The basic inputs to HEC-RAS include the river discharge, channel, floodplain geometry, and channel resistance.

With HEC-RAS, it is possible to simulate water surface profiles for gradually varied steady flow and the effects of various obstructions such as culverts, bridges, structures in the overbank region and weirs. Several applications of HEC-RAS model to determine the extent of inundation resulting from a given flood have been reported: Al-Qudah K.A (2010), Manandhar B. (2010) and Solaimani K. (2011). Abdelbasset (2015) applied HEC-RAS model to calculate water surface profiles corresponding to selected flood events downstream of Al Wahda Dam in Sebou basin in Northern Morocco.

In HEC-RAS model the following activities were employed to develop flood inundation map of the study area for the selected study area. After Setup for HEC-RAS preprocessing was completed by HEC-GeoRAS extension, the geometry data created in Hec-GeoRAS and ArcGIS were exported into HEC-RAS. This procedure allows the geometric data transfer from ArcGIS into HEC-RAS. Figure 3.8 shows the imported geometric data in HEC-RAS

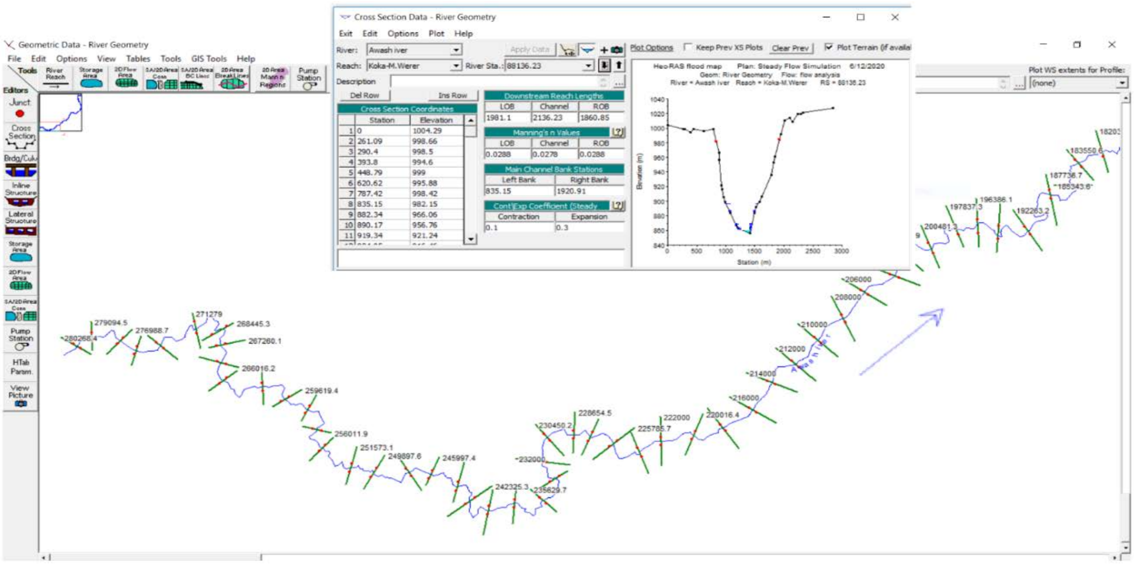


Figure 3. 8 Imported geometric data in HEC-RAS

Manning's roughness coefficient values were applied to the cross sections using land cover data obtained from figure 3.2 and Ethiopian Awash basin Authority. The Manning's n values based on land use landcover according to Chow was shown in table 3.5. This Manning's roughness values represent the roughness of the channel surface, which can influence the overall flow rates and velocities in the channel.

Table 3. 5 Manning's n values based on land use landcover according to Chow

S/No	Reach Name	Land use landcover	Manning's n value	Average Manning's n value
1	Awash below Koka dam station–Awash @Wanji station	- Agricultural land on level or slightly undulating ground - Irrigated agricultural land - Woodland and savanna	- 0.045 - 0.034 - 0.04	0.0595
2	Awash @Wanji station–Awsh @Nura Hera Station	- Agricultural land on sloping ground - Open woodland and bush - Woodland and savanna	- 0.03 - 0.031 - 0.04	0.0337
3	Awash @Nura Hera station–Methara station	- Shrub savanna - short grass savanna - Woodland and savanna	- 0.05 - 0.035 - 0.035	0.425
4	Metehara station – Awash Melka Sadi	- Woodland and savanna - short grass savanna - Badlands	- 0.04 - 0.035 - 0.0113	0.0288
5	Awash Melka Sadi Station – Awash Melka Werer station	- Dense forest (montane and riparian) - Shrub savanna - short grass savanna/shrubland	- 0.25 - 0.035 - 0.05	0.1117

Source: Haile Arefayne Shishaye and Asmlash Tilahun, 2020 and Ethiopia, Awash basin Authority

Then the steady flow data along with the boundary conditions was supplied to the stead flow editor and flow inputs were conducted to all reaches of the selected study area. The estimated flood peaks for different return period (obtained by the best fit probability distribution and L-moment ratio diagram from flood frequency analysis) were used as an input to HEC-RAS to model steady flow analysis. This is because to estimate water surface profiles and extent of inundation under different flood intensities, peak discharges for different return periods are required. The peak flows are input into HEC-RAS to find the expected flood levels along river reaches extending through populated areas of the basin. Flood hazard mapping was done for the six model scenarios of: 25, 50, 100, 200, 500, and 1000-years of flood and upstream and downstream reach boundary conditions as critical depth of low.

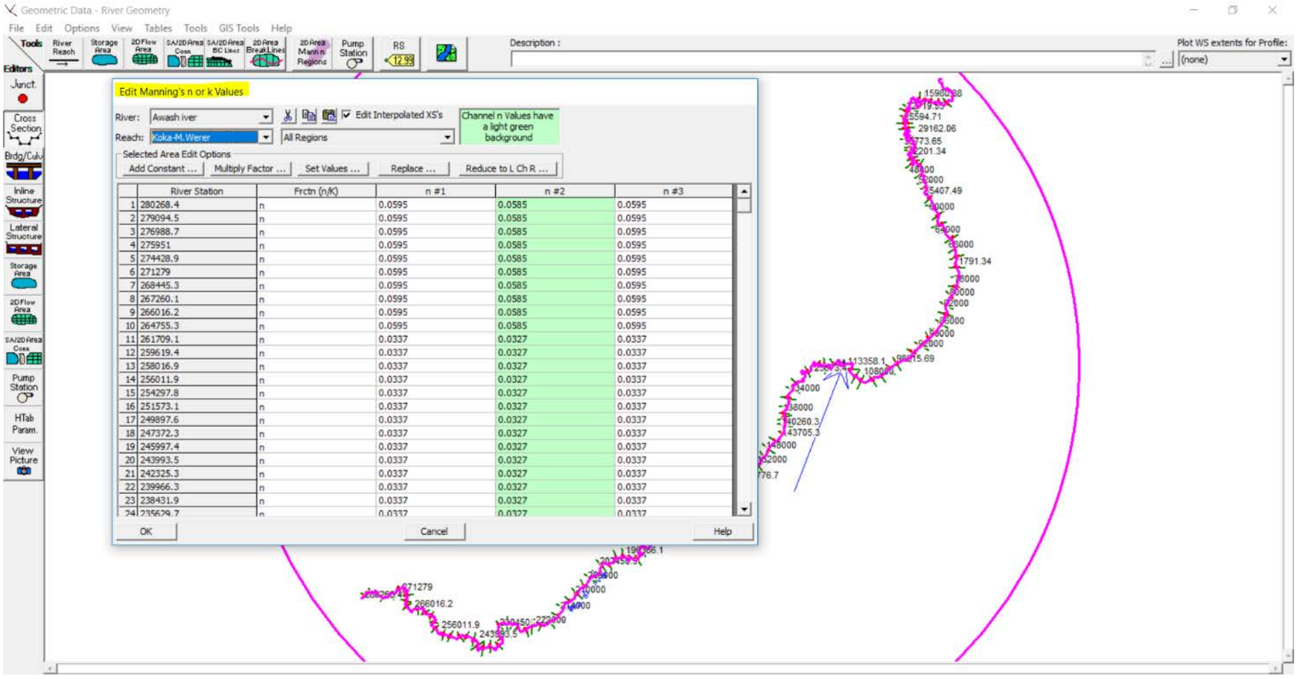
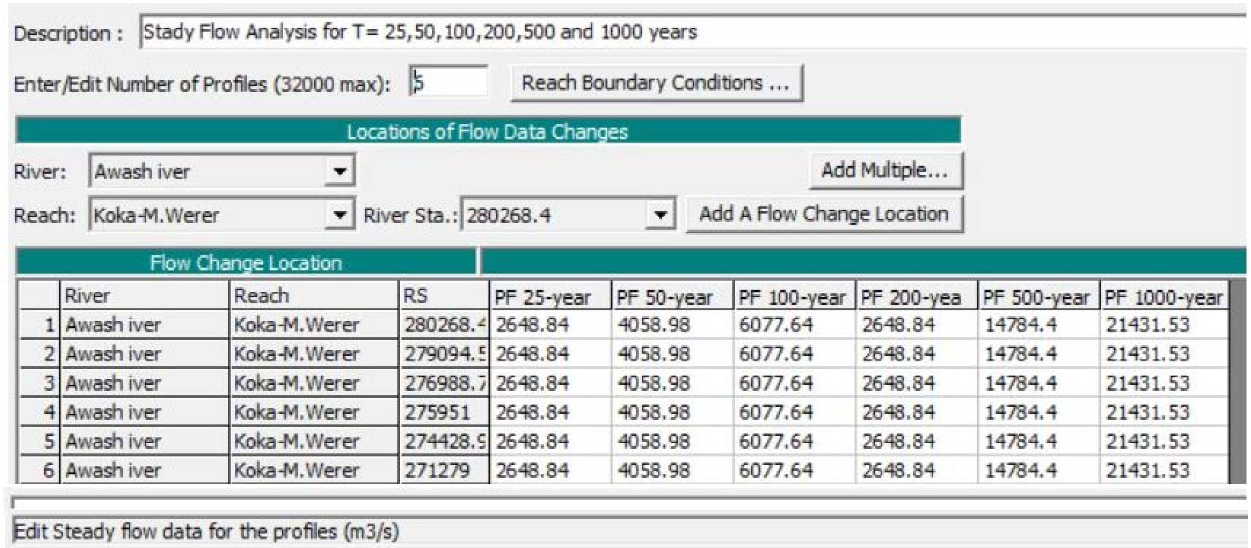


Figure 3. 9 Applied Manning's roughness coefficient



	River	Reach	RS	PF 25-year	PF 50-year	PF 100-year	PF 200-yea	PF 500-year	PF 1000-year
1	Awash iver	Koka-M.Werer	280268.4	2648.84	4058.98	6077.64	2648.84	14784.4	21431.53
2	Awash iver	Koka-M.Werer	279094.5	2648.84	4058.98	6077.64	2648.84	14784.4	21431.53
3	Awash iver	Koka-M.Werer	276988.7	2648.84	4058.98	6077.64	2648.84	14784.4	21431.53
4	Awash iver	Koka-M.Werer	275951	2648.84	4058.98	6077.64	2648.84	14784.4	21431.53
5	Awash iver	Koka-M.Werer	274428.9	2648.84	4058.98	6077.64	2648.84	14784.4	21431.53
6	Awash iver	Koka-M.Werer	271279	2648.84	4058.98	6077.64	2648.84	14784.4	21431.53

Figure 3. 10 Applying estimated flood peak of selected return period for steady flow analysis

Upon the completion of data inputs, the HEC-RAS model was executed to perform a detailed analysis of the flow. A detailed report of the analysis showing the depth of flow, discharge at each cross section, water surface profile, rating curves and other details was generated by the model. Finally, flood inundation mapping of the study area was developed. This can be done using the following techniques.

i) The final results of HEC-RAS model are exported to ArcGIS in the form of RASGIS Export File. The RAS GIS Export File is imported into ArcGIS after generating water surface and floodplain delineation, and a raw floodplain map will be obtained.

ii) Using RAS Mapper of HEC-RAS model, TIN or DEM of the study area is imported to create the terrain of study region on which flood inundation map is to be developed.

In this study the second technique was used to compute flood inundation mapping of the study area. After mapping flood, the inundation area was computed using Hec-RAS Mapper. A new geometry of flood inundation area was created to edit the geometry of the storage area by delineating the floodplain already mapped.

3.7 General Conceptual Framework of the Methodology

In general, the conceptual work flow to be undertaken throughout the study was given as in figure below.

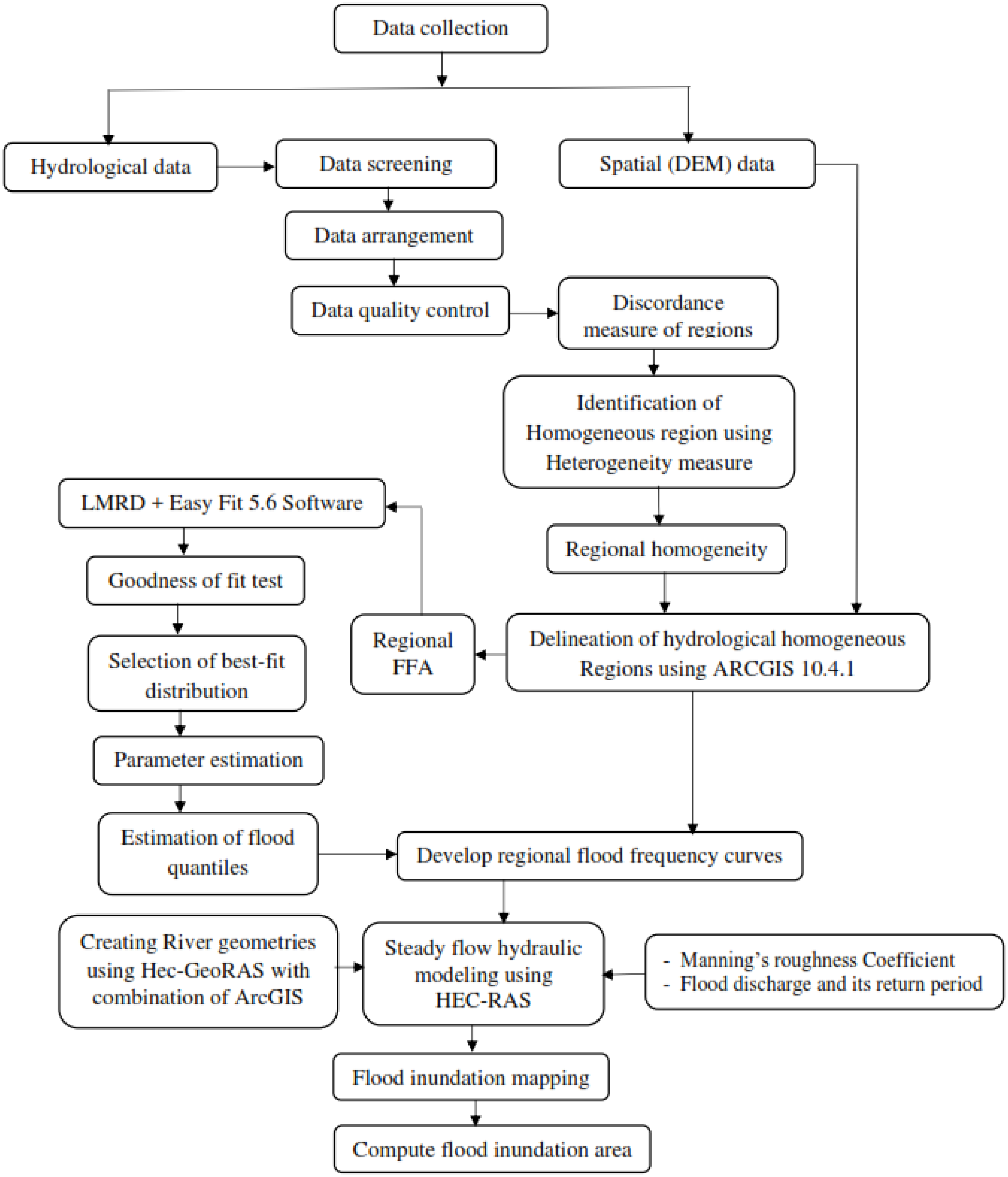


Figure 3. 11 Flowchart of the methodology

4. RESULT AND DISCUSSIONS

Annual maximum peak flow data of six gauging stations with 16 years record length of the study area have been provided for the analysis of regional flood frequency analysis and its mapping. This chapter involves the following stages: checking quality of data, identification of homogeneous regions, choice of an appropriate probability distribution, parameter and quantile estimation and finally flood mapping (Hosking and Wallis 2005).

4.1 Checking Data Quality

The annual maximum flow series (AMF) for all selected stations of study area were checked for their independency and stationarity, homogeneity and stationarity, outliers and continuity with previous batch of data check the quality of data used. As shown in tables 4.1, 4.2 and 4.3 respectively, all tests showed that the AMF for all stations of the study area were independent and stationary, homogeneous and stationary and they were continuous and had no outliers.

Table 4. 1 Results of test for independence and stationarity

Station Name	R	$\bar{R}$	Var(R)	St. Dev	S_1	S_2	S_3	S_4	U
Awash below Koka	719131.1	680813.5	5735415347	75732.5	3350.5	1013714.7	358788970.9	136472671801	0.51
Awash@ Wanji	310856.7	296882.2	425925229.9	20637.9	2200.3	388029.4	80600248.4	18302943310	0.68
Awash @Nura Hera	737888.7	784273.3	646409560.3	25425.6	3558.9	901313.4	254708347.1	78656707654	-1.8
Metehara	490815.8	454507.6	216274165.9	14706.3	2708.5	518553.9	108359782.3	23962657126	2.5
Awash Melka Sadi	734557.5	697609.7	255085667.1	15971.4	3352.9	778053.7	201182359.7	57850402278	2.3
Awash Melka Werere	1514546.l	1520430.8	1374964614	37080.5	4949.6	1691919.5	632401868	255319554425	-0.2

The summarized result of the test for all stations show that the absolute vale of U is less than the critical value at 5% significance level of $U_{0.25}$ = 1.96. Hence the annual maximum flow series for all stations of the study area were independent and stationary.

Table 4. 2 Results of test for homogeneity and stationarity

Station Name	N	N-1	p	q	R	V	W	U	U_{mean}	$\sum T$	Var(U)	U
Awash below Koka	16	15	7	9	52	24	39	24	31.5	0	89.25	-0.79
Awash@ Wanji	16	15	7	9	55	27	36	27	31.5	0	89.25	-0.48
Awash @Nura Hera	16	15	7	9	61	33	30	30	31.5	0	89.25	-0.16
Metehara	16	15	7	9	38	10	53	10	31.5	0	89.25	-2.3
Awash Melka Sadi	16	15	7	9	46	18	45	18	31.5	0	89.25	-1.43
Awash Melka Werer	16	15	7	9	45	17	46	17	31.5	0	89.25	- 1.53

From table 3.4 above, the summarized result of the test for all stations show that the absolute vale of U is less than the critical value at 5% significance level of $U_{0.25}$ = 1.96. Hence the annual maximum flow series for all stations of the study area were homogeneous and stationary.

Table 4. 3 Results of test for high and low outliers

Station Name	N	$\bar{x}$	s	kN	QH	QL	Q_{max}	Q_{min}	Analysis result	
									High outlier	Low outlier
Awash below Koka	16	2.19	0.37	2.279	1106.4	22.2	475.24	24.5	No high outlier	No low outlier
Awash @Wanii	16	2.08	0.24	2.279	414.19	34.5	268.66	59.35	No high outlier	No low outlier
Awash @Nura Hera	16	2.32	0.17	2.279	508.62	84.46	407.7	106.2	No high outlier	No low outlier
Metehara	16	2.19	0.17	2.279	391.2	63.5	254.16	86.91	No high outlier	No low outlier
Awash Melka Sadi	16	2.3	0.14	2.279	411.4	96.9	398.56	104.59	No high outlier	No low outlier
Awash Melka Werer	16	2.47	0.15	2.279	633.2	136.06	547.76	164.74	No high outlier	No low outlier

4.2 Identifying and Delineating Homogeneous Region

4.2.1 Discordancy Measure (Di) of Regions

In the processes of screening of the data, it is standard practice to compute a discordancy measure (Di) which identifies unusual sites that are grossly discordant with the group as a whole. A large value of Di indicates the discordancy of site i with other sites and the definition of large depends on the number of sites in the study region. Thus, a site is regarded as discordant if Di exceeds the critical value given in tabular form (see Table 3.7) (Hosking and Wallis 1997).

The discordancy measures together with the sample L-moment ratios for the 6 sites of the study area are computed and given in Table 4.1 along with the catchment areas and statistical characteristics. It is observed that Di values of the 6 gauging sites are ranges from 0.07 to 0.66 which are less than the critical value for the discordancy statistic Di for the 6 sites, i.e., 1.648 shown in table 3.4. Thus, data from all the 6 river flow gauging sites are suitable for flood frequency analysis.

Table 4. 4 L-Moment Ratios and Discordancy Measure for selected stations

Station	Catchment Area (km2)	Sample size (years)	L-CV (t)	L Skewness (t3)	L Kurtosis (t4)	Discordancy measure (Di)
Awash below Koka	11219	16	0.3995	0.1371	-0.0610	0.28
Awash @Wanji	11690	16	0.3124	0.2399	-0.0517	0.19
Awash @Nura Hera	14173	16	0.2199	0.1385	0.1792	0.08
Metehara	16416.8	16	0.2198	0.0251	-0.1240	0.39
Awash Melka Sadi	21520	16	0.1775	0.2247	0.3703	0.66
Awash Melka Werer	31183	16	0.1895	0.1470	0.1553	0.07

4.2.2 Identifying Homogeneous Regions

To determine whether the region is homogeneous or not, the samples L-moment ratio of data were calculated. According to Hosking and Wallis (1997), if the heterogeneity measure $H < 1$ the region is acceptably homogeneous. If $1 \leq H < 2$, the region is possibly heterogeneous and if H ≥ 2 the region is definitely heterogeneous. The values of heterogeneity measures computed by carrying out of Monte Carlo simulations after estimating parameters of kappa distribution are H = 0.85. The H-test statistics indicates that the region is possibly acceptably homogeneous since H < 1, where thus used for the development of regional flood frequency curve.

Table 4. 5 Results of H-statistics heterogeneity measure

$\sum_{i=1}^{N} n_i \left(t^{(i)} - \bar{t}\right)^2$	$\sum_{i=1}^{N} n_i$	σ_V	μ_V	V_1	H
0.59064	96	0.4457	0.4585	0.0784	-0.8529

4.2.3 Delineation of Homogeneous Regions

The delineation homogeneous region of the study area was done using ArcGIS and shown as in figure 4.1

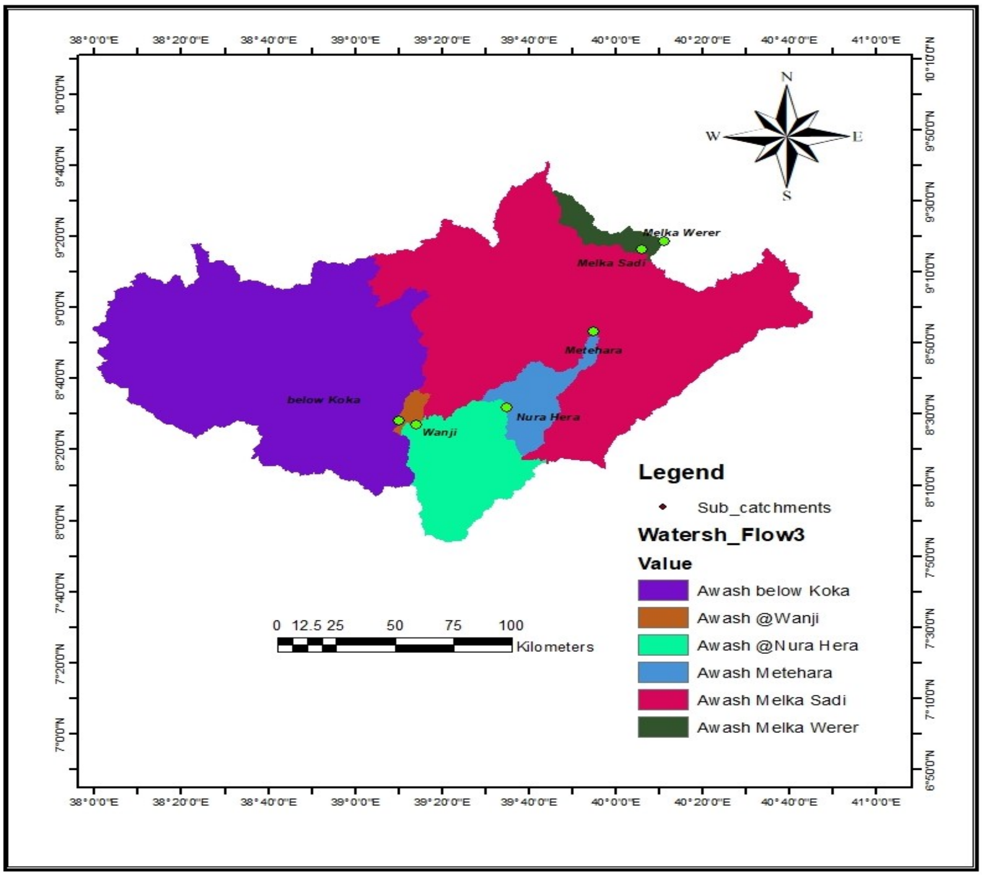


Figure 4. 1 Spatial distribution of delineated homogeneous regions

4.3 Selection of Best Fit Regional Probability Distribution

To obtain the best-fit distribution to the annual maximum flow series, L-moment ratio diagram and goodness-of-fit tests such as Kolmogorov-Smirnov, Anderson-Darling and Chi-squared and P-P and Q-Q plots were applied.

4.3.1 Using L-Moment Ratio Diagram

Using L-moment ratio diagram, the assessment of the best probability distribution was based on a visual indication of which distribution may be expected to give a good fit to a data points by plotting them on a graph of L-skewness versus L-kurtosis as shown in figure 4.2 below. The graph was generated by using data computed in tables 4.6 and 4.7

Table 4. 6 L-moment ratios for candidate distributions using equation (3.32)

L-C_s (Coefficient Skewness, t_3)	L-C_k (Kurtosis Coefficient, t_4				
	Generalized extreme value (GEV)	Generalized logistic (GLO)	Generalized Pareto (GPA)	Lognormal (LN)	Log Pearson type 3 (LP-III)
-0.1	0.104	0.175	-0.010	0.131	0.126
-0.05	0.104	0.169	-0.008	0.125	0.123
0	0.107	0.167	0.000	0.123	0.122
0.1	0.127	0.175	0.029	0.131	0.126
0.15	0.143	0.185	0.051	0.140	0.130
0.2	0.163	0.200	0.077	0.154	0.136
0.25	0.187	0.219	0.107	0.172	0.145
0.3	0.215	0.242	0.142	0.193	0.157
0.35	0.247	0.269	0.180	0.219	0.173
0.4	0.283	0.300	0.222	0.250	0.193
0.45	0.322	0.335	0.268	0.284	0.218
0.5	0.366	0.375	0.318	0.323	0.249
0.55	0.413	0.419	0.372	0.366	0.287
0.6	0.464	0.467	0.429	0.413	0.331
0.65	0.518	0.519	0.489	0.466	0.383
0.7	0.576	0.575	0.553	0.523	0.444
0.75	0.638	0.635	0.620	0.585	0.512
0.8	0.703	0.700	0.690	0.653	0.589

Table 4. 7 L-moment ratios for data samples of stations

Gauging Stations	L-C_v (t)	L-C_s (t3)	L-C_k (t4)
Awash below Koka	0.399533	0.13715	-0.061
Awash @Wanji	0.312354	0.23993	-0.0517
Awash @ Nura Hera	0.219945	0.13846	0.17918
Metehara	0.219769	0.02514	-0.124
Awash Melka Sadi	0.177518	0.22474	0.37026
Awash Melka Werer	0.189549	0.14704	0.15532

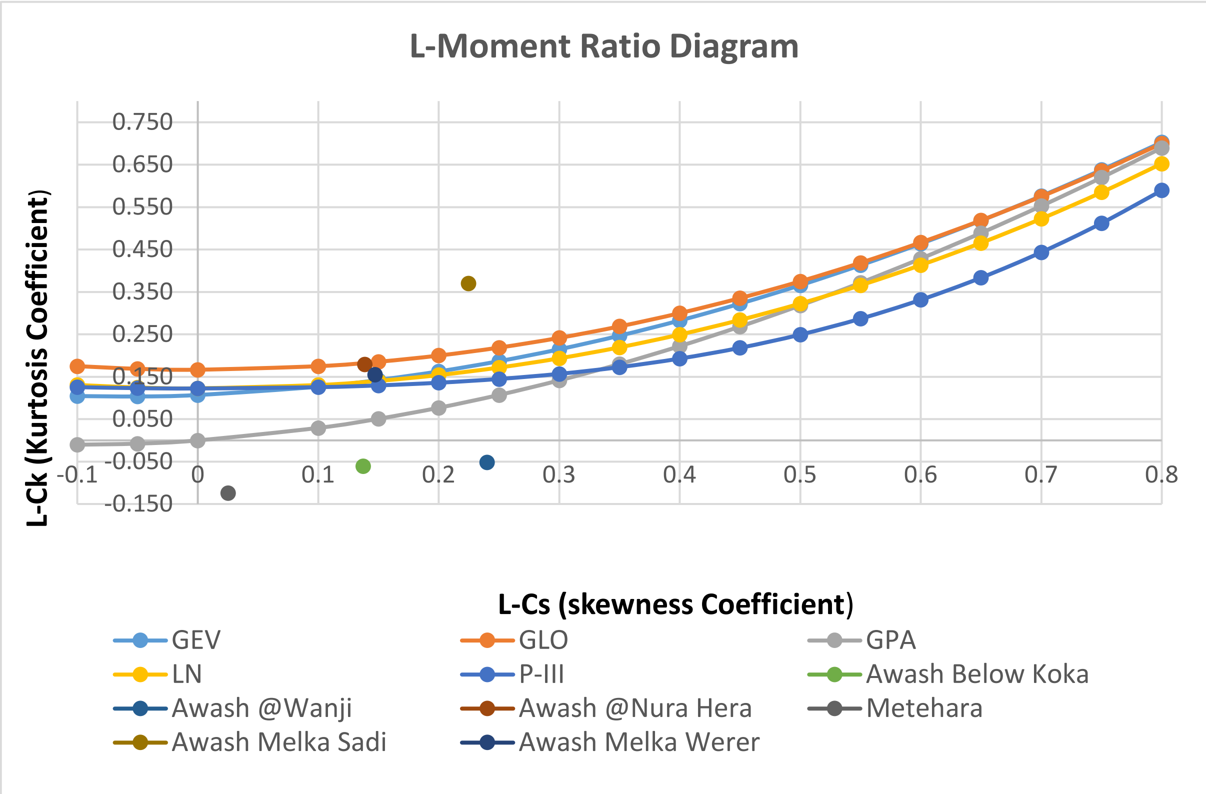


Figure 4. 2 L-moment ratio diagram of the selected basin for various distributions

4.3.2 Using Goodness-of-Fit Test

Using goodness-of-fit tests, the assessment of the best probability distribution was based on the total rank obtained from all the tests and visual indication of which distribution may be expected to give a good fit to P-P and Q-Q plots. Ranks ranging from one to six (1–6), as shown in table 4.8, are given to each distribution model based on the criteria that the distribution(s) with the highest total score is or are chosen as the best distribution model(s) for the data of a particular site.

Table 4. 8 Goodness of fit-test summery from EasyFit

Goodness of Fit - Summary for Awash below Koka Station

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.15565	2	0.51432	5	2.9751	6
2	Gen. Pareto	0.12738	1	0.34841	1	0.01149	2
3	Log-Pearson 3	0.15773	3	0.39706	2	0.19894	3
4	Logistic	0.20894	6	0.85827	6	2.4642	5
5	Lognormal	0.17245	4	0.47373	3	9.1305E-4	1
6	Lognormal (3P)	0.17609	5	0.4888	4	0.20208	4

Goodness of Fit - Summary for Awash @Wanji Station

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.14734	4	0.55147	4	1.9096	4
2	Gen. Pareto	0.1405	2	0.41612	1	0.02393	2
3	Log-Pearson 3	0.14625	3	0.50262	3	2.2702	5
4	Logistic	0.23921	6	1.046	6	1.4921	3
5	Lognormal	0.16223	5	0.5996	5	2.4367	6
6	Lognormal (3P)	0.13625	1	0.43871	2	0.00432	1

Goodness of Fit - Summary for Awash @Nura Hera Station

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.12652	1	0.2185	1	4.1164	6
2	Gen. Pareto	0.14683	5	0.30212	6	3.6749	3
3	Log-Pearson 3	0.12665	2	0.23013	2	3.9683	5
4	Logistic	0.15641	6	0.30041	5	0.38024	1
5	Lognormal	0.13767	4	0.26986	4	1.8253	2
6	Lognormal (3P)	0.12748	3	0.24875	3	3.8625	4

Goodness of Fit - Summary for Metehara Station

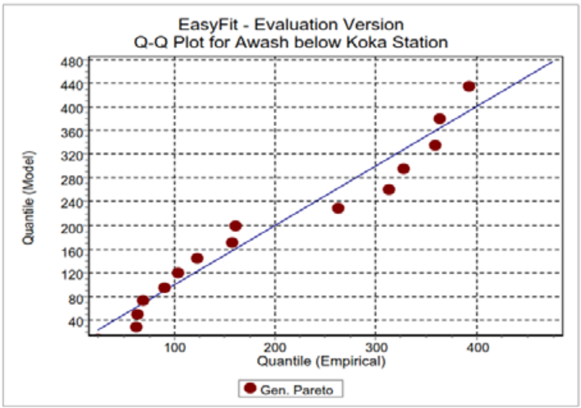
#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.12256	1	0.47553	2	0.6334	2
2	Gen. Pareto	0.12473	2	0.36198	1	0.372	1
3	Log-Pearson 3	0.14511	3	0.54189	3	2.003	4
4	Logistic	0.16292	6	0.75843	6	1.2651	3
5	Lognormal	0.15988	5	0.61832	5	2.2072	6
6	Lognormal (3P)	0.15834	4	0.61685	4	2.2062	5

Goodness of Fit - Summary for Awash Melka Sadi Station

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.17361	1	0.46517	2	1.4924	2
2	Gen. Pareto	0.18846	5	4.3131	6	N/A	
3	Log-Pearson 3	0.17597	2	0.46563	3	3.3962	5
4	Logistic	0.22667	6	0.72662	5	3.5672E-4	1
5	Lognormal	0.18033	4	0.46756	4	3.3511	4
6	Lognormal (3P)	0.17606	3	0.46259	1	3.3402	3

Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.17793	4	0.40479	1	1.3645	3
2	Gen. Pareto	0.21164	5	0.63977	6	0.05825	1
3	Log-Pearson 3	0.17186	2	0.40599	2	1.4063	4
4	Logistic	0.21881	6	0.50741	5	1.0062	2
5	Lognormal	0.17364	3	0.43669	4	1.4985	6
6	Lognormal (3P)	0.17089	1	0.43262	3	1.4719	5



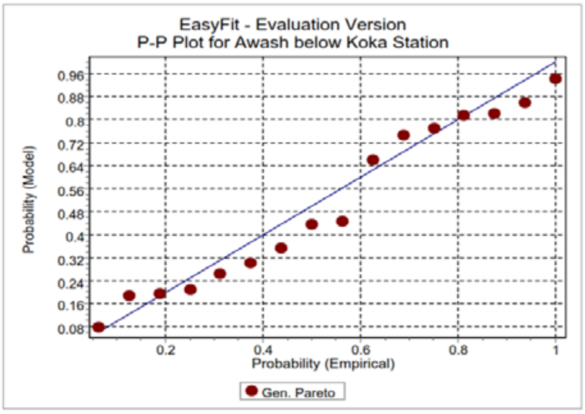


Figure 4. 3 Q-Q and P-P plots (observed data plotted against the fitted distribution)

The summery of goodness-of-fit test results and the P-P and Q-Q plots are shown under the **Appendix-A, B** and **C** respectively. According to the L-moment ratio diagram, goodness-of-fit test and P-P and Q-Q plots analyzed above, it was found that Generalized Pareto (GPA) distribution is best fit for Awash below Koka, Awash @Wanji and Metehara sites and Generalized Extreme Value (GEV) distribution is best for Awsh @Nura Hera, Awsh Melka Sadi and Awsh Melka Werer. This best distribution is used to determine the extreme flow quantiles for the corresponding stations.

Table 4. 9 Selected distributions for observed AMF by L-Moment ratio diagram and Goodness-of-Fit tests

		Goodness of Fit			
No.	**Stations**	**Kolmogorov Smirnov**	**Anderson Darling**	**Chi Squared**	**Selected Distribution (Best Fit)**
1	Awash below Koka	GPA	GPA	LN (3p)	GPA
2	Awash @Wanji	LN (3p)	GPA	LN (3p)	GPA
3	Awash @Nura Hera	GEV	GEV	GLO	GEV
4	Metehara	GEV	GPA	GPA	GPA
5	Awash Melka Sadi	GEV	LN (3p)	GLO	GEV
6	Awash Melka Werer	LN (3p)	GEV	GPA	GEV

Parameter Estimation for Selected Distribution

At-site parameters estimation of selected distributions for observed annual maximum flow (AMF) of six stations were done by using EasyFit statistical software. The software is applying the method of L-moments for the selected probability distributions: GPA and GEV. The results of estimated parameters for the selected distribution are shown in Table 4.10 below.

Table 4. 10 Estimated parameters of selected distributions for respective gauging stations

Site	Best Fit Distribution	Parameters
Awash below Koka	Generalized Pareto (GPA)	k = -0.51757, σ = 319.65 μ = -1.2257
Awash @Wanji	Generalized Pareto (GPA)	k = -0.226, σ = 117.23 μ = 41.902
Metehara	Generalized Pareto (GPA)	k = -0.90189, σ = 205.33 μ = 61.322
Awash @Nura Hera	Generalized Extreme Value (GEV)	k = -0.04957, σ = 73.737 μ = 183.32
Awash Melka Sadi	Generalized Extreme Value (GEV)	k = 0.08352, σ = 49.381 μ = 176.63
Awash Melka Werer	Generalized Extreme Value (GEV)	k = -0.03594, σ = 87.357 μ = 261.93

Where; k = shape parameter, σ = scale parameter and μ = location parameter

4.4 Estimating Extreme Flow Quantiles and Developing Flood Frequency Curves

4.4.1 Estimating Extreme Flow Quantiles

After estimating parameters for the selected best-fit distributions, the extreme flow quantile (Q_T) corresponding to different return periods, T in year were computed from the statistics of the adopted distribution. The return period is related to the probability of non-exceedance (F) by the relation, F = 1-1/T where F = F(X_T) is the probability of having a flood of magnitude equal or less than Q_T. An annual maximum event has a return period (or recurrence interval) of T years if its magnitude is equaled or exceeded once on the average of every T years. The reciprocal of T is the exceedance probability, 1- F, of the event, that is, the probability that the event is equaled or exceeded in any one year. The results of estimated extreme flow quantiles from observed AM discharge for six selected stations in the sub-basin are shown in Table 4.11 for recurrence intervals of 2, 5, 10, 20, 25, 50, 100, 200, 500, 1000, 2000, 5000, 10000 years. The result indicated that Awash Metehara and Awash below Koka stations revealed high quantile estimates.

Table 4. 11 Estimated extreme flow quantiles of different return period

T	F	Awash below Koka	Awash @Wanji	Awash @Nura Hera	Metehara	Awash Melka Sadi	Awash Melka Werer
		Estimated extreme flow quantiles, Q_T (m^3/s)					
		GPA	GPA	GEV	GPA	GEV	GEV
2	0.5	265.29	129.87	210.59	259.05	194.45	294.16
5	0.8	801.78	269.46	298.14	805.72	246.25	396.56
10	0.9	1414.82	396.02	358.86	1649.96	277.94	466.68
20	0.95	2292.43	544.04	419.28	3227.44	306.52	535.75
25	0.96	2648.84	596.84	438.90	3984.03	315.24	558.04
50	0.98	4058.98	778.92	500.75	7588.68	341.07	627.85
100	0.99	6077.64	991.88	564.31	14324.00	365.24	698.91
200	0.995	8967.45	1240.95	629.87	26909.05	387.97	771.51
500	0.998	14784.40	1636.18	719.89	61702.63	416.00	870.12
1000	0.999	21431.53	1994.51	790.70	115436.58	435.81	946.82
2000	0.9995	30947.19	2413.62	863.96	215839.14	454.49	1025.44
5000	0.9998	50101.44	3078.66	964.72	493419.72	477.59	1132.40
10000	0.9999	71989.37	3681.62	1044.04	922105.26	493.92	1215.67

4.4.2 Developing Flood Frequency Curves

Using table 4.11 above, regional flood frequency curves were developed by drawing return period T versus the estimated extreme flood quantile Q_T. The curves were indicated in figure 4.4. Flood frequency curves of all regions were generated separately and shown on **Appendix-D**

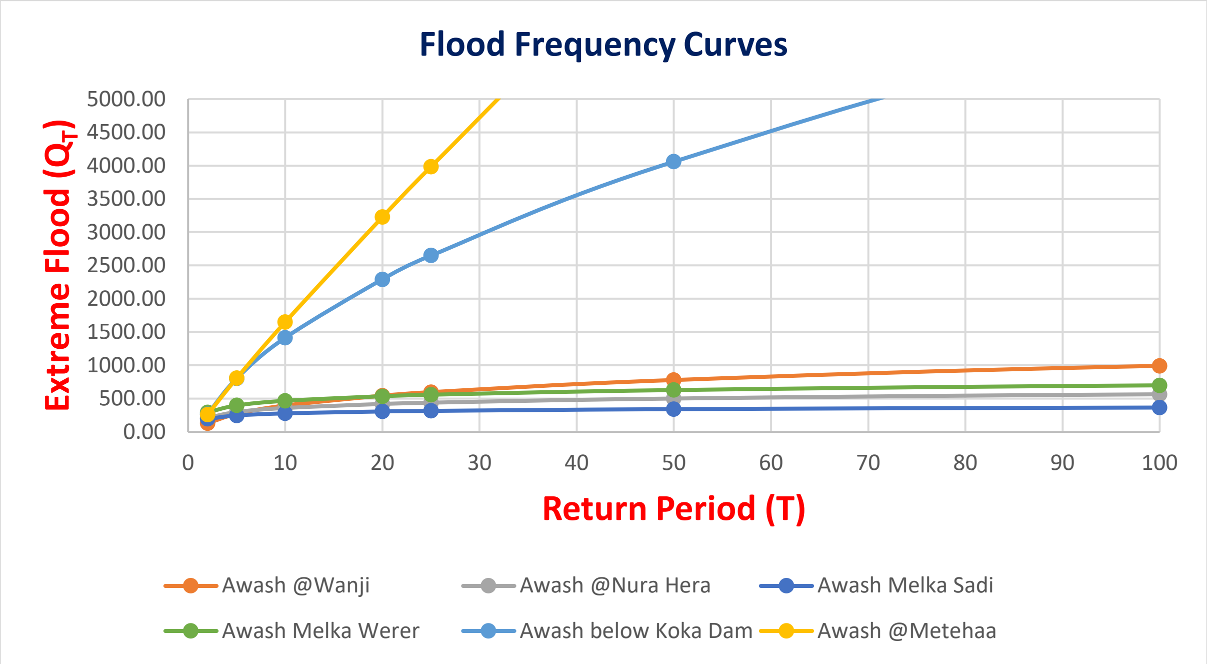


Figure 4. 4 Regional flood frequency curves

4.4.3 Developing of Regional Growth Curves

Depending on the selected distribution and estimated standardize flood quantiles shown in table 4.12 of each station, regional growth curves of the study region were derived and indicated in figure 4.6. The constructed regional growth curves indicated that all curves have different flood characteristics. Awash river reaches from Metehara up to Awash Melka Sadi and from Koka dam up to Awash Wanji respectively cause the first and second most extensive flood in the study region. Regional growth curves of all regions were generated separately and shown on **Appendix-E**

Table 4. 12 Estimated standardize flood quantiles of stations

T	Gumble's reduced variate	Growth Factor (Q_T/Q_{ave})					
		Awash below Koka Q_{ave}=209.41	Awash @Wanji Q_{ave}=137.52	Awash @Nura Hera Q_{ave}=222.43	Metehara Q_{ave}=169.56	Awash Melka Sadi Q_{ave}=209.56	Awash Melka Werer Q_{ave}=309.35
		GPA	GPA	GEV	GPA	GEV	GEV
2	0.37	1.48	1.08	1.24	1.69	1.05	1.74
5	1.50	4.46	2.23	1.75	5.25	1.33	2.34
10	2.25	7.87	3.28	2.11	10.76	1.51	2.76
20	2.97	12.75	4.51	2.47	21.04	1.66	3.16
25	3.20	14.74	4.94	2.58	25.97	1.71	3.30
50	3.90	22.58	6.45	2.95	49.47	1.85	3.71
100	4.60	33.81	8.22	3.32	93.38	1.98	4.13
200	5.30	49.89	10.28	3.71	175.42	2.10	4.56
500	6.21	82.25	13.55	4.24	402.24	2.25	5.14
1000	6.91	119.23	16.52	4.65	752.53	2.36	5.59
2000	7.60	172.17	19.99	5.08	1407.04	2.46	6.06
5000	8.52	278.73	25.50	5.68	3216.58	2.59	6.69
10000	9.21	400.50	30.50	6.14	6011.16	2.67	7.18

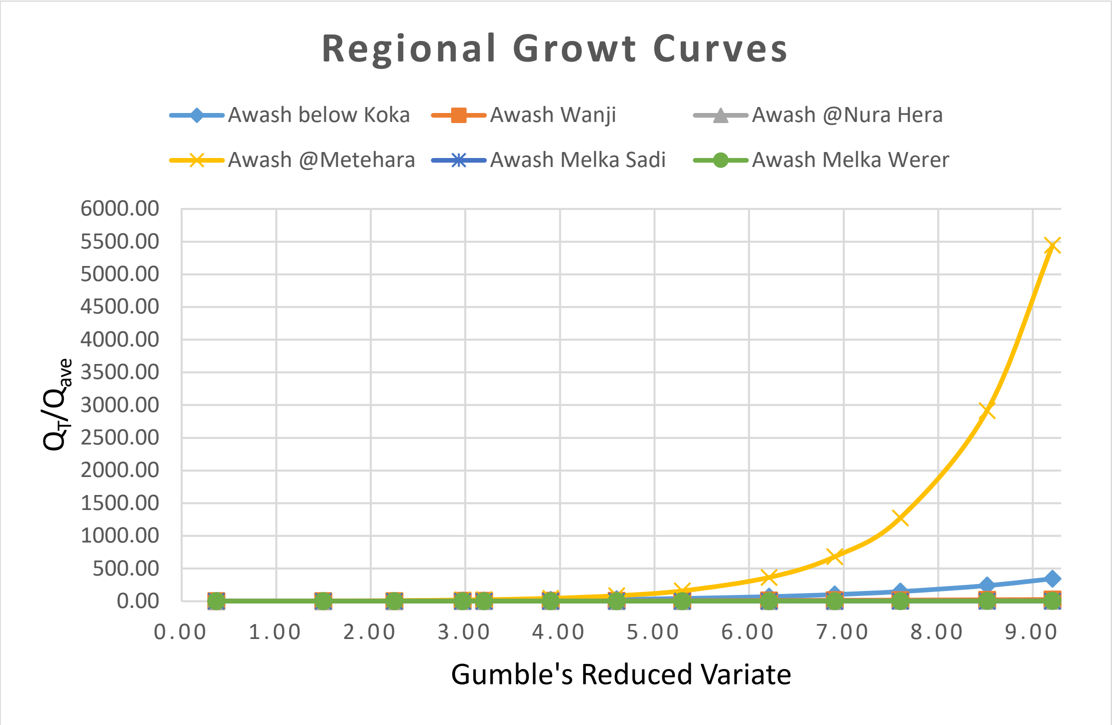


Figure 4. 5 Regional growth curves

4.5 Mapping Flood Risk Area for Selected Return Period

The analysis of the extent of the area inundated under different return-period floods is driven by the peak flows obtained from frequency analysis conducted using best fit probability distribution and L-moment ratio diagram available at the selected river gauging stations. Peak flood for return periods of 25, 50, 100, 200, 500 and 1000 years were used as steady flow inputs in HEC-RAS at all selected stations

4.5.1 River cross section

River cross-section represents the river geometry like river center line, bank line, flow path line and XS Cut line along the flood plain. For this study about 130 XS Cut lines were drawn using HEC-Geo RA in combination with Arc GIS. Then, HEC-Geo RAS with the help of Arc GIS assigns the river stations for every XS Cut lines and also calculated the channel length, left and right over bank length. In the figure 4.6, the blue line indicated the river center line along the flood plain, the red points shown the left and right banks line points and the green line represented the river XS Cut lines drawn across the river center line which is normal to the flow paths.

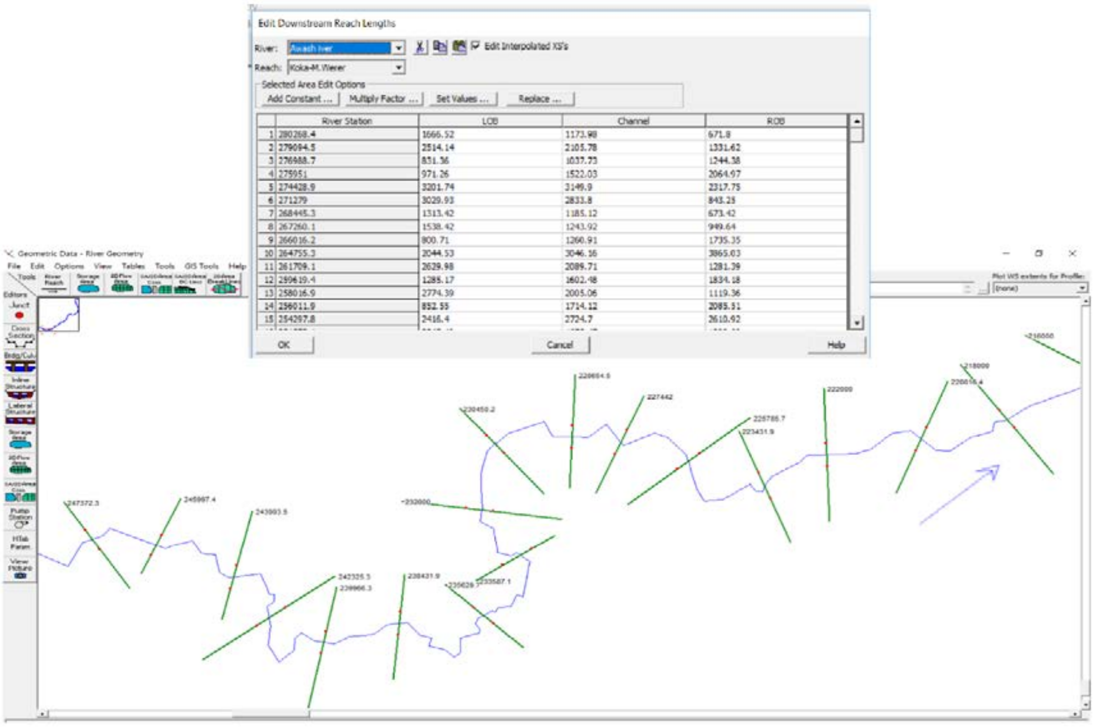


Figure 4. 6 River station and cross section network exported from Arc GIS to HEC-RAS

Water depth in the river cross-section at some distance downstream of Awash below Koka station under 25, 50, 100, 200, 500 and 1000-year return-period floods was shown in figure 4.7 below. Water depths in the river cross-section at some distance downstream of all selected stations under 25, 50, 100, 200, 500 and 1000-year return-period floods were shown in **Appendix-F**.

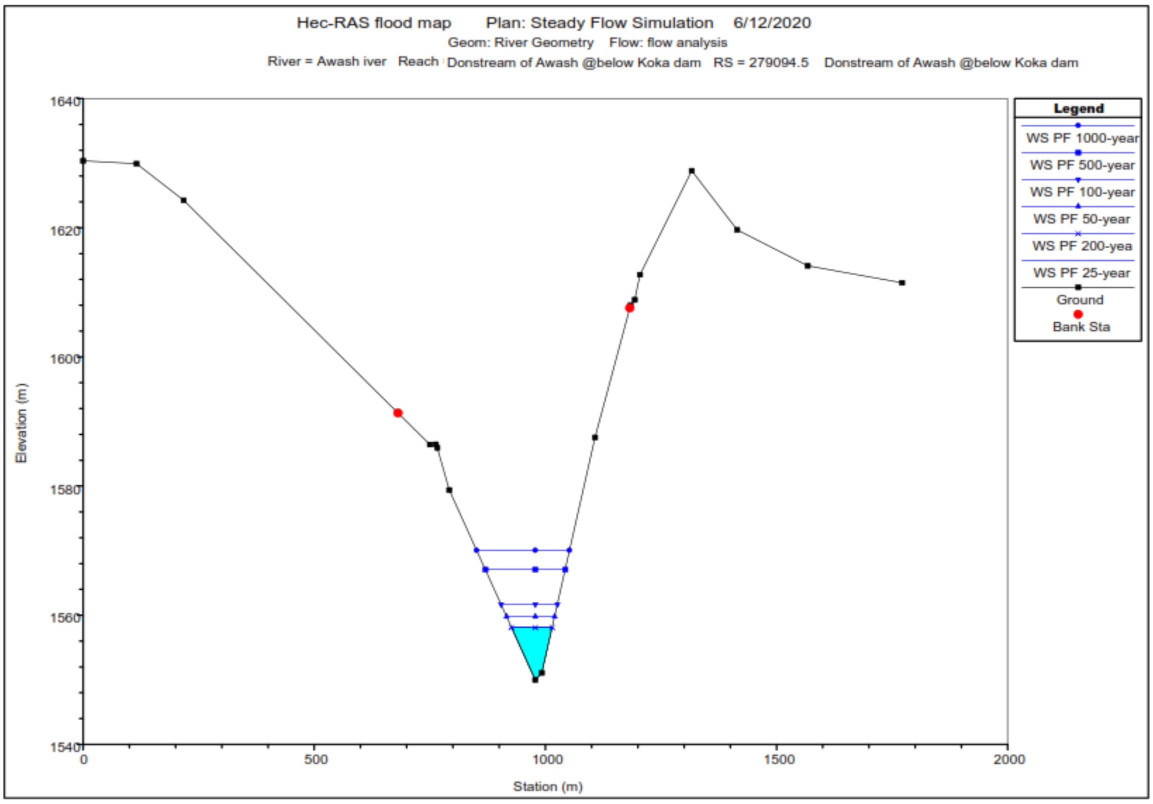


Figure 4. 7 Water surface elevations downstream of Awash @below Koka station

4.5.2 Water Surface Profile

Water surface profile of river reach from Awash Metehara station up to Awash Melka Sadi station under 25, 50, 100, 200, 500 and 1000-year return-period floods was shown in figure 4.8 below. Water surface profile of the whole river reach (from Awash below Koka station up to Awash Melka Werer station) under 25, 50, 100, 200, 500 and 1000-year return-period floods were shown in **Appendix-G**.

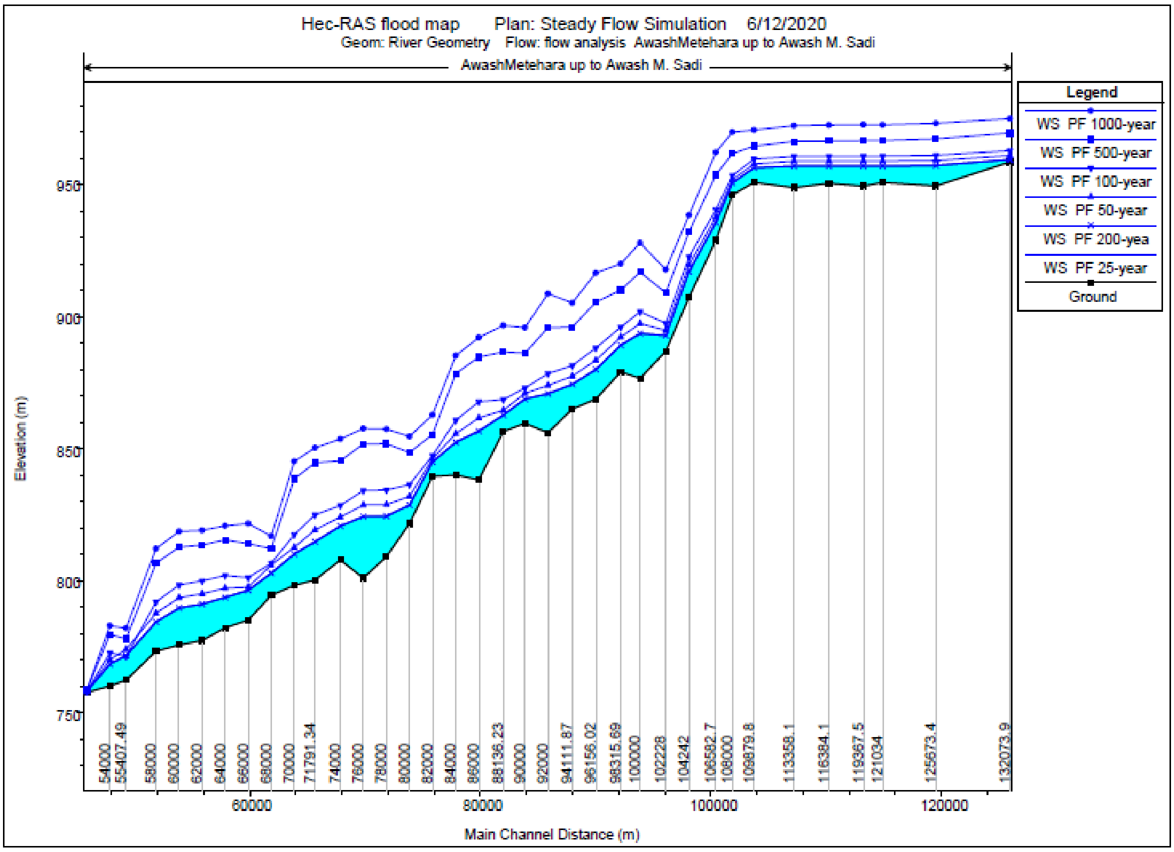


Figure 4. 8 Water surface profile of Awash rive reach from Metehara up to Awash Melka Sadi

The 3-D view of water surface profile of Awash river reach from Awash @Nura Hera up to Awash Metehara stations for 25, 50, 100, 200, 500 and 1000 years return period as a result of steady flow analysis in HEC-RAS is shown in figure 4.9 below. The 3-D view of water surface profile of all river reach of the study area for 25, 50, 100, 200, 500 and 1000 years return period as a result of steady flow analysis in HEC-RAS were shown in **Appendix-H**.

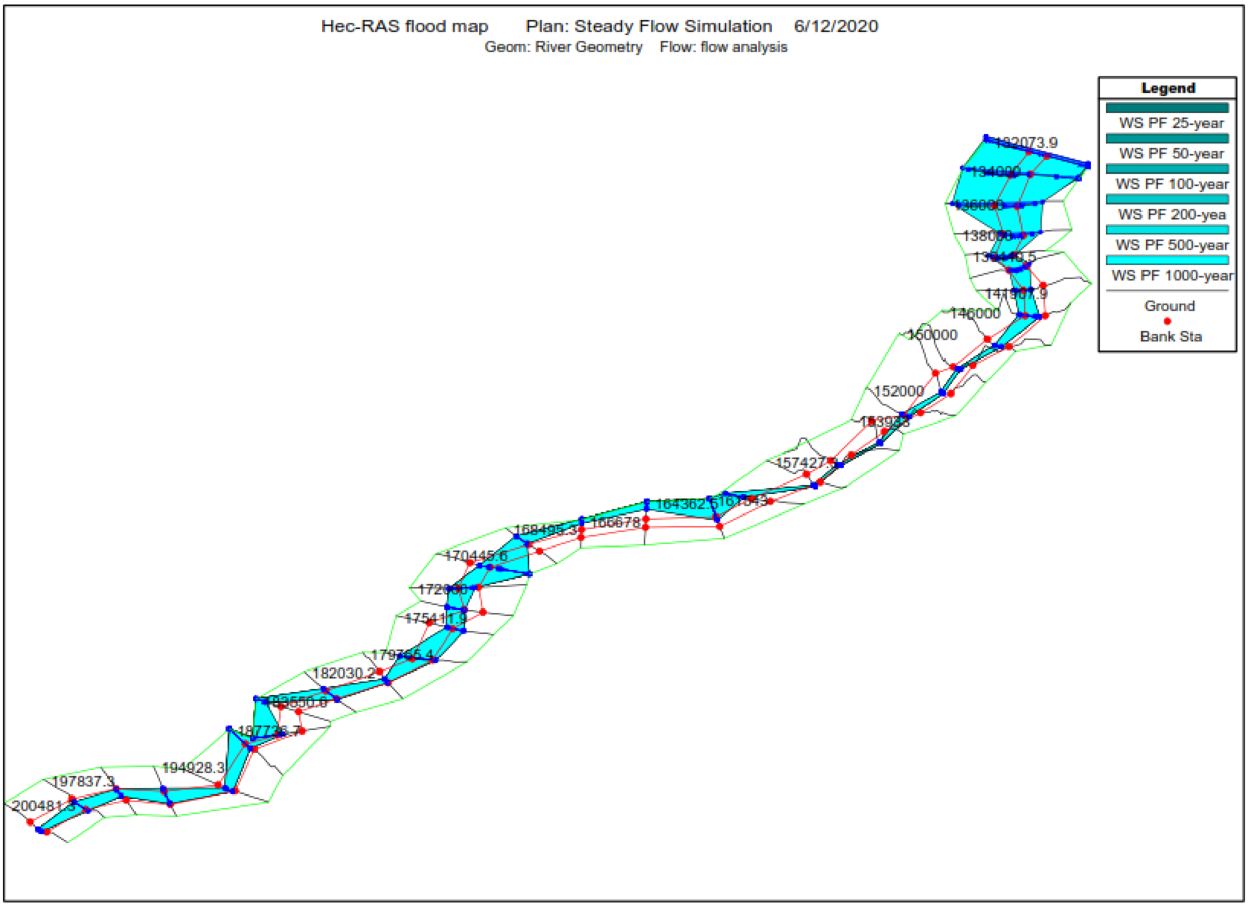


Figure 4. 9 3-D view of water surface profile of Awash rive from Nura Hera up to Metehara

4.5.3 Flood Mapping

Using RAS Mapper of HEC-RAS model, TIN of the study area was imported and a new terrain of the study region was created. Using the water surface dada extracted from HEC-RAS model and TIN created for the basin, the flooded area under different return period floods was delineated. Although floodplain maps were prepared for various return periods, namely 25, 50, 100, 200, 500 and 1000 years, but only the 50, 100, 200 and 500-year floodplain maps were presented. Figure 4.10 below shows a 100-year floodplain map of Awash river reach selected for this particular study. Floodplain map of the 50,200 and 500-year were shown in **Appendix-K**. Separate floodplain maps were presented for each reach. For instance, figure 4.11 below shows the regions that are likely to be inundated under 100-year return period flood of Awash river reach from Awash @Metehara station up to Awash Melka Sadi station. Results of simulation under different return period flood indicate that the most severely affected reaches of the river are all parts of Awash below Koka dam up to Awash @Wanji, some parts of Awash Nura Hera up to Awash @Metehara and some parts of Awash @Metehara up to Awash @Melka Werer. A 100-year floodplain maps of the remaining river reaches were shown in **Appendix-K.**

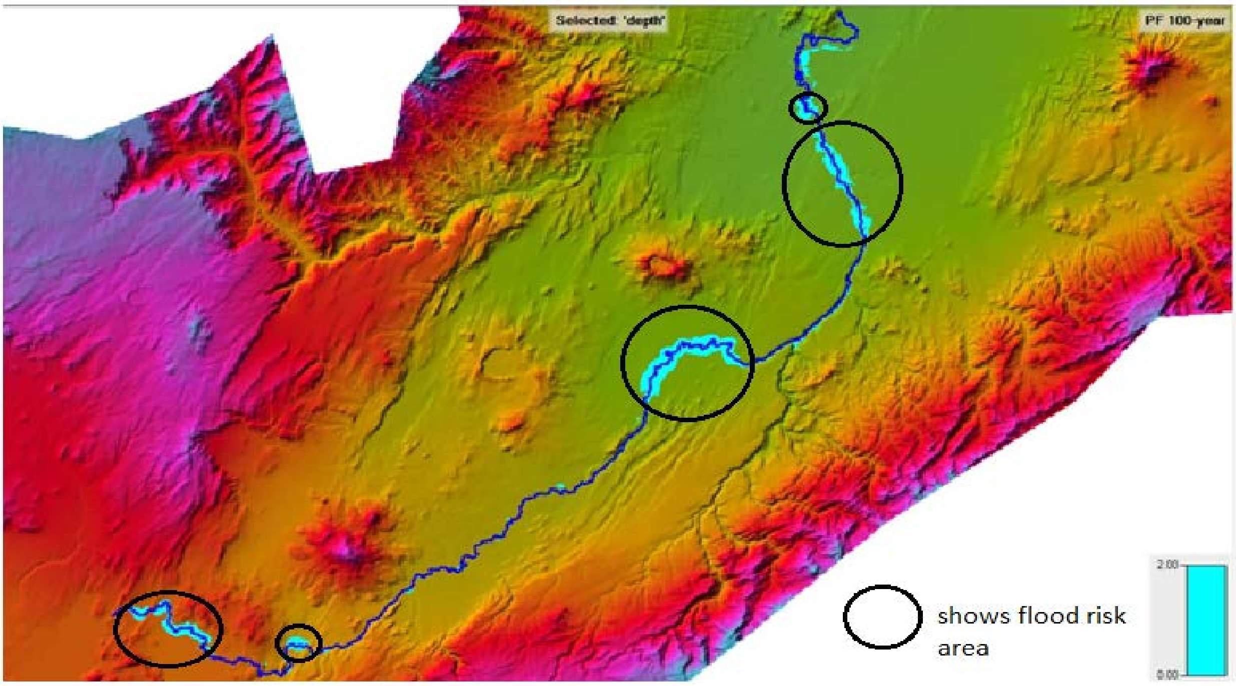


Figure 4. 10 100-year floodplain map for Awash River of study region

Figure 4. 11 100-year floodplain map for Awash River from Metehara up to Awash Melka Sadi.

4.5.4 Flood Inundation Area

Using Hec-RAS Mapper a new geometry of flood inundation area was created to edit the geometry of the storage area by delineating the floodplain already mapped. Then, the inundated area for each river reach under different return period were computed and can be displayed from storage area Attribute Table as shown in Figure 4.12 below. For instance, Table 4.13 below shows flood inundation area of each river reach flood plain under a 100-year flood.

Table 4. 13 Flood inundated area of all river reach under a 100-year flood

Name of River Reach	Peak Flow (m^3/s)	Flood inundation area (Km^2)
Awash @Koka to Awash @Wanji	6077.64	15.19
Awash @Wanji to Awash @NuraHera	991.88	50.97
Awash @Nura Hear to Awash @Metehara	564.31	61.73
Awash @Metehara to Awash @Melka Sadi	14324	126.54
Awash @Melka Sadi to Awash @Melka Werer	365.24	55.58

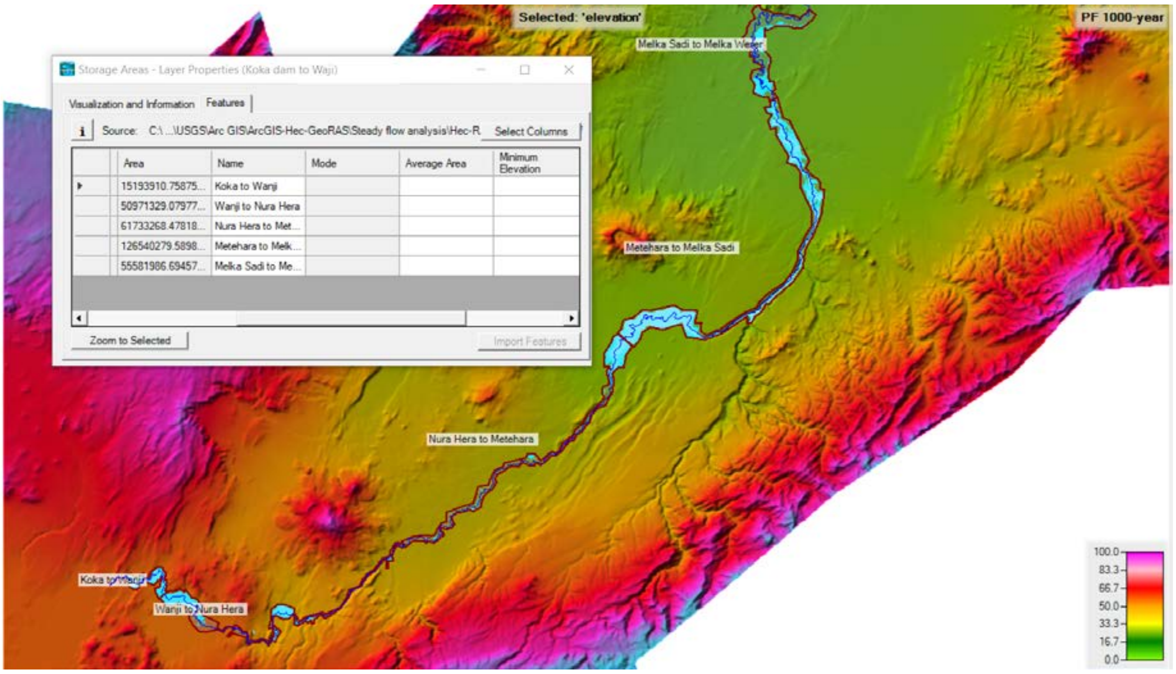


Figure 4. 12 100-year floodplain map for Awash River of study region

5. CONCLUSIONS AND RECOMMENDATIONS

Conclusions

The major focus of this study was to carry out the regional flood frequency analysis using annal maximum series flow data for the period of 1999-2014 of six river gauging stations of middle part of Awash River Basin, from Koka dam up to Melka Werer, and determining the extent of flood inundation using Hec-GeoRAS and HEC-RAS models under different return-periods.

The annual maximum flow series of all gauging stations were fitted by five of the most frequently used distributions in the analysis of hydrologic extreme variables. These distributions were generalized extreme value (GEV), general logistic (GLO), generalized pareto (GPA) lognormal (LN) and log-Pearson type 3 (LP-III) distributions. Comparison of best fit distribution were done based on the use goodness-of-Fit tests such as Kolmogorov-Smirnov, Anderson-Darling and Chi-Squared tests and L-moment ratio diagram.

The goodness-of-Fit tests and L-moment ratio diagram analysis indicated that Generalized Pareto (GPA) is the best fit distribution for Awash below Koka, Awash @Wanji and Metehara gauging stations and Generalized Extreme Value (GEV) is best distribution for Awsh @Nura Hera, Awsh Melka Sadi and Awsh Melka Werer gauging stations. Using these distributions, the values of 2, 5 ,10, 20, 25, 50, 100, 200, 500, 1000, 2000, 5000 and 10000-year return period floods and their regional growth curves were computed and shown on table 4.5 and table 4.6 respectively.

Using the Hec-GeoRAS and HEC-RAS in models, in combination of ArcGIS, an analysis of the extent of areas likely to be inundated under different return-period floods was carried out. Triangular irregular network (TIN) was used to extracting river geometries like cross-sections, bank-lines, flow-paths and river center-lines using Hec-GeoRAS in combination with ArcGIS. The extracted river geometries were used to run steady flow analysis in HEC-RAS model for computation of water surface elevation, water surface profile and flood inundation area mapping for 25, 50, 100, 200, 500 and 1000-year return period floods. The result of flood inundation area for the river reach from U/S to D/S respectively under a 100-year return period are 15.19, 50.97, 61.73, 126.54 and 55.58 Km2. The result indicated that all parts of Awash below Koka dam up to Awash @Wanji, some parts of Awash Nura Hera up to Awash @Metehara and some parts of Awash @Metehara up to Awash @Melka Werer river reaches were under the impact of 25, 50 100, 200, 500 and 1000-year flood problem.

5.1 Recommendations

For hydraulic model development, HEC-RAS needs river geometry and river cross-section, manning's roughness coefficient for precise computations of water surface elevation, water depth in the river, velocity of flow and flood inundation area. For this study, river geometry was extracted from TIN of the specified flood plain and manning's roughness coefficient was taken from land use land cover from the previous study on the river basin. But the river geometry was fluctuating from time to time due to different land use management. Therefore, it is more profitable if some will conduct proper field survey in order to collect river geometry and carried out some experiment so as to investigate the actual manning's roughness coefficient so that more accurate model can be developed.

For accurate extraction of river geometry and flood inundation mapping, HEC-RAS requires high precision digital terrain model. So, it is more preferable if someone uses high precision digital terrain model during hydraulic model development and flood inundation mapping using HEC-RAS. In this study, although flood hazard mapping was prepared for various return periods, namely 25, 50, 100, 200, 500 and 1000 years, appropriate return period can be selected for flood risk management.

This study has shown that large areas found closer to the selected river reach are inundated by the 25, 50, 100, 200, 500 and 1000-year floods. Hence, there should be flood management strategic plan to reduce disaster risks and potential damages caused through comprehensive and coordinated interventions in the region. So, this study can assist policy makers and planners in the development of flood mitigation measures for the river reach where the impact of flood with a required return period occurs.

Having the aforementioned recommendations in mind further researching can be done in order to investigate flood risks and its management techniques.

REFERENCES

Abdelbasset, M.; Abderrahim, L.; Chaouni, A.A.; Abdellah, B.; Lahcen, B.; Laila, B.: Integration of GIS and HEC-RAS in floods modeling of the Ouergha river, Northern Morocco. Eur. Sci.J. 11(2), 1857–7881 (2015)

Alaghmand, S., Abdullah, R., Abustan, I., Vosoogh, B., 2010. GIS-based river flood hazard mapping in urban area (a case study in Kayu Ara River Basin, Malaysia). Int. J. Eng. Technol. 2 (6), 488–500.

Alam, J., Muzzammil, M. and Khan, M.K. (2016). Regional Flood Frequency Analysis: Comparison of L-moment and Conventional Approaches. Journal of Hydraulic Engineering. Vol. 22 (3).

Ali El-Naqa and Mohammad Jaber, 2018. Floodplain Analysis using ArcGIS, HEC-GeoRAS and HEC-RAS in Attarat Um Al-Ghudran Oil Shale Concession Area, Jordan

Al-Qudah, K.A.: Floods as water resource and as a hazard in arid regions: a case study in Southern Jordan. Jordan J. Civ. Eng. 5(1), 148–161 (2011)

Amalina, N.M, Shabri, A. and Badyalina, B. (2016). Selecting Probability Distribution for Regions of Peninsular Malaysia Streamflow. AIP Conference Proceedings 1750, 060014. the American Institute of Physics.

Amirataee, B., Montaseri, M. and Rezaei, H. (2014). Assessment of Goodness of Fit Methods in Determining the Best Regional Probability Distribution of Rainfall Data. International Journal of Engineering. Vol. 27 (10), 1537-1546.

Ashraful A. M., Emura, K., Farnham, C. and Yuan. (2018). Best-Fit Probability Distributions and Return Periods for Maximum Monthly Rainfall in Bangladesh. The Journal of climate. Vol.6(9).

Athulya, P. S. and James, K.C. (2017). Best Fit Probability Distributions for Monthly Weather Data. International Journal of Advances in Management, Technology and Engineering Sciences. Vol.7 (12).

Awash Basin Authority (2017). Awash River Basin Flood and Drought Management Strategic Plan

Badreldin, G. and Fengo, P. (2012). Regional Rainfall Frequency Analysis for the Luanhe Basin using L-moment and Cluster Techniques.

Betul Saf Fatih B. and Mutlu Y. (2009). Regional flood frequency analysis of the lower west Mediterranean subregion of Turkey with L-moments

Bhagat, N. (2017). Flood Frequency Analysis Using Gumbel's Distribution Method: A Case Study of Lower Mahi Basin. Journal of Water Resources and Ocean Science. Vol. 6(4),5154.

Bocchiola, D., De Michele, C. & Rosso, R. (2003). Review of recent advances in index flood estimation. Hydrol. Earth Sys. Sci., 7(3), 283–296.

Burn, D.H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. Water Resour. Res., 26(10), 2257–2265.

Chavoshi, S.B. and Azmin, W.S. (2009). Development of L-Moment Based Models for Extreme Flood Events. Malaysian Journal of Mathematical Sciences 3(2), 281- 296.

Capesius, J.P. and Stephens, V.C. (2009). Regional Regression Equations for Estimation of Natural Streamflow Statistics in Colorado: U.S. Geological Survey Scientific Investigations, Report. 5136, 46.

Chavoshi, S.B. and Azmin, W.S. (2009). Development of L-Moment Based Models for Extreme Flood Events. Malaysian Journal of Mathematical Sciences 3(2), 281- 296.

Crochet, P. (2012a). Estimating the flood frequency distribution for ungauged catchments using an index flood procedure. Application to ten catchments in Northern Iceland. Icelandic Met. Office report No. VÍ 2012-005, 59pp.

Cunnane, C. (1989). Statistical Distributions for Flood Frequency Analysis. WMO operational Hydrology Report No. 33. World Meteorological Organization, Giuseppe Motta. No. 5, Geneva, Switzerland.

Das, S. and Simonovic, S.P. (2012). Assessment of Uncertainty in Flood Flows Under Climate Change-the Upper Thames River basin (Ontario, Canada). Water Resources Research Report.

Demissie, M. and Michael, Y. (2008). Regional Flood Frequency Analysis for Upper Awash Sub-basin (Upstream of Koka). MSc. Thesis Submitted to Addis Ababa University.

De Paola, Francesco, Giugni, Maurizio, Topa, Maria Elena, Bucchignani, Edoardo, 2014. Intensity-Duration-Frequency (IDF) Rainfall Curves, for Data Series and Climate Projection in African Cities. Springer plus, pp. 133. Open access., 3.

Dessalegn, B., Ayalew, S. and Hailu, D. (2016). Evaluation of Extreme Flow Quantiles Estimated from Global Reanalysis Runoff data. A Case Study on Blue Nile River Basin. MSc. Thesis Submitted to Addis Ababa University.

Dubey, A. (2014). Regional Flood Frequency Analysis Utilizing L-Moments: A Case Study of Narmada Basin: International Journal of Engineering Research and Applications. Vol. 4(2):155-161.

Frequency Analysis. Journal of Advances in Water Resources.105, 144-153.

Ganora, D. and Laio, F. (2016). A Comparison of Regional Flood Frequency Analysis Approaches in a Simulation Framework. Water Resources Research Report. 52, 56445661.

Gebeyehu, A. (1989). Regional Flood Frequency Analysis, PhD. Dissertation Submitted to Stockholm University. Hydraulics Laboratory, the Royal Institute of Technology. Florida: CRC Press LLC.

Gedefa, T.W. and Seleshi, Y. (2009). Regional Flood Frequency Analysis for Upper OmoGibe Sub-basin; MSc. Thesis Submitted to Addis Ababa University.

Ghosh, S., Roy, M.K., and Biswas, S.C. (2016). Determination of the Best Fit Probability Distribution for Monthly Rainfall Data in Bangladesh, American Journal of Mathematics and Statistics. Vol. 6(4), 170-174.

Hailegeorgis, T.T and Alfredsen, K. (2017). Regional Flood Frequency Analysis and Prediction in Ungauged Basins including Estimation of Major Uncertainties for MidNorway. Journal of Hydrology: Regional Studies. Vol.9, 104–126.

Haile Arefayne Shishaye and Asmelash Tilahun Asfaw (2020). Analysis and evaluation o the spatial and temporal variabilities of river water quality parameters

Hosking, J. R. M., and Wallis, J. R. (1997). Regional Frequency Analysis: An Approach Based on L-moments. Cambridge University Press, New York, USA.

Hosking JRM, Wallis JR (2005) Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press.

Hussein, B. and Wagesho, N. (2016). Regional Flood Frequency Analysis for Abaya-Chamo Sub-Basin, Rift-Valley River Basin, Ethiopia: Journal of Resources Development and Management. An International Peer-Reviewed Journal: Vol.24

Hossein Mehrannia*1 & Alireza Pakgohar[2] (2014). Using Easy Fit Software for Goodness-Of-Fit Test and Data Generation

Huxley, Chris, Kuch, Anthony, 2016. Urban drainage modeling: tools for improving urban planning and design decision making. FMA (Flood Manage. Assoc.) newsletter 26 (1).

H. M. Raghunath. Hydrology Principles. Analysis. Design

IPCC, 2007. Fourth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland p 104.

Irwin, S.K, Srivastav, R. and Simonovic, S. P. (2014). Instruction for Watershed Delineation in an ArcGIS Environment for Regionalization Studies: Water Resources Research Report No: 87

Jalayer, F., De Risi, R., Kyessi, A., Mbuya, E., Yonas, N., 2015. Vulnerability of built environment to flooding in African cities. Urban Vulnerability and Climate Change in Africa. Springer International Publishing, pp. 77–106.

Kachroo, R.K., Mkhandi, S.H. and Parida, B.P. (2000). Flood Frequency Analysis of Southern Africa: I. Delineation of Homogenous Regions. Journal of Hydrological Science. 45 (3), 437–448.

Kanti, K. K., Sung-Kee Y., Jun-Ho L. and Khan, K. (2017). Regional Flood Frequency Analysis for Using L-moments Approach in Jeju Island, Korea. Geo-environmental Disasters. Vol.4(18).

Karagiozi, E., Fountoulis, L., Konstantinidis, A., Andreadakis, E., Ntou, K., 2011. Flood hazard assessment based on geomorphological analysis with Gis tools – the case of Laconia (Peloponnesus, Greece). GIS Ostrava 23–26, Ketsela, H., Tadele, K. and Temam, D. (2017). An Assessment of Predictive Accuracy for Regional Flood Frequency Distributions and Analysis Estimation Methods on Awash River basin. MSc. Thesis Submitted to Jimma University.

Kumar R. and Chatterjee C. (2011). Development of Regional Flood Frequency Relationships for Gauged and Ungauged Catchments Using L-Moments. Journal of Hydrology, In Extremis. Descriptive Events and Trends in Climate and Hydrology. Berlin, Heidelberg.

Lan-Fen, Chu, 2012. Statistical Modelling of Extreme Rainfall in Taiwan. Institute of Economics Academia Sinica, Taiwan.

Malekinezhad, H., Nachtnebel, H.P. and Klik, A. (2011). Regionalization Approach for Extreme Flood Analysis Using L-moments. Journal of Agricultural Science and Technology. Vol.13, 1183–1196.

Manandhar, B.: Flood plain analysis and risk assessment of Lothar Khola. Master of Science Thesis in Watershed Management. Tribhuvan University Institute of Forestry Pokhara, Nepal (2010)

Mehrannia, H. and Pakgohar, A. (2014). Using Easy Fit Software for Goodness-of-Fit Test and Data Generation. International Journal of Mathematical Archive: Vol. 5(1), 118124.

Mishra, B.K., Takara, K., Yamashiki, Y. and Tachikawa, Y. (2009). Selection of Regional Distribution Using Simulated Flood Data. Annual Disaster Prevention Research Institute, Kyoto University, No.52B.

Mkhandi S.1, Opere A.O.2, Willems P.3 Comparison Between Annual Maximum and Peaks over Threshold Models for Flood Frequency Prediction

Mengistu Demissie: Regional Flood Frequency Analysis for Upper Awash Sub-Basin

Nirman Bhagat. Flood Frequency Analysis Using Gumbel's Distribution Method: A Case Study of Lower Mahi Basin, India. Journal of Water Resources and Ocean Science.

Nobert, J., Mugob, M. and Gadain, H. (2014). Estimation of Design Floods in Ungauged Catchments Using a Regional Index Flood Method. A Case Study of Lake Victoria Basin in Kenya: Journal of Physics and Chemistry of the Earth.

Noto, L.V. and Loggia, G.L., (2009). Use of L-Moments Approach for Regional Flood Frequency Analysis in Sicily, Italy. Journal of Water Resources Management.Vol.3, 2207–2229.

O’Brien, J., 2011. FLO-2D Reference Manual, Version 2009.

Parida, B. P., Kachroo, R. K. and Shrestha, D. B. (1998). Regional Flood Frequency Analysis of Mahi Sabarmati Basin (Subzone-3a) Using Index Flood Procedure with L-Moments. Water Resources Management. Vol. (12), 1–12.

Patil, S. and Stieglitz, M. (2012). Controls on Hydrologic Similarity: Role of Nearby Gauged Catchments for Prediction at Ungauged Catchment. Hydrological Earth System Science. Vol. (16), 551–562.

Rabba, Zeinu Ahmed (2017). Flood estimation in developing countries with case studies in Ethiopia.

Rahmana, B, K. Haddad and Kuczerac, G. (2015). Features of Regional Flood Frequency Estimation Model in Australian Rainfall-Runoff 21st International Congress on Modeling and Simulation, Gold Coast, Australia.

Rao, A.R. and Srinivas, V.V. (2008). Regionalization of Watersheds by Hybrid Cluster Analysis. Journal of Hydrology. The Netherlands.

Rao, A.R. and Hamed, K.H. (2000). Flood Frequency Analysis. CRC Press, Washington.

R. Kidson and K.S. Richards (2005). Flood frequency analysis: Assumptions and Alternatives

Romali, N.S and Yusop, Z. (2017). Frequency Analysis of Annual Maximum Flood for Segamat River. MATEC Web of Conferences 103, 04003.

Schendel, T and Thongwichian, R. (2017). Considering Historical Flood Events in Flood 10.11648/j.wros.20170604.11

Schittkowski, K. (2002). Easy-Fit: A Software System for Data Fitting in Dynamic Systems.Vol. 23(2), 153-169.

Sifan Abera, 2012: Flood Early Warning Using Lisflood And Remote Sensing Derived Products from Geonetcast; Case Study of The Awash River Basin, Ethiopia

Sine, A. and Ayalew, S. (2004). Identification and Delineation of Hydrological Homogeneous Regions. The Case of Blue Nile River Basin. Lake Abaya Research Symposium Proceedings. Vol.4

Sitotaw Haile Erena, Hailu Worku, Francesco De Paola, (2018). Flood hazard mapping using FLO-2D and local management strategies of Dire Dawa city, Ethiopia

Solaimani, K.: Flood forecasting based on GIS and hydraulic model. J. Adv. Dev. Res. 1(1), 125–131 (2011)

Tadesse, A.H., Gottschalk, L. and Tallaksen, L.M. (2011). Regional Flood Frequency Analysis in Southern Africa, MSc. Thesis Submitted to the University of Oslo.

US Army Corps of Engineering (USACE), 2016. HEC-RAS user’s manual.

US Army Corps of Engineers (2011) HEC-GeoRAS- GIS tools for support of HEC-RAS using ArcGIS, user’s manual, Institute for water resources, Hydrologic engineering center report CPD-83, Davis, California, USA.

Vanneuville, W., Kellens, W., De Maeyer, Ph., Reniers, G., Witlox, F., 2011. Is’ Flood Risk Management’ Identical to’ Flood Disaster Management’? Earthzine, March 21. ICEO/IEEE, Washington DC 13 pp.

Ven Te Chow David R. Maidment Larry W. Mays. Applied Hydrology Vol. 6, No. 4, 2017, pp. 51-54. doi

Vivekanandan, N. (2015). Flood Frequency Analysis using Method of Moment and L-moment Probability Distribution. American Institute of Science. International Journal of Mathematics and Computational Science. Vol. 2 (3). 141-146

Zaman, M. A., Rahman, A. and Haddad, K. (2012). Regional Flood Frequency Analysis in Arid Regions: A Case Study for Australia. Journal of Hydrology. 475,74–83.

Zelenhasic, E. 1970. Theoretical Probability Distributions for Flood Peaks. Colorado University Press, Colorado.)

APPENDIXES

Appendix-A: Goodness fit test results and descriptive statistics for selected distribution of gauging stations

EasyFit- Evaluation Version - **Awash below Koka station**

Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.15565	2	0.51432	5	2.9751	6
2	Gen. Pareto	0.12738	1	0.34841	1	0.01149	2
3	Log-Pearson 3	0.15773	3	0.39706	2	0.19894	3
4	Logistic	0.20894	6	0.85827	6	2.4642	5
5	Lognormal	0.17245	4	0.47373	3	9.1305E-4	1
6	Lognormal (3P)	0.17609	5	0.4888	4	0.20208	4

Goodness of Fit - Details [hide]

Gen. Pareto [#2]					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	16 0.12738 0.92848 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.25778	0.29472	0.32733	0.36571	0.39201
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	16 0.34841 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value Rank	1 0.01149 0.91465 2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.6424	2.7055	3.8415	5.4119	6.6349
Reject?	No	No	No	No	No

EasyFit- Evaluation Version - **Awash @Wanji station**

Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.14734	4	0.55147	4	1.9096	4
2	Gen. Pareto	0.1405	2	0.41612	1	0.02393	2
3	Log-Pearson 3	0.14625	3	0.50262	3	2.2702	5
4	Logistic	0.23921	6	1.046	6	1.4921	3
5	Lognormal	0.16223	5	0.5996	5	2.4367	6
6	Lognormal (3P)	0.13625	1	0.43871	2	0.00432	1

Goodness of Fit - Details [hide]

Gen. Pareto [#2]					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	16 0.1405 0.868 2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.25778	0.29472	0.32733	0.36571	0.39201
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	16 0.41612 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value Rank	2 0.02393 0.98811 2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Reject?	No	No	No	No	No

EasyFit - Evaluation Version - **Awash @Nura Hera station**

Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.12652	1	0.2185	1	4.1164	6
2	Gen. Pareto	0.14683	5	0.30212	6	3.6749	3
3	Log-Pearson 3	0.12665	2	0.23013	2	3.9683	5
4	Logistic	0.15641	6	0.30041	5	0.38024	1
5	Lognormal	0.13767	4	0.26986	4	1.8253	2
6	Lognormal (3P)	0.12748	3	0.24875	3	3.8625	4

Goodness of Fit - Details [hide]

Gen. Extreme Value [#1]					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	16 0.12652 0.93178 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.25778	0.29472	0.32733	0.36571	0.39201
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	16 0.2185 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value Rank	2 4.1164 0.12769 6				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Reject?	Yes	No	No	No	No

EasyFit - Evaluation Versin - **Awash @Metehara station**

Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.12256	1	0.47553	2	0.6334	2
2	Gen. Pareto	0.12473	2	0.36198	1	0.372	1
3	Log-Pearson 3	0.14511	3	0.54189	3	2.003	4
4	Logistic	0.16292	6	0.75843	6	1.2651	3
5	Lognormal	0.15988	5	0.61832	5	2.2072	6
6	Lognormal (3P)	0.15834	4	0.61685	4	2.2062	5

Goodness of Fit - Details [hide]

Gen. Pareto [#2]					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	16 0.12473 0.93833 2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.25778	0.29472	0.32733	0.36571	0.39201
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	16 0.36198 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value Rank	2 0.372 0.83027 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Reject?	No	No	No	No	No

EasyFit- Evaluation Version - **Awash @Melka Sadi station**

Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.17361	1	0.46517	2	1.4924	2
2	Gen. Pareto	0.18846	5	4.3131	6	N/A	
3	Log-Pearson 3	0.17597	2	0.46563	3	3.3962	5
4	Logistic	0.22667	6	0.72662	5	3.5672E-4	1
5	Lognormal	0.18033	4	0.46756	4	3.3511	4
6	Lognormal (3P)	0.17606	3	0.46259	1	3.3402	3

Goodness of Fit - Details [hide]

Gen. Extreme Value [#1]					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	16 0.17361 0.65831 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.25778	0.29472	0.32733	0.36571	0.39201
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	16 0.46517 2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value Rank	2 1.4924 0.47416 2				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	3.2189	4.6052	5.9915	7.824	9.2103
Reject?	No	No	No	No	No

EasyFit - Evaluation Versin - **Awash @Melka Werer station**

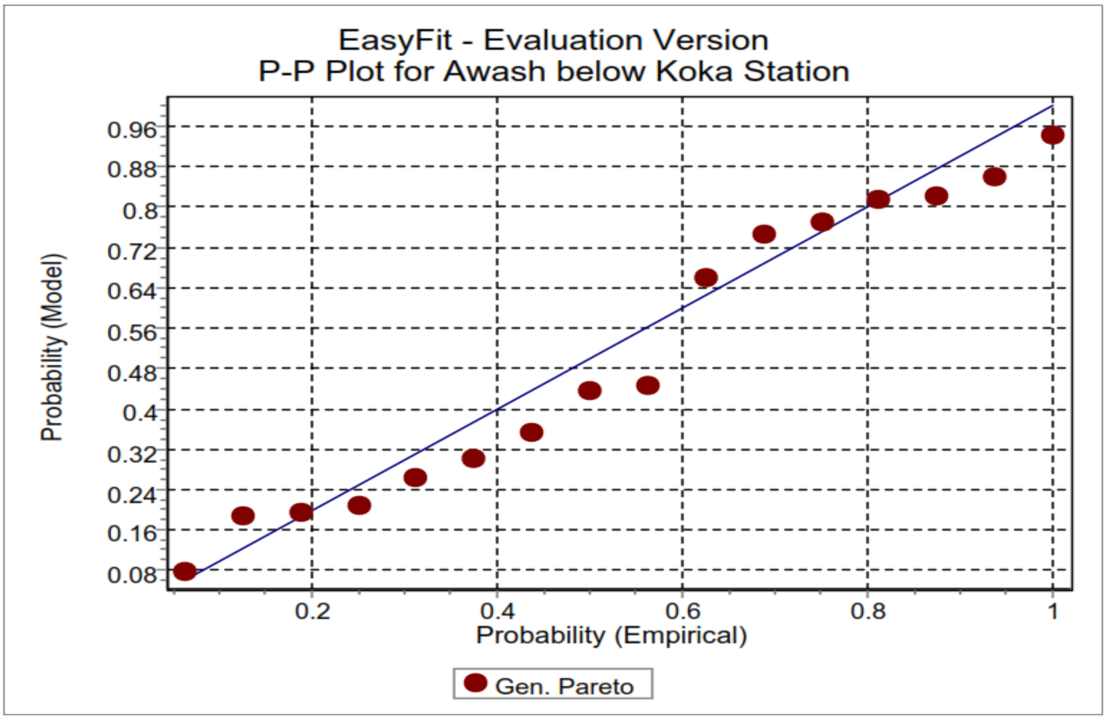
Goodness of Fit - Summary

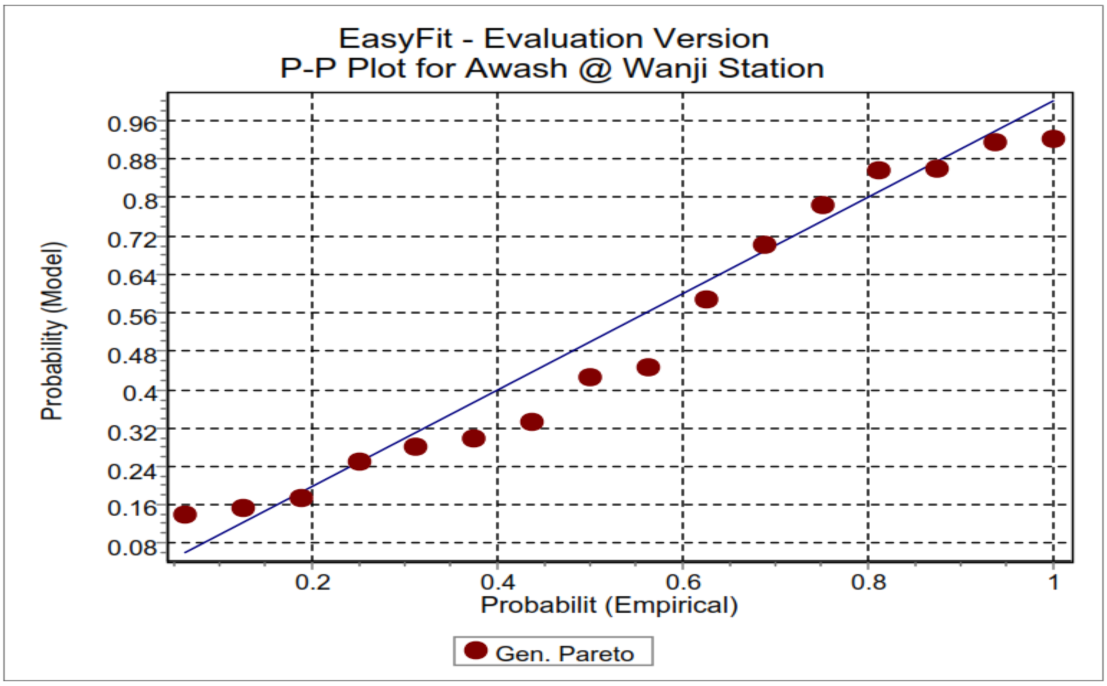
#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.17793	4	0.40479	1	1.3645	3
2	Gen. Pareto	0.21164	5	0.63977	6	0.05825	1
3	Log-Pearson 3	0.17186	2	0.40599	2	1.4063	4
4	Logistic	0.21881	6	0.50741	5	1.0062	2
5	Lognormal	0.17364	3	0.43669	4	1.4985	6
6	Lognormal (3P)	0.17089	1	0.43262	3	1.4719	5

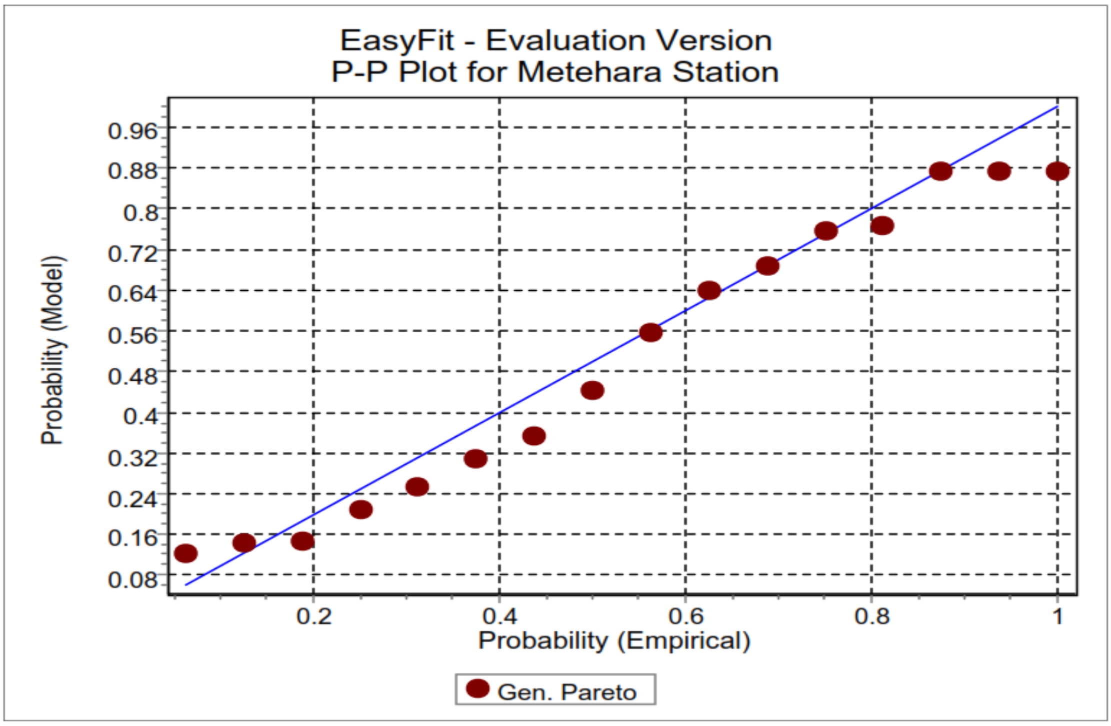
Goodness of Fit - Details [hide]

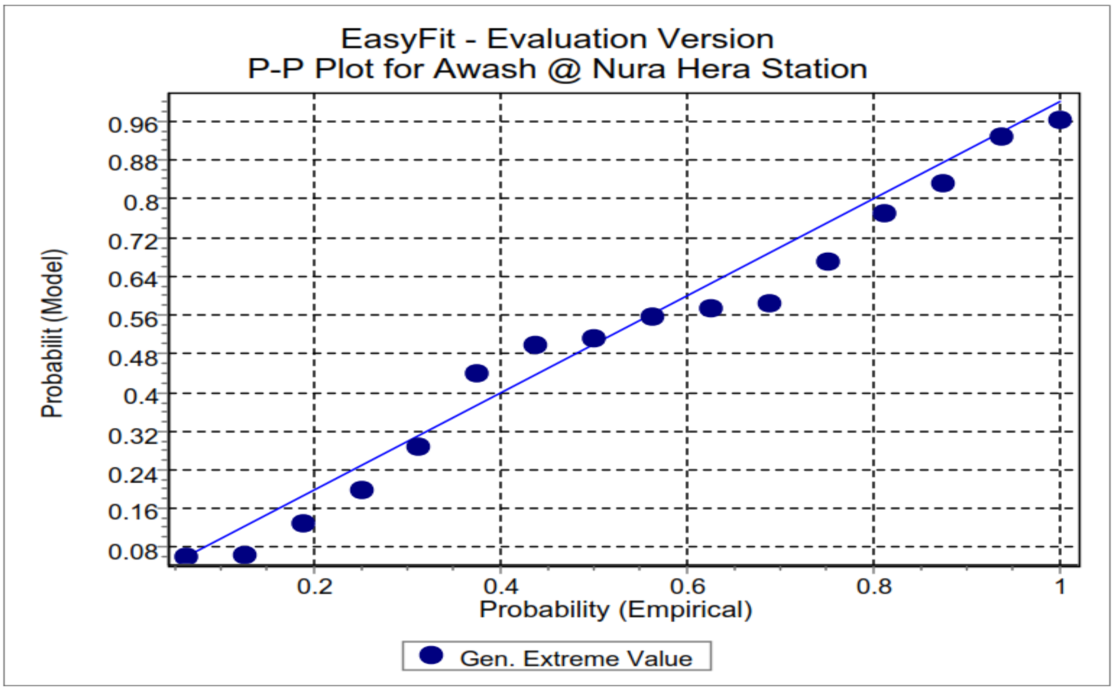
Gen. Extreme Value [#1]					
Kolmogorov-Smirnov					
Sample Size Statistic P-Value Rank	16 0.17793 0.62881 4				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.25778	0.29472	0.32733	0.36571	0.39201
Reject?	No	No	No	No	No
Anderson-Darling					
Sample Size Statistic Rank	16 0.40479 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No
Chi-Squared					
Deg. of freedom Statistic P-Value Rank	1 1.3645 0.24276 3				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.6424	2.7055	3.8415	5.4119	6.6349
Reject?	No	No	No	No	No

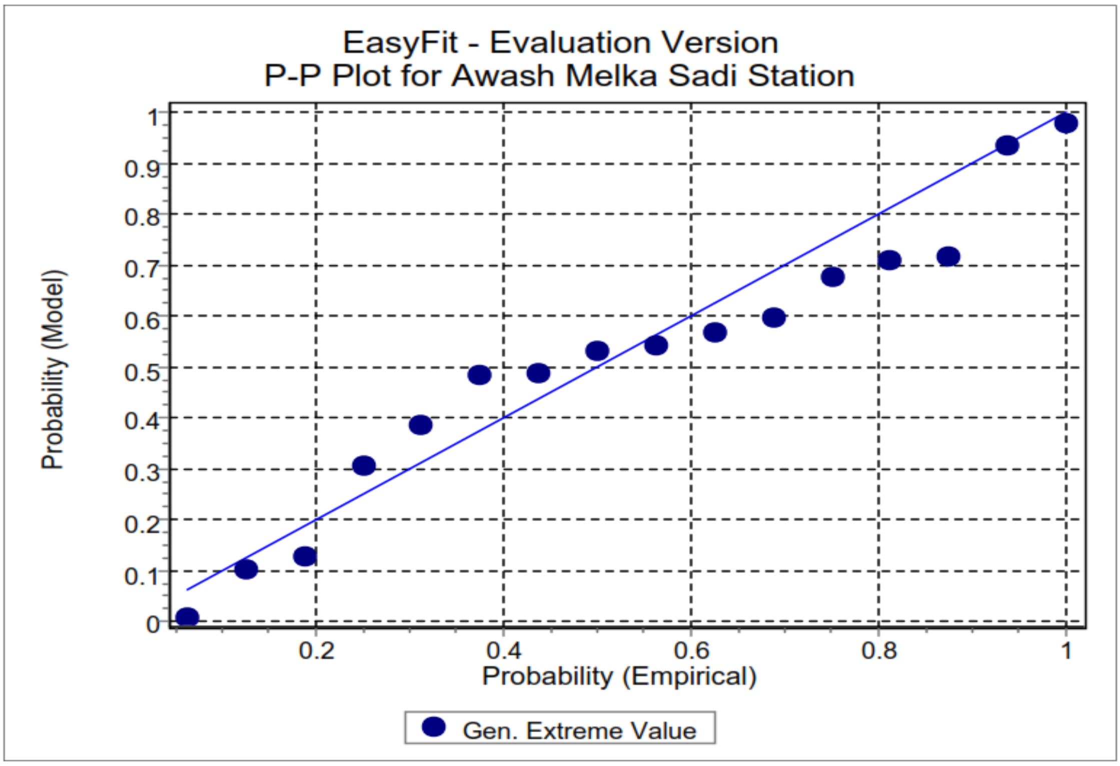
Appendix-B: Probability-Probability plots of stations with their best fit distribution

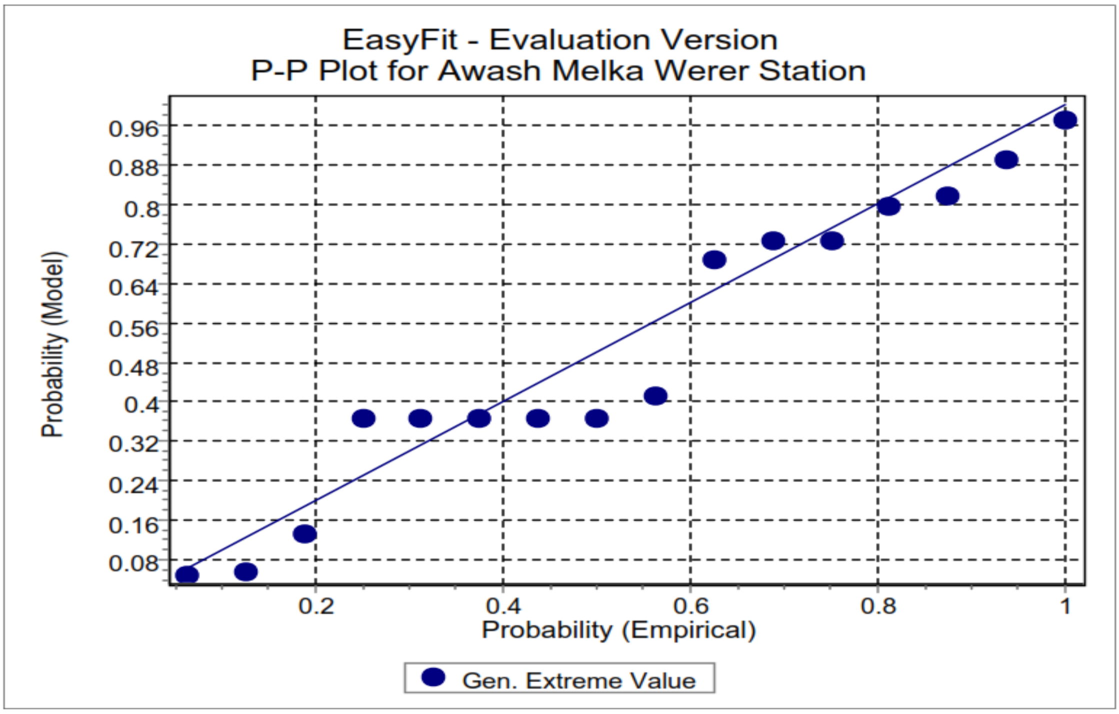




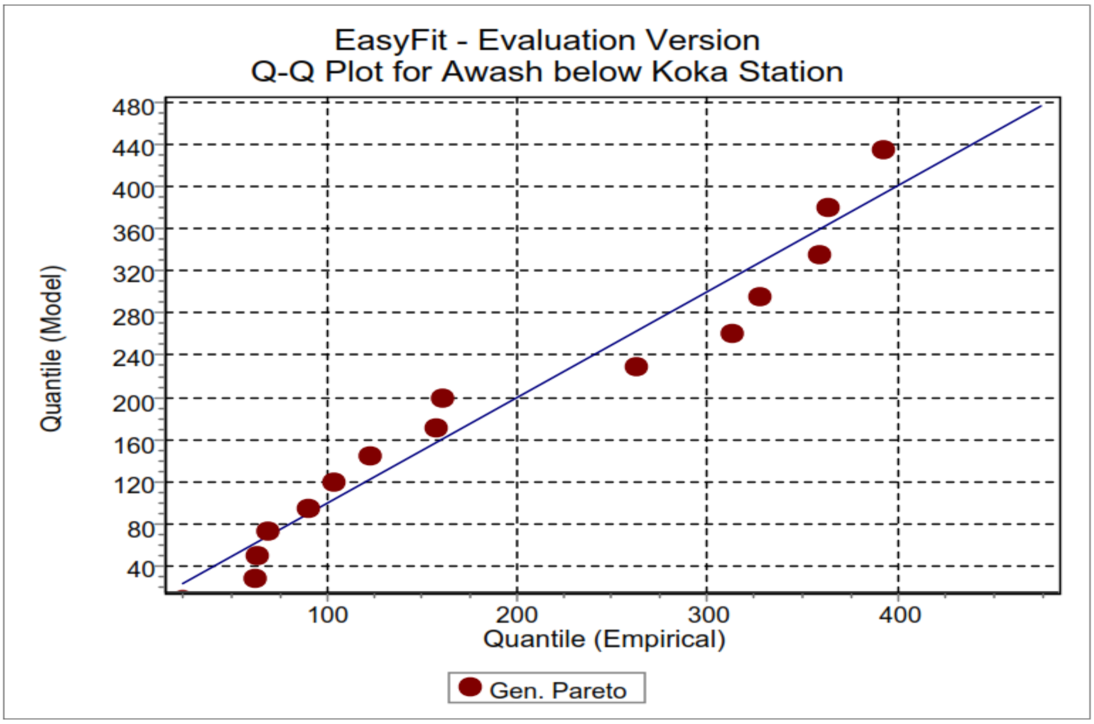

EasyFit - Evaluation Version
P-P Plot for Metehara Station
Probability (Model)
0.96
0.88
0.8
0.72
0.64
0.56
0.48
0.4
0.32
0.24
0.16
0.08
0.2
0.4
0.6
0.8
1
Probability (Empirical)
Gen. Pareto

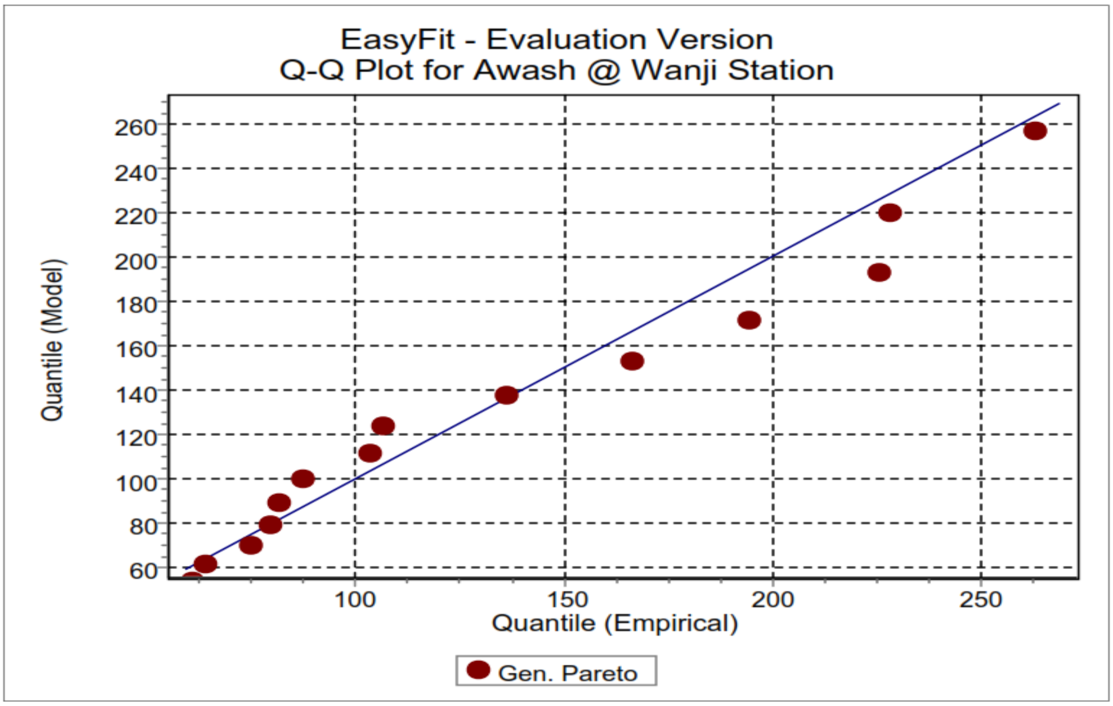

EasyFit - Evaluation Version
P-P Plot for Awash @ Nura Hera Station
Probabilit (Model)
0.96
0.88
0.8
0.72
0.64
0.56
0.48
0.4
0.32
0.24
0.16
0.08
0.2
0.4
0.6
0.8
1
Probability (Empirical)
Gen. Extreme Value

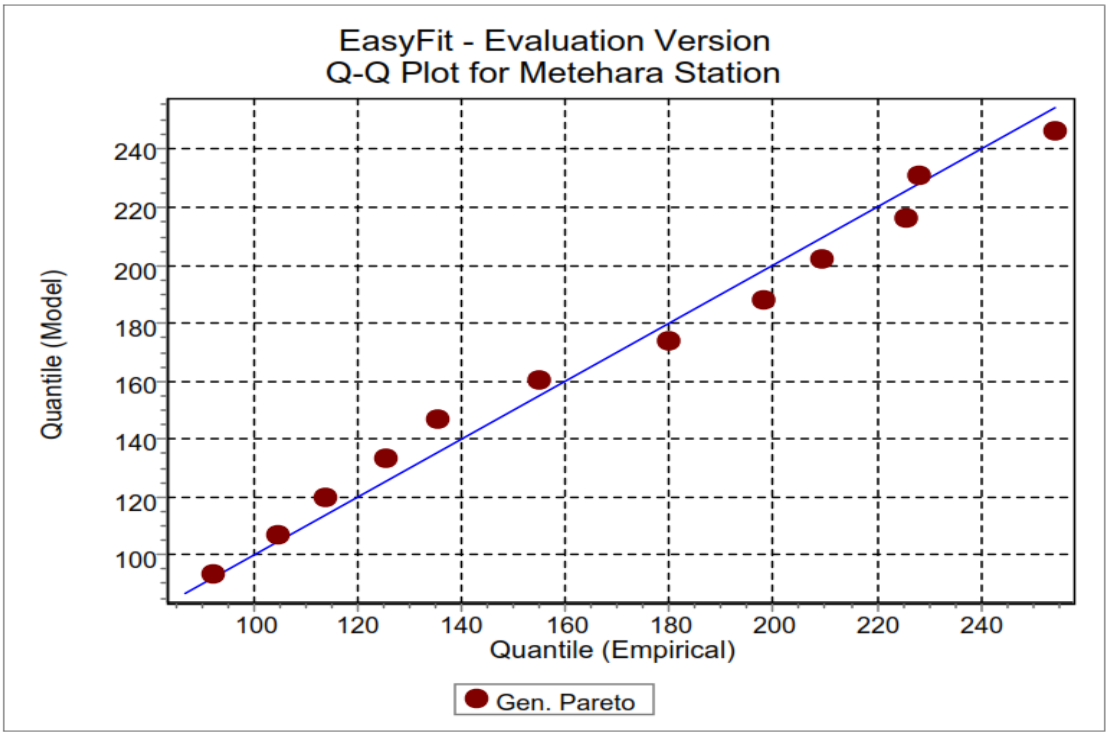

EasyFit - Evaluation Version
P-P Plot for Awash Melka Sadi Station
Probability (Model)
Probability (Empirical)
Gen. Extreme Value

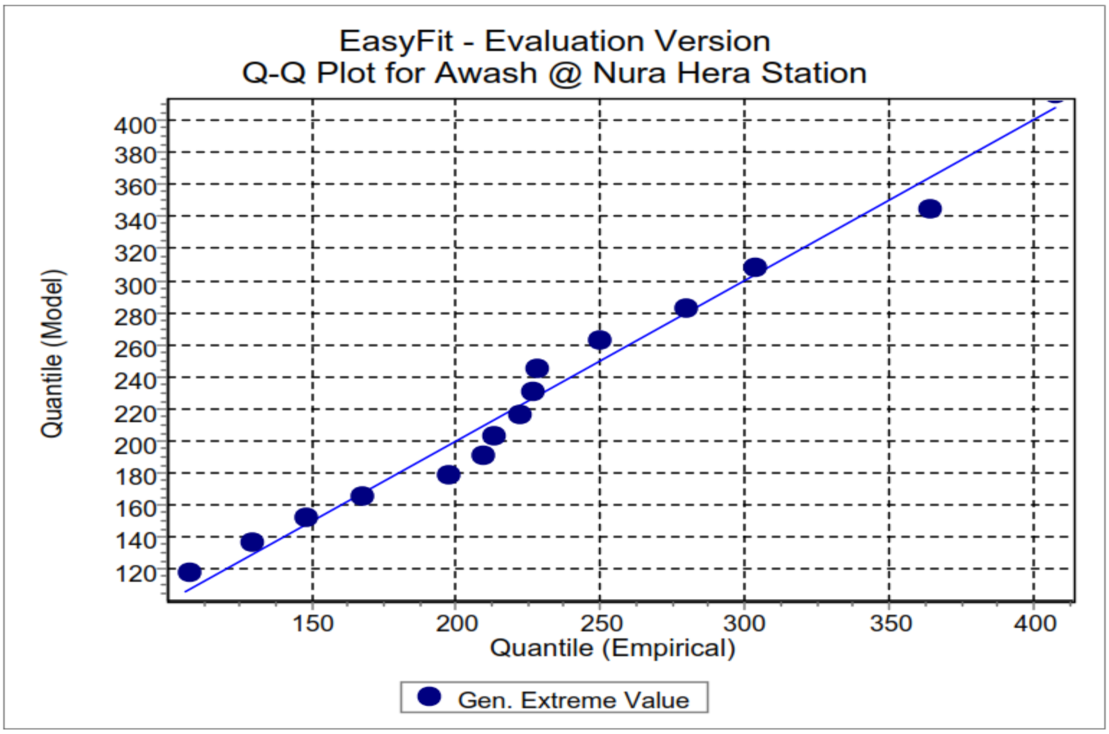

EasyFit - Evaluation Version
P-P Plot for Awash Melka Werer Station
Probability (Model)
Probability (Empirical)
Gen. Extreme Value

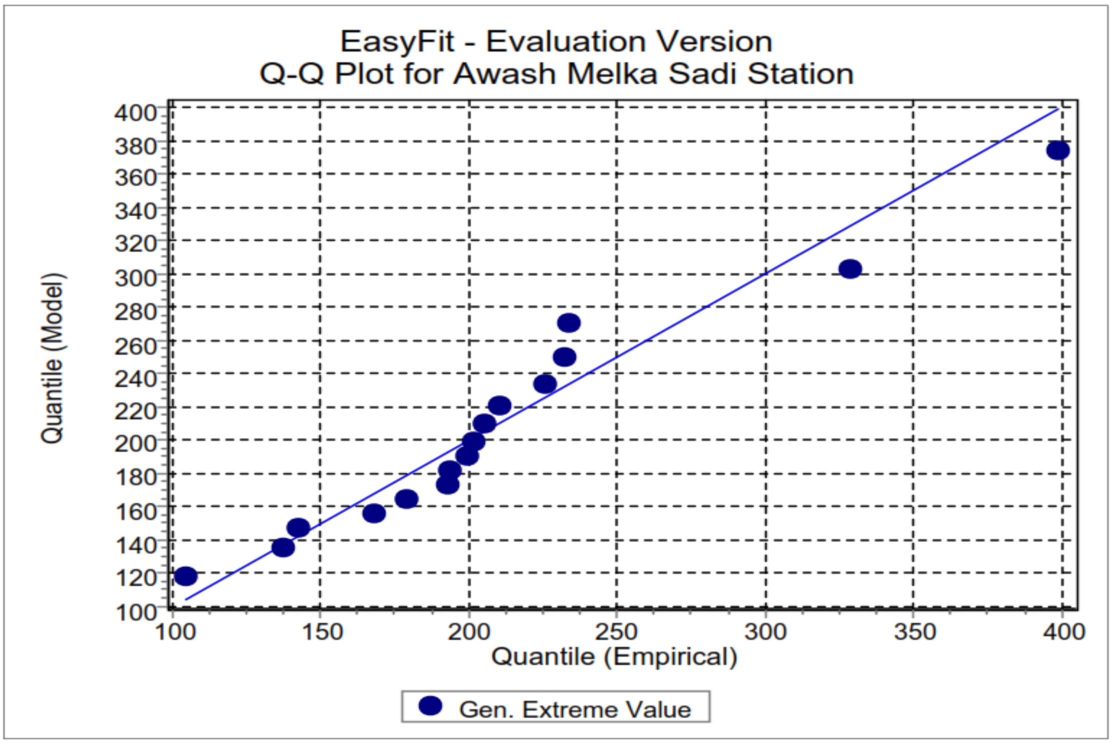
Appendix-C: Quantile-Quantile plots of stations with their best fit distribution

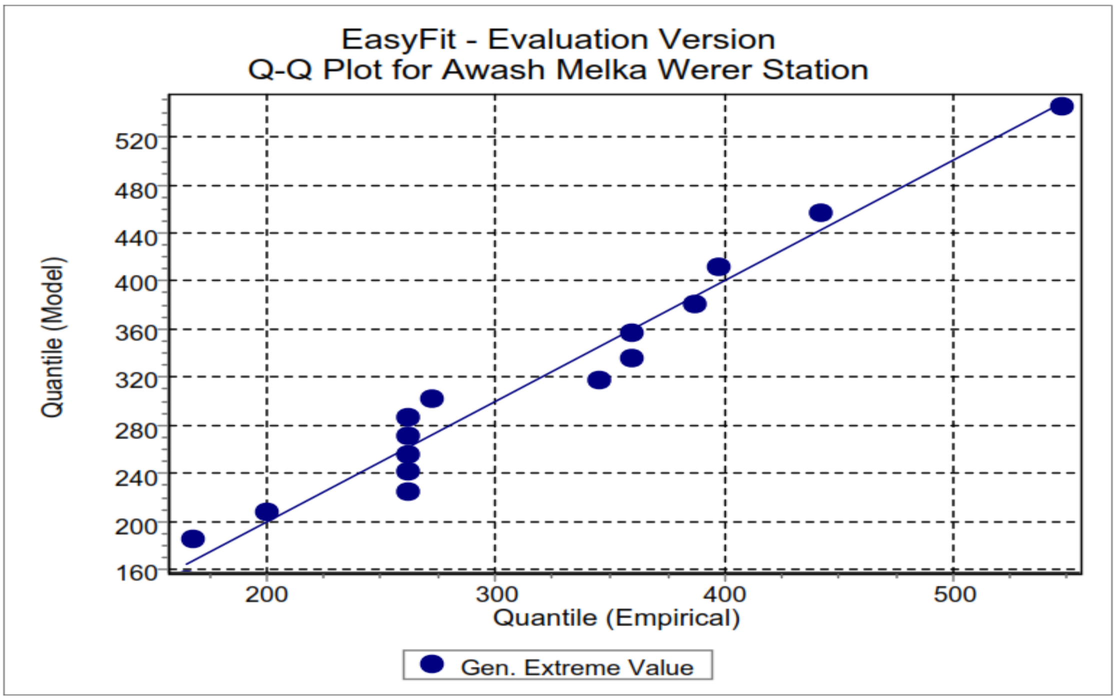




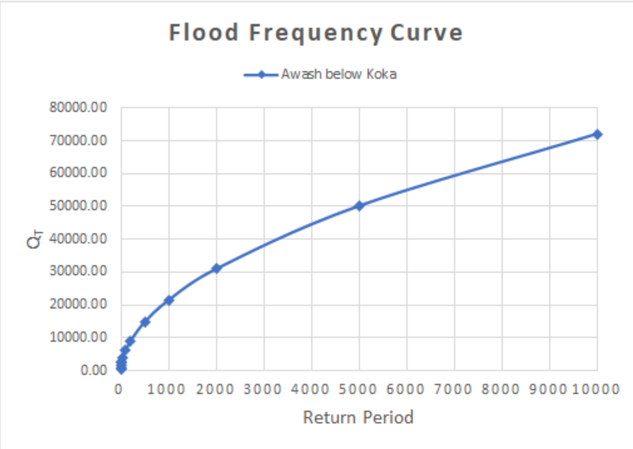

EasyFit - Evaluation Version
Q-Q Plot for Metehara Station
Quantile (Model)
240
220
200
180
160
140
120
100
100
120
140
160
180
200
220
240
Quantile (Empirical)
Gen. Pareto

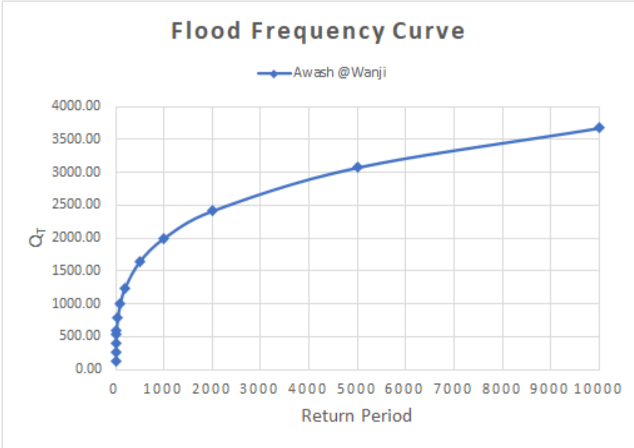

EasyFit - Evaluation Version
Q-Q Plot for Awash @ Nura Hera Station
Quantile (Model)
400
380
360
340
320
300
280
260
240
220
200
180
160
140
120
150
200
250
300
350
400
Quantile (Empirical)
Gen. Extreme Value

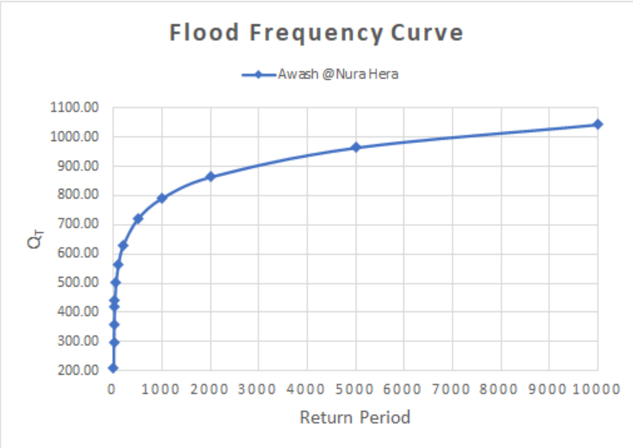

EasyFit - Evaluation Version
Q-Q Plot for Awash Melka Sadi Station
Quantile (Model)
Quantile (Empirical)
Gen. Extreme Value

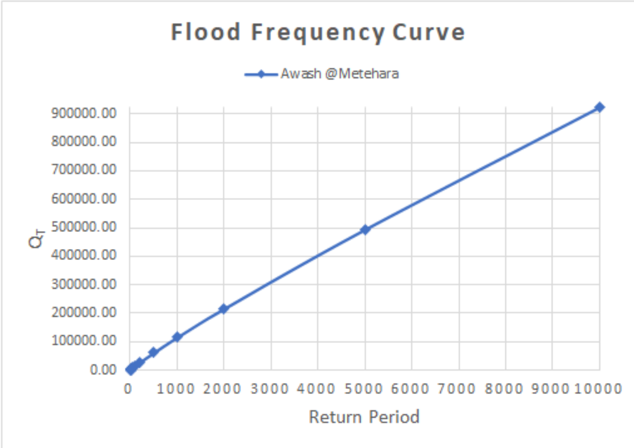

EasyFit - Evaluation Version
Q-Q Plot for Awash Melka Werer Station
Quantile (Model)
Quantile (Empirical)
Gen. Extreme Value

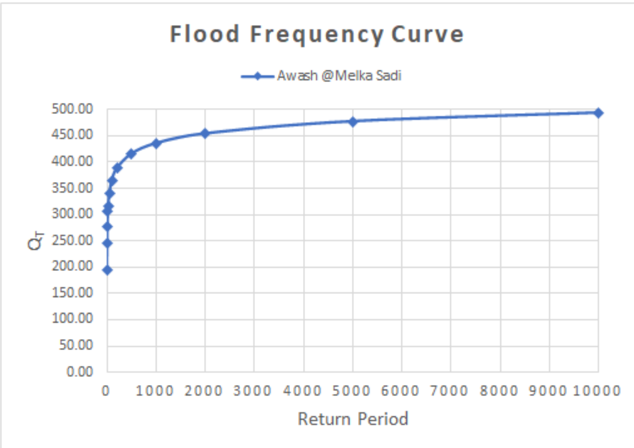
Appendix-D: Regional Flood frequency curves of the study regions

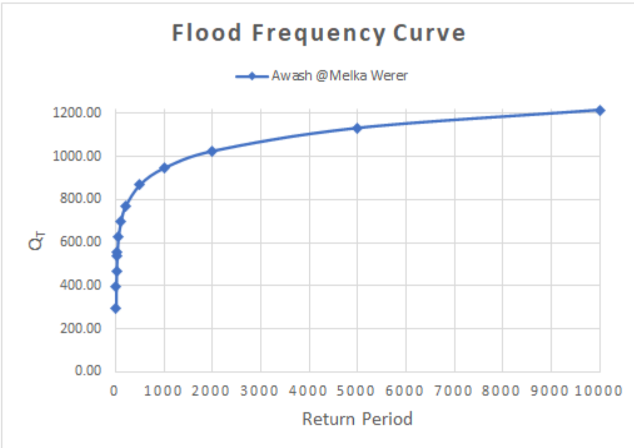




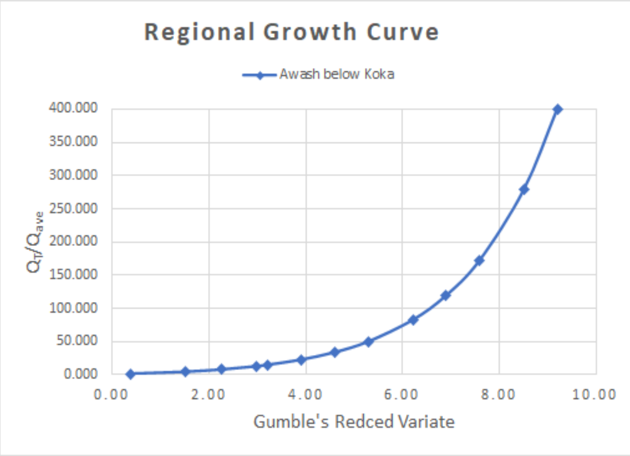


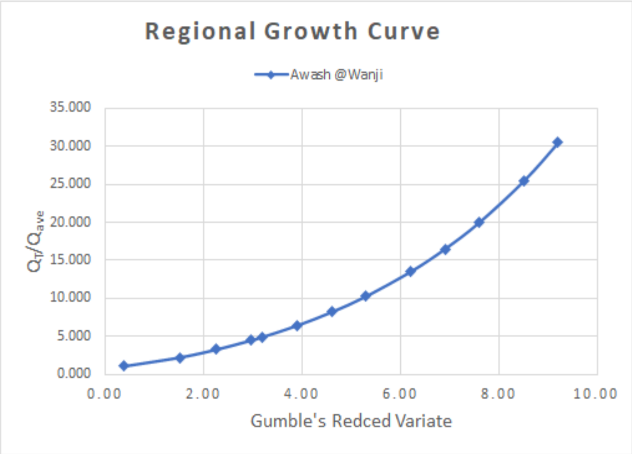


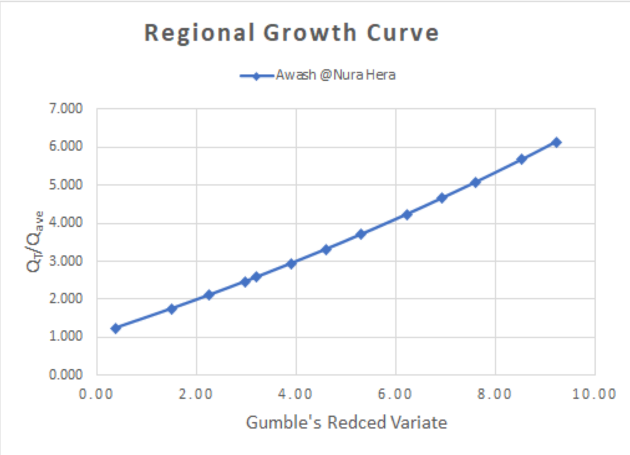


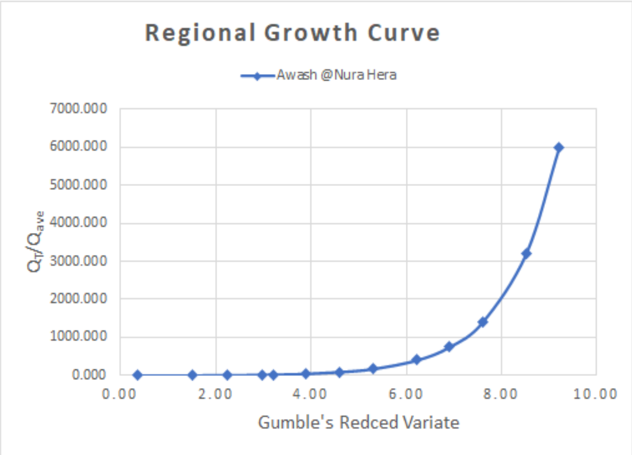


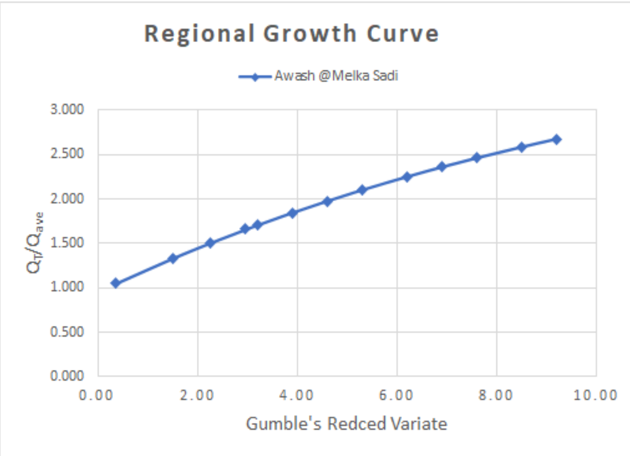
Appendix-E: Regional Growth curves of the study regions

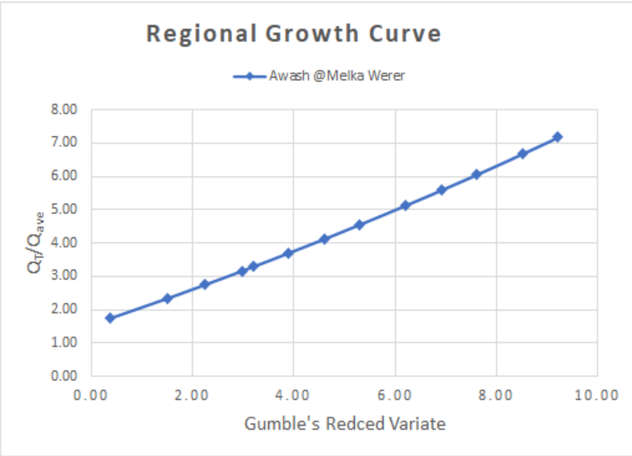




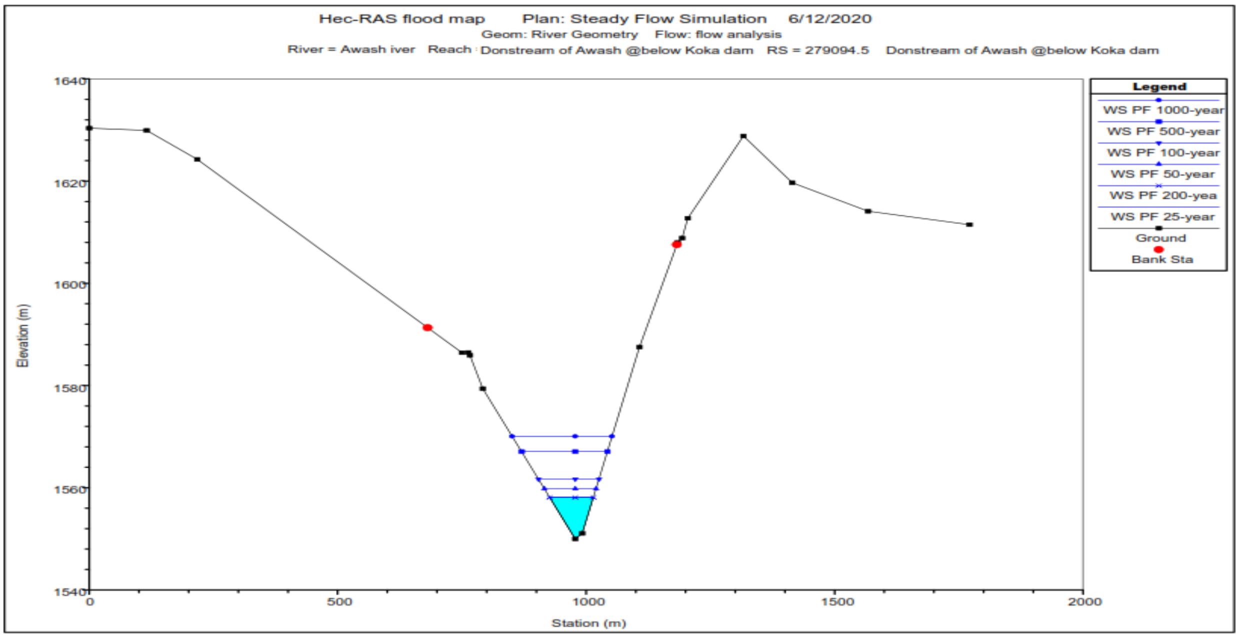




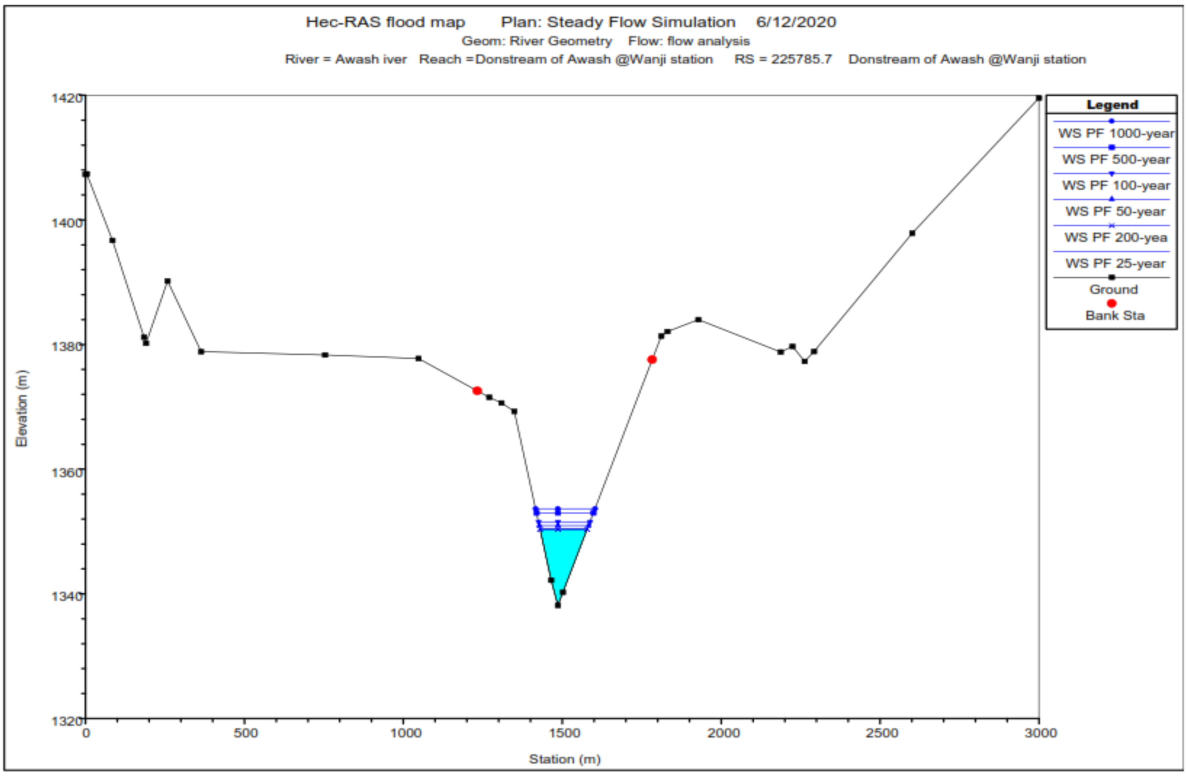




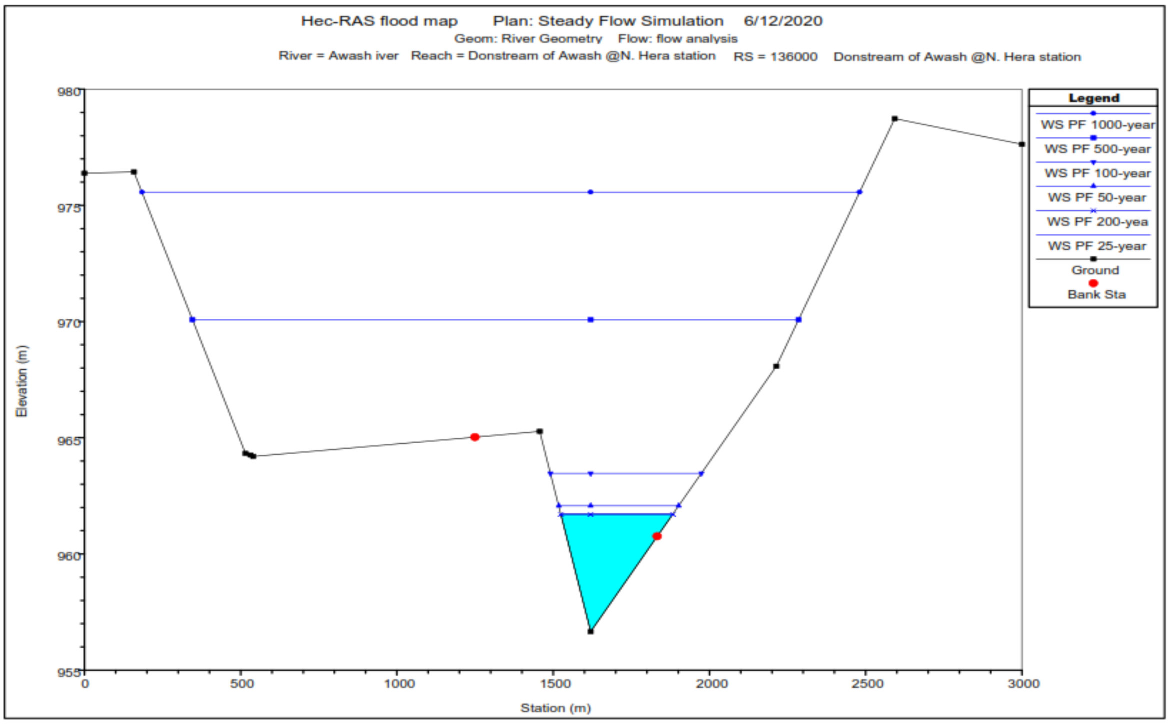
Appendix-F: Water surface elevations Awash river at some distance downstream of all stations



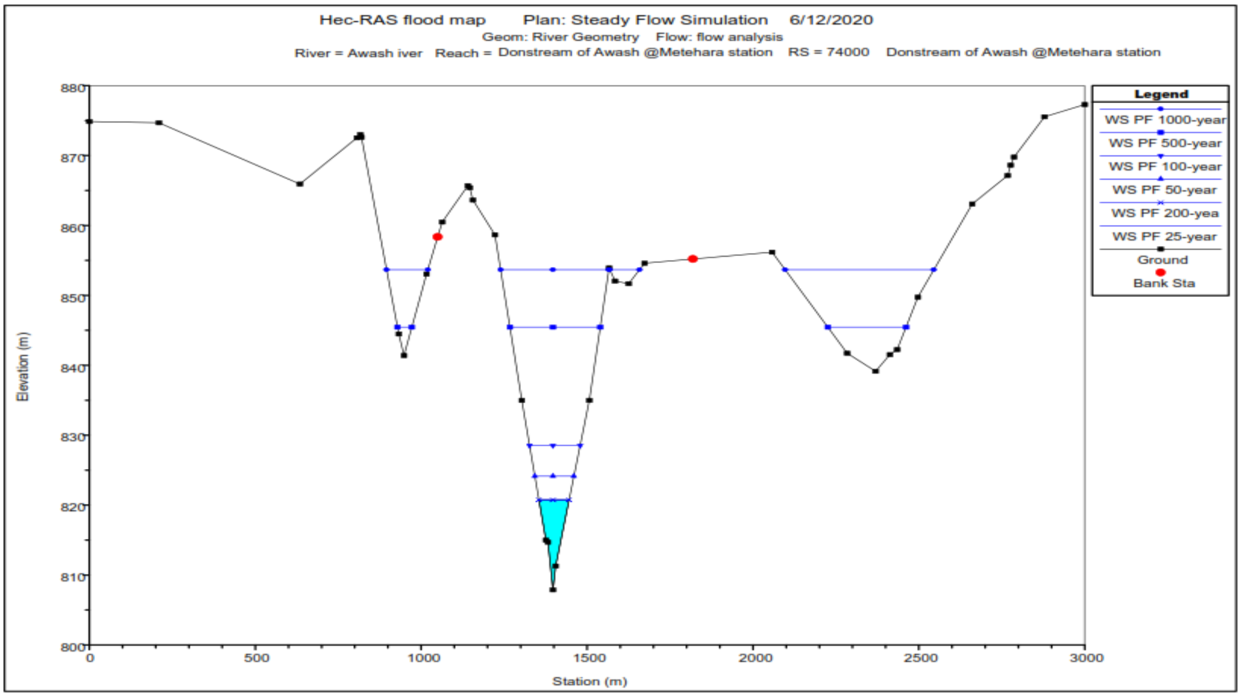
Water surface elevations at some distance downstream of Awash @below Koka station



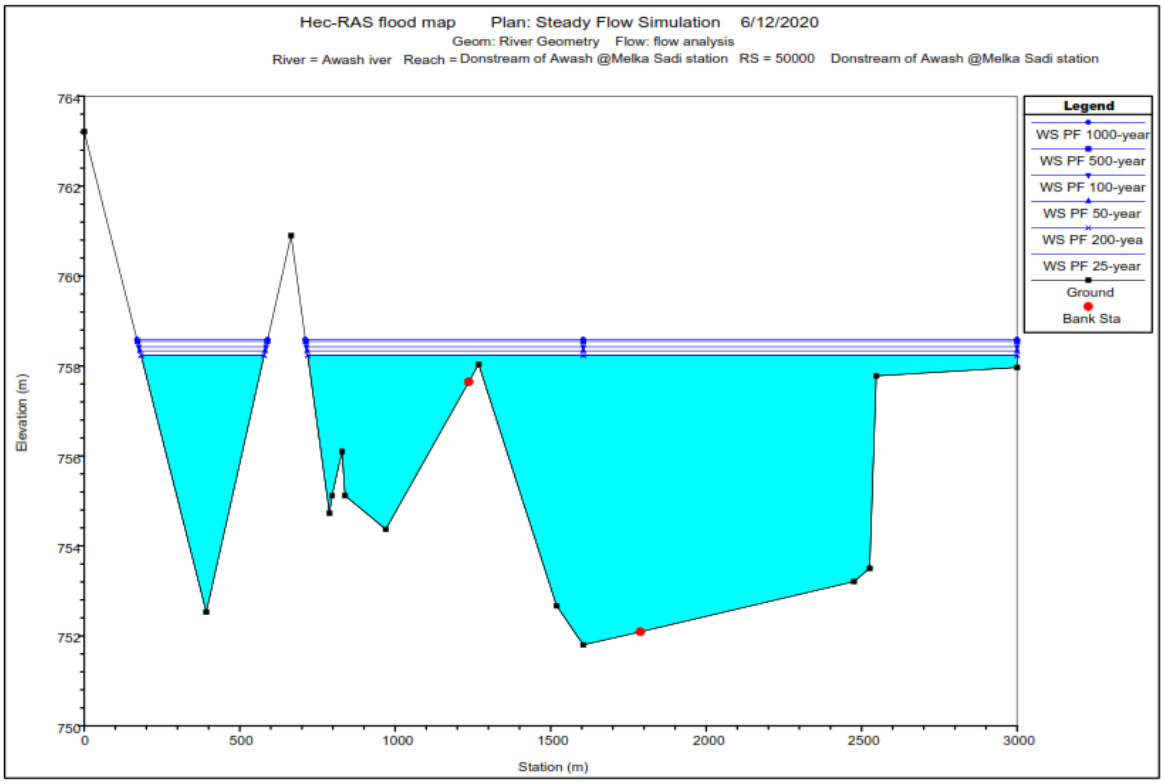
Water surface elevations at some distance downstream of Awash @Wanji station



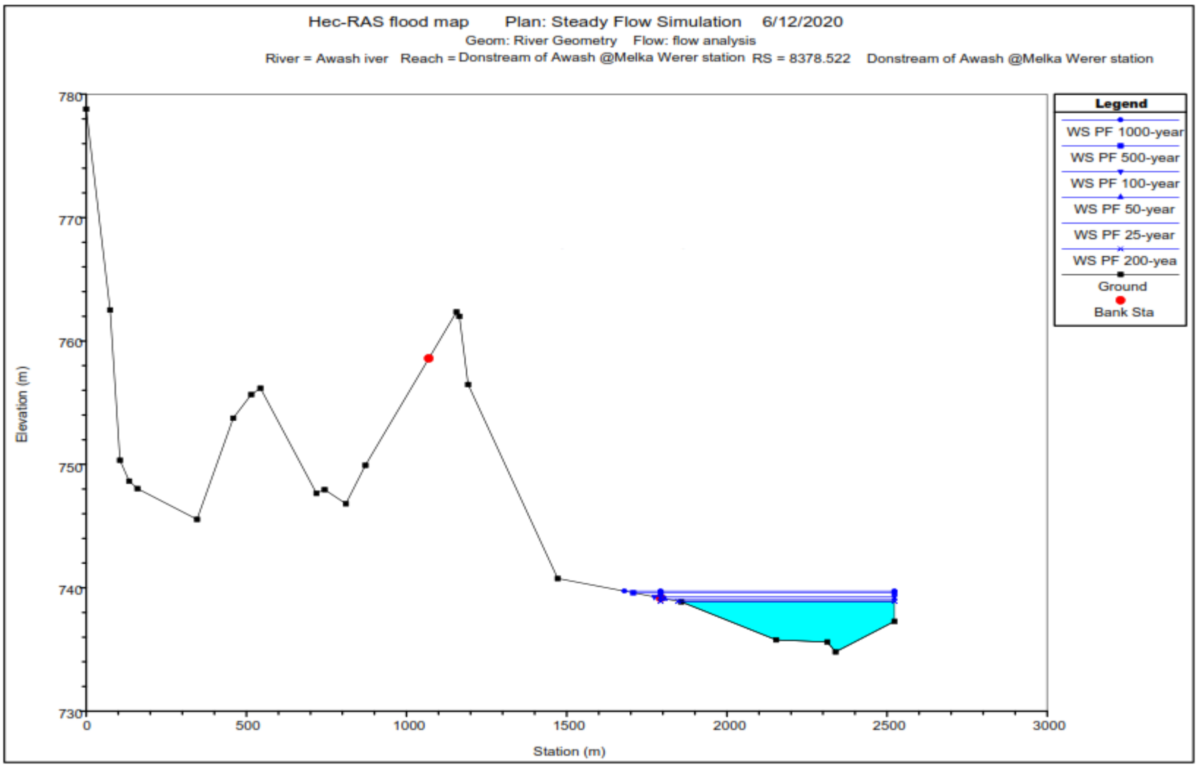
Water surface elevations at some distance downstream of Awash @N. Hera station



Water surface elevations at some distance downstream of Awash @Matahara station

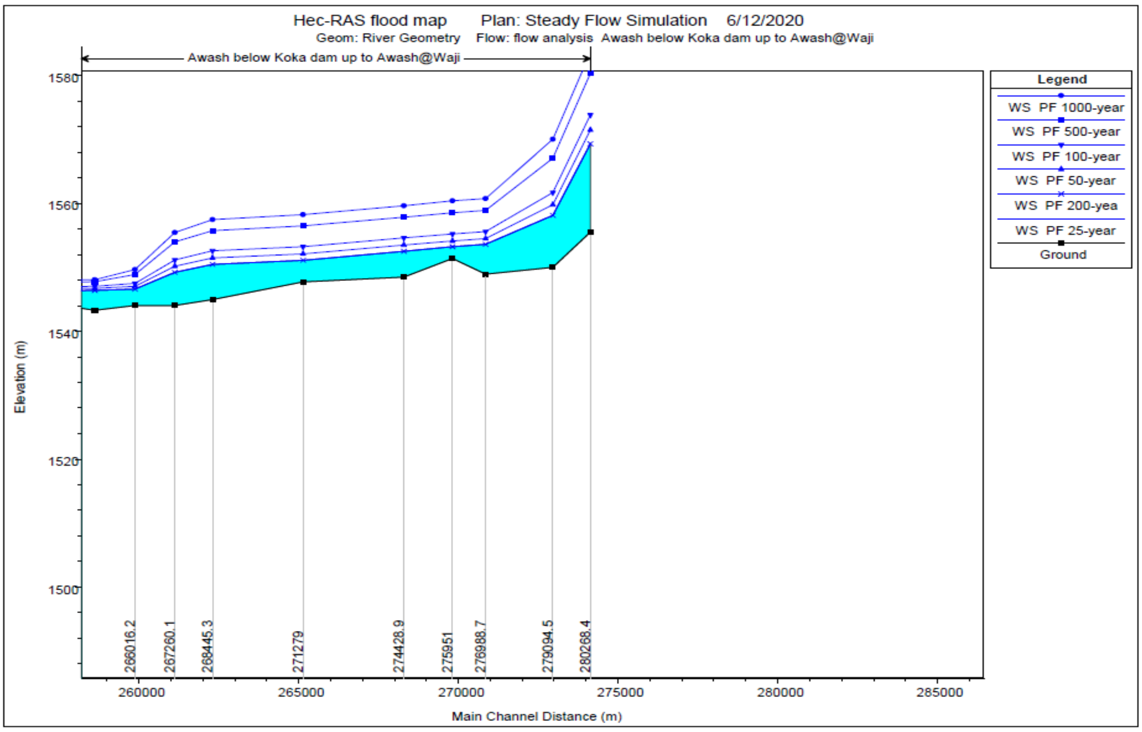


Water surface elevations at some distance downstream of Awash @Mela Sadi station

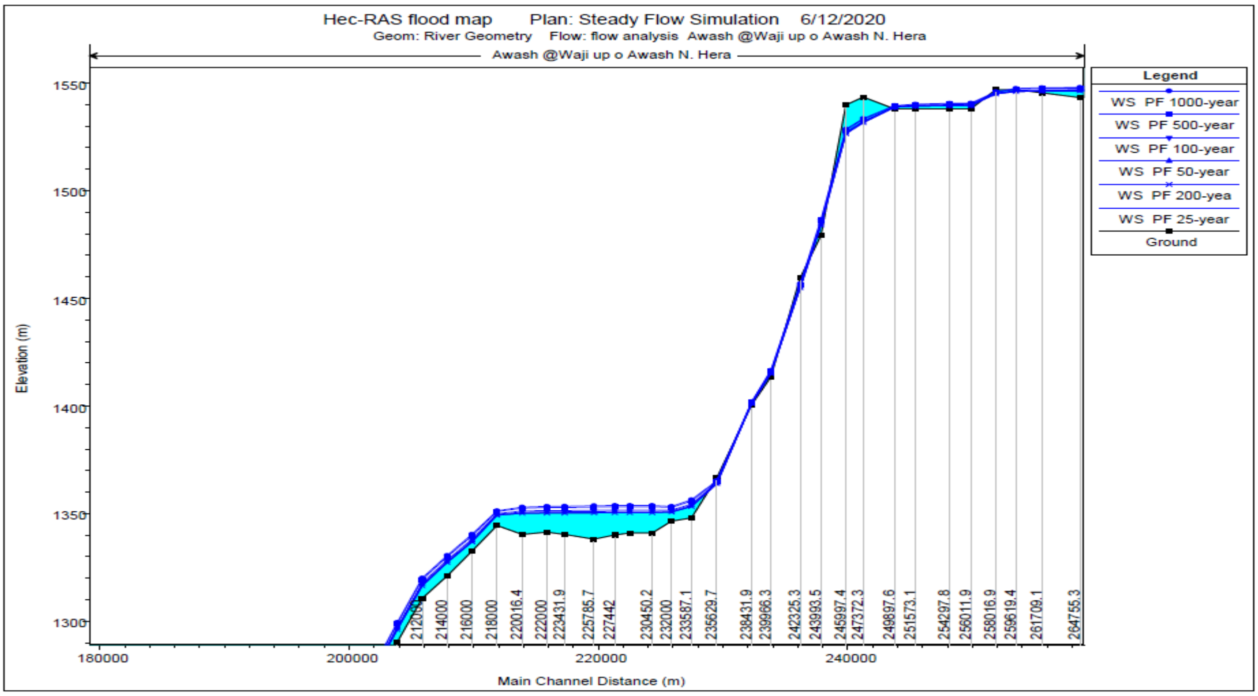


Water surface elevations at some distance downstream of Awash @Melka Werer station

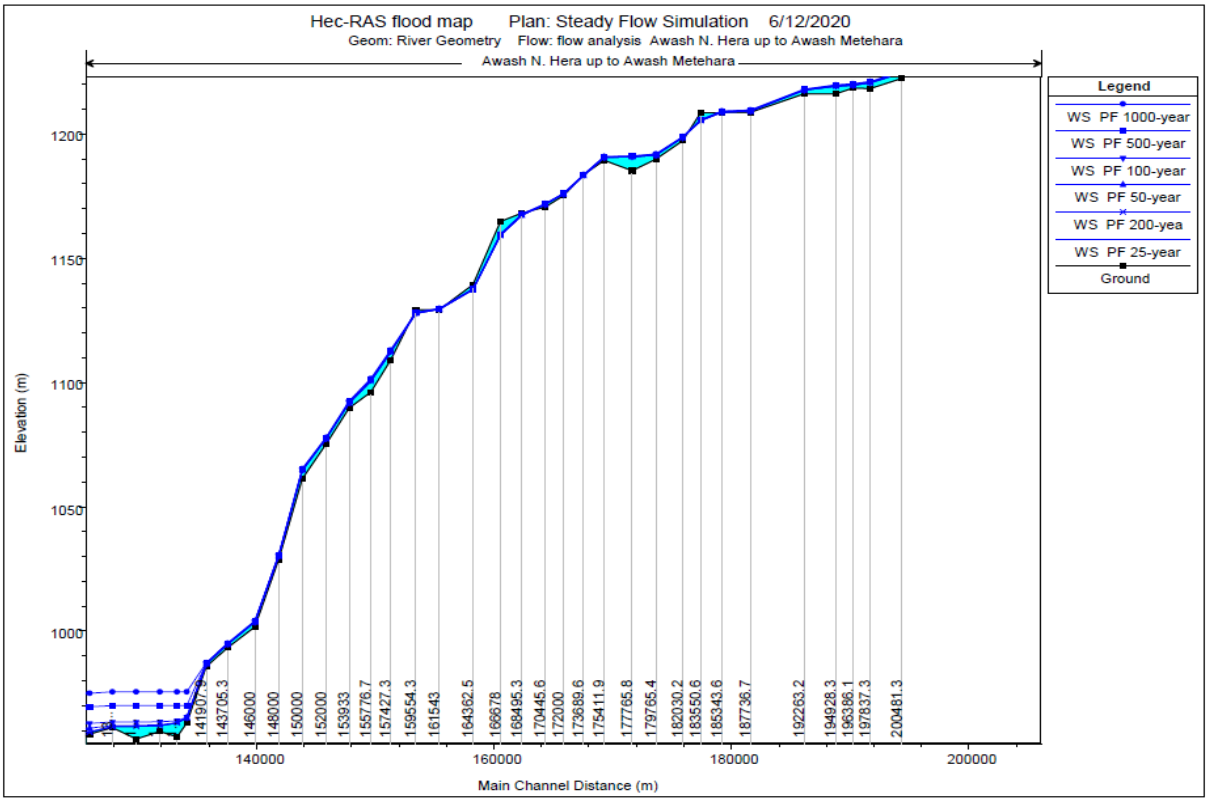
Appendix-G: Water surface profile of Awash rive reach



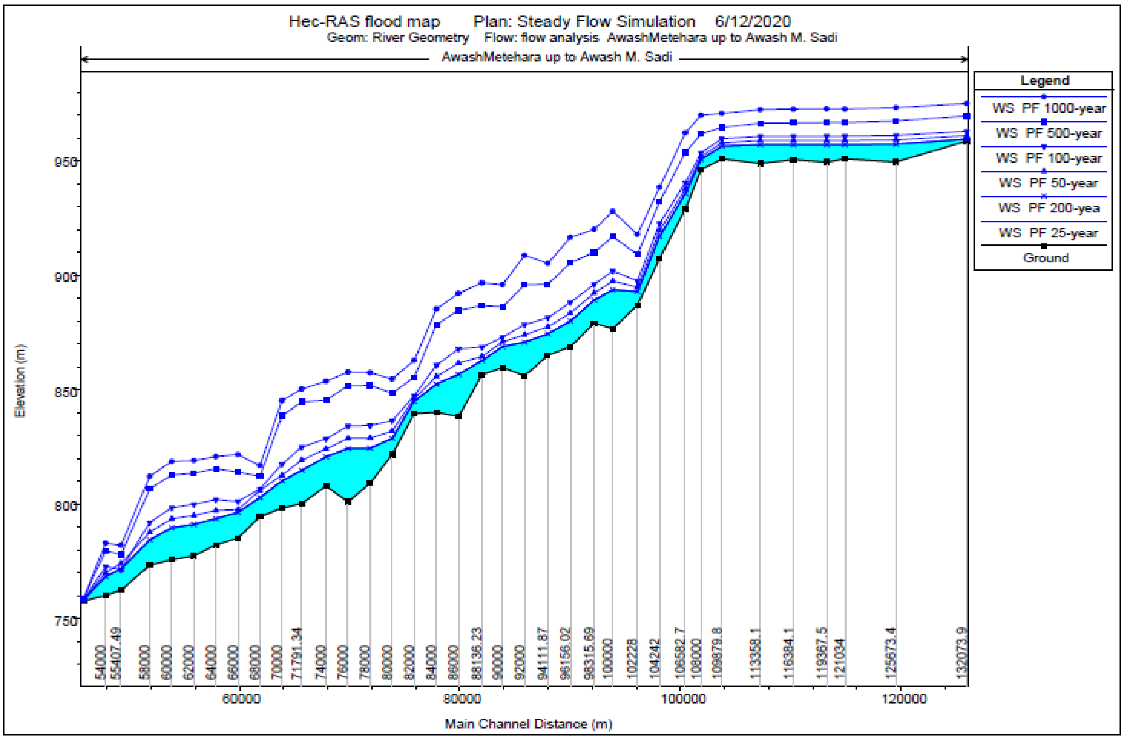
Water surface profile of Awash rive reach from Awash below Koka dam – Wanji



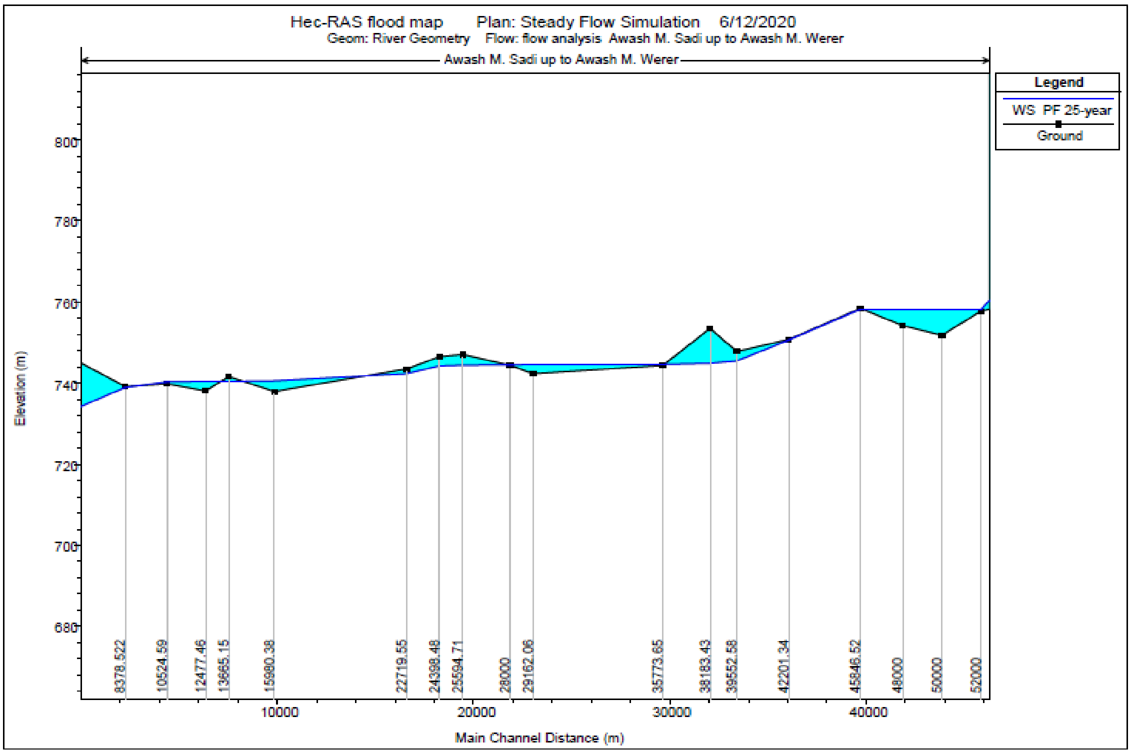
Water surface profile of Awash rive reach from Awash Wanji – Awash Nura Hera



Water surface profile of Awash rive reach from Awash Nura Hera – Awash Matahara

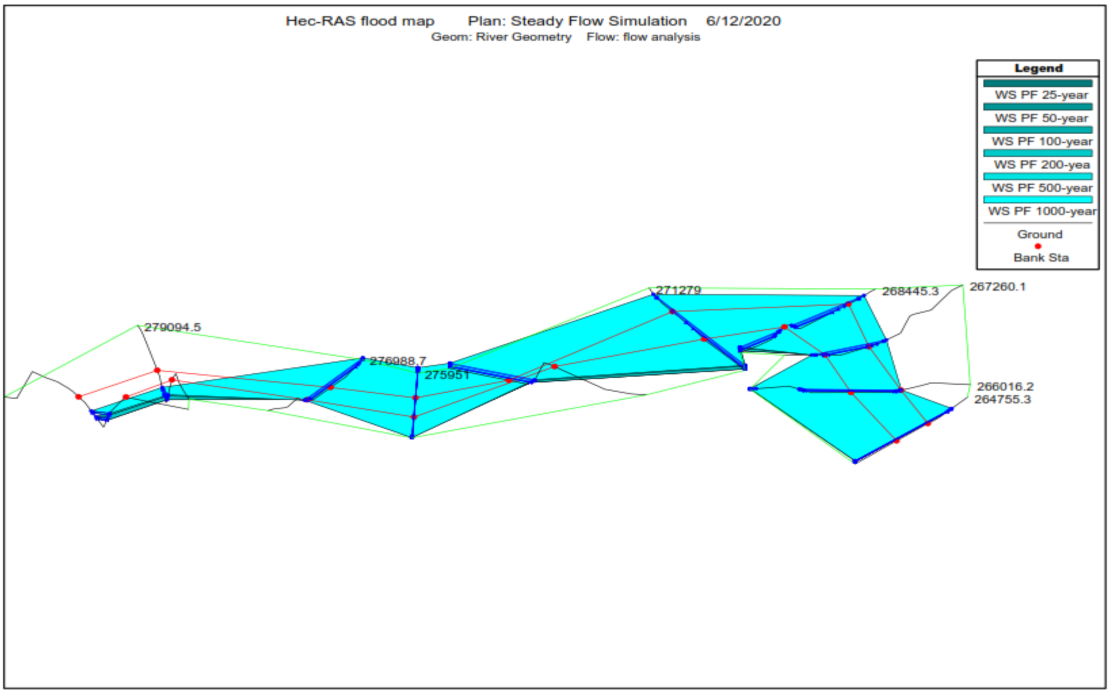


Water surface profile of Awash rive reach from Awash Matahara–Awash Melka Sadi

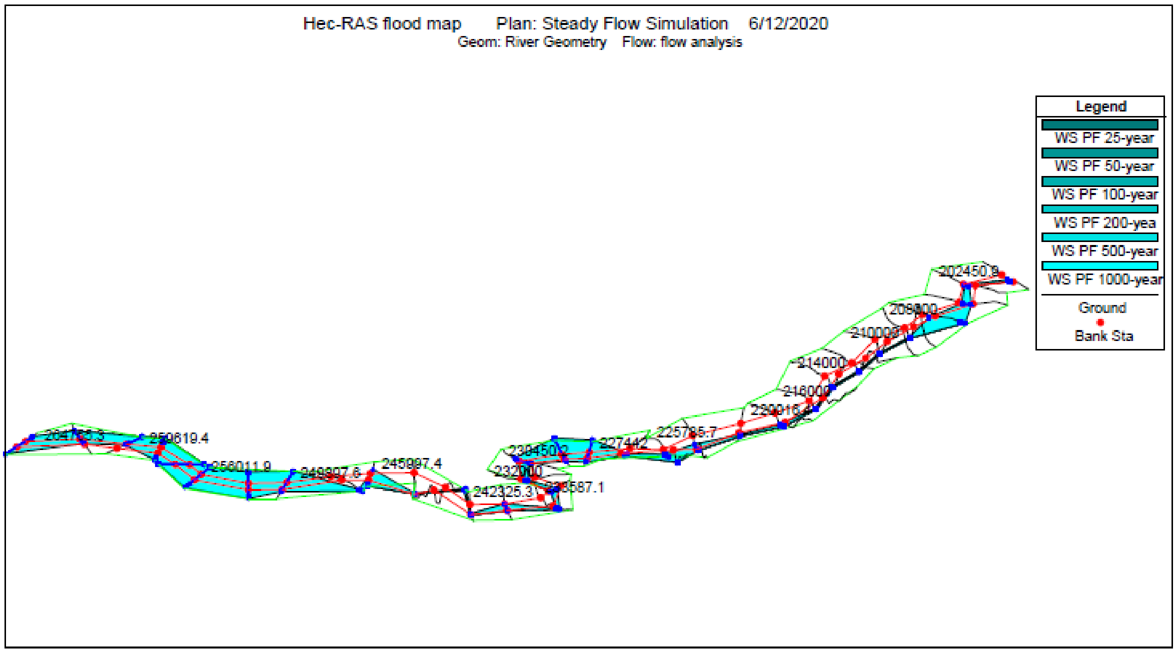


Water Surface profile of Awash rive reach from Awash Melka Sadi - Awash Melka Werer

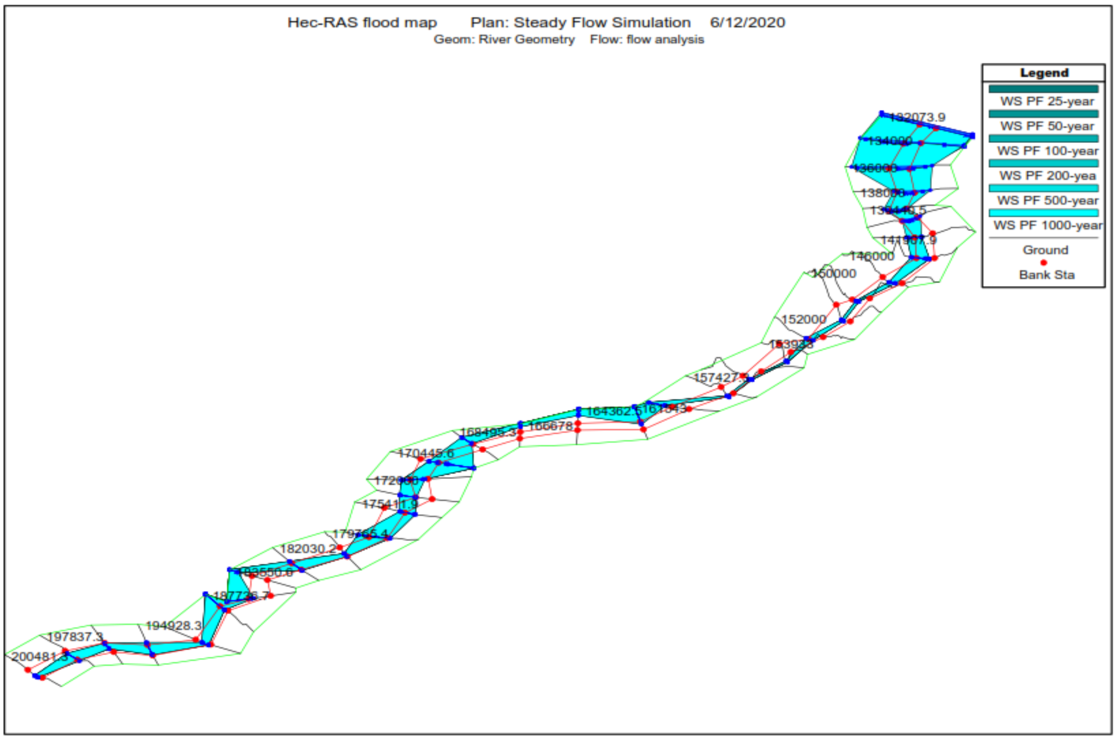
Appendix-H: 3-D view of water surface profile of Awash rive



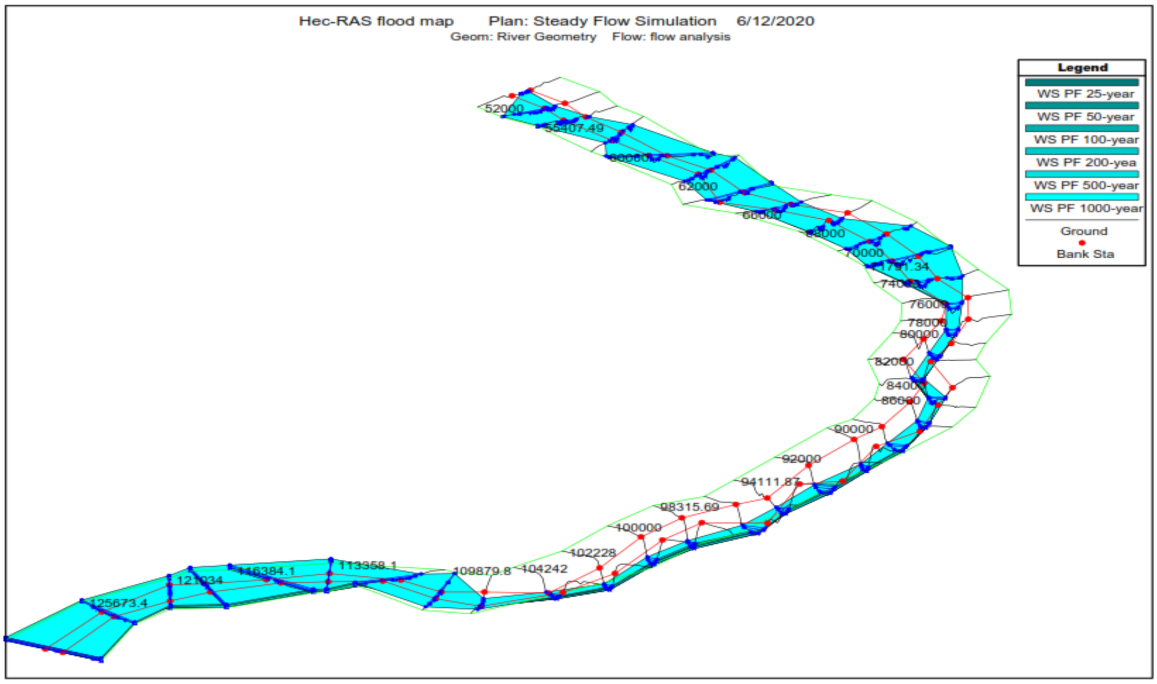
3-D view of water surface profile of Awash rive from Awash below Koka dam to Wanji



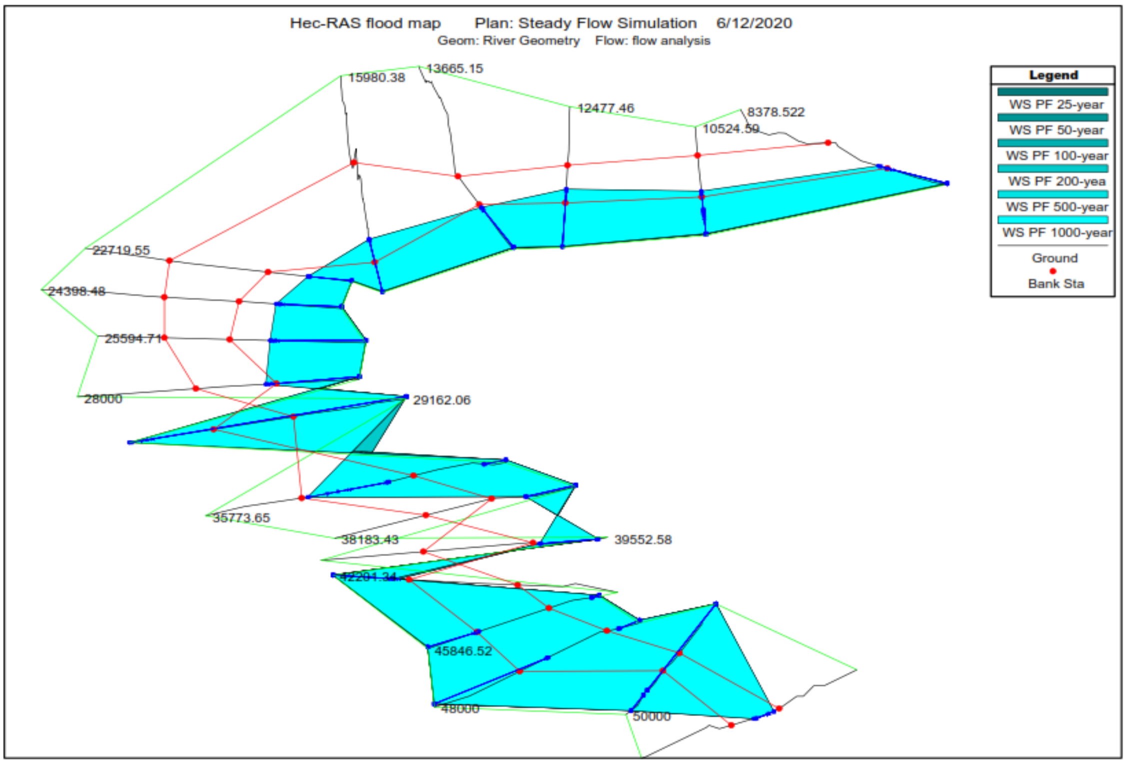
3-D view of water surface profile of Awash rive from Awash @Wanji to Nura Hera



3-D view of water surface profile of Awash rive from Awash @Nura Hera to Matahara



3-D view of water surface profile of Awash rive from Awash @Matahara to Melka Sadi



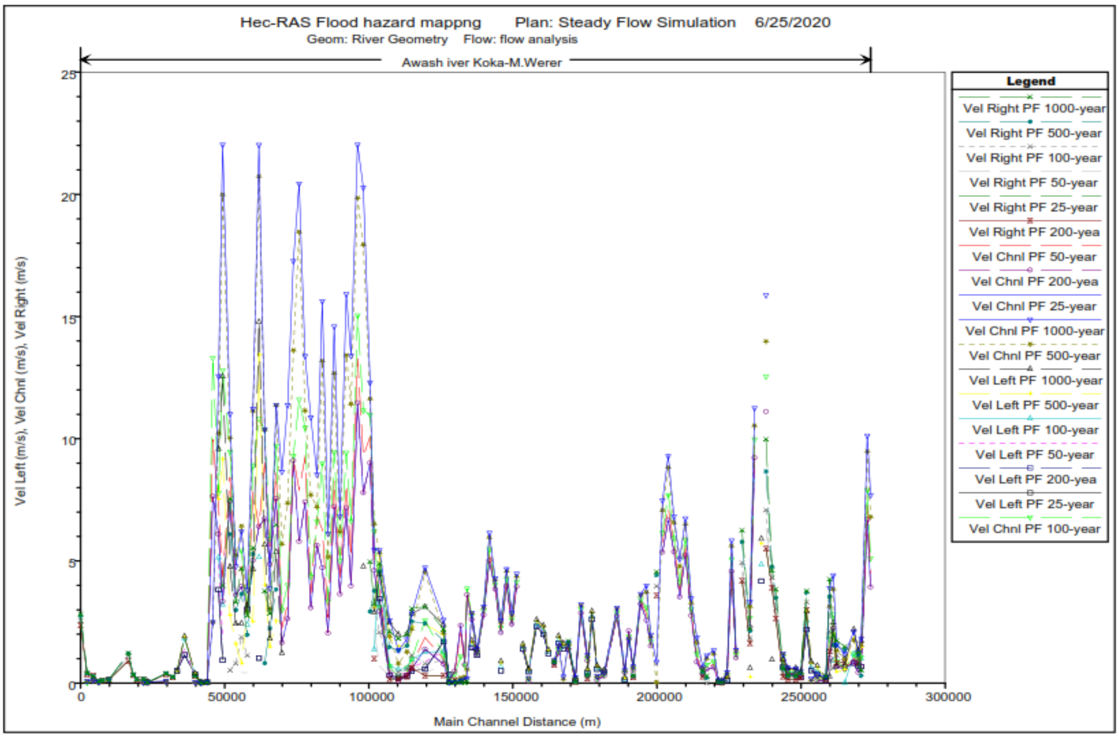
3-D view of water surface profile of Awash rive from Awash @Melka Sadi to Melka Werer

Appendix-I: Sample of profile summery table of HEC-RAS model output of the river reach for different return period

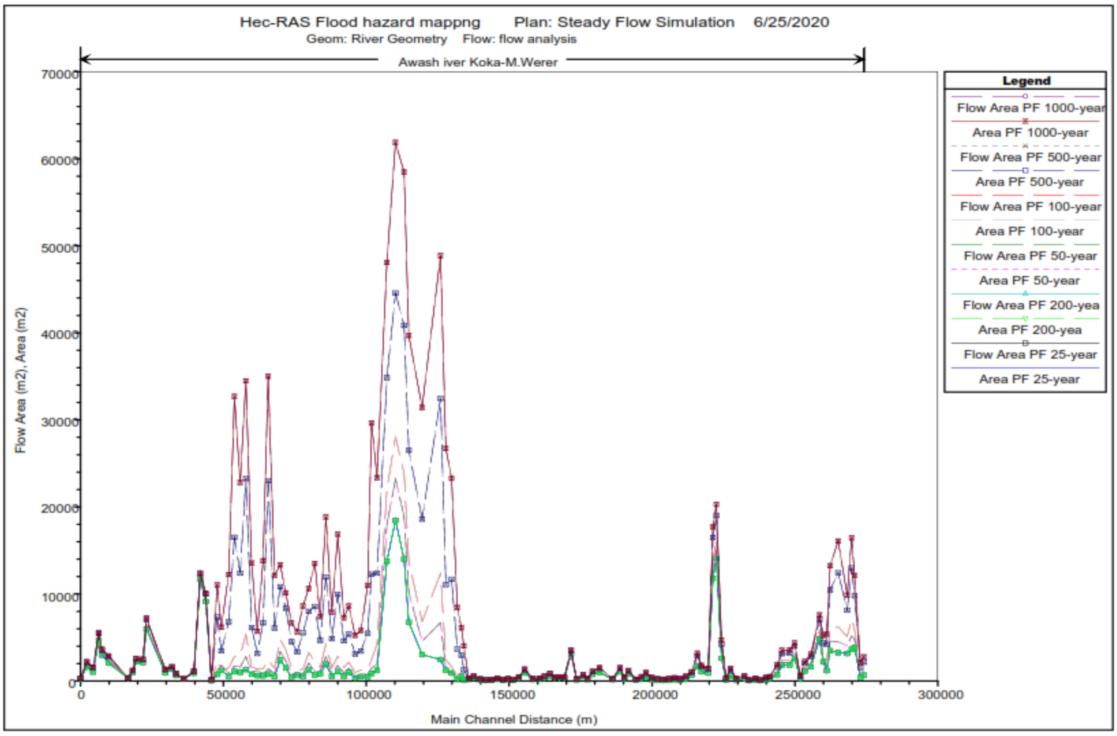
HEC-RAS Plan: T=25,50,100,200,500&1000 yrs River: Awash iver Reach: Koka-M.Werer

Reach	River Sta	Profile	Q Total	Min Ch El	W.S. Elev	Crit W.S.	E.G. Elev	E.G. Slope	Vel Chnl	Flow Area	Top Width	Froude # Chl
			(m3/s)	(m)	(m)	(m)	(m)	(m/m)	(m/s)	(m2)	(m)	
Koka-M.Werer	280268.4	PF 25-year	2648.84	1555.51	1569.31	1565.76	1570.10	0.004266	3.94	673.13	97.57	0.48
Koka-M.Werer	280268.4	PF 50-year	4058.98	1555.51	1571.49	1567.69	1572.52	0.004578	4.50	902.93	113.00	0.51
Koka-M.Werer	280268.4	PF 100-year	6077.64	1555.51	1573.85	1569.83	1575.18	0.004924	5.11	1189.22	129.68	0.54
Koka-M.Werer	280268.4	PF 200-yea	2648.84	1555.51	1569.31	1565.76	1570.10	0.004266	3.94	673.13	97.57	0.48
Koka-M.Werer	280268.4	PF 500-year	14784.40	1555.51	1580.33	1575.98	1582.68	0.005809	6.79	2177.20	175.47	0.62
Koka-M.Werer	280268.4	PF 1000-year	21431.53	1555.51	1583.58	1579.26	1586.60	0.006329	7.69	2785.28	198.46	0.66
Koka-M.Werer	279094.5	PF 25-year	2648.84	1550.00	1558.13	1558.13	1560.41	0.021134	6.69	396.16	87.99	1.01
Koka-M.Werer	279094.5	PF 50-year	4058.98	1550.00	1559.80	1559.80	1562.52	0.019988	7.30	556.14	103.81	1.01
Koka-M.Werer	279094.5	PF 100-year	6077.64	1550.00	1561.69	1561.69	1564.88	0.018819	7.90	769.06	121.71	1.00
Koka-M.Werer	279094.5	PF 200-yea	2648.84	1550.00	1558.13	1558.13	1560.41	0.021134	6.69	396.16	87.99	1.01
Koka-M.Werer	279094.5	PF 500-year	14784.40	1550.00	1567.07	1567.07	1571.64	0.016779	9.47	1561.48	172.74	1.01
Koka-M.Werer	279094.5	PF 1000-year	21431.53	1550.00	1570.05	1570.05	1575.27	0.015647	10.12	2116.89	200.92	1.00
Koka-M.Werer	276988.7	PF 25-year	2648.84	1548.93	1553.60	1550.08	1553.62	0.000258	0.69	3873.11	970.73	0.11
Koka-M.Werer	276988.7	PF 50-year	4058.98	1548.93	1554.50	1550.43	1554.54	0.000323	0.87	4780.40	1026.07	0.13
Koka-M.Werer	276988.7	PF 100-year	6077.64	1548.93	1555.60	1550.87	1555.65	0.000382	1.06	5936.63	1092.53	0.14
Koka-M.Werer	276988.7	PF 200-yea	2648.84	1548.93	1553.60	1550.08	1553.62	0.000258	0.69	3873.11	970.73	0.11
Koka-M.Werer	276988.7	PF 500-year	14784.40	1548.93	1558.89	1552.35	1559.01	0.000483	1.56	9819.99	1227.72	0.17
Koka-M.Werer	276988.7	PF 1000-year	21431.53	1548.93	1560.76	1553.25	1560.92	0.000511	1.83	12137.08	1261.91	0.18
Koka-M.Werer	275951	PF 25-year	2648.84	1551.39	1553.22		1553.25	0.000584	0.44	3656.69	1698.24	0.13
Koka-M.Werer	275951	PF 50-year	4058.98	1551.39	1554.13		1554.16	0.000460	0.58	5211.42	1718.43	0.13
Koka-M.Werer	275951	PF 100-year	6077.64	1551.39	1555.23		1555.27	0.000384	0.71	7115.36	1742.84	0.13
Koka-M.Werer	275951	PF 200-yea	2648.84	1551.39	1553.22		1553.25	0.000584	0.44	3656.69	1698.24	0.13
Koka-M.Werer	275951	PF 500-year	14784.40	1551.39	1558.54		1558.61	0.000329	1.07	13009.40	1816.34	0.13
Koka-M.Werer	275951	PF 1000-year	21431.53	1551.39	1560.41		1560.50	0.000326	1.26	16445.26	1849.78	0.14
Koka-M.Werer	274428.9	PF 25-year	2648.84	1548.47	1552.52		1552.55	0.000461	0.86	3175.91	907.86	0.15
Koka-M.Werer	274428.9	PF 50-year	4058.98	1548.47	1553.48		1553.53	0.000487	1.02	4056.49	917.74	0.15
Koka-M.Werer	274428.9	PF 100-year	6077.64	1548.47	1554.61		1554.68	0.000520	1.19	5096.38	929.27	0.16
Koka-M.Werer	274428.9	PF 200-yea	2648.84	1548.47	1552.52		1552.55	0.000461	0.86	3175.91	907.86	0.15
Koka-M.Werer	274428.9	PF 500-year	14784.40	1548.47	1557.84		1558.00	0.000671	1.76	8145.77	955.52	0.20
Koka-M.Werer	274428.9	PF 1000-year	21431.53	1548.47	1559.63		1559.87	0.000753	2.11	9859.76	960.71	0.22
Koka-M.Werer	271279	PF 25-year	2648.84	1547.73	1551.08		1551.12	0.000579	0.58	3260.08	1185.46	0.14
Koka-M.Werer	271279	PF 50-year	4058.98	1547.73	1552.09		1552.14	0.000511	0.57	4514.30	1344.22	0.14
Koka-M.Werer	271279	PF 100-year	6077.64	1547.73	1553.23		1553.29	0.000473	0.57	6223.80	1662.71	0.13
Koka-M.Werer	271279	PF 200-yea	2648.84	1547.73	1551.08		1551.12	0.000579	0.58	3260.08	1185.46	0.14
Koka-M.Werer	271279	PF 500-year	14784.40	1547.73	1556.48		1556.56	0.000379	0.99	12427.71	2019.41	0.14
Koka-M.Werer	271279	PF 1000-year	21431.53	1547.73	1558.26		1558.36	0.000365	1.19	16093.09	2090.59	0.14
Koka-M.Werer	268445.3	PF 25-year	2648.84	1544.95	1550.46		1550.49	0.000397	0.68	3478.76	1043.03	0.13
Koka-M.Werer	268445.3	PF 50-year	4058.98	1544.95	1551.48		1551.52	0.000419	0.78	4594.69	1153.24	0.14
Koka-M.Werer	268445.3	PF 100-year	6077.64	1544.95	1552.59		1552.64	0.000452	0.90	5944.20	1273.85	0.15
Koka-M.Werer	268445.3	PF 200-yea	2648.84	1544.95	1550.46		1550.49	0.000397	0.68	3478.76	1043.03	0.13
Koka-M.Werer	268445.3	PF 500-year	14784.40	1544.95	1555.75		1555.86	0.000501	1.33	10478.33	1561.32	0.17
Koka-M.Werer	268445.3	PF 1000-year	21431.53	1544.95	1557.47		1557.61	0.000527	1.58	13268.11	1678.67	0.18

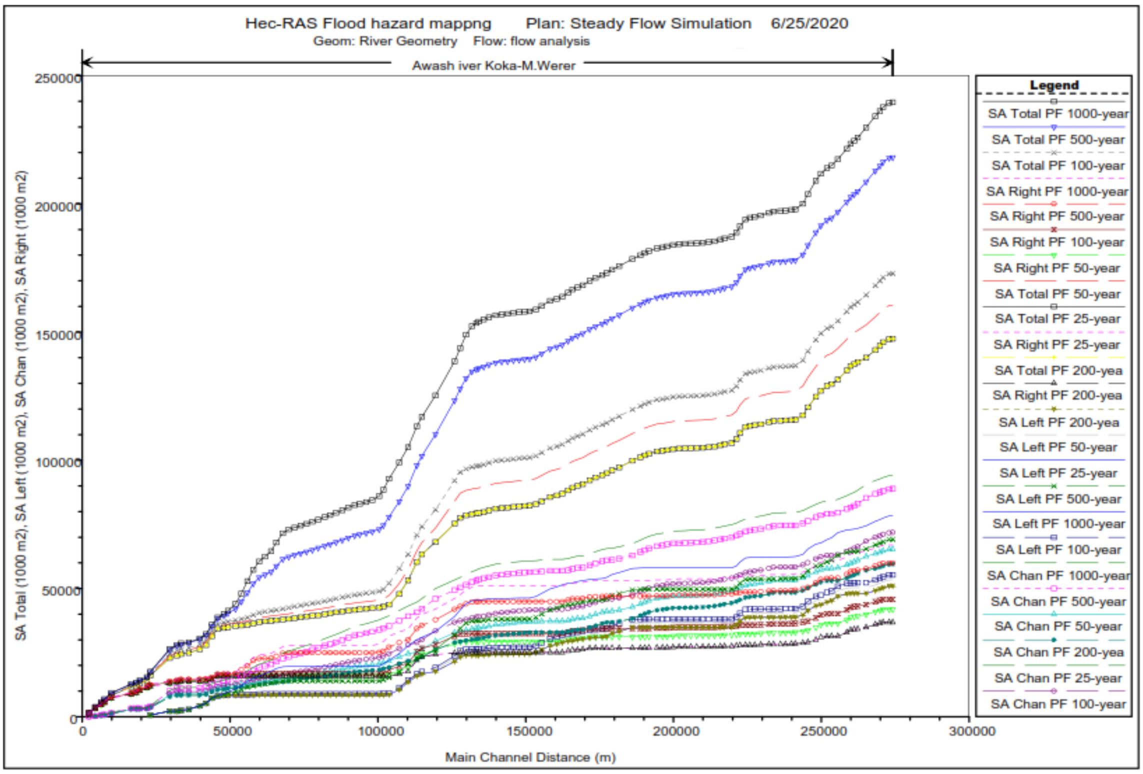
Appendix-J: General profile plot



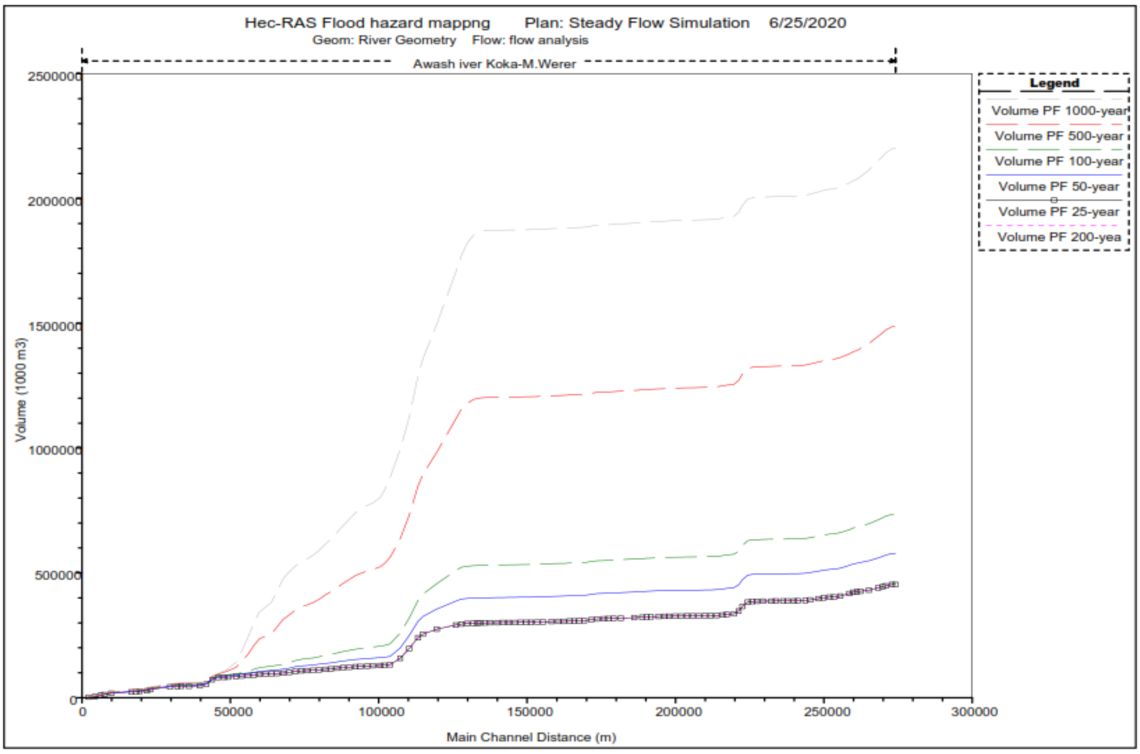
General profile plot- Velocity of flow in the river reach for different return period



General profile plot- Flow area of channel in the river reach for different return period

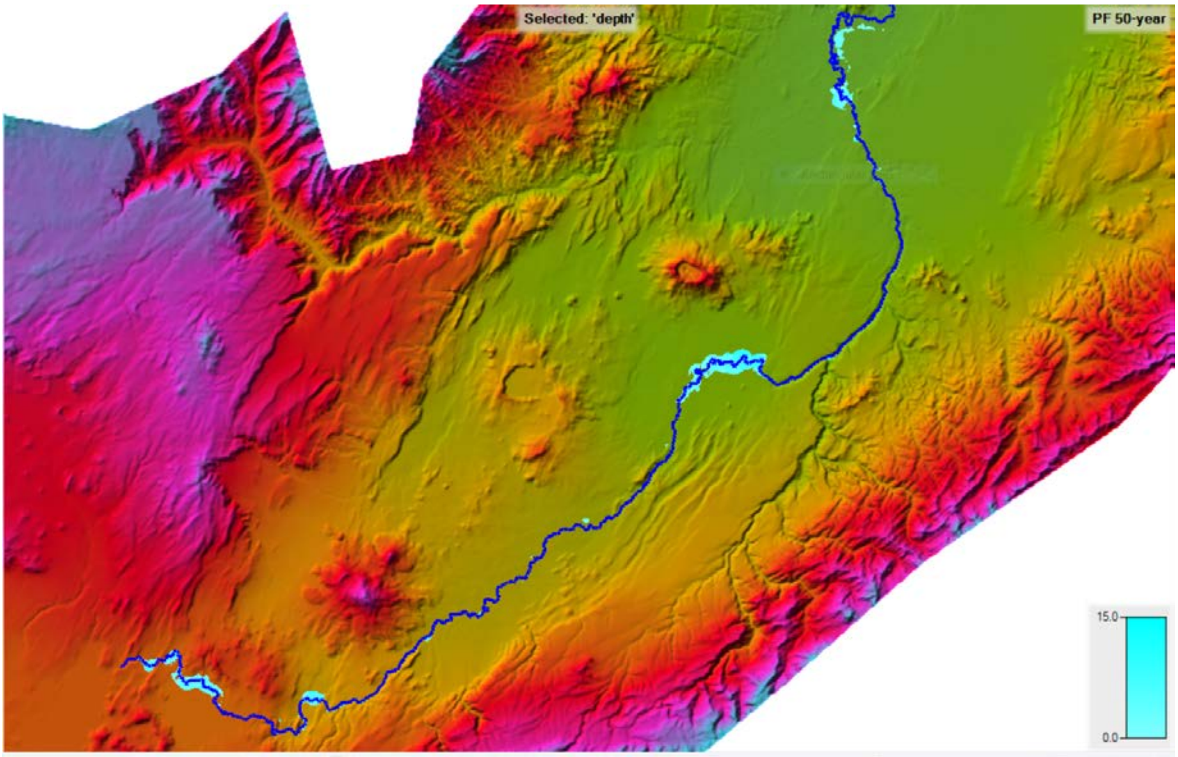


General profile plot- Surface area of channel in the river reach for different return period

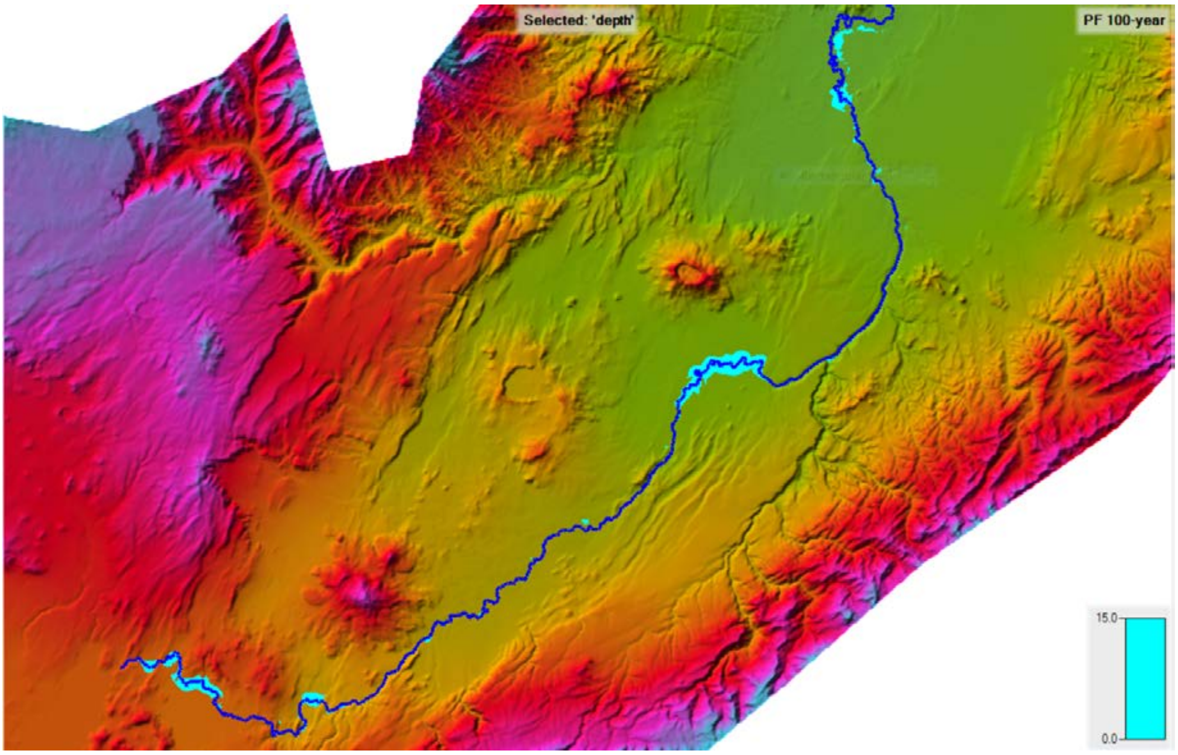


General profile plot- Volume of water in the river reach for different return period

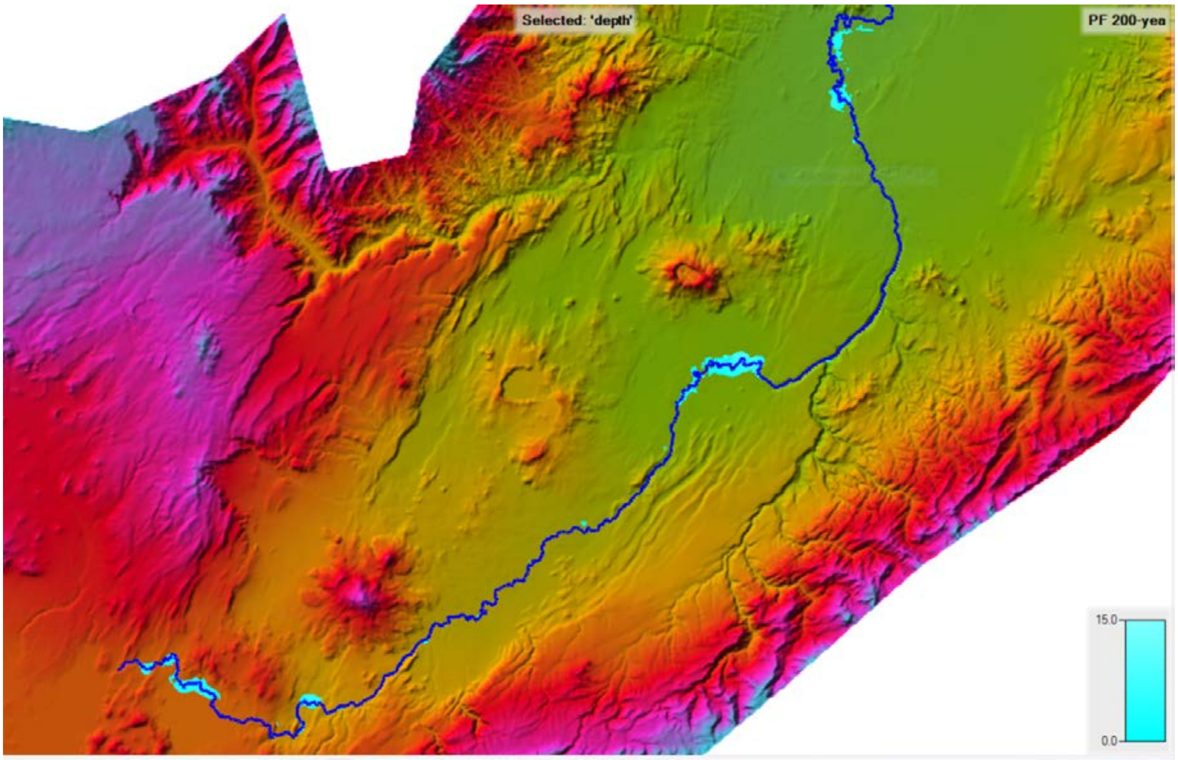
Appendix-K: 50, 100, 200, 500 and 1000-year floodplain map for the study region



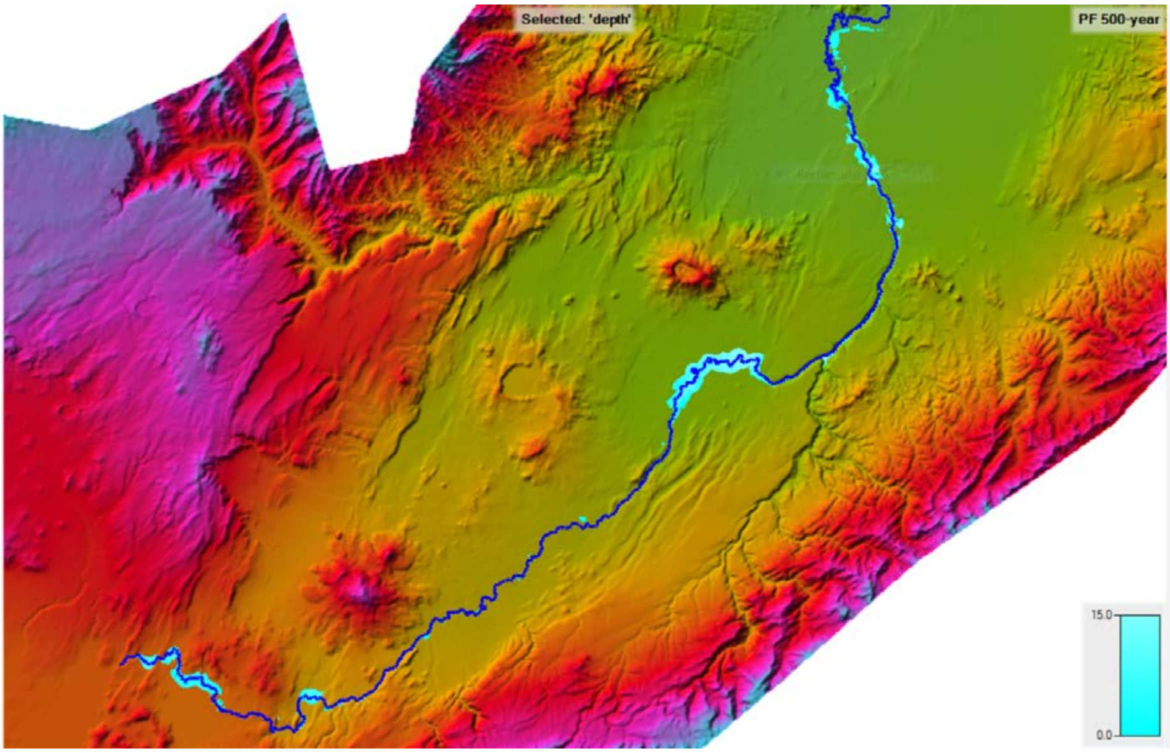
50-year floodplain map for Awash River of study region



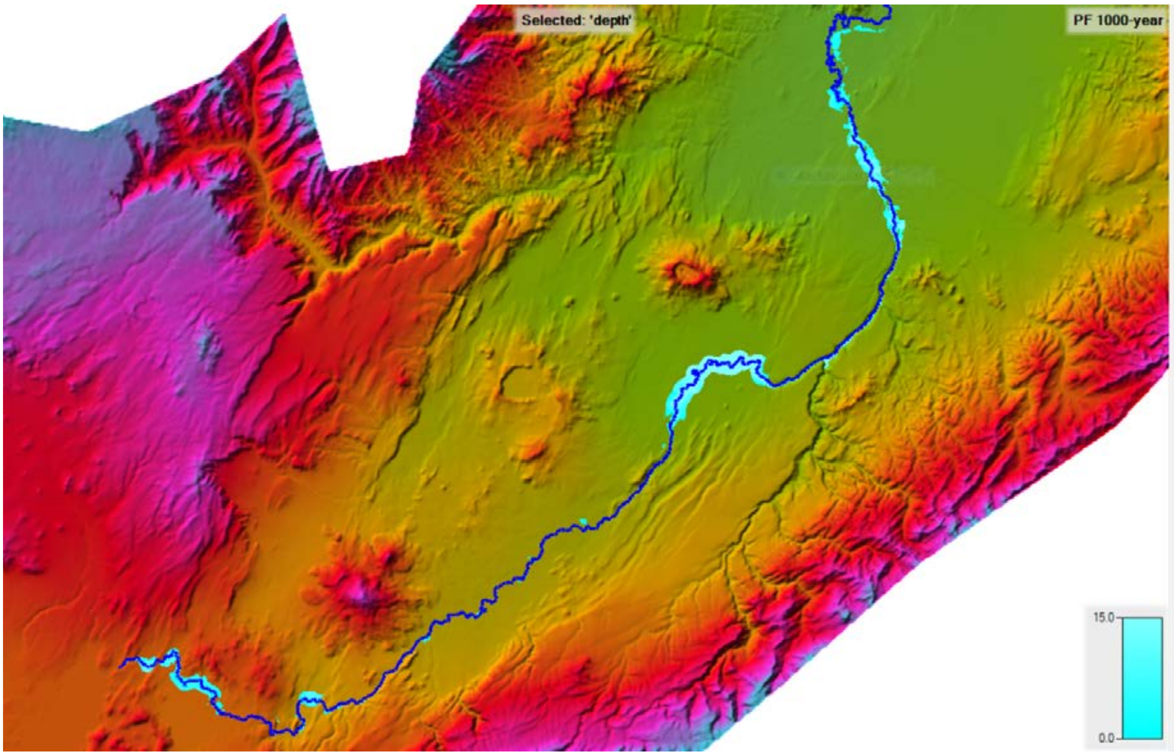
100-year floodplain map for Awash River of study region



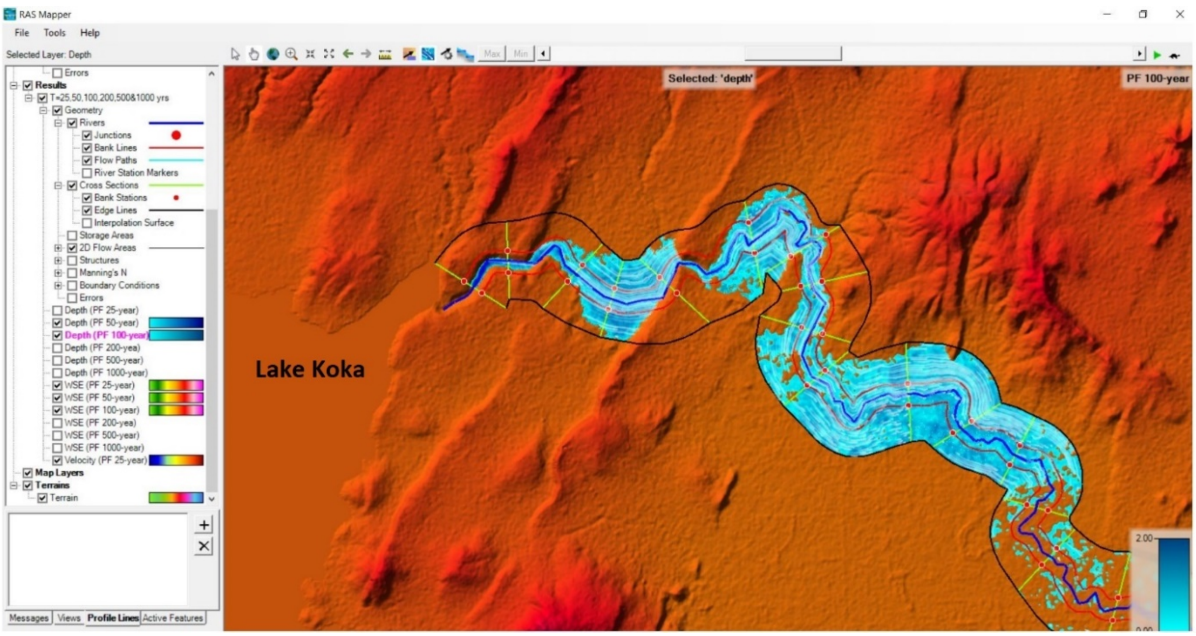
200-year floodplain map for Awash River of study region



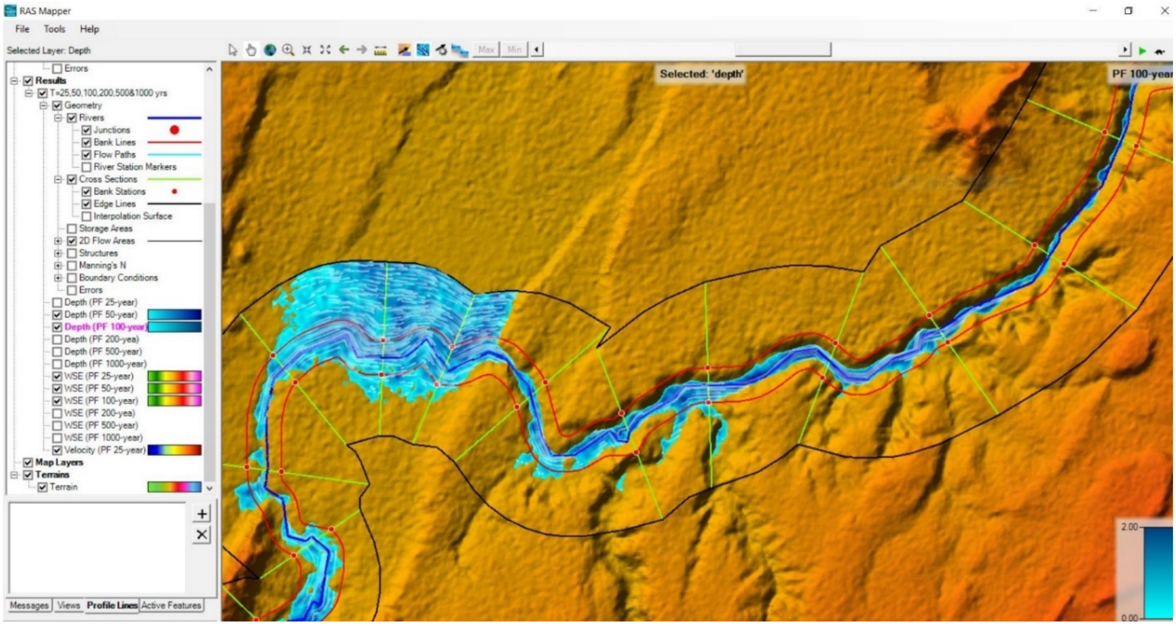
500-year floodplain map for Awash River of study region



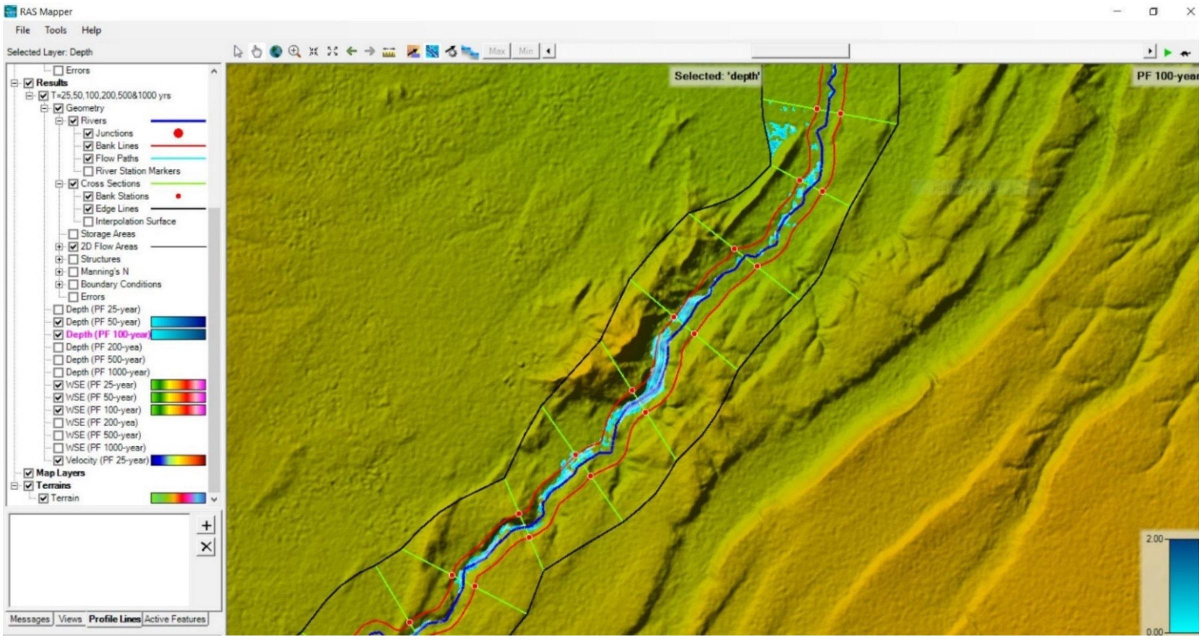
1000-year floodplain map for Awash River of study region



100-year floodplain map for Awash River from Awash @below Koka dam station up to Awash @Wanji station.

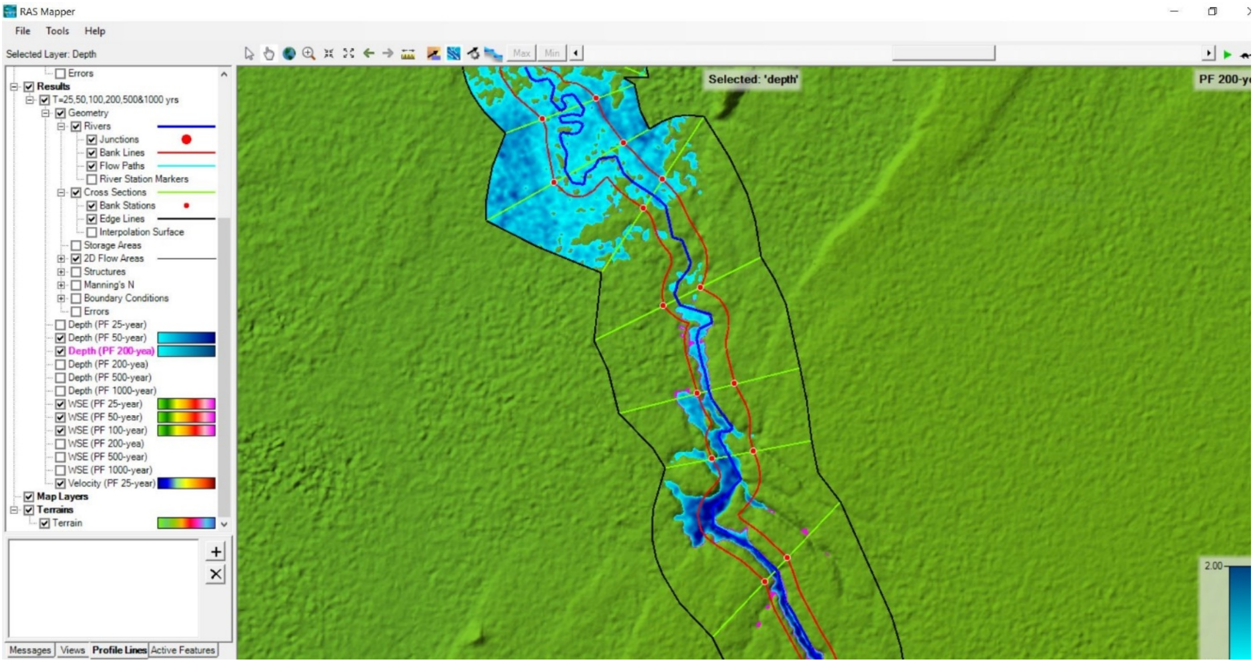


100-year floodplain map for Awash River from Awash @Wanji station up to Awash @Nura Hera station.



100-year floodplain map for Awash River from Awash @Nura Hera station up to Awash @Metehara station.

100-year floodplain map for Awash River from Awash @Metehara station up to Awash @Melka Sadi station.



100-year floodplain map for Awash River from Awash @Melka Sadi station up to Awash @Melka Werer station.

Appendix-L: Annual and Monthly maximum flow time series of selected stations

Year	Annual Maximum Flow Series (m3/s)					
	Awash belw Koka Dam	Awash @WAnji	Awash @Nura Hera	Awash @Metehara	Awash @Melka Sadi	Awash @Melka Werer
1999	327.776	166.19	106.2	125.69	398.56	199.537
2000	157.284	59.35	303.22	155.14	328.823	345.778
2001	104.167	81.69	129.3	91.22	142.262	261.351
2002	90.073	87.22	364.02	104.86	104.589	359.202
2003	123.061	103.37	107.7	86.91	178.851	261.351
2004	62.545	79.36	249.9	114.05	168.375	261.351
2005	69.652	106.78	228.3	92.10	192.986	261.351
2006	312.777	225.61	213.01	228.15	225.384	164.739
2007	359.12	262.99	167.2	225.38	210.306	359.202
2008	475.238	268.66	279.4	179.93	193.378	272.425
2009	24.499	64.11	222	197.95	137.128	261.351
2010	363.746	194.39	407.5	135.396	205.564	387.126
2011	262.774	60.936	197.6	209.3	199.628	547.763
2012	392.296	228.398	147.7	254.155	231.962	397.243
2013	161.419	136.388	226.2	254.155	201.671	167.131
2014	64.083	74.844	209.6	254.16	233.472	442.683
Q_{ave}	209.41	137.52	222.43	169.28	209.56	309.35

Month	Monthly Maximum Flow Series (m3/s)					
	Awash below Koka	Awash @Wanji	Awash @Nura Hera	Metehar	Awash Melka Sadi	Awash Melka Werer
Jan	79.892	70.17	121.4	233.53	137.92	156.173
Feb	69.652	73.81	109.2	59.06	139.25	114.367
Mar	98.985	87.42	163.6	85.22	216.37	184.548
Apr	71.297	69.17	149.76	136.51	205.33	205.947
May	97.968	75.54	364.02	155.14	179.75	156.173
Jun	97.968	81.69	156.4	88.90	143.90	148.767
Jul	230.116	100.45	228.3	254.16	234.50	397.243
Aug	392.296	228.40	356.1	209.30	398.56	442.683
Sep	475.238	268.66	407.5	249.51	311.83	547.763
Oct	120.734	110.17	180.2	254.16	348.73	261.351
Nov	87.22	69.24	160	115.62	153.34	261.351
Dec	68.031	73.77	85.25	232.779	233.472	211.553