Jimma University
School of Graduate Studies
Jimma Institute of Technology
Faculty of Civil and Environmental Engineering
Hydrology and Hydraulic Engineering Chair
Master of Science Program in Hydraulic Engineering

Regional Low Flow Analysis: A Case Study on Upper Awash River Sub-Basin, Ethiopia

By: Desye Asaye

A Thesis Submitted To School of Graduate Studies of Jimma University in Partial Fulfillment of the Requirements for Degree of Master of Science in Hydraulic Engineering

December, 2019
Jimma, Ethiopia

Jimma University
School of Graduate Studies
Jimma Institute of Technology
Faculty of Civil and Environmental Engineering
Hydrology and Hydraulic Engineering Chair
Masters of Science Program in Hydraulic Engineering

Regional Low Flow Analysis: a Case Study on Upper Awash River Sub-Basin, Ethiopia

A Thesis Submitted to School of Graduate Studies of Jimma University in Partial Fulfillment of the Requirements for Degree of Masters of Science in Hydraulic Engineering

By: Desye Asaye

Main Advisor: Prof., Dr. - Ing. Esayas Alemayehu

Co-Advisor: Mr. Wakjira Takala (PhD. Fellow)

December, 2019
Jimma, Ethiopia

DECLARATION

This thesis is my original work and has not been presented for a degree in any other university. The thesis research entitled “Regional Low Flow Analysis; a case study on Upper Awash River Basin, Ethiopia.”

Submitted by	Signature	Date
Desye Asaye	----------------	---------------
Student		
Approved by	Signature	Date
Prof. Dr. Ing. - Esayas Alemayehu	---------------	---------------
Main Advisor		
Mr. Wakjira Takala	Signature	Date
Co-advisor	---------------	---------------

APPROVAL

The thesis entitled “Regional Low Flow Analysis: A Case Study on Upper Awash Sub-River Basin, Ethiopia” submitted by Desye Asaye is approved and accepted as a Partial Fulfillment of the Requirements for the Degree of Masters of Science in Hydraulic Engineering at Jimma Institute of Technology.

Name of Main Advisor	Signature	Date
Prof. Dr. Ing. - Esayas Alemayehu	------------------	-----------------
Name of Co-advisor	Signature	Date
Mr. Wakjira Takala	-------------------	-----------------
Name of chairman	signature	Date
Mr. Fayera Gudu (MSc.)	-------------------	-----------------
Name of Internal Examiner	signature	Date
---------------------------------	-------------------	------------------
Name of External Examiner	signature	Date
---------------------------------.	-------------------	------------------

ABSTRACT

Prediction of low flow of a river in magnitude as well as in frequency is necessary for the planning and design of water resource projects since low flow affects significantly in water supply, water quality and river ecological status. However, there is a high level of gap in the country Ethiopia in understanding the regional low flow and frequency analysis compared to flood studies. Thus, the main objective of this study was to perform appropriate regional low flow analysis on Upper Awash river basin of Ethiopia. This paper discusses regionalization of the Upper Awash river basin using three approaches: flow duration curve, low flow frequency analysis and base flow index. L- Moment ratio diagrams were used for identifying and grouping of stations in to hydrological homogeneous regions and hence the underlying statistical distribution. Accordingly, the basin was delineated in to two homogeneous regions, Region one and Region two covering 34.66% and 65.34% of the sub basin respectively. Discordancy measure of regional data of the L-moment statistics was performed using MATLAB R2013a. Both regions have shown satisfactory results for discordance measures and homogeneity tests. For the candidate distribution from l-moment ratio diagram, best fit distribution is selected by using Easy Fit software. All the three goodness of fit test used in this study shows a first rank for Generalized Pareto and Generalized Extreme Value for region one and two respectively. For the selected distribution parameter estimation technique was selected by performing a Robustness Assessment. Accordingly, GPA with PWM selected for region one and GEV with MOM was selected for region two. Based on best-fit distributions for the two regions, regional low flow frequency curves were constructed for the return periods of 2, 5, 10, 20, 50 and 100 years resulting 0.125, 0.084, 0.070, 0.064, 0.059 and 0.038 m^3/s quantile for region one and 1.206, 0.964, 0.807, 0.659, 0.611 and 0.570 m^3/s for region two . Base flow separation has been made using BFI+ software using a local minimum method. Accordingly 40% of the stations has BFI ranging from 0.15-0.39, 50% of the station constituting most part of the Region two shows BFI of 0.4-0.65. Similarly, 10% of the stations has BFI >0.7. The results indicate that the region one in Upper Awash river basins shows a high diminished in low flow quantile relative to the region two.

Keywords: Base flow index, Homogenous region, L-moment, low flow frequency analysis Parameter estimation methods, Regionalization

ACKNOWLEDGMENT

First of all, I am grateful to the Almighty God for helping me to complete this work and for his endless mercy, grace and wisdom upon me during all these days in all my life. I am greatly thankful to my employer Assosa University for their sponsorship of the MSc. program and for financial support made through the program during both my course work and research time. Also my gratitude goes to Jimma University in their great support during learning process.

With great pleasure and a deep sense of indebtedness, I extend my gratitude and appreciation to my advisors Prof., Dr.-Ing Esayas Alemayehu and Wakjira Takala (PhD. Fellow) for their persistent, support, constructive guidance and advice throughout the work.

My honest gratitude goes to all staff members of Ministry of Water and Irrigation, mainly Hydrology department, GIS department and Library, for their significant support in providing me hydrological data, digitized map and other reference materials.

I would like to express my warm feeling of appreciation and tank to my friends especially to Mr. Tarekegn Dejene who helped me in all stages for all challenges, knowledge sharing and helping me during the whole study period.

Last, but not the least by any means, special and great thanks go to my parents who have been providing with the Necessary help during my academic careers.

TABLE OF CONTENT

DECLARATION i
APPROVAL ii
ABSTRACT iii
ACKNOWLEDGMENT iv
TABLE OF CONTENT v
LIST OF TABLE viii
LIST OF FIGURE ix
ABBREVIATIONS AND ACRONYMS x
1. INTRODUCTION 1
  1.1. Background 1
  1.2. Statement of the problem 3
  1.3. Objective 4
    1.3.1. General Objective 4
    1.3.2. Specific Objectives 4
  1.4. Research Question 4
  1.5. Significance of the study 5
  1.6. Scope of the study 5
  1.7 Thesis Out line 5
  1.8. Limitation 6
2. LITERATURE REVIEW 7
  2.1. Low Flow Hydrology 7
  2.2. Definition of low flow 7
  2.3. Importance of Low flow Studies 7
  2.4. Flow Duration Curve 8
  2.5 Low Flow Frequency Analysis 8
  2.6 Low Flow Frequency Models 9
    2.6.1 Annual Minimum Series Model 9
    2.6.2 Partial duration series model 10
  2.7 Low Flow Statistics 10

2.7.1. Statistical Requirements 10

2.8. Low flow Regionalization 11

2.8.1. Identification and Delineation of Homogenous Region 11

2.8.2. Regional Homogeneity Tests 12

2.9. Statistical Fitting Distributions 13

2.9.1. Goodness of Fit Test 13

2.9.2. Method of L-moment ratio diagram 14

2.10. Selection of Parameter Estimation Method 14

2.11. Quantile Estimation of Low Flow 16

2.12. Base Flow 16

2.12.1. Definition and Introduction 16

2.12.2. Base Flow Separation Methods 16

2.13. Previous Studies on RLFFA in Ethiopian River Basin 17

3. MATERIALS AND METHODS 19

3.1. Description of the Study Area 19

3.1.2. Climate and Hydrology 20

3.1.3. Land use and Soil type 21

3.2. Materials Used 22

3.3. Methodology and procedure 23

3.4. Source and availability of data 23

3.6. Method of data analysis 25

3.6.1. Missed data filling 25

3.6.2. Data quality Analysis 26

3.8. Low flow Regionalization of Upper Awash River Basin 29

3.8.1. Identification and Delineation of Homogeneous Region 29

3.8.2. Regional Homogeneity Test 30

3.8.3. Fitting the probability distributions 33

3.8.4. Goodness of fit test 34

3.8.5. Evaluation of the Performance of Fitting distribution 36

3.8.6 Parameter and low flow Quantile Estimation 37

3.9. Standard error of estimates 38

3.10. Derivation of Regional Frequency Curve for the Homogeneous Regions 39

3.10.1 Estimation of Index Low flow.......... 39
3.11. Base Flow Separation Methods.......... 39
4. RESULT AND DISCUSSION.......... 41
4.1. Flow Duration Curve Analysis.......... 41
4.1.1 Low flow Indices.......... 43
4.1.2. Zero Flow condition Of FDC.......... 43
4.2. Low flow Regionalization.......... 44
4.2.1. Identification of homogeneous region.......... 44
4.2.2 Regional Homogeneity test.......... 46
4.2.3. Delineation of Homogenous Region.......... 48
4.3. Identification of Regional frequency distribution for low flow Analysis.......... 49
4.3.1. Selection of distribution by LMRD.......... 50
4.3.2. Goodness of fit tests.......... 51
4.3.3. Estimation accuracy of selected distribution.......... 52
4.4. Parameter Estimation.......... 53
4.5. Standard error and Low flow quantile.......... 54
4.6. Base flow Separation.......... 57
4.6.1. Result of BFI.......... 57
5. CONCLUSION AND RECCOMENDATION.......... 60
5.1 Conclusion.......... 60
5.2. Recommendation.......... 61
REFERENCE.......... 62
APPENDIX.......... 67

LIST OF TABLE

Table 3.1 General characteristics of the station in the Upper Awash River Sub-basin 25
Table 3.2 Filling accuracy and regression equation of the station 26
Table 3.3 Independence test of each station 27
Table 4.1 Group One Low flow indices 43
Table 4.2 Group Two low flow indices 43
Table 4.3 preliminarily identified homogenous region 45
Table 4.4 Results of CC-based homogeneity test for Region One 46
Table 4.5 Results of CC-based homogeneity test for Region Two 47
Table 4.6 Results of discordance measure test for Region One 48
Table 4.7 Results of discordancy measure for region two 48
Table 4.8 Selected candidate distribution for each region 51
Table 4.9 Goodness of fit measure for Region One 51
Table 4.10 Goodness of fit measure for Region Two 52
Table 4.11 Parameter summary of each region 54
Table 4.12 Standard error and quantile estimation for Region One 55
Table 4.13 Standard error and quantile estimation for Region Two 55
Table 4.14 Base flow index for all station 57

LIST OF FIGURE

Figure 3.1 Study area map 20

Figure 3.2 Flow chart of methodology 23

Figure 3.3 Location of Gauging station 24

Figure 3.4 Relationship of N with Base flow index....................40

Figure 4.1 Daily Flow Duration Curve for all station....................41

Figure 4.2 Daily Flow Duration Curve for Group One station....................42

Figure 4.3 Daily FDC for Group Two station....................42

Figure 4.4 1, 7, 14 and 30 day FDC of Bello Station 42

Figure 4.5 Lcs Vs Lck moment ratio diagram of each station....................45

Figure 4.6 Delineated Region in the sub-basin....................49

Figure 4.7 Regional average of LMRD for Upper Awash River Basin....................50

Figure 4.8 Probability-Probability plots of Region One....................53

Figure 4.9 Quantile-Quantile plots of Region One....................53

Figure 4.10 Regional growth curve for delineated homogenous region....................56

Figure 4.11 Typical hydrograph and base flow hydrograph....................58

Figure 4.12 Typical base flow hydrograph....................58

Figure 4.13 Separated base flow hydrograph 59

ABBREVIATIONS AND ACRONYMS

AM	Annual Minimum
AMNS	Annual Minimum Series
AMXS	Annual Maximum series
DEM	Digital Elevation Model
FDC	Flow Duration Curve
GIS	Geographic Information System
IHR	Identification of Homogenous Region
ITCZ	Inter-Tropical Convergence Zone
Lck	Linear coefficient of kurtosis
LCL	Lower Confidence Limit
Lcv	Linear coefficient of variation
Lcs	Linear coefficient of skewness
LFFA	Low Flow Frequency Analysis
LFFC	Low Flow Frequency Curve
LMRD	L-moment ratio diagram
MAM7	7-day Mean Annual Minimum flow
MML	Maximum livelihood method
MOM	Method of Moments
MoWIE	Ministry of Water Irrigation and Energy
PBT	Peaks Below Threshold
POT	Peaks over Threshold
PWM	Probability weighted moments
SEE	Standard Error of Estimates
T	Return period
UARB	Upper Awash River Basin
UCL	Upper Confidence Limit
US	United States
USA	United States of America
WMO	World Metrological Organization
XT	Low flow quantile

1. INTRODUCTION

1.1. Background

With an increased attention towards surface water management, information about the estimates of d day, T year low flows are routinely required for the maintenance of water quality standards. Such statistics describing low flows are commonly used in waste load allocation, waste treatment plants, issues governing minimum downstream release requirements for irrigation, hydropower and water supply, etc. Low flow information can be quantified in a variety of ways depending on the type of data available and the output information desired (Joshi and Hillaire, 2013)

The general complexity and diversity of low-flow processes and the multidisciplinary nature of low-flow studies make it often difficult to find a common ground for exchange in scientific societies. The Public Utility Board Low-Flow Workshop, held in Quebec City, April 12-13, 2007, was intended to provide a forum for discussion to build up a common cross-disciplinary language and share methodologies. Such a forum is also important for the improvement of the general level of understanding of low-flow processes in natural and regulated river environments. The systematic cross-disciplinary summary, evaluation of current results and existing methodologies in this area provided by this special issue can guide scientists in new research directions that can lead to significant improvements in low-flow estimation capabilities (Taha *et al*., 2008).

Low-flows are an important part of the natural flow regime of rivers and streams (Curran et al., 2012). The spatial and temporal variability of river low-flow characteristics can also be considerable. Smakhtin (2001) presented a comprehensive review of low-flow hydrology covering such issues as generating mechanisms, estimation methods and applications. The availability of reliable low-flow occurrence and magnitude estimates is crucial for a wide array of engineering applications such as aquatic ecosystem modeling, environmental impact analysis, water supply assessment for potable and irrigation purposes, liquid waste effluent dilution, river navigation planning and water quality management (Assefa et al., 2018).

Flow duration curves (FDC) give a relationship between magnitude and frequency of stream flow discharges and can be constructed for different time periods: annual, monthly and daily.

These curves constructed for daily time series enable a detailed examination of the duration characteristics of a river. For curves constructed for n-day and n-month average flow time series, moving average approach is used. From the perspective of low flows, the section of FDC below Q_{60} (discharge equaled or exceeded 60% of the time) is considered vita (Deepti and Hilaire, 2013).

Regional low flow frequency analysis is one of the practical means providing low flow information at sites with little or no local data. In the regional approach, available low flow data series from hydrological homogenous region are pooled in dimensionless form and a frequency distribution is fitted to combined data. However, the main problem of low flow frequency analysis is the determination of probability distribution that can provide a curve that defines the regional average relation between standardized flow magnitudes (or) and return period (T) (Tegenu, 2007).

Regionalization of stream flow characteristics is based on the premise that catchments with similar geology, topography, climate, vegetation, and soils would have similar stream flow responses. It consists of the identification of regional laws, applicable over a more or less wide area, a region, which generally use catchment characteristics as independent variables (Santhi et al., 2008). A region is considered to be homogeneous if all sites included in the region have some common characteristics (Ahmad et al., 2016).

Base flow is an important component of the ground water system. It is the component of stream flow that is attributed to ground water discharge and other delayed sources such as snow melt into streams. Neglecting base flow as a nutrient source to streams leads to misinterpretation of data. Knowledge on base flow availability is important in development of water management strategies (especially for drought conditions), estimation of small to medium water supplies and water quality, management of salinity, algal blooms and others (Santhi *et al*., 2008).

Therefore, the general goal of this study is to characterize the regional low flow in Upper Awash sub-basin to provide the necessary information about the low flow of the sub-basin.

1.2. Statement of the problem

Low-flows analysis and the resulting long-term droughts Prediction is associated with a high economic value. It should be mentioned that droughts have more severe consequences and are often more costly than flood events. Damage accounting to approximately US$40 billion occurred during the USA droughts of 1988-1989(Demuth, 2005).

Due to increased high population pressure and food insecurity in Ethiopia, it is envisaged that there would be huge demand for irrigated agriculture by the farmers, investors and as the matter of policy priority by the government at all level. This will undoubtedly create huge demand on the water resources particularly during lean season. This extra demand on the river may create undesirable environmental as well as upstream downstream conflict (Tatek, 2015).

Recently there are observations on drying out of some rivers, where the low flow is becoming no flows. Decreasing in low flow would impact the environmental flow in a given ecosystem and affect multi-purpose operations which depend on that water system such as river and lakes (Assefa *et al*., 2018). Reduced low flow may also cause an impairment of water quality, and affect river ecological status and navigation and power supply sectors (Middlekoop *et al*., 2001). To understand the causes and take remedial action for the sustainable utilization of the low flows, the dynamics in low flows in a river system should be evaluated, which could include quantifying the low flow quantiles, and developing regional curves is a very important approach for proper management of the water resources (Assefa *et al*., 2018).

Alternatively there are several ongoing and planned water resource projects in the Awash basin mainly in the Upper valley which is moreover known by population density and suitable irrigation potential land. These works require a reliable estimation of low flow quintiles using reliable flow records measured at gauging stations. However, most of the catchments in developing countries, like Ethiopia are poorly gauged, which hinders the country’s water resources management and flood prediction (Rabba *et al*., 2018).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. This 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 (Saf, 2009).

Due to the scarcity of the required data at or near the site of interest, professionals responsible for the design of water resources projects have had to depend on unsatisfactory sources of information for their input parameters (Gebeyehu, 1989). This enforces to adopt a more conservative approach in their design techniques with the obvious implication of higher costs on the projects, which is indeed a burden on the financial resources of the country.

Upper Awash river sub-basin have a good potential in water resources development including hydropower, small scale irrigation, water supply, aquatic ecosystem and etc. (Henock et al., 2008). But little or no research has been done in the low flow characteristics of the Upper Awash river sub-basin so far. Most of all the availability and quality of information is not adequate, so further development of any water resource project within the region is difficult and unreliable unless and otherwise the low flow characteristics is well known. Therefore, for informed decision making and appropriate policy recommendation on minimum environmental and downstream requirements, analyzing and characterizing low flow of the Upper Awash river sub basin is indispensable.

1.3. Objective

1.3.1. General Objective

The general objective of this study is to analyze the regional low flow characteristics of Upper Awash river basin.

1.3.2. Specific Objectives

The specific objectives of this thesis are:

- To perform low flow regionalization of the entire sub-basin
- To identify the best-fit statistical distribution and parameter estimation for low flow analysis of each region
- To estimate the regional low flow quantile and draw low flow frequency curve of each region in the sub-basin

1.4. Research Question

The research questions, which address this particular study are:

1. How hydrologically homogenous regions based on low flow characteristics are identified

and delineated?

2. What are the best-fit statistical distribution and parameter estimation for low flow analysis of each homogenous region in the sub-basin?

3. How much is the regional low flow quantile of the sub-basin?

1.5. Significance of the study

The result of this study will be expected to become valuable up to date information to analyze regional low flow in the selected river basin. The research findings can also bring many benefits to understand the characteristics of hydrological droughts and water availability during low flow periods in Upper Awash basin, which is crucial for the optimization of water resources allocation and planning in the region.

1.6. Scope of the study

Generally, the study deals the probability of low flow occurrence and its magnitude that might take place depending on the hydrological response of the selected sub-basin. There are many types of regional low flow analysis, but this study is limited to the Flow duration curve analysis, frequency analysis and base flow separation of each gauging station in the study area.

1.7 Thesis Out line

The thesis comprises five Chapters; the first chapter explains the general background of the study, problem statement, objective of the study, scope and significance of the study including the questions to be answered by the thesis.

Chapter two consist of Literature review of low flow analyses including concept of low flow, importance of low flow study and methods of low flow analysis such as flow duration curve, frequency analysis and base flow separation.

Chapter three merely focused on the materials used and the procedures followed to accomplish the study including description about the study area, sources and quality of data, method of filling missed data and the chapter also elaborates the way how the flow duration curve, the low flow frequency analysis and the base flow separation were executed in this study.

Chapter four is all about the result and discussion of the study. This chapter discusses the analysis of Regional low flow frequency including the other methods of low flow analysis and the main output of the study. Conclusion and recommendation are discussed in the last chapter five.

Finally, references and appendixes in the form of tables and figures serving as a supporting document to this study are attached to make the work complete.

1.8. Limitation

World Meteorological Organization recommended as minimum record length of data of 10 years duration or more for estimation and prediction of low flow frequency analysis. Owing to short record length of data, this study considered only 10 gauging stations that fulfill the recommended base period which may have an influence on the regional low flow frequency analysis in the sub-basin.

2. LITERATURE REVIEW

2.1. Low Flow Hydrology

Low flow hydrology differs from the analysis of the flood events. Different methods are available for characterizing or defining low flows. Many researchers have identified number of different types of analysis which are used in analysis low flows. The term low flow measure is used to describe the many ways that have been developed for summarizing the low flow regime of a river (Smakhtin, 2001).

2.2. Definition of low flow

The term low flow may mean different things to different interest groups. To many, it may be considered as the actual flows in a river occurring during the dry season of the year, others may be concerned with the length of time and the conditions occurring between flood events (example, in erratic and intermittent semi-arid flow regimes). Yet others may perceive 'low flows' not only as discharges occurring during a dry season, but as a reduction in various aspects of the overall flow regime.

International glossary of hydrology (WMO, 2008) defines low flow as 'flow of water in a stream during prolonged dry weather'. This definition does not make a clear distinction between low flows and droughts. A low flow is a seasonal phenomenon, and an integral component of a flow regime of any river. Drought, on the other hand, is a natural event resulting from a less than normal precipitation for an extended period of time (Smakhtin, 2001).

2.3. Importance of Low flow Studies

Population growth and the associated expansion in domestic, industrial, and agricultural use of water have placed an increasing demand on water resources throughout the world. This growth has been accompanied by the increased use of watercourses for the disposal of industrial and domestic effluent, which has inevitably led to competing demands on river systems. awareness of the ecological, recreational and amenity benefits of the river corridor and concerns over changes in river regimes in the longer term have added to the pressures on water resources. It is in times of drought or in other words in low flow periods, that river systems are most stressed and thus an understanding of the natural variability of drought conditions in time and space is fundamental to a wide range of water management problems (Hewa, 2001).

2.4. Flow Duration Curve

The flow-duration curve (FDC) is a graph of river discharge plotted against exceedance frequency and is normally derived from the complete time series of recorded river flows. It is simple to construct and used in many different water resources applications over the entire range of river flows. It effectively reorders the observed hydrograph from one ordered by time to one ordered by magnitude. The percentage of time that any particular discharge is exceeded can be estimated from the plot (WMO, 2008).

FDC can also be constructed for different time periods: annual, monthly, seasonal and daily. For curves constructed for n-day and n-month average flow time series, moving average approach is used. From the perspective of low flows, the section of FDC below Q_{60} (discharge equaled or exceeded 60% of the time) is considered (Smakhtin, 2001). The record length required to determine FDCs with an acceptable sampling error would depend on the natural flow variability. In temperate climates, a 10-year period or more is recommended (WMO, 2008).

Tatek (2018) have used daily flow duration curve to group stations in the Dedessa river basins. Accordingly, he has classified the region in to two groups, group one and two, and Joshi and Hillaire (2013) in his low flow frequency analysis of three rivers in Eastern Canada has also used it to obtain low flow indices using 1, 7, 10 and 30 days moving discharges.

2.5 Low Flow Frequency Analysis

Low flow information can be quantified in a variety of ways depending on the type of data available and the output information desired. Low flow frequency analysis (LFFA) is a stochastic approach for characterizing low flow events. Low flow frequency curve; LFFC can be constructed on the basis of annual flow minima (daily or monthly minimum discharges) and seasonal minimum values (winter or summer low flows).

According to Smakhtin (2001), an analysis made on a time series of 7-day average flows is less sensitive to measurement errors. The 7-day period reduces the day-to-day variations in the artificial component of the river flow.

Low-flow frequency indices are widely used in drought studies, design of water supply systems, estimation of safe surface water withdrawals, classification of streams potential for waste dilution (assimilative capacity), regulating waste disposal to streams, maintenance of certain in-stream discharges, etc. (Joshi and Hillaire, 2013).

2.6 Low Flow Frequency Models

In low flow frequency Analysis, the objective is to determine a Q-T relationship at any required site along a river. 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 (Desalegn *et al*., 2016). The following two models are available for this purpose.

2.6.1 Annual Minimum Series Model

Flood frequency analysis is generally performed on a data series comprising of single highest peak in a year, known as the Annual Maximum Series (AMXS). For low flows, Annual Minimum Series (AMNS) is considered. Annual minimum series (AMNS) involves selecting single lowest value in each year. The value of low-flow frequency analysis can be improved by considering 7- day or 10- day moving averages of flow. AMNS in that case would involve annual minimum 7- day or 10- day flow (Joshi and Hillaire, 2013). Assefa *et al* (2018), Ni llar and Khin (2014), Taha *et al* (2008) and other researchers has used Annual minimum series model to quantify low flows in different regions.

In this study, the minimum 7-day low flow is used for the analysis. The 7-day low-flow index was chosen for three reasons:

(a) The 7-day low-flow is the most widely used index in the USA, UK and many other countries. The 7-day period covered by MAM7 eliminates the day-to-day variations of the river flow.

(b) Previous studies, as reviewed by Smakhtin (2001), have shown that, compared with 1-day low flow, an analysis based on a time series of 7-day average flows is less sensitive to measurement errors.

(c) Practically, the 7-day low flow better represents the drought conditions of concern and can be used more effectively in water management.

d) 7-day low flows are not very different from 1-day low flows (Smakhtin, 2001). Averaging over some days also allows smoothing out some human influences on flows such as variation of hourly flows due to hydro peaking and little abstraction from farmers.

2.6.2 Partial duration series model

Certain flows (for example, channel-forming flows, flows that move the substrate) occur more than once in a year and annual maximum series do not account for these flows. An appropriate technique in such cases is the Partial Duration Series approach. PDS involves selecting those values that lie above (Peaks over threshold; POT) and below (Peaks below threshold; PBT) a threshold level, chosen for its relevance to the issue for which the analysis is being carried out. Therefore, to avoid the problem of data dependency, the annual minimum flow series model is selected. In addition to this, AMN series is widely and universally used model by different researchers for the purpose of low flow frequency analysis (Desalegn *et al.*, 2016).

2.7 Low Flow Statistics

Preference to L-moment statistics (Lcv, L_{CS}, and Lck) over conventional moments is given because of the fact that conventional moments exhibit substantial bias and variance for the small samples encountered in hydrological applications. L-Moment method is a powerful and efficient method to compute any statistical parameters, Also it cannot be influenced with the presence of outliers (Rao and Hamed, 2000). The statistical parameters computed includes:-

- Mean
- Standard deviation
- Coefficients of Skeewness
- Coefficients of Variations and Coefficients of Kurtosis

2.7.1. Statistical Requirements

The governing requirements for the statistical treatment of low flows are similar to those usually identified for flood frequency analysis. One difference between low-flow and flood frequency analyses is that the data for low- flow analyses consist of the annual events that have the lowest average flow of the required duration D during each water year of record. Thus, the records of flow for each water year are evaluated to find the period of D-days during which the average flow was the lowest; these annual values are used as the sample data (McCuen, 1998).

The major differences are listed here:

1. Instead of using the exceedence probability scale, the non-exceedence scale (that is, the scale at the bottom of the probability paper) is used to obtain probabilities. The non-exceedance scale is important because the T-year event is the value that will not be exceeded.

2. The data are ranked from low to high, with the smallest sample magnitude associated with a Weibull probability of l/(n+l) and the largest magnitude associated with a probability of n/(n+l); any other plotting position formula could be used in place of the Weibull. However, plotting position probabilities are non-exceedance probabilities.

2.8. Low flow Regionalization

Regionalization of stream flow characteristics is based on the premise that catchments with similar geology, topography, climate, vegetation, and soils would have similar stream flow responses. It consists of the identification of regional laws, applicable over a more or less wide area, a region, which generally use catchment characteristics as independent variables (Santhi *et al*., 2008). It refers to grouping catchments into homogenous regions. The resulting regions were assumed homogeneous in terms of hydrologic response. This assumption actually is not true as it may have very different relief and stations within the same geographic region, which have high correlation that will cause some bias in the regionalization. The availability of data is an important aspect in frequency analysis. Large variation associated with small sample size cause the estimate to be unrealistic. 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 (Rao and Hamed, 2000).

Regionalization can be done based on geographic proximity, physiographic and climatic characteristics of the catchments. Gebeyehu (1989) further advancement in the field of flood frequency analysis has led to a better approach other than geographic proximity. Researchers made their division of homogeneity by analyzing the statistical characteristics of flow data of different stations within the basin (Rao and Hamed, 2000).

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 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).

2.8.1. Identification and Delineation of Homogenous Region

The most prudent step in RFFA is formulating homogeneous regions. A region is considered homogeneous if all sites included in the region have some common characteristics. There are different grouping methods available in literature used for this purpose, e.g., geographical convenience, subjective partitioning, objective partitioning, and cluster analysis (Ahmad *et al*., 2011). 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). Hosking and Wallis (1997) regard cluster analysis of site characteristics as the most practical method of forming regions from large data sets. However, they noted that the output of this analysis should not be considered final and it needs subjective decisions at several stages. In addition, they provided insight into the maximum and minimum size of the regions to be formed by this procedure for use with the index flow method.

2.8.2. Regional Homogeneity Tests

Once a set of physically plausible regions has been defined, it is desirable to assess whether the regions are meaningful. This involves testing whether a proposed region may be accepted as being homogeneous and whether two or more homogenous regions are sufficiently similar that they should be combined into a single region (Hussen and Wagesho, 2016). In other words, the hypothesis of the homogeneity test is that the at-site frequency distributions are identical except for a site-specific scale factor

Some of the most commonly used statistical homogeneity tests are discordance measure test and CC– based homogeneity test. Both these homogeneity test are employed to check regional homogeneity of the proposed stations in the study 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). Discordancy test is also used to examine the appropriateness of the data and to screen out the data from unusual stations (called discordant stations) which have different probabilistic behaviors compared with other stations in a given region (Hosking and Wallis, 1997).

In order to evaluate the regional homogeneity, Wiltshire used a non-parametric jack-knife procedure to estimate the at-site distribution, unlike Dalrymple who assumed Gumbel distribution as the parent distribution at each site. Hosking and Wallis (1993) proposed the next important statistical test for homogeneity test based on the sample L-moments ratios. Chowdhury *et al*. (1991) suggested another statistical test based on L-moments, which were more powerful than previous tests; however, the most rigorous L-moment based test of homogeneity is that of Hosking and Wallis (1993). It compares the variability of the L-moment ratios of the sites within a region with the expected variability obtained from simulation from a collection of sites with the same record length as their real world counterparts.

2.9. Statistical Fitting Distributions

After formulating homogeneous regions, the next step is to choose the most robust frequency distribution for each homogenous region (Ahmad *et al*., 2016). The candidate distributions are usually evaluated for the accuracy of the quantile estimates for each site. The procedure includes trying to fit several theoretical distributions to the observed low flow data and selecting an appropriate distribution by using statistical tests. Many studies have attempted to ascertain suitable distributions for annual minima and those occurring at different averaging intervals. Despite several attempts, no fixed probability distribution for low flows has been agreed on. One of the crucial issues that most of LFFA and distribution fitting studies confront is the occurrence of zero values (Joshi and Hillaire, 2013).

Hydrological datasets for example, stream flow and precipitation, often have zero as a lower limit. Ignoring zero values may lead to an unreliable estimation of the concerned variable. However, distributions fit to zero values assign positive probabilities to negative values of the variable. In such cases, the distributions can be restricted to have a lower limit, which may give physically meaningless results along with challenging the flexibility of the distribution (Smakhtin, 2001).

Different regional frequency distributions were selected in several regional studies. However, Hosking and Wallis (1997) indicated that the L-moment diagrams is only a tool in selecting the candidate distributions and final distribution selection should be made using more objective test that reflects the robustness of the distribution.

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).

2.9.1. Goodness of Fit Test

For a given region that contains sites with similar statistical distribution and parameter values, the main aim of this test is to examine whether the candidate distribution fits to a data set better than the others (Tadesse *et al*., 2011). These tests calculate test-statistics, which are used to analyze how well the data fits the given distribution. These tests describe the differences between the observed data values, and the expected values from the distribution being tested (Millington *et al*., 2011).

There are several methods available for testing the goodness-of-fit of theoretical distribution for extreme events both at-site and/or regional average data. For example, the graphical (like histogram and probability plotting) and statistical tests such as chi square test (χ^2- test), Anderson Darling tests and Kolmogorov-Smirnov statistical methods are discussed in the literature by Gottschalk and Krasovskaia (2001) and Hosking and Wallis (1997).

2.9.2. Method of L-moment ratio diagram

One of the main applications of L-moments is identification of the probability distribution of the observed phenomena using the L-MRD (Assefa *et al*., 2018). This is a diagram of L-Skeewness and L-Kurtosis of the sample data set, which is plotted against constant lines and points of known statistical distributions of interest. This is a common technique used in Regional Low Flow Frequency Analysis, which uses the average values of L-Skeewness and L-Kurtosis from several stations in an area. The goodness of fit for the observed data is determined by comparing the values against the fitted regional data (Millington *et al*., 2011).

Many statistical distributions have predetermined relationships between L-Skeewness and L-Kurtosis (Z_3 and Z_4). These are useful and necessary for creating L-Moment Ratio Diagrams, to visually inspect which distribution has the best fit (Hosking and Wallis 1997). One of the main applications of L-moments is identification of the probability distribution of the observed phenomena using the L-MRD (Assefa *et al*., 2018).

A convenient way of representing the L-moments of different distributions is the L-moment ratio diagram whose axes are L-Skeewness and L-kurtosis. A two-parameter distribution plots as a single point on this diagram, three-parameter distributions as a line, and distribution with more than three-parameters generally cover two-dimensional areas on the graph (Hosking and Wallis, 1997). For a given region, the sample L-moment ratios Z_3 and Z_4 for each station as well as their regional average are plotted on the L-moment ratio diagram. A suitable parent distribution is that which averages the scattered points closely (Mishra *et al*., 2009).

2.10. Selection of Parameter Estimation Method

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 may not yield good estimates. Hence, some guidance is needed for estimation methods.

i) Method of Moments (MOM)

It is one of the most commonly used methods of estimating parameters of a probability distribution. The estimates of the parameters of a probability distribution function are obtained by equating the moments of the sample with the moments of the probability distribution function. It provides simple calculation, but higher order moment estimates are biases. Parameter estimation by MOM is known to be biased and inefficient especially with three-parameter distribution but it is more preferable for two parameter distribution types.

ii) Maximum Likelihood Method (MLM)

In this method, the parameter estimates are determined by maximizing the sample log likelihood function. The unknown parameters may be obtained by setting each of the partial derivatives with respect to each parameter equal to zero and solving the resulting equations simultaneously. The equations are usually complex as a result of this difficulty; the solution set may not properly found (Cunnane, 1989).

iii) Probability Weighted Moments (PWM) methods

PWMs are useful in deriving expression for the parameters of distributions whose inverse forms X=X (F) can be explicitly defined. In particular, they allow parameter estimates to be obtained for distributions. Methods of parameter estimation are obtained in this method by equating moment of the distribution with the corresponding sample moment of observed data. 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 method of maximum likelihood (ML). Indeed, in small samples PWM may be as efficient as ML. With a suitable choice of distribution PWM estimation also contributes to robustness and is attractive from that point of view (Cunnane, 1989).

2.11. Quantile Estimation of Low Flow

The estimated parameters for specific probability distributions are used to calculate quantiles of low flows for different return periods. This is carried out by using the distribution function, in which the parameters of the distribution were replaced by their estimates and the relationship between return periods, T and probability of exceedance, F (Assefa *et al*., 2018). 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).

2.12. Base Flow

2.12.1. Definition and Introduction

Base flow is an important component of a ground water system. It is the component of stream flow that is attributed to ground water discharge and other delayed sources such as snow melt into streams (Santhi *et al*., 2008). Base flow is generally considered that component of flow, which originates from stored sources within the catchments, principally groundwater aquifers, as distinct from that flow component which is rapidly transmitted to streams following rainfall events.

In order to determine the contribution from overland flow in a watershed to the streams in the watershed, it is necessary to separate out the base flow from stream gage data. Reay et al. (1992) found that neglecting base flow (shallow ground water discharge) as a nutrient source to streams leads to misinterpretation of data and mismanagement. Stuckey (2006) infers that studies estimating base flow contributions to streams are useful for watershed planners to determine water availability, water use allocations, assimilative capacity of streams and aquatic habitat needs (Santhi *et al*., 2007).

2.12.2. Base Flow Separation Methods

In general, base flow is estimated through hydrograph analysis by separating stream flow into surface runoff and base flow. The separation is often estimated by using standard analytical methodologies or tracer techniques or a mass balance approach (Santhi *et al*., 2007). Exact separation of the stream flow hydrograph into surface flow and ground water flow is difficult and time consuming, especially, if there is a need to deal with regional scale studies. In addition, while such separation methods are valuable in indicating regional trends in the base flow and surface flow, they require long-term continuous stream flow data without missing. A number of methods have been proposed for separating the direct runoff and base flow. The method selected for any one-watershed analysis will depend on the type and amount of measured data available, the necessary accuracy for the design problem, and the effort that the modeler wishes to expend (McCuen, 1998).

For example, in simple graphical approach, a hydrograph is plotted on a semi logarithmic scale and the groundwater recession curve can be identified as an approximately straight line, assuming groundwater flow can be approximated with linear reservoir concept. A simple straight line of this point with the time at the beginning of the flood event is used to separate the base flow during a flood event. This method is based on the assumption that the base flow response is significantly slower compared to the surface runoff, which is not always the case in mountainous areas (Gonzales *et al.*, 2009).

Pettyjhon and Henning (1979) formulated three base flow separation methods with the objective of processing long records of groundwater discharge data: (i) fixed interval (ii) sliding interval, and (iii) local minimum methods, which are also called filtering separation methods. Basically, these methods take the minimum values of the hydrograph within an interval by following different criteria and connect them; the discharge under the constructed line is defined as base flow accordingly.

2.13. Previous Studies on RLFFA in Ethiopian River Basin

On the researcher's level of knowledge, no research has been done on the same topic of regional low flow analysis in the Upper Awash river basin; rather Habtamu Ketsela (2017) established Regional Flood Frequency Analysis for Awash basin that can have a relation with this research paper in some degree. He used the application of index-flood method and regionalizes the basin into five regions. Tatek (2015) has done estimation of low flow quantiles in ungauged river catchments on Dedessa river basin. He has regionalize the basin in to two regions as Region one and two and select Wakeby and Generalized Pareto distributions as the best fit distribution for Region one and two respectively.

A title called Low flow frequency analysis for Abay River basin was done by Tegenu zerfu, in 2009 with the main objective of delineating the Abay river basin into hydrological homogeneous regions, which would form the basic units to form and develop frequency curves for each region. Low flow frequency analysis was done for the Abay river basin using 7-day annual minimum flow series and six homogeneous regions were delineated using a simple test based on the variability of at-site values of Cv. In this research, statistics derived from observed low flow data from different stations are used for regionalizing and delineation of the basin into homogeneous region.

Regional Low flow analysis has been done for the Abaya- Chamo river sub-basin by Melaku in 2008, using statistical distribution technique. Accordingly, three homogeneous regions were delineated using a simple test based on the variability of at-site values of Cv.

3. MATERIALS AND METHODS

3.1. Description of the Study Area

The Awash River basin with a total area of 110,000 km^2 drains the northern part of the rift valley in Ethiopia. The river rises at an elevation of about 3000m in the central highlands, West of Addis Ababa and flow north-east wards along the Rift Valley. The main river length is about 1200 km. The Awash basin is geographically located in between 38^{o}E to 43.50^{o}E longitudes and 8^{o}N to 12.2^{o}N latitudes (Mengistu and Sivakumar, 2017).

The Upper Awash Basin lies in the Ethiopian highland plateau in elevation ranging from 1500 to 3000m above sea level. It is located upstream of Koka dam which covers the river section from its source up to Koka Reservoir. The Upper Awash River drains a catchment area close to 11,700 km^2 and the length of the river up to Koka is around 220km. The major tributaries to the upper Awash are Akaki and Mojo River. Akaki River starts from the mountainous areas of the northern part of Addis Ababa and join the main Awash River between Melka-Kunture and Melka-Hombole gauging stations. Mojo River, the other main tributary to Awash, originates from the high lands northeast of Addis Ababa. It drains a catchment area close to 1,900 km^2 and travels a total length of about 105 km before joining Awash. The other major tributaries of Awash above Koka are Akaki, Holeta, Berga and Legedadi rivers, contributes a significant proportion of the flow in the river. It is located approximately at 8°12’59.39’’N to 9°18’00.64’’N, latitude and 37°56’41.73’’E to 39°16’53.09’’E, longitude (Henock *et al*., 2008).

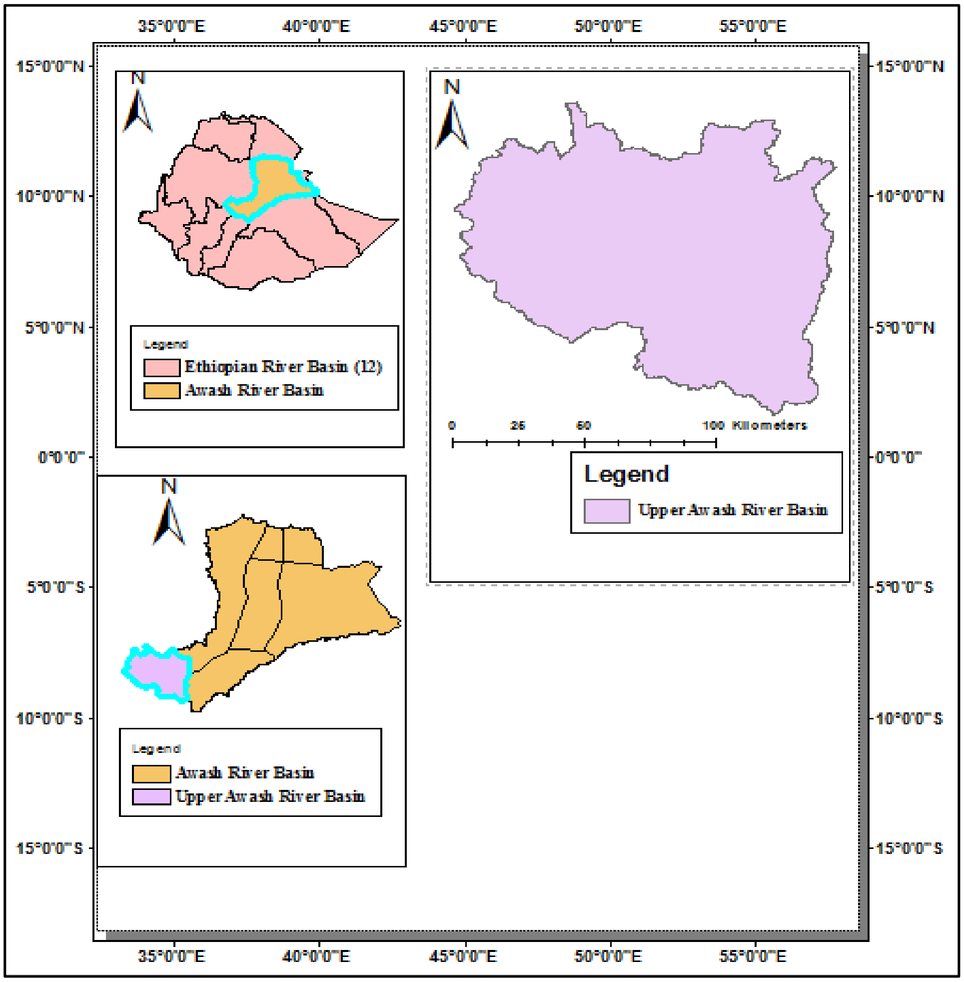


Figure 3.1 Study area map

3.1.2. Climate and Hydrology

The climate of the Upper Awash Basin comes under the influence of the Inter-Tropical Convergence Zone (ITCZ). This zone of low pressure marks the convergence of dry tropical easterlies and the moist equatorial westerlies. The seasonal rainfall distribution within the basin results from the annual migration of the ITCZ. In March, the ITCZ advances across the Basin from the south, bringing the small or spring rains. In June and July it reaches its most northerly location beyond the Basin which then experiences the heavy or summer rains. It then returns southwards during August to October, restoring the drier easterly airstreams which prevails until the cycle repeats itself in March (Halrcow, 1989). The altitude of Upper Valley of the Awash River is about 3 km, and rises on the high plateau to the west of Addis Ababa. Generally, plateaus between 3000m and 2,500m receive 1,400-1,800 mm/yr and regions with altitudes ranges from 1600 to 2500m receive 1000-1400mm/yr. The rainfall distribution is bimodal in this region, with a main rainy season from June to September and the short rainy period in March and April. Although the rainfall intensity is high in the region, the potential evapo-transpiration (PET) in the Upper Valley is higher, for instance at Koka is 1810 mm almost twice of the annual rainfall (Henock *et al*., 2008). The mean annual temperature ranges from 20.8°C to 29°C at Koka. The mean annual wind speed at Addis Ababa is 0.9m/s. The Awash basin experiences 2700 hours of sunshine annually. The monthly variation closely follows the rainfall pattern as would be expected with more sunshine hours in the dry months than in the wet months. Sunshine hours vary from a daily mean of 9.4 hours in December to 3 hours in July at Addis Ababa. The mean annual relative humidity of the basin is 60.2% measured at Addis Ababa. The monthly variation in relative humidity at Addis Ababa ranges from 50.9% in March to 78.5% in August (Gobena, 2010).

A watershed of the Upper Awash sub-basin (u/s of Koka) is usually a complex and heterogeneous system. Its characteristics vary in space. Hydrologic processes vary both in space and time. One way to account, at least partly, for spatial variability of governing hydrologic factors is to divide the watershed in to sub-catchments depending on the soil type, vegetation, land-use and topography that significantly affect stream flow. Mean annual runoff which enters to Lake Koka is approximately about 1660Mm^3. The mean annual stream flows is 1.9 m^3/s (Henock *et al*., 2008).

In the Koka sub basin there are about 21 gauging stations that records the flow and lake depth in the sub-basin out of this gauging stations only 10 are used in this research.

3.1.3. Land use and Soil type

The land use condition in the Upper Awash catchments includes mainly of cultivated agricultural land, grassland, and forest land, rural and urban settlements. It is estimated that 67% is intensively cultivated, 25.5% is moderately cultivated, 4.5% is bush land or shrub land or wooded grassland, and 3% is urban area and alpine vegetation. Main crops grown are teff, beans, wheat, barley and oil seeds. Other commercial farms produce fruits and vegetables (Henock *et al*., 2008). In the upper most part where there is high rainfall, land use is complete in May with barley and teff. Steeper slopes are heavily wooded with natural acacia and eucalyptus. On the lower most part, however, rainfall is too unreliable and the sparse dry acacia scrub gives way to wide stretches of bare ground with clumps of coarse grass and occasional thickets of acacia. The soil type in the Upper Awash sub-basin is diverse. The most common soil types are Clay, Sand, Clay-Loam, Silt-Clay-Loam, Sand-Clay and Silt-Clay.

3.2. Materials Used

The tools and some important Software that used for this particular study includes:-

a) Microsoft Excel Spread sheet and XLSTAT2018

MS-excel and XLSTAT2018 were used for transposing daily data, infilling missing data and to calculate the various statistical parameters of available hydrological data.

b) ArcGIS10.4.1

Were used to generate the study area map, for representing geographical location of gauging stations and delineate hydrologically homogeneous regions.

c) Easy fit5.6 Software

This software was used in performing statistical analyses of hydrologic data.

A statistical Easy fit software (trial version) was used for Selection of suitable probability distribution for each regions and to estimate goodness of fit and descriptive statistics of each region.

d) MATLAB R2013a

MATLAB R2013a program was used to test the discordancy of sites from the regions identified.

e) BFI+3.0

A program called BFI+3.0 is used to perform a separation of the base flow from the total stream flow and calculate the base flow index.

3.3. Methodology and procedure

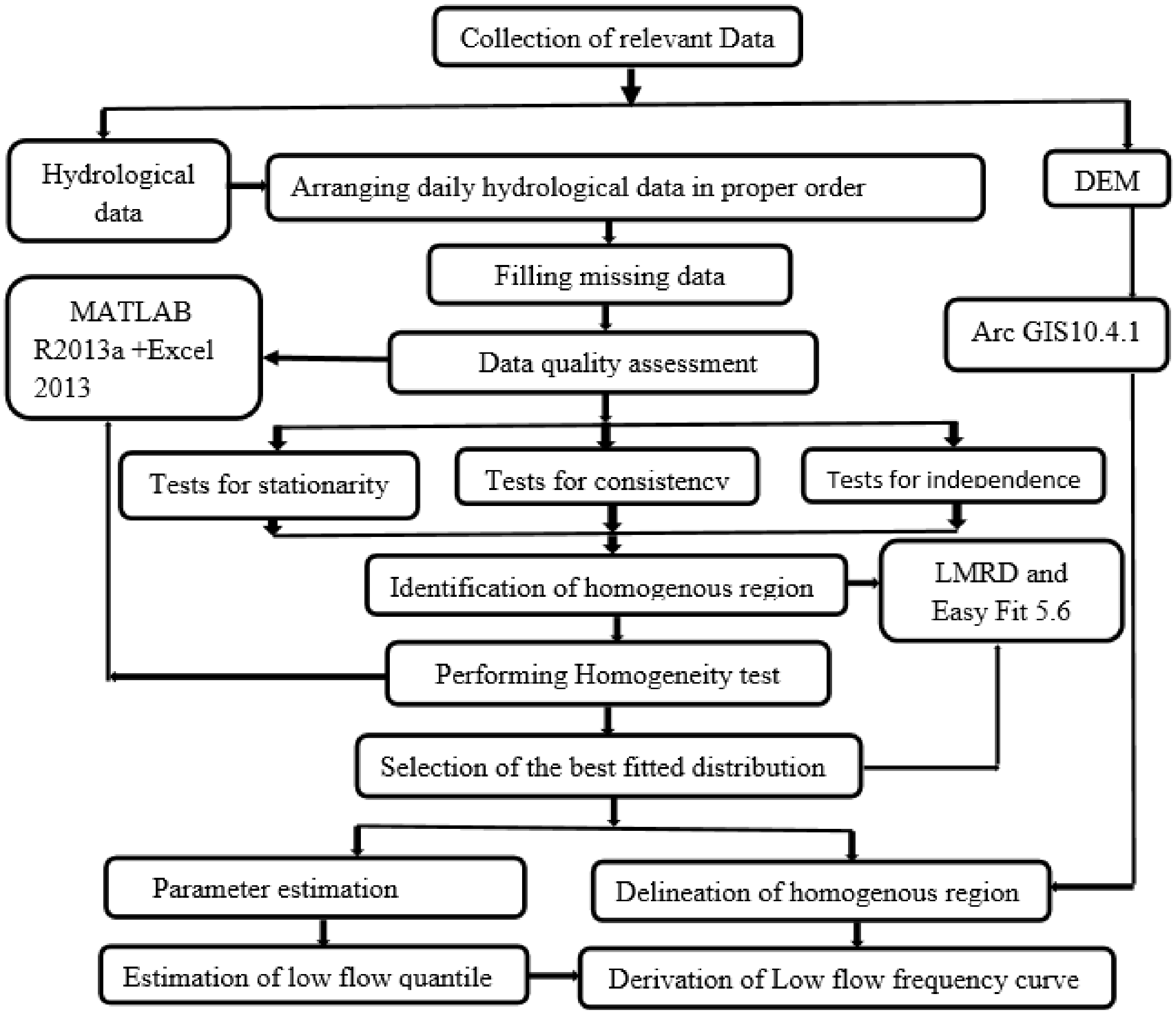


Figure 3.1 Flow chart of methodology

3.4. Source and availability of data

Data needed for this study were collected from different organization in order to use for regional frequency analysis. In the Awash basin there are about 67 gauging stations, out of these 21 stations are found in the Upper Awash sub-basin particularly in the upstream of Koka Dam. These all stations are not used for the analysis because some of the stations have a very short period of data due to the influence of different conditions. The stream flow data used for the analysis was obtained from the Ministry of Water Irrigation and Electricity (MoWIE), Hydrology department. Among the functional stream flow stations in the basin 10 river gaging stations were selected.

The selection was based on;

1) The long record of data, less missing records and functionality and
2) Consideration of the spatial distributions of gaging stations and sub basins in the basin.

The DEM size of 30x30 of the sub-basin is the most important input to generate different basin out puts and to regionalize the sub-basin; that is collected from MoWIE, GIS department.

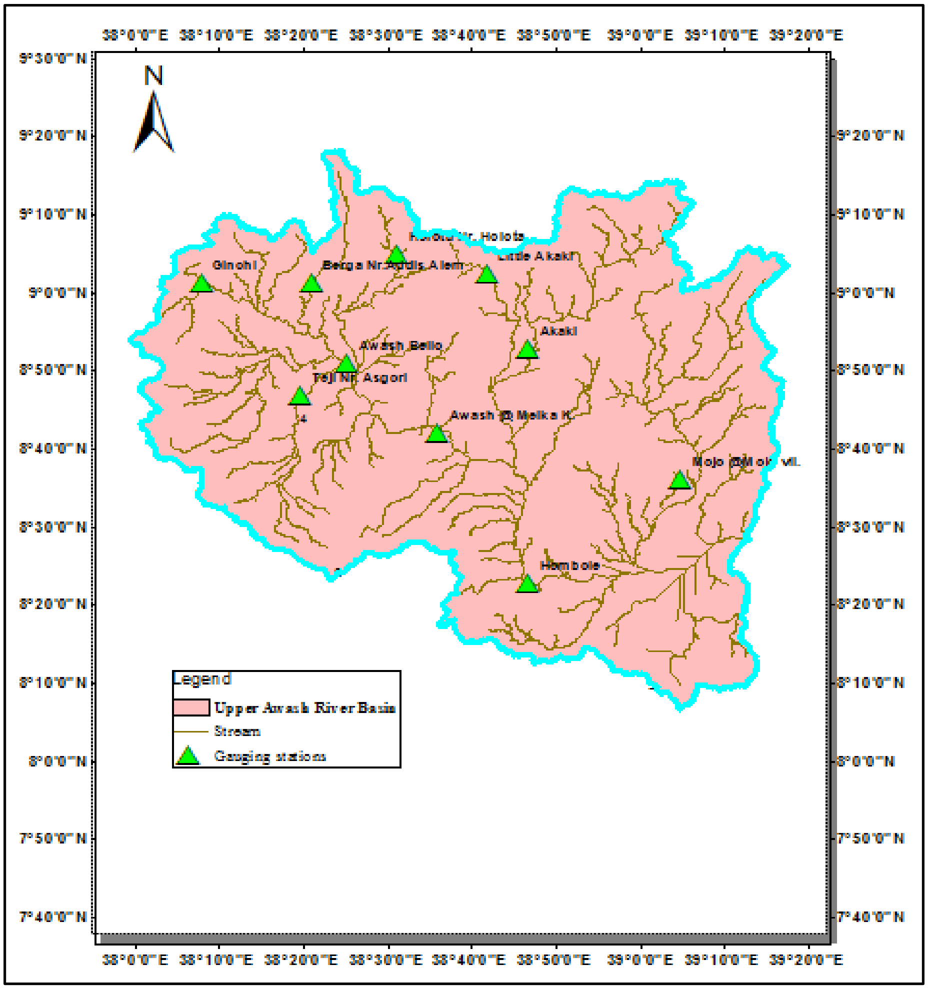


Figure 3.2 Location of Gauging station

The stations are summarized in table 3.1 and only ten (10) gauging stations with a minimum of 16 years record length and above were used for the analysis. The stations consist of daily flow series, which are used to produce different duration of low flows, flow duration curves, base flow, base flow index and also used for low flow frequency analysis for the whole region.

Table 3.1 General characteristics of the station in the Upper Awash River Sub-basin

Station Name	Station Code	River Length (km)	Catch. slope (%)	Lat.(N) (UTM)	Lon.(E) (UTM)	Year of Record
Berga Nr. A. Alem	31001	13.485	17	996,668	428,556	23
Holeta Nr. Holeta	31002	14.97	76	1,004,010	446,886	20
Teji @ Asgori	31003	24.142	61	970,876	426,678	18
Awash @ Bello	31020	32.383	35	978,231	435,855	25
Akaki @ Akaki vil.	31004	44.799	68	981,872	476,159	16
Mojo @Mojo vil.	31014	42.526	96	950,545	509,170	16
Awash @Hombole	31013	106.151	33	950,551	476,159	20
Awash @ M. Kun.	31012	57.353	30	961,622	455,998	20
Little Akaki	31021	21.34	56	466,795	999,047	16
Awash @ Ginchi	31033	58.36	24	404,697	997,173	18

3.6. Method of data analysis

3.6.1. Missed data filling

When undertaking an analysis of stream flow data from gauges where observations are made, it is often to find times where no observations are recorded at one or more gauges due to many reasons such as the absence of recorder, carelessness of the observer, break or failure of instruments. Therefore, it is often necessary to estimate these missing records (sine and Ayalew, 2004). Different methods such as arithmetic mean, graphical correlation method, normal ratio method and linear regression method are used for filling of missed flow data for a given gauging station. All selected 10 gauging stations have missed data and only Hombole and Teji stations have a missing of less than 0.05%; As a result, both regression and arithmetic mean analysis is used to fill the missing instantaneous daily flow data with satisfactory correlation coefficients.

The model performance can be good if the correlation coefficient (r) is between 0.6 and 1. The results of correlation between stations were indicated below and results all selected stations are well correlated and performed. It is an indication for the accuracy of the equation that was tested on different gauging stations of the basin.

Table 3.2 Filling accuracy and regression equation of the station

Station Name	Regretted with	Equation	R^2	Remark
Berga Nr. A. Alem	Holota	Q2=0.575Q1+0.012	0.70	WC
Holota Nr. Holota	Little Akaki	Q2=1.38Q1^2-0.494Q1+0.063	0.72	WC
Akaki@Akaki Vil.	Awash@M.Kunture	Q2=1.013Q1+1.069	0.68	WC
Hombole@Hombole		Filled with Arithmetic mean	0.94	WC
Mojo@Mojo vil.	Hombole@Hombole	Q2=-0.0694Q1+0.3407	0.62	WC
Awash@M.Kunture	Akaki@Akaki Vil.	Q2=0.666Q1-0.481	0.68	WC
Little Akaki	Akaki@Akaki Vil.	Q2=0.039Q1^2-0.127Q1+0.138	0.78	WC
Awash @Bello	Berga Nr. Addis Alem	Q2=-15.395Q1^2+6.264Q1-0.237	0.84	WC
Awash @Ginchi	Berga Nr. Addis Alem	Q2=-0.303Q1+0.127	0.60	WC
Teji Nr. Asgori		Filled with Arithmetic mean	0.92	WC

3.6.2. Data quality Analysis

Some errors may exist in the stream flow observation that is collected, such as misplaced decimal numbers, very huge unrealistic numbers and negative flow records in some cases. Checking the data quality of the K day sustained low flow data series was vital as it enhances the analysis. Some of the common methods to assess the data quality carried out before the low flow analysis includes outliers, stationarity, consistency and independency (Assefa *et al.*, 2018).

3.6.2.1. Randomness and Independence of data

One of the basic assumption in flood frequency analysis are the independence of the data series (Rao and Hamed, 2000). Specific Statistical test for independence are incorporated in various computation in frequency analysis. When applied to short series, however, commonly used tests can be misleading: they may indicate non-independence when the events are actually independent, or failed to indicate it when serial correlation long lags is in fact present. According to Guru and Jha (2016), the randomness test is needed to find independent AM series from all the data sets values at each station. For this study, the lag-1 correlation coefficient was used to verify the independence of data of the selected station. It describes the strength of the relationship between a value in a series and that preceding it by one time interval.

According to Dahmen and Hall (1990), the lag-1 serial correction coefficient, R1, is given by:

$$\mathrm{R1} = \sum_{i=1}^{n} \frac{(Xi-\dot{x})(\mathrm{Xi+1}-\dot{x})}{\sum_{i=1}^{n}(Xi-\dot{x})^{2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.1)$$

Where Xi and ẋ are an observation and the mean of observation

X_{i+1} is the next observation and n is the number of data

After the computation of R1, the test hypothesis is that H0: R1= zero against the alternative hypothesis, H1: R1<> 0. The null hypothesis, H0 determines that there is no correlation between two consecutive observations.

Anderson (1942) defines the critical region, R1 at the 5% level of significance as: (−1, (LCL) R1 (UCL), 1) where the upper confidence limit, UCL for R1 is given by

$$\mathrm{UCL(R1)} = \frac{(-1 + 1.96(N-2)^{0.5})}{N-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (3.2)$$

The lower confidence limit LCL, for R1 is also defined as;

$$\mathrm{LCL\ (R1)} = \frac{(-1-1.96(N-2)^{0.5})}{N-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots.(3.3)$$

To accept the hypothesis H0: R1=0, the value of R1 should fall between the UCL and LCL. Applying this condition, The result for the test: LCL (R1) <R1< UCL (R1) is satisfied for the nine stations and one station is found to be dependent.

Table 3.3 Independence test of each station

Station	R1	UCL(R1)	LCL(R1)	Remark
Berga Nr. Addis Alem	0.376	0.363	-0.454	dependent
Holota Nr. Holota	-0.108	0.385	-0.49	Independent
Akaki@Akaki Vil.	0.209	0.422	-0.556	Independent
Hombole@Hombole	0.347	0.370	-0.465	Independent
Mojo@Mojo vil.	0.211	0.422	-0.556	Independent
Awash@M.Kunture	0.134	0.385	-0.49	Independent
Little Akaki	0.294	0.422	-0.556	Independent
Awash @Bello	0.338	0.442	-0.415	Independent
Awash @Ginchi	-0.070	0.402	-0.52	Independent
Teji Nr. Asgori	0.400	0.402	-0.52	Independent

3.6.2.2. Consistency and Stationarity of data

Before conducting the analysis, the series should be scrutinized for possible errors or inconsistency and for any indication that contravene basic statistical assumption (Rao and Hamed, 2000). The F-test for stability of variance and t-test for stability of the mean verify not only the stationary of a time series, but also its absolute consistency and homogeneity (Dahmen and Hall, 1990). Thus, these two sets are adopted to check stream flow observations stationarity and consistency.

a) F-test for the stability of variance

The test statistics is the ratio of the variance of two split, non-overlapping subsets of the series (Dahmen and Hall, 1990). The 7-day mean annual minimum stream flow observations are divided into equal or nearly equal series and the variance of both time series is computed by:

$$Ft=\frac{Variance\ of\ time\ series\ 1}{Variance\ of\ time\ series\ 2}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..\ (3.4)$$

According to this method, the variance of the time series is stable if and only if:-

F {V_1, V_2, 2.5%} <Ft < F {V_1, V_2, 97.5%} Where V_1 =n_{1-1}, V_2=n_{2-1} and n_1= n_2=the number of observation point in each subsets.

b) T-Test for stability of mean

The test for stability of the mean involves computing and then comparing of the mean of non-overlapping subsets of the time series (Dahmen and Hall, 1990). Similar subsets with F tests will also used for the computation of this test.

$$Tt=\frac{\bar{x}series1-\bar{x}series2}{\left((n1-1)s1^2+(n2-1)s2^2*\frac{1}{n1+n2-2}*\left(\frac{1}{n1}+\frac{1}{n2}\right)\right)^{0.5}}\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots...\ldots\ldots..\ (3.5)$$

Where x̄, is the mean of the series; n is the number of yearly stream flow records and s_1 and s_2 are the standard deviation of series1 and 2 respectively.

According to this test, the mean of the time series is stable if and only if:

t {V, 2.5%} <Tt < t {V, 97.5%}

Note that, both t{V, 2.5%}, t {V, 97.5%} and F {V1, V2, 2.5%}, F {V1, V2, 97.5%} critical values for 5% significance levels are putted at an Appendix-B and Appendix C respectively. The respective result of both hydrological data quality test is presented on Appendix-D. Accordingly, all 10-selected station for computation shows stable mean and two stations, Berga Nr. Addis Alem and Awash @ Teji shows unstable variance.

C) Tests 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 outlier 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 (Rao and Hamed, 2000). Outliers can be excluded from the estimation procedure only if it is certain that annual minimum flows can be adequately modelled by a single distribution form (Cunnane, 1989).

Outlier test is not done in this study, but to avoid the effects of outliers, an efficient method of parameter estimation method like L-moment was employed.

3.7. Flow Duration Curves

The aim is to estimate how much flow is available at particular site and how much reliable. The degree of certainty of the available flow at a particular site could be well described by using flow duration curve. During the study, flow duration curves of 1,7,14 and 30 days mean annual minimum flow have been plotted. For low flow, the flow duration curve has been developed by ranking the data in ascending order regardless of the sequence in which they occurred and the probability of exceedence for each data has been developed using the most commonly used formula known as Weibull plotting position (Joshi and Hillaire, 2013).

$$\mathrm{Pi} = \frac{i}{N+1} * 100 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.6)$$

Where, i is the rank of each data, N is the number of data and P_i is the plotting position.

The flow measurement versus the probability of non-exceedence is plotted to get the flow duration curve.

3.8. Low flow Regionalization of Upper Awash River Basin

From many types of regionalization procedures available, an index low flow method described in the Rao and Hammed (2000) is used in this study which assumed the distribution of low flow at different sites in a region is the same except for a scale or index low flow parameter, which reflects a rainfall or runoff characteristics of each region. Statistical values of each station under analysis (Lcv, Lcs and Lck) for the flow is computed using L-moment. L-Moments are analogous to conventional moments but are estimated by linear combinations of an ordered data set, namely L-statistics (Rao and Hamed, 2000).

3.8.1. Identification and Delineation of Homogeneous Region

The Lcs and Lck of standardized flow values at each station have been plotted on the LMRD, together with various theoretical distribution functions, those stations close to a particular theoretical distribution line are considered to be homogeneous stations and are grouped together (Gebregiorgis *et al*., 2013). The regionalization process is further statistically verified by using discordant measures and homogeneity tests to assess the degree of variability within the pool.

Furthermore, the method is proven by comparing the measure of scale and dispersion value of both the L-moment (LCv) and conventional moment (Cv) of gauging stations that belong to different regions. The clustering of sites into homogeneous regions were carried out by applying the hierarchical geographic regionalization technique with the method of L-moments as a guideline for regionalization.

The delineation of homogenous regions is performed by using the geographical information system (GIS) software ArcMap10.4.1. As an input, DEM size of 30mx30m of Upper Awash River Basin (UARB) were used to identify the site characteristics and delineate the homogenous region. Delineations are normally done using Arc GIS software of Arc view catalogue, as it is difficult to perform it using Arc Map catalogue if the stations in the homogenous regions have no common outlet unlike in this study.

3.8.1.1. Characteristics of the site

Site characteristics are important inputs for preliminarily identify the homogenous region. Sites having the same or nearly the same characteristics will lie on the same region. In this study site characteristics such as latitude, longitude and elevation of the station are used.

3.8.2. Regional Homogeneity Test

The hypothesis of the homogeneity test is that the at-site frequency distributions are identical except for a site-specific scale factor. The preliminary identified regions have to be checked by various homogeneity tests. From the various homogeneity tests, the tests used in this study are measure of scale, dispersion based tests (Cv-based homogeneity test and LCv-based homogeneity test) and discordance measure tests. The most rigorous L- (Hosking and Wallis, 1997) moment based test of homogeneity is that of Hosking and Wallis (1993). It compares the variability of the L-moment ratios of the sites with in a region with the expected variability obtained from simulation from a collection of sites with the same record length as their real world counterparts.

3.8.2.1. Cv-based homogeneity test

To investigate whether the assumption in regionalization have been met, Many researchers such as Cunnane (1989) have used the values of the mean coefficient of variation (Cv) and the at-site COV (CC) of both conventional and L-moments of the proposed region. The criterion used to check for regional homogeneity was based on the value of CC. According to the researchers, the higher the values of Cv and CC, the lower the performance of the index flood method for the considered region. Hence, for better performance of the index flood method, CC should be kept low, usually less than 0.3.

For each site in a region calculate mean ($\bar{Q}$), standard deviation (δ) and coefficient of variation (CV) using Nobert *et al*. (2014).

$$\bar{Q} = \frac{1}{n}\sum_{i=1}^{n} Qi \tag{3.7}$$

$$\delta_i = \sqrt{\frac{\sum_{i=1}^{n}\left(Qi-\bar{Q}\right)^2}{n}} \tag{3.8}$$

$$Cv_i = \frac{\delta i}{\bar{Q}} \tag{3.9}$$

Where: Qi= the flow rate the site in region i, $\bar{Q}$= the mean flow rate at site i,

δi= Standard deviation for the region at site i, Cvi= Coefficient of variation of site i

Using the statistic calculated above, the regional mean, Cvi and finally the corresponding CC value using the following relation will be calculated.

$$\overline{C_v} = \frac{\sum_{i=1}^{n} Cvi}{n} \tag{3.10}$$

$$\delta_{cv} = \sqrt{\frac{\sum_{i=1}^{n}\left(Cvi - \overline{C_v}\right)^2}{N}} \tag{3.11}$$

The weighted regional Cvi of all the sites, CC is defined as;

$$CC = \frac{\delta cv}{\overline{C_v}} \tag{3.12}$$

Where: n=Number of site in a region, $\overline{C_v}$=average coefficient of at site Cv values

δcv = Standard deviation of at site Cv values

The region declared to be homogenous if CC<0.3

3.8.2.2. LCV-based Homogeneity test

The Homogeneity test based on L-CV, proposed by Hosking and Wallis (1997), is nowadays routinely used in regional analysis**.** L-CVs are used because between-site variation in L-CV has a much larger effect than variation in the L-CS or L-CK on the variance of the estimates of quantiles, except those in the far tail of the distribution. Sample estimates of L-Cv are so strong that they are not affected by the presence of outliers in the data set and characterize most of probability distributions than conventional moments. The procedural calculation is the same as that of the CV.

The unbiased estimator of β_0, β_1, β_2 and β_3 are defined by (Hosking, 1990) as

$$\beta_0 = \frac{1}{n}\sum_{i=1}^{n} Qi \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.13)$$

$$\beta_1 = \sum_{i=1}^{n-1} \frac{(j-1)(Qi)}{n(n-1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.14)$$

$$\beta_2 = \sum_{i=1}^{n-2} \frac{(j-1)(j-2)(Qi)}{n(n-1)(n-2)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.15)$$

$$\beta_3 = \sum_{i=1}^{n-3} \frac{(j-1)(j-2)(j-3)(Qi)}{n(n-1)(n-2)(n-3)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.16)$$

Where Q_i is the yearly flow data, N – is the number of year, J – is the Rank and β_o, β_1, β_2 and β_3- are L-moment Estimators.

The first few moments are:

$\lambda_1 = \beta_O$, $\lambda_2 = 2\beta_1 - \beta_O$, $\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_O$, $\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta O$

L-moment ratios are the quantiles

$z_2 = \frac{\lambda 2}{\lambda 1}$ and $z_r = \frac{\lambda r}{\lambda 2}$, for $r \geq 3$ (i.e $z_3 = \frac{\lambda 3}{\lambda 1}$ and $z_4 = \frac{\lambda 4}{\lambda 1}$)

In particular, $\lambda 1$ is the mean of the distribution, a measure of location; $\lambda 2$ is a measure of scale; Z_3 and Z_4 are measures of skewness and kurtosis, respectively. The Lcv, $Z_2 = \lambda 2/\lambda 1$, is analogous to the usual coefficient of variation. Using the formula above, we have;

$$\bar{L}_{Cv} = \frac{1}{n}\sum_{i=1}^{n} Lcvi \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.17)$$

$$\delta_{CV_i} = \sqrt{\sum_{i=1}^{n} \frac{\left(Lcvi - \overline{\mathrm{Lcv}}\right)^2}{n-1}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.18)$$

$$CC = \frac{\delta cvi}{\overline{\mathrm{Lcv}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.19)$$

The region declared to be homogenous if CC<0.3.

3.8.2.3. Discordancy Measure

The discordance measure is used to identify those sites that are grossly discordant with the group as a whole. The discordance measure D estimates how far a given site is from the center of the group based on statistical properties (Rao and Hamed, 2000). A suitable criterion to classify a station as discordant is when D is greater than or equal to 3.

If $U_i = [z_2^{(i)}, z_3^{(i)}$ and $z_4^{(i)}]$ is the vector containing the z_2, z_3 and z_4 values for site (i), then the group average for NS sites is given by;

$$\bar{U}i = \frac{1}{NS}\sum_{i=1}^{NS} Ui \qquad (3.20)$$

The sample covariance matrix is given by equation 3.21;

$$S = (NS-1)^{-1}\sum_{i=1}^{NS}(Ui-\bar{U}_i)(Ui-\bar{U}_i)^T \qquad (3.21)$$

The discordancy measure is defined by;

$$Di = \frac{1}{3}(U_i - \bar{U}_i)S^{-1}(U_i - \bar{U}_i)^T \qquad (3.22)$$

Where NS- is the total number of site, Di- Discordancy measure,

Ui- is a vector containing the L-moment ratio for site i

$\bar{U}$i- is the group average for NS sites, and S- is the sample covariance matrix of Ui.

For this study, MATLAB R2013a program was used to simplify the numerical calculation of discordancy index, as a simple matrix multiplication was difficult and quite complex. After determining the index, a critical value tabulated by Hosking and Wallis (1997) at a significant level of 10% was used to identify the discordant region. The Hosking and Wallis critical values and the programming code used to calculate the discordancy index were given on the Appendix-E and F respectively**.**

3.8.3. Fitting the probability distributions

The true probability distributions of low flows are unknown and the practical problem is to identify a reasonable functional distribution and estimate its parameters (Joshi and Hillaire, 2013).

According to Smakhtin (2001), a universally accepted distribution for low flow analysis is unlikely to be identified. In this study, 10 different distribution models are considered and the selection of the best-fitted distribution were executed using a plot of L-MRD and Easyfit software.

3.8.3.1. L-Moment Ratio Diagram (L-MRD)

L-MRD plot as fairly well separated groups and permit better discrimination between the distributions (Assefa *et al.*, 2018). L-MRD (Lcs versus L_{Ck}) was used to identify the appropriate distribution for each region. For a given region, the sample regional average moment ratios, Z_3, and Z_4 analogues to Lcs, and Lck, respectively, are plotted on the LMRD to select the suitable distribution model. In this study, the selected distributions are Generalized extreme value (GEV), Logistic, Generalized Pareto (GPA), Normal, Uniform, Weibull, Lognormal (LN) and Log Pearson type 3 (LPIII).

Step by Step L-Moment Ratio Diagrams

i. create a table containing L-Skewness and L-Kurtosis values for each data set

ii. Plot L-Skewness against L-Kurtosis of the observed data sets

iii. Plot L-Skewness against L-Kurtosis of the given distributions, and visually compare the plot

3.8.3.2 Easyfit Software for distribution fitting

In this study, selection of the best-fit distribution from candidate distribution obtained from the above-mentioned analysis (L-MRD) is done by using software, which is called Easy Fit. Within this software all goodness of fit test such as Chi-Square, Kolmogorov Smirinove and Anderson Darling tests are done and the best-fit distribution from the candidate is displayed automatically.

3.8.4. Goodness of fit test

Many distributions have been suggested for AMN series models. Some of them are used for this particular study. When a theoretical distribution has been assumed, the validity of the assumed distribution may be verified or disproved statistically by goodness of test.

The results of the goodness of fit tests are used to select a distribution for frequency analysis of stations. To check the fit of probability distributions in this study, the three most commonly used tests of goodness of fit namely Chi-Square (CS) test, Anderson Darling test and Kolmogorov-Smirnov (KS) tests were applied to the data series. All test statistics were defined and carried out at 5% significance level (Ashraful et al., 2018). The goodness of fit tests were executed in the downloadablesoftware Easyfit, available at http://www.mathwave.com/easyfit-distribution-fitting.html.

i). Kolmogorov-Smirnov Test

The test statistic in the Kolmogorov-Smirnov test is extremely simple; it is now the maximum vertical distance among the empirical cumulative distribution functions of the two samples. The empirical cumulative distribution of a sample is the proportion of the sample values that are a lesser amount or equal to a known value. Kolmogorov-Smirnov (KS) test is a different and commonly used goodness-of-fit moreover Chi-square test. The advantage of the

Kolmogorov-Smirnov tests over the Chi-square test it that it is not essential to divide the data into bins; hence, the problems associated with the chi-square approximation for small number of intervals would not appear with the Kolmogorov-Smirnov test (Mengistu and Sivakumar, 2017).

The statistic Dn is evaluated by observing the deviation of the sample distribution function P(x) from the completely specified continuous hypothetical distribution function $P_0(x)$, such that

$$Dn = \underbrace{Max}_{1 \le i \le n} |P(xi) - Po(xi)| \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots. (3.23)$$

If the computed statistic is smaller than the critical value, it indicates that the distribution fits the data well and the distribution can be accepted (Assefa *et al*., 2018). The fixed values of α (0.01, 0.05) are generally used to evaluate at various significance levels. A value of 0.05 is typically used for most applications.

ii). Chi-Square test

Chi-Square goodness of fit test is a non-parametric test that is used to get exposed how the observed value of a particular phenomenon is considerably unlike from the estimated value. In Chi-Square goodness of fit test, the word goodness of fit is used to contrast the observed sample distribution with the estimated probability distribution. Chi-Square goodness of fit test determines how fine theoretical distribution (such as normal, binomial, or Poisson) fits the experimental distribution. In Chi Square 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 statistic is calculated by

$$\chi^2 = \frac{(Oj - Ej)^2}{Ej} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots... (3.24)$$

Where χ^2 -is the chi-square goodness of fit test.

O_j is the observed value in the jth class interval and E_j is the value that would be expected from the theoretical distribution. The hypothesis will be accepted if the considered value of χ^2 is less than the table value.

iii). Anderson Darling test

The Anderson-Darling test compares an observed CDF to an expected CDF. 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 Anderson-Darling test are dependent on the specific distribution that is being tested. The test rejects the hypothesis regarding the distribution level if the statistic obtained is greater than a critical value at a given significance level (α). The significance level most commonly used is α=0.05, producing a critical value of 2.5018. This number is then compared with the test distributions statistic to determine if it can be rejected or not.

3.8.5. Evaluation of the Performance of Fitting distribution

The result, which have been obtained from statistical analysis are essentially uncertain, and to be trustful, methods of uncertainty assessment should be applied (Hosking and Wallis, 1997). The authors also illustrate that "the assessment of the accuracy of the estimate should therefore take into account the possibility of misspecification of the frequency distribution and statistical dependence between observations at different sites to an extent that is consistent with the data". Hence, for this analysis, two methods of uncertainty assessments were performed. These are the plotting position i.e., the probability-probability plot and quantile-quantile plot using Easyfit software. The distribution that has the most number of points nearby to the line taken as the best-fitted probability distribution.

i. Probability-Probability (P-P) plot

P-P Plots is the variable's cumulative magnitude in opposition to the cumulative magnitude of any of a number of trial distributions. Probability plots are commonly used to determine whether the distribution of a variable matches a given distribution. If the selected variable matches the test distribution, the points come together approximately a straight line (Mengistu and Sivakumar, 2017). Probability plot are extremely helpful for visually instructive the character of a data set. Plots are a valuable way to see what the data look like and to determine it fitted distribution appears reliable with the data.

ii. Quantile-Quantile (Q-Q) Plots

Quantile-Quantile (Q-Q) plots are plots of two quantiles against each other. A quantile is a small part where certain values fall below that quantile. The purpose of Q-Q plots is to get out if two sets of data come from the same distribution. The performance of the best distribution model identified for the respective regions were evaluated by comparing observed values with simulated values. The argument was that the values that obtained by Easyfit software simulations should be matched to the particular characteristics of the data (i.e., the intersection of the values should be closed to the line 1:1). The best frequency distribution was subjected to randomly simulate the same size as observed series. Thus, the quantiles of the normalized stream flows and simulated values are plotted on one graph that represents on the x-axis and y-axis, respectively.

3.8.6 Parameter and low flow Quantile Estimation

After selecting the best-fit probability distribution for each region, the parameters of probability distribution could be estimated in a number of estimation techniques such as Method of Moments (MOM), the Maximum Likelihood Method (MLM) and the Probability Weighted Moments (PWM) methods. The simulated quantiles of each region based on candidate distributions with the selected distribution/estimation procedure are compared with the actual flow at each station. This helps to identify the most robust procedure to develop the frequency curves.

In this study, Standard error of estimate (SEE) was used as a measure of selecting the most robust parameter estimation methods. It helps to identify the best and most efficient PEM for a given distribution model. The best and most efficient PEM usually provides the lowest SEE value. Accordingly, the estimated parameters for specific probability distributions were used to calculate quantiles of low flows for different return periods. This was carried out by using the distribution function, in which the parameters of the distribution were replaced by their estimates and the relationship between return periods (T) and probability of exceedance (F) (Assefa *et al*., 2018).

In low flow frequency analysis, the assumption for the relationship between the low flows quantiles with return period was based on the non-exceedance probability indicated in Equation (3.26).

$$F = \frac{1}{T} \text{ (For flood)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.25)$$

$$F = 1 - \frac{1}{T} \text{ (For low flow)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.26)$$

For each region; the scale, location and shape parameters were calculated using the formula and then it is possible to determine the low flow quantile with different return periods using different equations for different distributions.

For GPA distribution, the flow quantile $X_{T,}$ can be estimated as-

$$x_T = \mu + \frac{\delta}{k}\left(1 - \left(1 - \frac{1}{T}\right)^{k}\right), for\ k \neq 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.27)$$

$$\mathrm{x_T} = \mu + \delta\left(\ln\left(1-\frac{1}{T}\right)\right), for\ k = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.28)$$

The low flow quantile, X_T can also be estimated by the GEV distribution as-

$$\mathrm{x_T} = \mu + \frac{\delta}{k}\left[1-\left(-\ln\left(\frac{1}{T}\right)\right)^{k}\right], for\ k \neq 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.29)$$

$$\mathrm{x_T} = \mu + \delta\left[\ln\left(-\ln\left(\frac{1}{T}\right)\right)\right], for\ k = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.30)$$

Where, δ= Scale parameter, μ= Location parameter and k = Shape parameter

3.9. Standard error of estimates

In order to have a reliable estimate of the regional quantile values, a robustness assessment has to done among the various available parameter estimation procedures. The combination of all the available parameter estimation provides us a lot of procedures, where as it will be difficult and cumbersome to fit all these procedures to a certain regional data set. Therefore, the robustness of only those distributions selected from the descriptive analysis so far from (from L-moment and Goodness of fit) and those commonly used parameter estimate are used in this paper. The proposed procedures selected for robustness assessment are GPA/MOM, GPA/MML, GPA/PWM, GEV/MOM, GEV/MML and GPA/PWM. In this study, standard error of estimate was applied for selecting the best parameter estimation method. Hussen and Wagesho (2016) used the standard error of estimate as a measure of the accuracy of predictions to perform robustness assessment. It is calculated by using the equation:

$$SEE = \sqrt{\frac{\sum_{i=1}^{n}(E-O)^2}{n-1}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3.31)$$

Where, n=the number of yearly data, E= estimated value,
SEE=standard error of estimate, O= observed value.

The performance of the parameter estimation method under test is then compared through the magnitude of SEE. The parameter estimation procedure giving the smallest values for SEE was selected to be the best parameter estimation procedure.

3.10. Derivation of Regional Frequency Curve for the Homogeneous Regions

The average of the regional growth curves was determined to represent the frequency curves of regions. For each region, the regional flow frequency curve was developed to estimate the standardized flow variations of different return periods.

3.10.1 Estimation of Index Low flow

In this study, the index low flow L-moment approach of regionalization was applied in order to carry out the regional low flow analysis. Initially the homogeneous group of stations were identified by using the L-MRD and different homogeneity test. The best probability distribution for each homogenous group was fitted by using similar procedures as discussed in above sections. Using this probability distribution, the long term standardized quantiles was estimated for various return periods such as 2, 5, 10, 20, 50 and 100 years. Plots of X_T against the Gumbel reduced Variate (-ln(-ln(1-1/T)) known as growth curves, were generated for each region and used in the derivation of the regional growth curves.

The dimensionless regional growth factor, X_T can be determined using equation (3.32) below.

$$X_T = \frac{Q_i}{\bar{Q}} \qquad (3.32)$$

Where Q_i is the low flow for T-years of return period, $\bar{Q}$ is the 7-day average annual low flow and X_T is the standardized quantile for T years return period.

3.11. Base Flow Separation Methods

To make the base-flow separation process less tedious and more consistent, a BFI+ Version 3.0 software prepared by department of Hydrogeology and Geothermal Energy, Geological Survey of Slovakia was used in this study. The methodic for module developing is based on Thallakson and Van Lannen (2004) and others listed on the user manual reference section. The module contains a local minimums and 10 other base flow separation methods to estimates the annual base-flow volume of rivers and streams and thus to computes an annual base-flow index (BFI, the ratio of base flow to total flow volume for a given year) for multiple years of data at one or more gage sites.

Among different methods discussed by Sloto & Crouse (1996) for the base flow separation, a method called local minimum method was used in this study due to the reason discussed in the literature above. According to Gonzales (2009) in his comparison of different base flow separation methods in Netherland, local minimum methods shows significant result than other methods under comparison. The BFI program uses a set of procedures in which the water year is divided into N-day period and the minimum flow during each N-day period is identified. Each fixed period minima is then compared to adjacent minima to determine turning points on the base flow hydrograph.

The local-minimum method checks each day to determine if it is the lowest discharge in one half the interval minus 1 day [0.5(2N*-1) days] before and after the day being considered. If it is, then it is a local minimum and is connected by straight lines to adjacent local minimums (Sloto & Crouse, 1996). N refers to the number of days over which a minimum flow is determined. It is the connection of these minimum points that determines the base flow.

Although the procedure developed by the department of hydrogeology recommend 5-day minimums, the author in the user manual of BFI+ on the other side recommend that N=3 is an appropriate value. To determine which N to use, in this study, a table of BFI Vs N of sample station was plotted to determine the critical value of N is more efficient than using a formula. Accordingly, N=4 is used for Awash at Bello station for example. N=3 up to 6 are used for other stations.

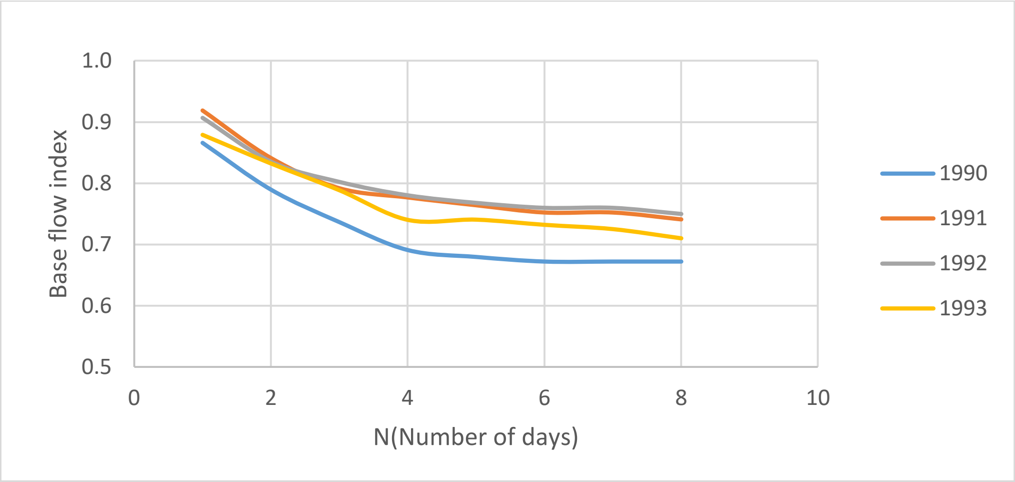


Figure 3.3 Relationship of N with Base flow index

The other parameter in the software to be determined by the user is f (turning point factor). The value of f is said to be insensitive so this can remain as a default value of 0.9.

4. RESULT AND DISCUSSION

4.1. Flow Duration Curve Analysis

In this study, the flow measurement was plotted versus the corresponding percentage of exceedence using equation (3.6) to obtain flow duration curve. During the study, flow duration curves of 1, 7, 14 and 30 days have been plotted to assess the flow regime of the gauged Rivers.

According to the nature of the FDC in the figure 4.1 and the low flow indices as in the table 4.1 and 4.2, Upper Awash river basin is categorized into group one and two. Stations located mainly in the lower parts of the sub basins such as Akaki, Hombole, Melka Kunture, Bello and Mojo having high low flow contribution are categorized as group One. Stations in the upper regions of the sub basins such as; Berga, Holota, Little Akaki, Ginchi and Teji having minimum or zero low flow contribution are categorized as Group Two.

1-day flow duration curves of all stations are presented in the figure 4.1. Each station of 1-day, 7-day, 14-day and 30-day FDC are attached in appendix D.

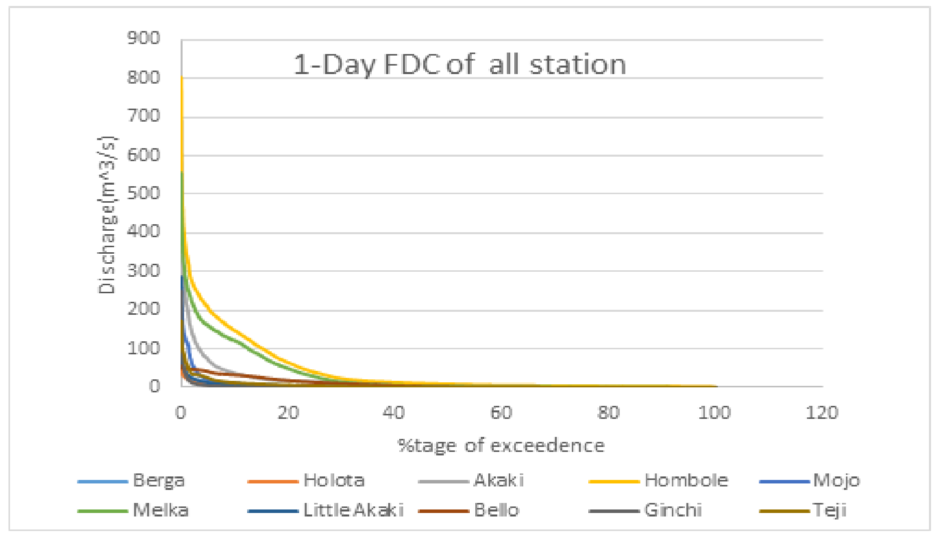


Figure 4.1 Daily Flow Duration Curve for all station

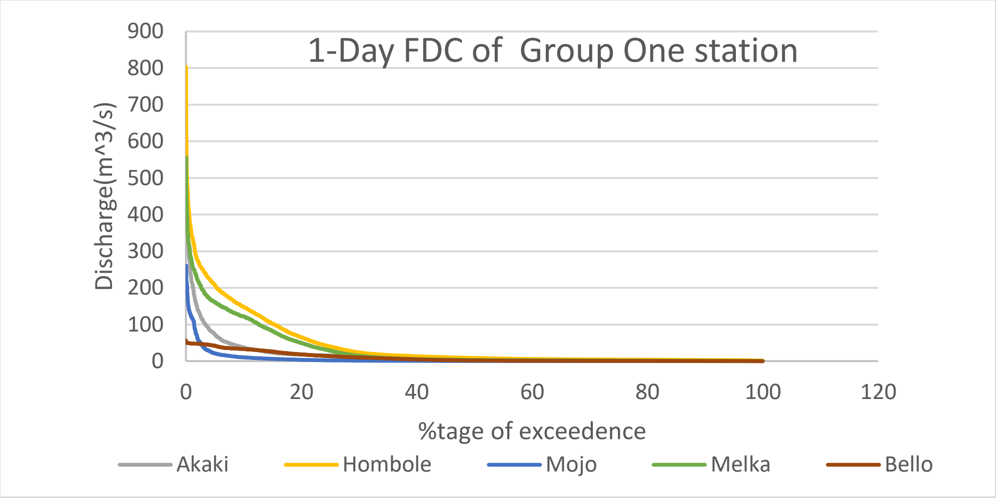


Figure 4.2 Daily Flow Duration Curve for Group One station

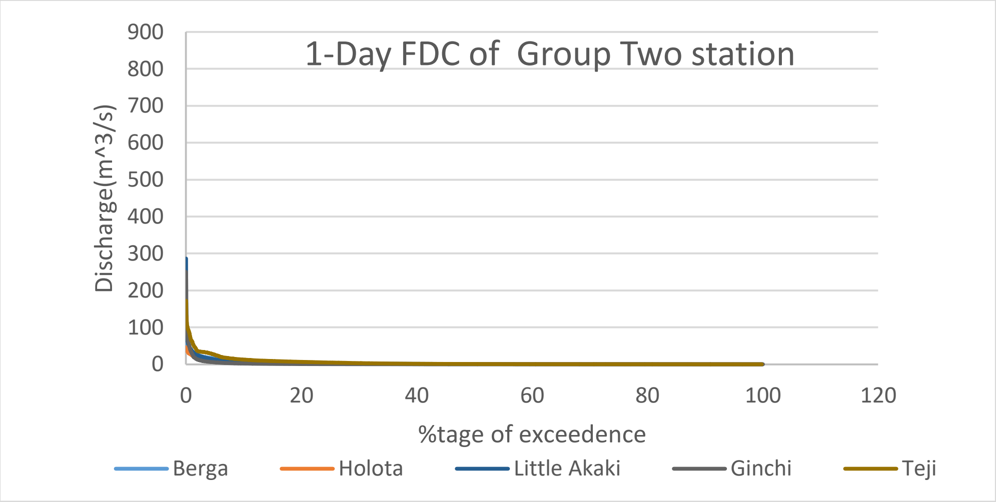


Figure 4.3 Daily FDC for Group Two station

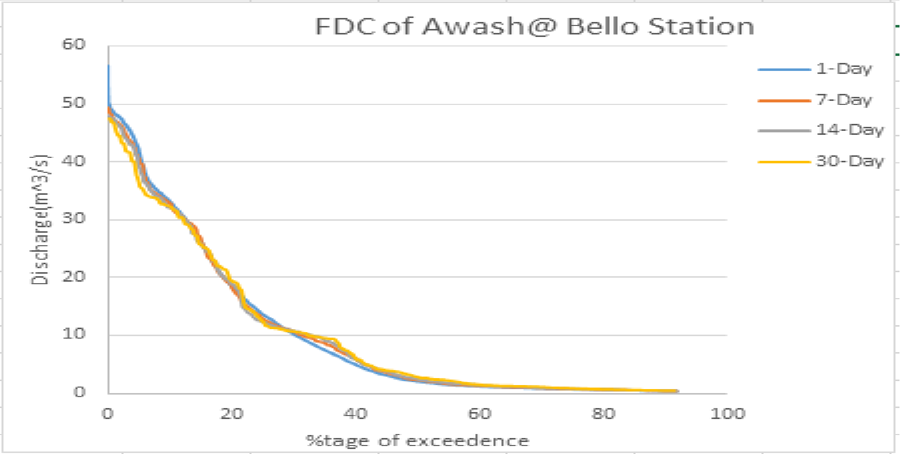


Figure 4.4 1, 7, 14 and 30 day FDC of Bello Station

4.1.1 Low flow Indices

Low flow indices derived from FDC on the basis of daily flow are the percentage which indicate a high frequency of exceedance and therefore present the low flow period of regime. Common practice used as the low flow are Q_{60}, Q_{70}, Q_{80}, Q_{90}, Q_{95} and Q_{100} (Joshi and Hilaire, 2013). In this study, 60%, 70%, 80%, 90%, 95% and 100% dependable flow of all 10 hydrological stations in the sub-basin have been analyzed and presented in the Table 4.1 and 4.2 showing that the magnitude of flow of each station having 60, 70, 80, 90, 95 and 100 percentage of exceedence resulted from the flow duration curve analysis . A 60% low flow indices for example is the magnitude of flow which is available 60% of the year.

Table 4.1 Group One Low flow indices

Station Name	60%	70%	80%	90%	95%	100%
Akaki	6.181	5.438	4.448	4.145	3.367	1.750
Hombole	6.635	5.005	3.892	3.441	3.069	2.650
Melka Kunture	2.591	2.031	1.487	1.143	0.907	0.701
Mojo	0.797	0.690	0.506	0.366	0.312	0.298
Bello	1.595	1.032	0.645	0.469	0.387	0.204

Table 4.2 Group Two low flow indices

Station Name	60%	70%	80%	90%	95%	100%
Berga	0.412	0.345	0.289	0.260	0.236	0.135
Holota	0.329	0.291	0.245	0.217	0.196	0.098
Little Akaki	0.312	0.229	0.167	0.110	0.080	0.024
Ginchi	0.148	0.107	0.080	0.069	0.051	0
Teji	0.414	0.316	0.195	0.166	0.140	0.068

4.1.2. Zero Flow condition Of FDC

The percentage of time the stream is at zero-flow conditions illustrates the degree of stream intermittency and represents either the percentage of zero-flow days or percentage of zero-flow months in a complete record. The longest recorded period of consecutive zero-flow days may be perceived as an indication of the most extreme drought, but the information content of this index is greatly dependent on the length of the record (Joshi and Hillaire, 2013).

From the results of flow duration curve analysis and low flow indices in Table 4.1 and 4.2, it is indicated that all rivers except, Holota and Ginchi have no zero record period which is an indication of there is no extreme drought in the catchment/ sub catchment. On other hand Holota and Ginchi have period of consecutive zero flow days even though it is for short period of record (only for Q_{100} for Holota and Q_{95} and Q_{100} for Ginchi) which shows a sign of drought in the catchment.

4.2. Low flow Regionalization

4.2.1. Identification of homogeneous region

Various approach have been proposed to select stations having the same parent distribution, where as in this paper the L-moment ratio diagram is employed to do so. The clustering of sites was carried out by hierarchical geographic regionalization procedure. LMRD were then used to group stations to confirm the hierarchical clustering. To use the LMRD some of the statistical parameters such as Lcs and Lck are first computed and drawn on the LMRD (see fig. 4.5) and those stations lie close to a single line or distribution on the diagram are then supposed to come from the same parent distribution and are considered to be in the same region.

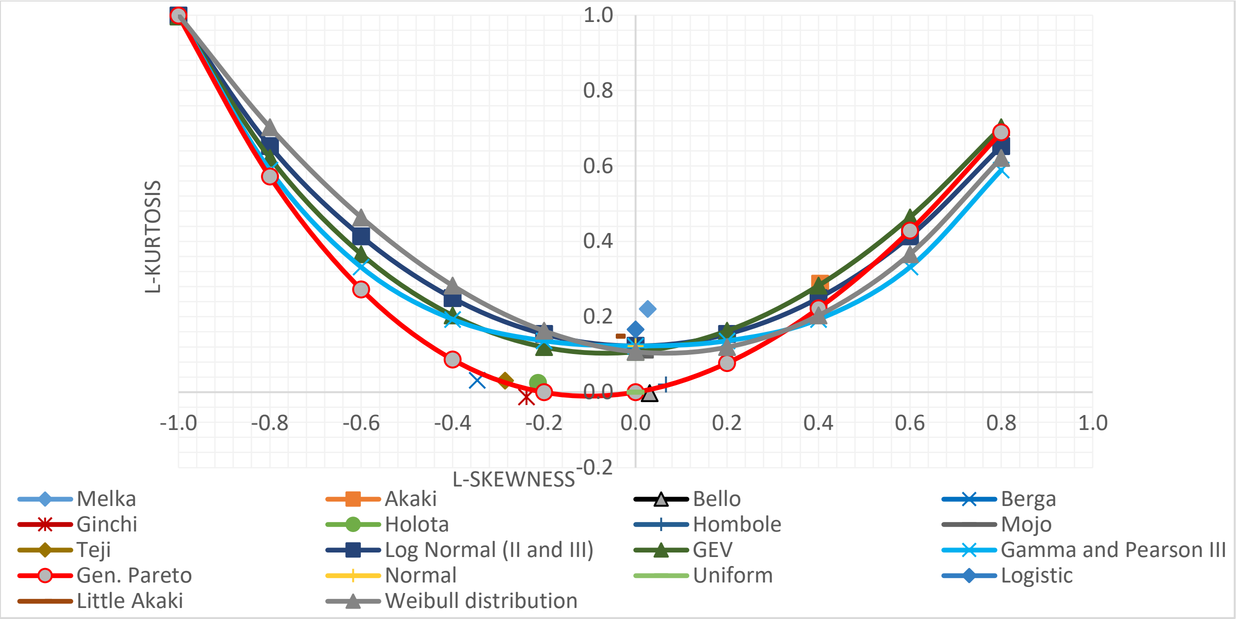


Figure 4.5 Lcs Vs Lck moment ratio diagram of each station

From the LMRD above, a homogenous region can be preliminarily identified based on the theory discussed until confirmed by different homogeneity tests. Accordingly, two homogenous regions are identified as in the table 4.3.

Table 4.3 preliminarily identified homogenous region

Region Name	Station Name	Candidate distribution from the LMRD
Region one	Berga	GPA
	Holeta	GPA
	Bello	GPA
	Ginchi	GPA
	Teji	GPA
Region two	Akaki	GEV/Lognormal (II&III)
	Melka	GEV/Lognormal (II&III)
	Hombole	GPA
	Mojo	GEV/Weibull
	Little Akaki	GEV/Weibull/Lognormal(II&III)

As it can be seen from the L-MRD, Little Akaki station for low flow analysis is grouped under region two but Mengistu Demissie on his flood analysis included it as upper region. This is due to having different statistical parameter (Lcs, Lcv and Lck) nature of the site for the flood and low flow characteristics as the two extreme event (flood and low flow) should be analyzed during the wettest and driest season of the year respectively. The geographic proximity nature of this site also confirms for region two. The methodology used to regionalize the basin may have also effect on the result (as the researchers doesn't used L-MRD to group the station).

4.2.2 Regional Homogeneity test

This involves testing whether a preliminarily identified region may be accepted as being homogeneous and whether two or more homogeneous regions are sufficiently similar that they should be combined into a single region.

4.2.2.1. CC-based regional homogeneity test

The statistical properties of the station can express the nature of the station. In this test the site-to-site coefficient of variation of the (CC) of both conventional and L-moments of the proposed region are used.

To investigate whether the homogeneity of the region has been met or not many researchers as Cunnane (1989), Ketsela and Sivakumar (2017), Hussen and Wagesho (2016) and others have used the values of mean coefficient of variation (CV) and the site – to – site coefficient of variation (CC) of both conventional and L – moments of the proposed region. According to the researchers the higher the values of CV and CC the lower the performance of index flood method for the considered region. This is due to the dominance of the flood quantile estimation variance by the variance of the at – site sample mean.

Thus, from the results described in the Table 4.4 and 4.5, it can be concluded that all regions were hydrologically homogeneous for both Cv and LCv based homogeneity tests since the CC values were less than 0.3. With regard to the results obtained above, all stations grouped as homogeneous were satisfied the stated homogeneity test criteria. As a result, it can be concluded that all regions were reasonably homogeneous.

Table 4.4 Results of CC-based homogeneity test for Region One

Station Name	Cv	Cs	Ck	Lcv	Lcs	Lck	Remark
Berga	0.894	0.760	-0.301	-0.584	-0.346	0.032	Homogenous
Holeta	0.755	0.699	-0.641	-0.425	-0.214	0.025	
Bello	0.506	-0.060	-0.707	-0.256	0.031	-0.003	
Ginchi	0.930	0.780	-1.036	-0.524	-0.238	-0.013	
Teji	0.765	0.922	-0.527	-0.582	-0.285	0.031	
Mean	0.770	0.620	-0.642	-0.474	-0.210	0.014	

Stand. dev	0.170	0.389	0.269	0.138	0.144	0.021	
CC	0.221			-0.291			

Table 4.5 Results of CC-based homogeneity test for Region Two

Station Name	Cv	Cs	Ck	Lcv	Lcs	Lck	Remark
Akaki	0.179	-1.241	1.350	-0.091	0.404	0.287	Homogenous
Melka	0.244	-0.353	0.963	-0.194	0.026	0.221	
Hombole	0.206	-0.112	-0.940	-0.118	0.066	0.020	
Mojo	0.361	-0.301	0.053	-0.209	0.020	0.095	
Little Akaki	0.372	0.073	-0.121	-0.209	-0.041	0.149	
Mean	0.272	-0.387	0.261	-0.164	0.095	0.155	
Stand. dev.	0.080	0.506	0.909	0.056	0.177	0.104	
CC	0.294			-0.341			

4.2.2.2. Discordance measure test

The discordancy measure, Di were computed for the site in the study region to find out whether any site were grossly discordant from the other sites. If the Di statistic for a site is more than the determined critical value, the data at such site have to be examined for possible problems. In this study, L-statistics of a preliminary identified two regions were examined for overall gross errors for 7-day minimum annual flow data sets. The test was carried out using a MATLAB program using equation 3.22. The computed Di values at each station for Upper Awash river basin are presented for each region in Table 4.6 and 4.7.

From the results below in the table, Di values range from 0.377 to 1.308. The high Di values always warrant a careful scrutiny of the data at the respective stations. These Di values in each of the two region are actually below the critical values given by Hosking and Wallis (1997). Therefore, data within these two groups are not discordant and they are suitable for applying the regional low flow frequency using their L-moments.

Table 4.6 Results of discordance measure test for Region One

Station Name	Lcv	Lcs	Lck	D_i	Remark
Berga	-0.584	-0.346	0.032	0.439	Homogenous
Holeta	-0.425	-0.214	0.025	0.995	Homogenous
Bello	-0.256	0.031	-0.003	1.219	Homogenous
Ginchi	-0.524	-0.238	-0.013	1.308	Homogenous
Teji	-0.582	-0.285	0.031	1.040	Homogenous

Table 4.7 Results of discordancy measure for region two

Station Name	Lcv	Lcs	Lck	D_i	Remark
Akaki	-0.091	0.404	0.287	1.300	Homogenous
Melka	-0.194	0.026	0.221	1.281	Homogenous
Hombole	-0.118	0.066	0.020	1.254	Homogenous
Mojo	-0.209	0.020	0.095	0.750	Homogenous
Little Akaki	-0.209	-0.041	0.149	0.377	Homogenous

4.2.3. Delineation of Homogenous Region

After organizing and assembling the data set, some of the important statistical parameters have been computed and interpolation of these statistical values (LCs, LCk) in collaboration with catchment characteristics are then used to come up to the following result of delineation of Upper Awash sub-basin on the Arc GIS interface. Generally depending on the result obtained from both tests two main regions region one and region two has been delineated.; the first region includes the Berga, Holeta, Awash bello Ginchi and Teji station and the second region have Akaki, Hombole, Melka kunture, and little akaki stations. The stations topographical and geographical proximity, altitude and other external fixture of the catchments have been visualized.

The regions have covered an area of 4061 and 7655 km^2 for Region one and two respectively. The sub basin can be divided in to two varying elevation and low flow characteristics as the firs region having higher elevation characterizing relatively small amount of flow in the dry season and in the second region, which has a very plate nature exhibit a good amount flow even in the dry season of the year. Generally, the delineated region is shown in the figure 4.6 below.

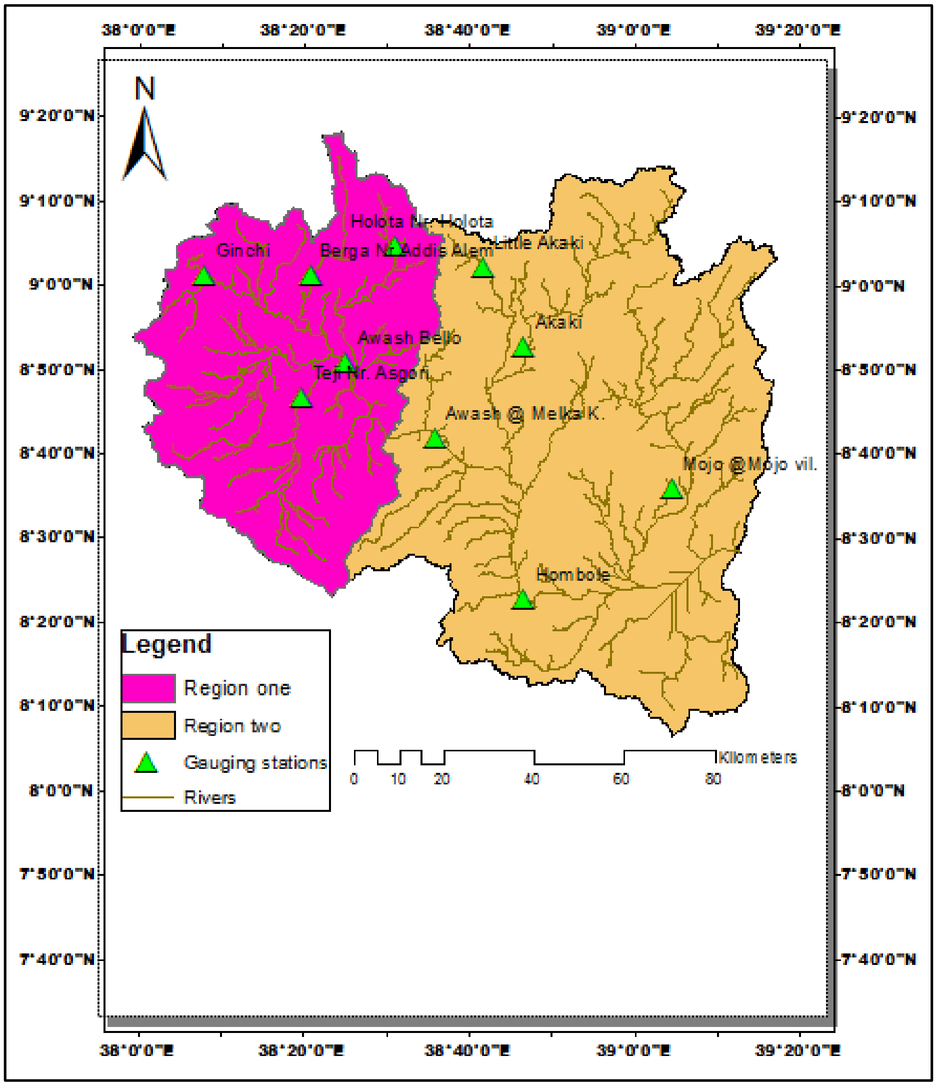


Figure 4.6 Delineated Region in the sub-basin

4.3. Identification of Regional frequency distribution for low flow Analysis

Once the homogeneous regions have been delineated, an appropriate distribution has to be selected as the regional frequency distribution. When several distributions fit the data sufficiently, any of them is a sensible choice for use in the final analysis, and the best choice among them will be the distribution that is most strong.

In this study, L-moment ratio diagram were used for preliminary selection of distribution for the regions of Upper Awash sub basin and Easyfit software confirm the choice of candidate distribution.

4.3.1. Selection of distribution by LMRD

For a given region, the sample regional average moment ratios, L_{Cv}, Lcs, and $L_{C}k$, are plotted on the LMRD to select the suitable distribution model (Fig. 4.7). Based on the regional average L-moments, candidate distributions are selected. The best parent distribution is the one that the average value of the point (Lcs, Lck) of all stations with in the region gets close to one of the drawn LMRD of the parent distribution.

Figure 4.7 below indicates that the points representing the regional average L-moment ratios Z_3^R=-0.211 and Z_4^R=0.014 for Region one and Z_3^R =0.095 and Z_4^R=0.155 for Region two lie close to the Generalized Pareto and General Extreme Value/ Lognormal (II&III) respectively, which can further be analyzed by the goodness of fit test to confirm them as the regional best-fit distribution of low flow analysis for the two region.

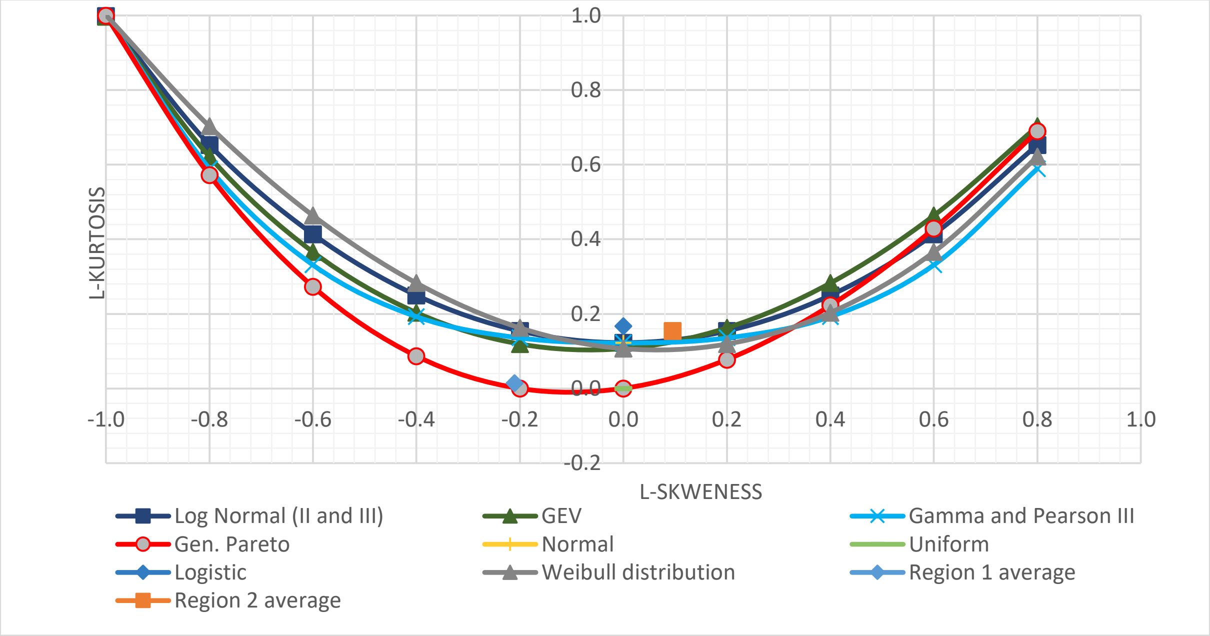


Figure 4.7 Regional average of LMRD for Upper Awash River Basin

The most possible underlying candidate distributions from the average regional L-moment diagram close to a specific distribution in the figure 4.7 are summarized on the table 4.8.

Table 4.8 Selected candidate distribution for each region

Region Name	Average regional L-moments (Z_2,Z_3,Z_4)	Selected candidate distribution from LMRD
Region one	(-0.474, -0.211, 0.014)	Generalized Pareto
Region two	(-0.164, 0.095, 0.155)	GEV/Lognormal(II&III)

4.3.2. Goodness of fit tests

In this study, selection of the best-fit distribution from candidate distribution obtained from the above-mentioned analysis is done by using software, which is called Easyfit. Within this software all goodness of fit test such as Chi-Square, Kolmogorov Smirinove and Anderson Darling tests were done and the best-fit distribution from the candidate is displayed automatically. This particular characteristic feature of the software is very useful for comparing fitted models to each other. Kolmogorov-Smirinove test can be named as the most used Goodness-of-Fit test (Mehrannia and Pakgohar, 2014) which is also confirmed in this study.

From the result of goodness of fit test in the table 4.9, the Generalized Pareto distribution is presented at rank one for region one under all the three tests and the General extreme value is also best fitted to region two for all of this tests, which supports the results of L-MRD. Based on the result, it can be concluded that GPA and GEV distributions are the best distribution to represent the regional model for both Region one and two respectively and all the three tests gives a reasonable estimation of low flow quantile in Upper Awash river basin. The justification of results was summarized in Table 4.9 for Region one and Table 4.10 for Region two.

Table 4.9 Goodness of fit measure for Region One

Fitting distribution	Kolmogorov- Smirnov		Anderson-Darling		Chi-Square	
	Statistics	Ranks	Statistics	Ranks	Statistics	Ranks
Generalized Pareto	0.1127	1	0.4461	1	0.0655	1
Uniform	0.1397	2	1.0794	7	2.2302	4
Gen. Extreme Value	0.1454	3	0.7125	4	3.2841	8
Weibull	0.1496	4	0.7356	5	1.4144	2
Log Pearson III	0.1511	5	0.6547	2	2.7825	5
Lognormal	0.1562	6	0.6953	3	2.9385	6

Normal	0.1598	7	1.022	6	1.88	3
Logistic	0.1748	8	1.3095	8	2.9608	7

Table 4.10 Goodness of fit measure for Region Two

Fitting distribution	Kolmogorov-Smirnov		Anderson-Darling		Chi-Square	
	Statistics	Ranks	Statistics	Ranks	Statistics	Ranks
Gen. Extreme Value	0.1239	1	0.5035	1	0.5279	1
Uniform	0.1455	2	11.281	8	N/A	
Pearson III	0.1539	3	0.5337	2	2.6303	6
Weibull	0.1625	4	0.5358	3	2.5989	4
Normal	0.1653	5	0.5913	4	2.6104	5
Generalized Pareto	0.1667	6	8.0452	7	N/A	
Logistic	0.1750	7	0.6958	5	0.9663	3
Lognormal	0.1982	8	0.9071	6	0.6706	2

From the regional average L-MRD in figure 4.7 above and the three goodness of fit tests to confirm the candidate distribution, Generalized Pareto shows absolute results for low flow analysis of region one. But comparing to the flood analysis in the region, Mengistu Demissie has selected Log normal II as best fit distribution. This is due to the difference in flood and low flow characteristics in the study region. The software he used to select the distribution may also have effect on the result.

4.3.3. Estimation accuracy of selected distribution

The P-P and Q-Q plots are plots for graphical assessment of fitting distribution, which are generally performed by plotting the observations so that they would fall approximately on a straight line if a postulated distribution were the true distribution from which the observations were drawn. Easy Fit displays the suggestion crossways line along which the graph points must fall. Both plots for region one are presented in the figure below and results both fits were well fits to the line. The plots for Region Two is demonstrated under appendix-J and K.

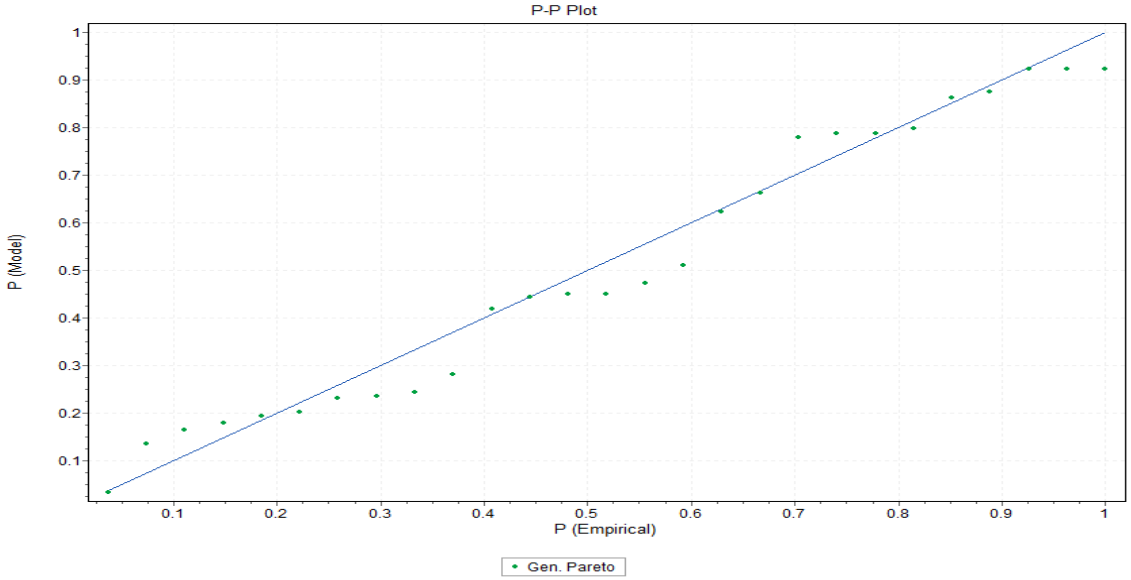


Figure 4.8 Probability-Probability plots of Region One

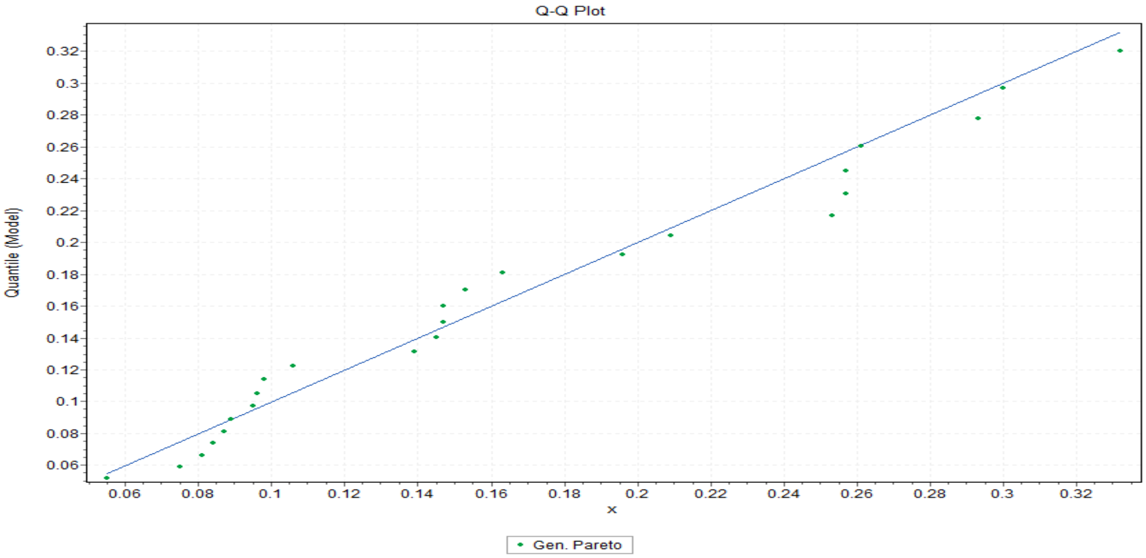


Figure 4.9 Quantile-Quantile plots of Region One

4.4. Parameter Estimation

As discussed in the methodology section, the parameter for the selected probability distributions has been estimated by using the method of moment (MOM), method of maximum likelihood (MML) and probability weighted moments (PWM) which are also the most commonly used methods by many researchers. The parameter estimates that maximize the likelihood function are computed by partial differentiation with respect to each parameters and setting these partial derivatives equal to zero and finally solve the resulting set of equations simultaneously.

The summary of parameter estimation of each region for selected distributions using all three estimation methods are shown in Table 4.11 below using hand calculation by Rao and Hamed (2000).

Table 4.11 Parameter summary of each region

Region name	Selected distribution	MOM		MML		PWM	
		Parameter	Value	Parameter	Value	Parameter	Value
Region one	Generalized Pareto	k	-0.122	K	0.111	k	3.813
		δ	0.132	Δ	0.149	δ	-2.255
		µ	0.055	Μ	0.054	µ	0.674
Region two	General Extreme Value	k	0.569	K	-0.528	k	0.025
		δ	0.291	Δ	0.280	δ	-0.219
		µ	1.109	Μ	1.110	µ	1.286

4.5. Standard error and Low flow quantile

As Hussen and Wagesho (2014), Rao and Hamed (2000) and other researchers, the standard error of estimate is a measure of the accuracy of predictions. In this study, the most efficient methods of parameter estimation was selected by computing the standard error of estimates using equation 3.31 were determined for the three estimation techniques. The performance of the parameter estimation method under test is then compared through the magnitude of SEE and the parameter estimation procedure giving the smallest values for SEE was selected to be the best parameter estimation methods. The result of standard error and low flow quantile with their respective estimation techniques is presented in the Table 4.10. The low flow quantile was estimated using equation 3.27 and 3.30.

Table 4.12 Standard error and quantile estimation for Region One

Return period (T)	Region one					
	Generalized Pareto					
	(MOM)		(MML)		(PWM)	
	SEE	Q_T	SEE	Q_T	SEE	Q_T
2	0.205	0.151	0.211	0.153	0.198	0.125
5	0.224	0.085	0.233	0.087	0.220	0.084
10	0.229	0.069	0.244	0.070	0.229	0.070
20	0.234	0.061	0.249	0.062	0.232	0.064
50	0.237	0.057	0.261	0.057	0.236	0.059
100	0.248	0.041	0.261	0.041	0.243	0.038

Table 4.13 Standard error and quantile estimation for Region Two

Return period (T)	Region two					
	General Extreme Value					
	(MOM)		(MML)		(PWM)	
	SEE	Q_T	SEE	Q_T	SEE	Q_T
2	0.271	1.205	0.273	1.219	0.271	1.206
5	0.321	0.992	0.323	0.987	0.337	0.964
10	0.366	0.920	0.373	0.910	0.453	0.807
20	0.398	0.876	0.405	0.867	0.581	0.659
50	0.428	0.816	0.436	0.828	0.755	0.611
100	0.446	0.709	0.454	0.806	0.887	0.570

From the results above, it is observed that PWM methods shows a relatively less standard error of estimates than others for region one and same MOM for region two. Therefore, it can be concluded that GPA/PWM and GEV/MOM are the most efficient distribution and parameter estimation procedures for region one and two respectively for low flow analysis in UARB.

4.5.1 Estimation of index-low flow for standardization

If used with care and with understanding of the underlying assumption, the regional frequency curves may be sufficient for design project. To represent the low flow frequency curve of the region, an average regional growth curve were used in this study. Using Index low flow Procedure, regional growth curve, a dimensionless quintile function, which is identical for homogeneous region has been computed as presented in the appendix-M.

The results of estimated standardized quantiles for specified region in the sub-basin are shown below in Figure 4.10 for recurrence intervals of 2, 5, 10, 20, 50 and 100 years expressed in Gumbel reduced Variate. From the result, it is observed that the constructed regional frequency curves from the two regions reflect both curves have different low flow characteristics. The higher variations of regional curves may be due to the considerable spatial fluctuations of elevations with their spatially undulating mountainous topography of regional boundaries. Region A consisting of the upper parts of the sub basin like Berga, Holota, Bello, Ginchi and Teji represent a high diminishing in magnitudes of low flows although the diminishing factor is not rapid. This needs water management plans to counteract with the upcoming extreme events to supplement the productivity with irrigation as irrigation in dry season requires substantial base flows of the river. In addition, it also needs watershed management to increase the recharge to sustain the environmental flows, which are related to the low flow situation in the region. However, none of the region has resulted “no flow” up to 100 years return period.

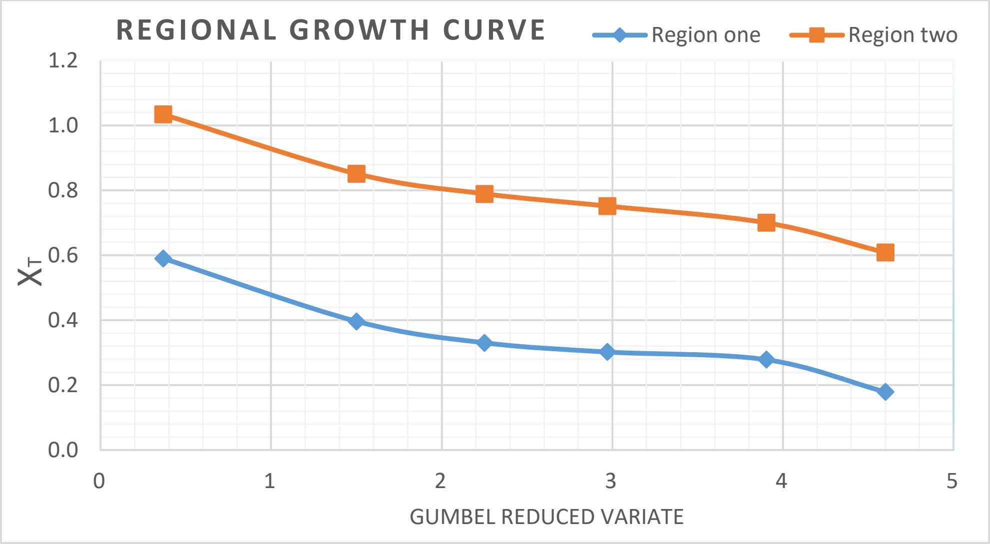


Figure 4.10 Regional growth curve for delineated homogenous region

4.6. Base flow Separation

In this study, using prescribed base flow separation methods makes the process less tedious and more consistent. The hydrograph is separated in to base flow and total flow for all 10 stations in the study area.

Base flow separation computer program estimates the annual base flow volumes of rivers/streams and computes an annual base flow index (the ratio of base flow to total flow of a given year) for multiple years of data. The average of this annual base flow index of multiple years of data was used as the average BFI of the station.

4.6.1. Result of BFI

Base flow index of those rivers having a value nearest to 1 indicates that its source of water is from the ground water recharge rather than runoff and also for base flow index value far less than 1, the stream flow contributes from overland flow than the base flow. From the result observed in Upper awash sub basin 40% of the stations representing most parts of Region One in the sub basin has BFI ranging from 0.15-0.39, 50% of the station constituting most part of the Region Two such as Akaki, Hombole, Little akaki, Melka and Awash at bello has BFI of 0.4-0.65. Similarly, 10% of the stations has BFI >0.7, which indicates more contribution of the stream flow obtained from the base flow than over land flow. In general, it can be concluded that Region One of the sub-basin have low base flow contribution for stream flow. This may be due to the steep nature of the region related to generate runoff than base flow. As it can be seen in the table below, similar or nearly similar BFI value can also tell us geologic similarities of sites which can also be the base for regionalization. The annual base flow discharge and base flow index for all stations are presented in the table 4.14.

Table 4.14 Base flow index for all station

Station Name	Average total flow (m^3/s/year)	Average base flow (m^3/s/year)	BFI
Berga Nr. Addis Alem	1407.118	539.943	0.384
Holeta Nr. Holeta	621.212	214.472	0.345
Teji @ Asgori	1937.799	690.647	0.356
Awash @ Ginchi	557.025	82.904	0.149
Awash @ Bello	3597.883	2200.330	0.612
Akaki @ Akaki village	6619.807	3268.121	0.494

Mojo @Mojo village	1756.087	909.546	0.518
Awash @M. Hombole	15187.302	10700.806	0.705
Awash @ M. Kunture	11529.240	7320.835	0.635
Little Akaki	1019.357	484.393	0.475

The total flow and base flow hydrograph has been plotted by excel spreadsheet using the value obtained from the base flow index and one typical output has been presented for demonstration in figure 4.11 and the remaining hydrograph are attached in the Appendix-O

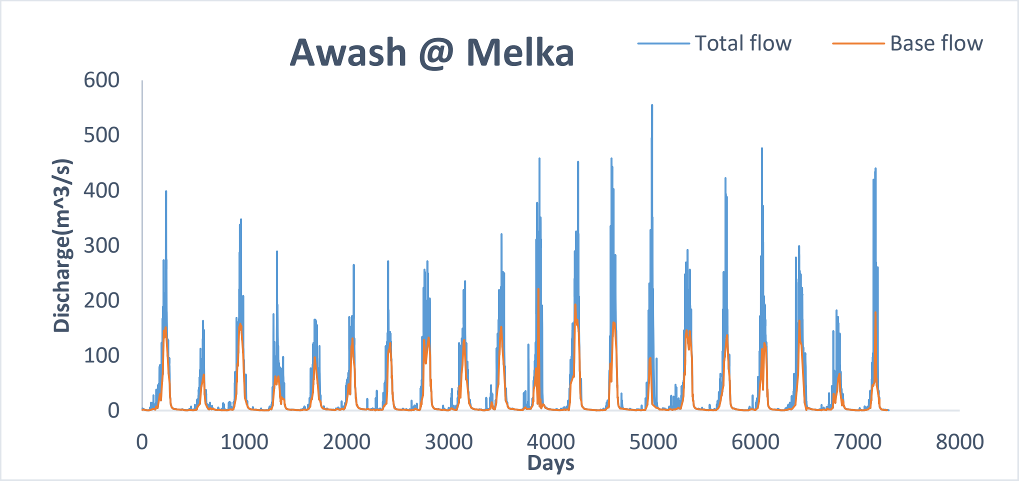


Figure 4.11 Typical hydrograph and base flow hydrograph

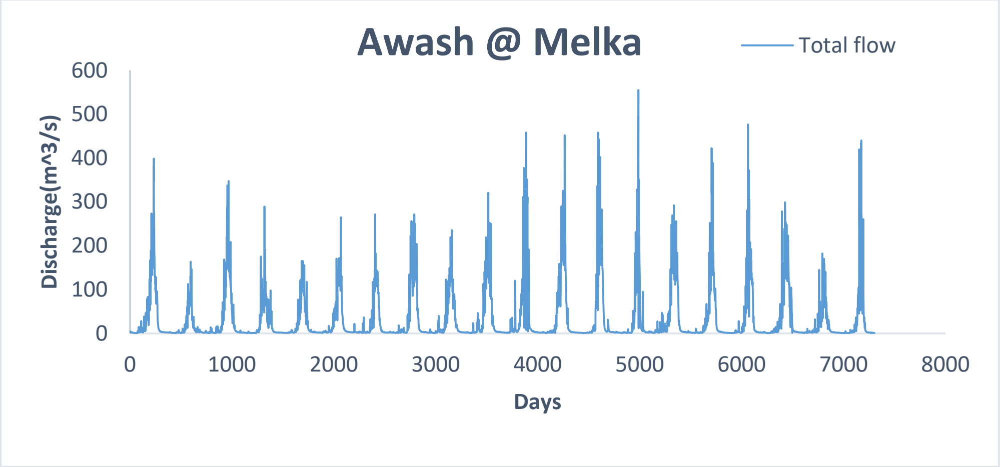


Figure 4.12 Typical base flow hydrograph

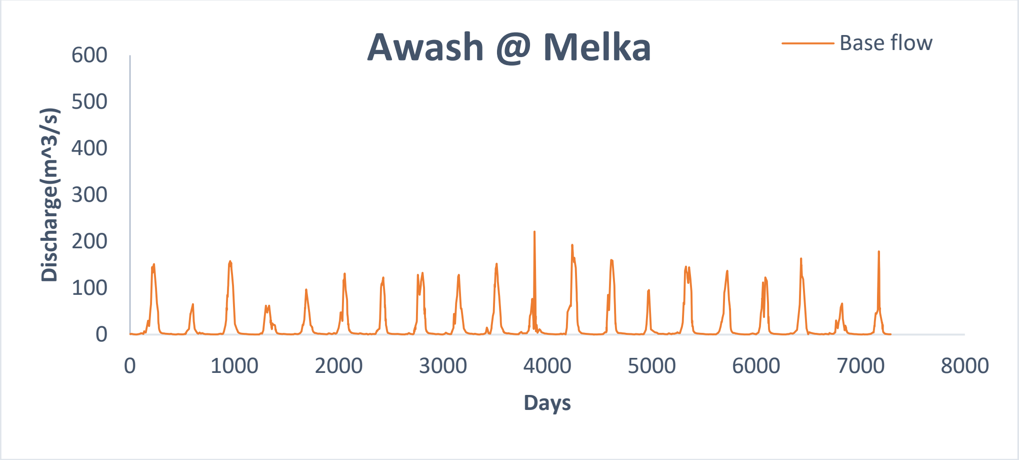


Figure 4.13 Separated base flow hydrograph

5. CONCLUSION AND RECCOMENDATION

5.1 Conclusion

In this study, a regional approach for conducting low flow analysis based on L-moment theory has been used for the rivers in Upper-Awash sub basin. The study focused on reliable low-flow estimation from stream flow time-series, which include frequency analysis of extreme low-flow events, regional curves plotting for low-flow prediction and base flow separation.

Homogenous regions have been delineated for UARB by using site data with statistical methods. Parameters such as LMRD and the regional average L-moment are used to group stations to same region. Accordingly, two regions are identified and delineated. Region two comprises the largest portion of the basin, covering 65.3% of the area. Both of the two regions have satisfied the homogeneity tests(discordance measure test, Cv and LCv based homogeneity tests).

Average values of $L_{C}s$ and L_{Ck} were used to select the best-fit statistical distribution of each region and goodness of fit test by using the software Easy Fit is used to approve the best-fit distribution. Robustness assessment was conducted to select a superlative method of parameter estimation for the selected distribution. For region one Generalized Pareto (GPA) with PWM was selected. For region two, Generalized Extreme Value with MOM was selected and these distributions with method of parameter estimations are finally used to develop a regional growth curve of each homogeneous region. The study concluded that LMRD and Easy Fit Statistical Software were acceptable methods for selecting best-fit distribution in UARB.

Regional low flow frequency curves were constructed for the return periods of 2, 5, 10, 20, 50 and 100 years resulting 0.125, 0.084, 0.070, 0.064, 0.059 and 0.038 m^3/s quantile for region one and 1.206, 0.964, 0.807, 0.659, 0.611 and 0.570 m^3/s for region two. Using the growth curve and the relation between low flows with catchment characteristics could help for estimating the low flows quantiles for water resource planning and management.

Furthermore, base flow separation using Base flow index have been computed, resulting 40 % of the stations have BFI in the range of 0.15-0.39 which indicates more contribution of the stream flow obtained from over land flow than the base flow. About 50% of stations in Upper-Awash sub basins have BFI in the range of 0.4-0.65 which indicates the stream flow gets equal contribution with surface run off . Similarly, 10% of the station have BFI >0.7 which indicate the stream gets more base flow than over land flow.

5.2. Recommendation

- Low flow analysis for many zero flow data should be treated in Special way not the usual way.
- Usually low flow frequency analysis is done by using statistical distribution technique (i.e. L-Moment method). But, this method is not always the best and efficient method, trying to use another methods or soft ware's is important in some cases. For example: Hyfran plus software.
- Identification of homogeneous region based on statically parameter may not be satisfactory but it should also consider the catchments characteristics.
- The estimated low flow quantile should use as an input for proper water resource planning, management and drought analysis of delineated homogeneous regions separately.
- Base flow separation can be done by graphical or other deterministic methods. Carrying out comparison of the result with the BFI+ Base flow separator computer program helps to know the accuracy of the program.
- More research is required to determine how aspects of climate change can be incorporated into low flow frequency analysis for design and planning purposes.

REFERENCE

Ahmad, U.N., Shabri, A. and Zakaria, Z.A. (2011). Flood Frequency Analysis of Annual Maximum Streamflow Using L-Moments and TL-Moments. Applied Mathematical Sciences, PP:243-253.

Ahmad, I., Fawad, M., Akbar, M., Abbas, A. and Zafar, H. (2016). Regional Frequency Analysis of Annual Peak Flows in Pakistan Using Linear Combination of Order Statistics. Journal of Environmental Studies, Vol. 25 (6). PP:1-10.

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), PP: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. Journal of climate. Vol.6(9).

Assefa, K. and Moges, M.A. (2018). Low Flow Trends and Frequency Analysis in the Blue Nile Basin, Ethiopia. Journal of Water Resource and Protection, PP:182-203.

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

Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). Applied Hydrology. McGraw Hill Book Company, USA .

Cunnane. (1989). Statistical Distribution for Flood Frequency Analysis,World Meteorological Organization Operational Hydrology Report No. 33. Geneva, Switzerland:

Dahmen, E.R. and Hall, M.J. (1990). Screening of Hydrological Data: Tests for Stationarity and Relative Consistency: International Institute for Land Reclamation and Improvement: Wageningen, Netherlands

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.

Demuth, S. (2005). Low flows and droughts-A European perspective. IAHS Newsleter 82, PP:7-8.

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): pp:155-161.

Durrans, S. (1996). Low-flow analysis with a conditional Weibull tail model.

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

Gebregiorgis, S. Moges, A., and Awulachew, B. (2013). Basin Regionalization for the Purpose of Water Resource Development in Limited Data Situation: Case of Blue Nile River Basin, Ethiopia . Journal of Hydrologic Engineering , PP:1-13.

Gedefa, T.W. and Seleshi, Y. (2009). Regional Flood Frequency Analysis for Upper Omo Gibe 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), PP:170-174.

Gobena, S. (2010). Daily rainfall-runoff modeling of Upper Awash Sub basin Using conceptual Rainfall Runoff model: Addis Abeba universsity. Addis Ababa, Ethiopia

Gonzales, A. L., Nonner, J., Heijkers, J., and Uhlenbrook S. (2009). Comparison of different base flow separation methods in a lowland catchment . Journal of Hydrology and Earth System Sciences Discussions,6, PP:3483–3515.

Halrcow. (1989). Masterplan for the development of surface water resources in the Awash basin. Addis Ababa: a Report on Climate and Hydrology.

Henock, B., Paul, V.E., and Tegenu, Z. (2008). Background Information for a Program Approach,Challenges and Possible Cooperation between Dutch and Ethiopian counterparts. Central Ethiopia: Integrated Water Resource Management on Upper Awash River Basin.

Hewa, G. A. (2001). A Methodology for Regional Low Flow . Ph.D. Thesis and Dissertation submitted to Galwey University.

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

Hussein, B. and Wagesho, N. (2016). Regional Flood Frequency Analysis for Abaya – Chamo Sub Basin, Rift Valley Basin, Ethiopia . Journal of Resources Development and Management , pp:15-28.

Joshi, D., and Hillaire, A.S. (2013). Low Flow Frequency analysis of three rivers in Eastern Canada. Statistical Hydroclimatology Research Group, PP:1-33.

Mengistu, K.H and Sivakumar, K. (2017). An Assessment Of Predictive Accuracy For Regional Flood Frequency Distribution Estimation Methods On Awash River Basins . International Journal of Latest Trends in Engineering and Technology Vol.(11)Issue(1) , PP:030-039.

McCuen, R. H. (1998). Hydrologic Design and Analysis. Library of Congress Cataloging-in-Publication Data.

Middelkoop, H., Daamen, K., Gellens, D., Grabs, W., Kwadijk, J.C.J., Lang, H., Parmet, B., Schadler, B., Schulla, J. and Wilke, K. (2001). Impact of Climate Change on Hydrological Regimes and Water Resources Management in the Rhine Basin. Journal of Climate Change , 49, PP:105-128.

Millington, N., Das, S. and Simonovic, S. P. (2011). The Comparison of GEV, Log-Pearson Type 3 and Gumbel Distributions in the Upper Thames River Watershed under Global

Climate Models, Water Resources Research Report: The University of Western Ontario London.

Mishra, B.K., Takara, K., Yamashiki, Y. and Tachikawa, Y. (2009). Selection of Regional frequency distribution using simulated flood data. Annuals of dissaster prevention and Resolution Institution:Kyoto University,Report No. 52B.

NILAR,. W., AND KHIN, M.W. (2014). Low Flow Frequency Analysis on Selected Rivers in Malaysia. International journal of Earth science and Engineering. ISSN 0974-5904, Volume 07, No. 04, PP:1315-1321

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, pp:1-8.

Pakgohar, A. (2014). Using Easy Fit Software for Goodness-of-Fit Test and Data Generation. International Journal of Mathematical Archive-5(1),7, PP:118-124.

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

Rabba et al. (2018). Applications of the PYTOPKAPI model to ungauged catchments. Centre for Research in Environmental, Coastal and Hydrological Engineering, University of KwaZulu-Natal, South Africa.

Rabba, Z.A., Fatoyinbo, B.S. and Stretch, D. D. (2018). Applications of the PyTOPKAPI model to ungauged catchments. Centre for Research in Environmental, Coastal and Hydrological Engineering, University of KwaZulu-Natal, South Africa.

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

Rosbjerg, D. (1985). Estimation in partial duration series with independent and dependent peak values. Journal of Hydrology.

Saf, B. (2009). Regional Flood Frequency Analysis Using L-moment for the West Mediterranean Region of Turkey. Journal of Water Resources Management, Vol.23(3), pp: 531-551.

Santhi, C., Allen, P. M., Muttiah, R. S., Arnold, J. G., Tuppad, P. (2008). Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. Journal of Hydrology(351), PP:139-153.

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.

Smakhtin, V. (2001). Low flow hydrology-A review. Journal of Hydrology 240, PP:147-186.

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.

Taha B.M.J.Quarda, C.H. (2008). Statistical model and estimation of low flows. Canadian water resource journal, PP:195-206.

Tatek, W. (2015). Estimation of low flow in ungauged river catchement. A thesis and dissertation submitted to Addis Ababa University .

Tegenu, Z. (2007). Regional Low Flow Frequency Analysis of Abay River Basin. MSc. Thesis report submitted to Arbaminich University.

World Meteorological Organization WMO (2008) Manual on Low-Flow: Estimation and Prediction. Operational Hydrology Report No. 50. Switzerland.

APPENDIX

APPENDIX–A: Critical values of the Grubbs T test Statistic as a function of number of Observations and Significance level.

N	5%	2.5%	1%	n	5%	2.5%	1%
3	1.15	1.15	1.15	20	2.56	2.71	2.88
4	1.46	1.48	1.49	21	2.58	2.73	2.91
5	1.67	1.71	1.75	22	2.60	2.76	2.94
6	1.82	1.89	1.94	23	2.62	2.78	2.96
7	1.94	2.02	2.10	24	2.64	2.80	2.99
8	2.03	2.13	2.22	25	2.66	2.82	3.01
9	2.11	2.21	2.32	30	2.75	2.91	
10	2.18	2.29	2.41	35	2.82	2.98	
11	2.23	2.36	2.48	40	2.87	3.04	
12	2.29	2.41	2.55	45	2.92	3.09	
13	2.33	2.46	2.61	50	2.96	3.13	
14	2.37	2.51	2.66	60	3.03	3.20	
15	2.41	2.55	2.71	70	3.09	3.26	
16	2.44	2.59	2.75	80	3.14	3.31	
17	2.47	2.62	2.79	90	3.18	3.35	
18	2.50	2.65	2.82	100	3.21	3.38	
19	2.53	2.68	2.85				

(Source: Grubbs, 1969)

Appendix-B: Percentile Points of the t-distribution t {V, p for the 5% level of Significance (Two-Tailed)}

P= P(t < = tp)	0.025	0.975	P= P(t < = tp)	0.025	0.975
4	-2.78	2.78	16	-2.12	2.12
5	-2.57	2.57	18	-2.10	2.10
6	-2.54	2.54	20	-2.09	2.09
7	-2.36	2.36	24	-2.06	2.06
8	-2.31	2.31	30	-2.04	2.04
9	-2.26	2.26	40	-2.02	2.02
10	-2.23	2.23	60	-2.00	2.00
11	-2.20	2.20	10	-1.98	1.98
12	-2.18	2.18	160	-1.97	1.97
14	-2.14	2.14	∞	-1.96	1.96

(Source: Dahmen and Hall, 1990)

Appendix–C: Percentile Points of the F-Distribution F {V1, V2, P} for the 5 % level of Significance (Two-Tailed)

P=P(F<FP		V1:4	5	6	7	8	9	10	11	12	14	16
0.025 0.975	V2:5	.107 .739	.140 7.16	.169 6.98								
0.025 0.975	6		.143 5.99	.172 5.82	.195 5.70							
0.025 0.975	7			0.176 5.12	.20 4.99	.221 4.90						
0.025 0.975	8				.204 4.53	.226 4.43	.244 4.36					
0.025 0.975	9					.230 4.10	.248 4.03	.265 3.96				
0.025 0.975	10						.252 3.78	.269 3.72	.284 3.66			
0.025 0.975	11							.273 3.53	.288 3.47	.301 3.43		
0.025 0.975	12								.292 3.32	.305 3.28	.328 3.21	
0.025 0.975	14									.312 3.05	.336 2.98	.355 2.92
		V1:14	16	18	20	24	30	40	60	100	160	∞
0.025 0.975	V2:16	.342 2.82	.362 2.76	.379 2.71								
0.025 0.975	18		.368 2.64	.385 2.60	.400 2.56							
0.025 0.975	20			.391 2.50	.406 2.46	.430 2.41						
0.025 0.975	24				.415 2.33	.441 2.27	.468 2.21					
0.025 0.975	30					.453 2.14	.482 2.07	.515 2.01				
0.025 0.975	40						.498 1.94	.533 1.88	.573 1.80			
0.025 0.975	60							.555 1.74	.600 1.67	.642 1.60		
0.025 0.975	100								.625 1.56	.674 1.48	.706 1.44	
0.025 0.975	160									.696 1.42	.733 1.36	
0.025 0.975	∞											1.00 1.00

Appendix-D: Result of hydrological data quality test for stationarity and consistency

Station Name	V	V1,V2	Ft2.5%	Ft	Ft97.5%	Tt2.5%	Tt	Tt97.5%
Berga	23	12,11	0.301	0.115	3.43	-2.060	0.016	2.06
Holota	20	10,10	0.269	0.773	3.72	-2.09	0.620	2.09
Teji	18	9,9	0.248	0.049	4.03	-2.10	0.042	2.10
Bello	25	12,12	0.305	0.329	3.28	-2.06	0.853	2.06
Akaki	16	8,8	0.226	0.228	4.43	-2.12	0.140	2.12
Mojo	16	8,8	0.226	0.894	4.43	-2.12	0.246	2.12
Hombole	20	10,10	0.269	0.829	3.72	-2.09	0.007	2.09
Melka	20	10,10	0.269	0.959	3.72	-2.09	0.366	2.09
Little Akaki	16	8,8	0.226	0.525	4.43	-2.12	0.056	2.12
Ginchi	18	9,9	0.248	0.643	4.03	-2.10	0.354	2.10

Appendix-E: Critical values of discordancy measure with N sites

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

Appendix-F: (Translated Matlab code for Discordancy Measure as provided by Hosking and Wallis, 1997)

U= number of gauging sites in the region (Enter the matrix of test statistics);

```
n=; %input ('enter the number of sites in the group :');
Ubar= [0; 0; 0];
For i=1: n
Ubar=ubar+1/n*(U (i, 1:3)');
end
S=zeros (3);
For i=1: n
```

```
S=S+ (U (i, 1:3)'-ubar)*(U (i, 1:3)'-ubar)';
end
For i=1: n
Di (i) =1/3*n*(U (i, 1:3)'-ubar)'*inv(S)*(U (i, 1:3)'-ubar);
end
```

APPENDIX-G Distributions used in L-moment statistics calculation in this study

Uniform: $z_3 = 0, z_4 = 0$

Logistic: $z_3 = 0, z_4 = \frac{1}{6}$

Normal: $z_3 = 0,\ z_4 = 0.1226$

Gumbel (EV1 (2)): $z_3 = 0.1699, z_4 = 0.1504$

Generalized Pareto

$$z_4 = 0.20196z_3 + 0.95924z_3^2 - 0.20096z_3^3 + 0.04061z_3^4$$

Generalized Extreme Value

$$z_4 = 0.10701 + 0.11090z_3 + 0.84838z_3^2 - 0.06669z_3^3 + 0.00567z_3^4 - 0.04208z_3^5 + 0.03763z_3^6$$

Gamma and Pearson III

$$z_4 = 0.1224 + 0.30115z_3^2 + 0.95812z_3^4 - 0.57488z_3^6 + 0.19383z_3^8$$

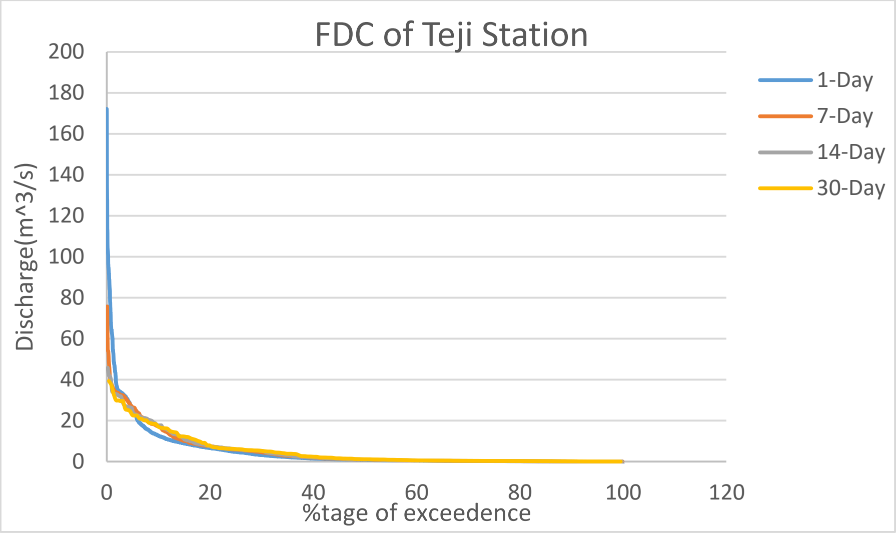
Lognormal (two and three parameters):

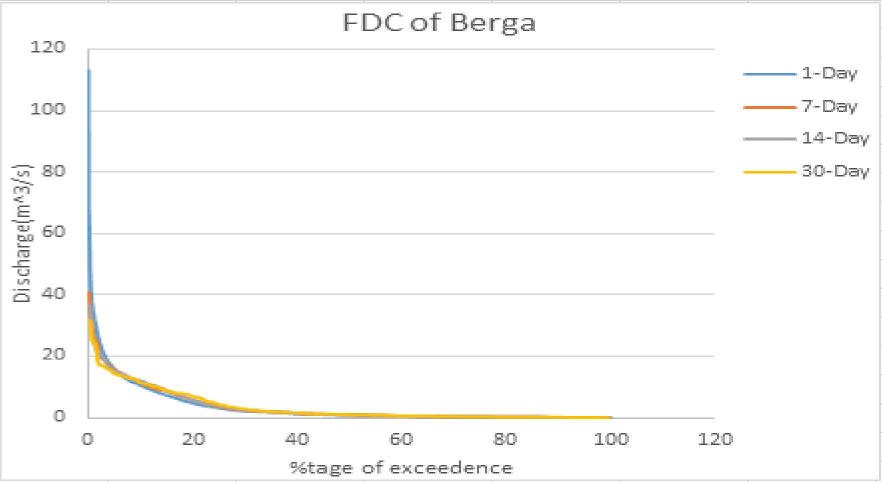
$$z_4 = 0.12282 + 0.77518z_3^2 + 0.12279z_3^4 - 0.13638z_3^6 + 0.11368z_3^8$$

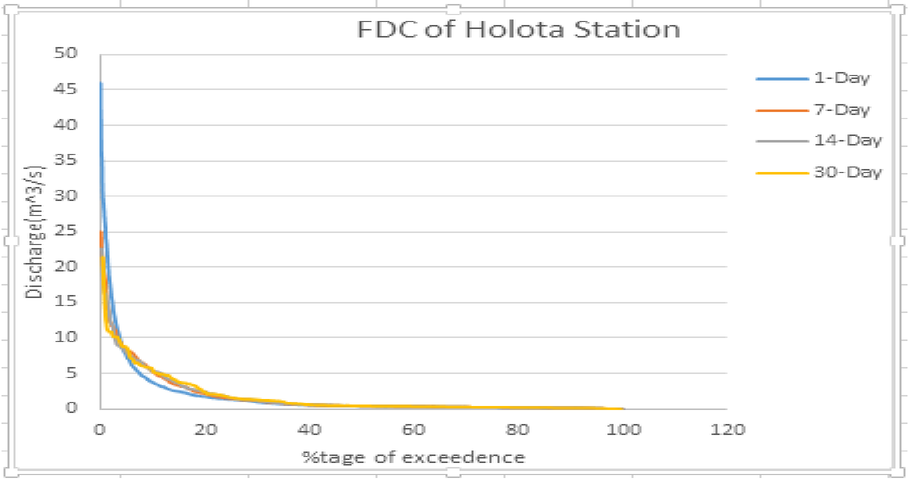
Weibull distribution

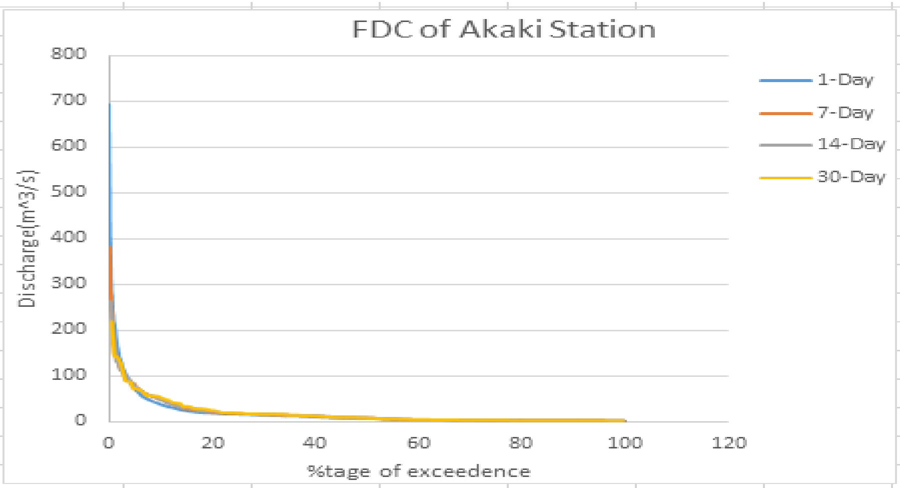
$$z_4 = 0.10701 - 0.11090z_3 + 0.84838z_3^2 + 0.06669z_3^3 + 0.00567z_3^4 + 0.04208z_3^5 + 0.03763z_3^6$$

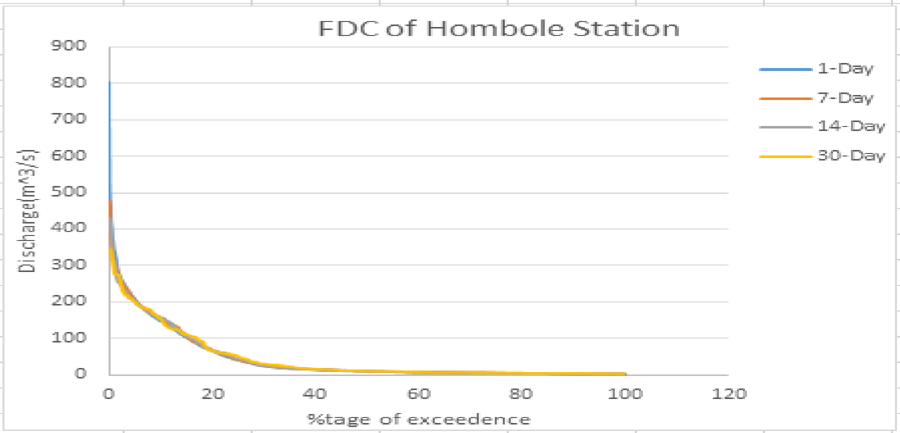
APPENDIX-H 1-Day, 7-Day, 14-Day and 30-Day FDC of all station

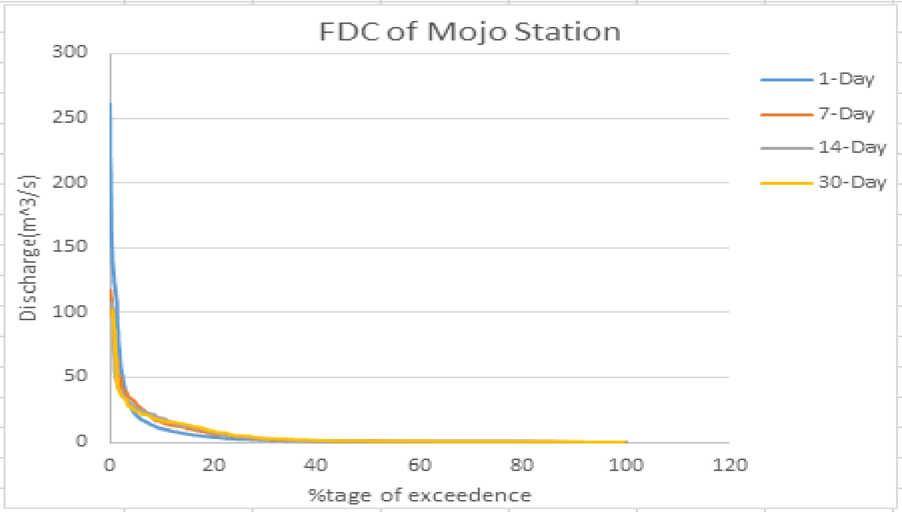


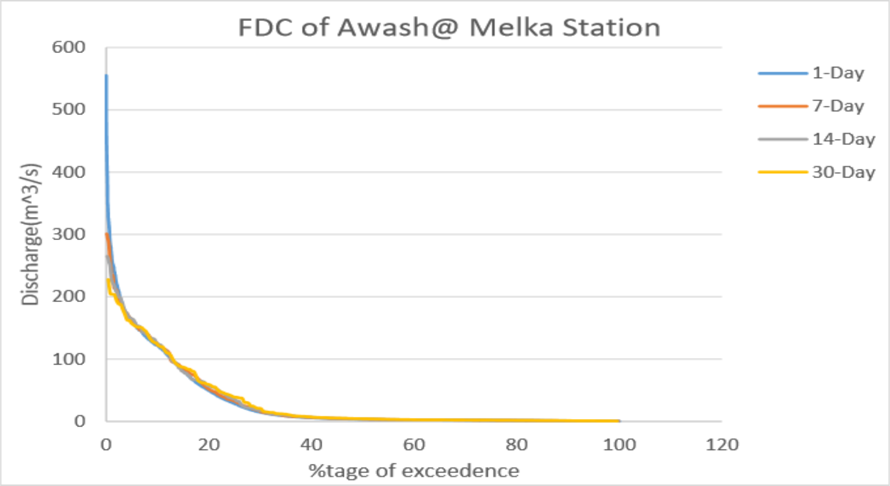


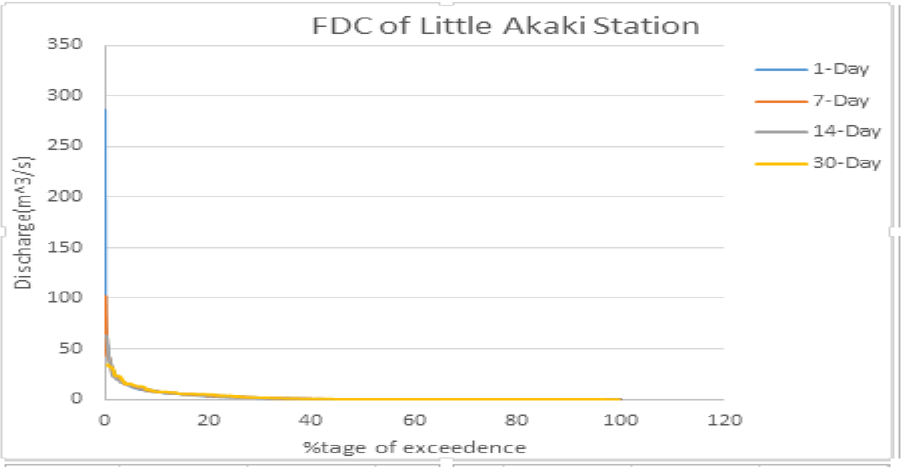


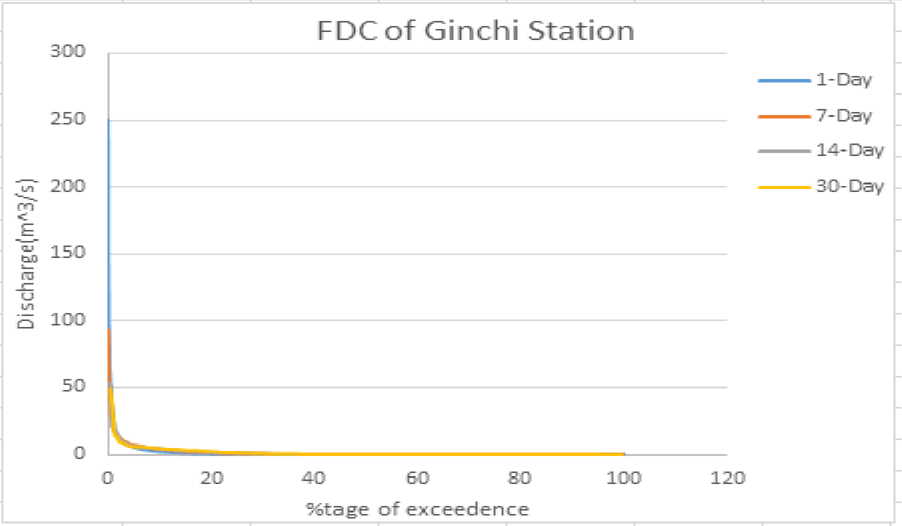

FDC of Akaki Station
800
700
600
500
400
300
200
100
0
Discharge(m^3/s)
0
20
40
60
80
100
120
%tage of exceedence
1-Day
7-Day
14-Day
30-Day


FDC of Hombole Station
900
800
700
600
500
400
300
200
100
0
Discharge(m^3/s)
0
20
40
60
80
100
120
%tage of exceedence
1-Day
7-Day
14-Day
30-Day

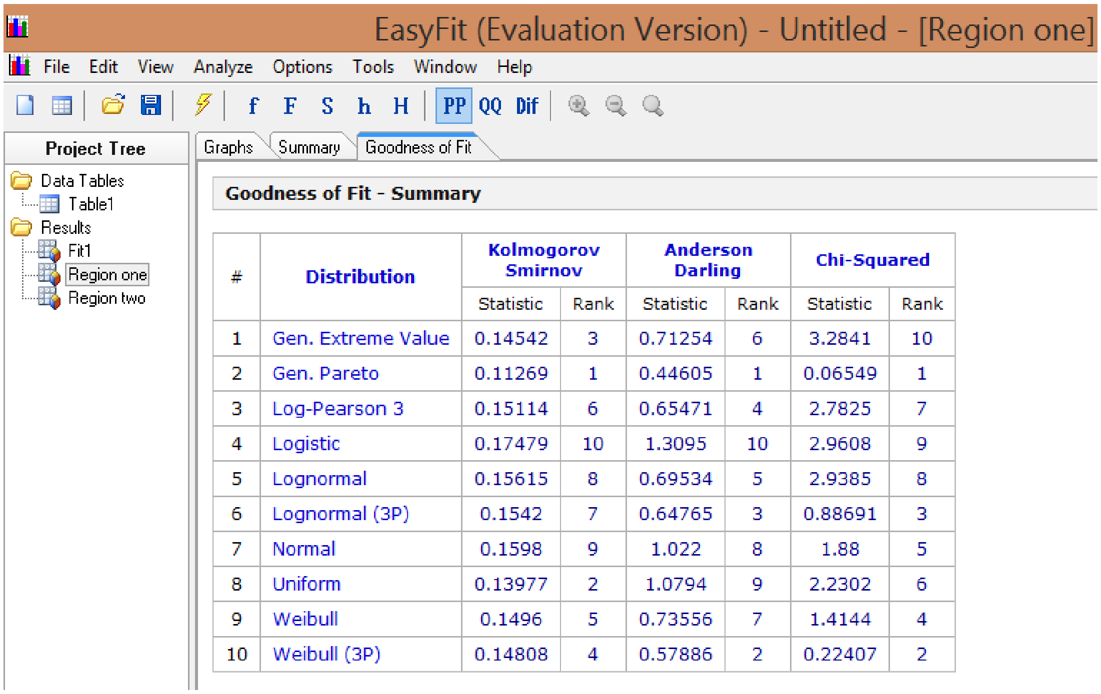

FDC of Mojo Station
300
250
200
150
100
50
0
Discharge(m^3/s)
0
20
40
60
80
100
120
%tage of exceedence
1-Day
7-Day
14-Day
30-Day


FDC of Awash@ Melka Station
Discharge(m^3/s)
600
500
400
300
200
100
0
0
20
40
60
80
100
120
%tage of exceedence
1-Day
7-Day
14-Day
30-Day


FDC of Little Akaki Station
Discharge(m^3/s)
350
300
250
200
150
100
50
0
0
20
40
60
80
100
120
%tage of exceedence
1-Day
7-Day
14-Day
30-Day

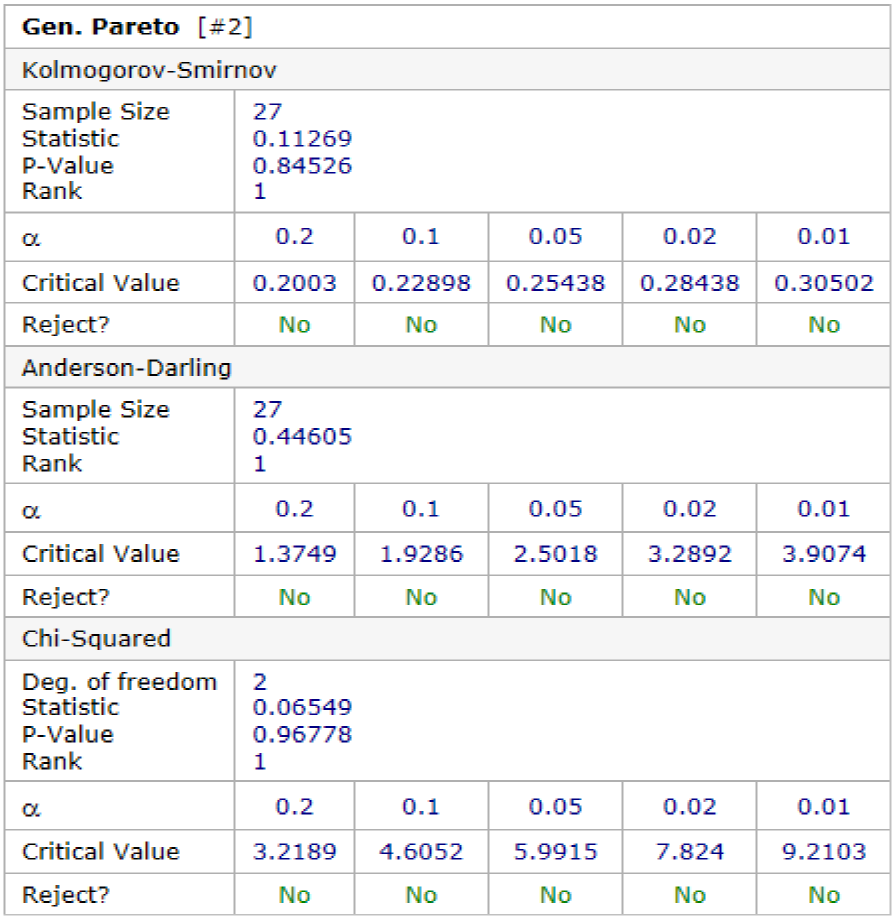

FDC of Ginchi Station
Discharge(m^3/s)
300
250
200
150
100
50
0
0
20
40
60
80
100
120
%tage of exceedence
1-Day
7-Day
14-Day
30-Day

APPENDIX-I Goodness of fit and descriptive Statistics of selected distribution



Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.14542	3	0.71254	6	3.2841	10
2	Gen. Pareto	0.11269	1	0.44605	1	0.06549	1
3	Log-Pearson 3	0.15114	6	0.65471	4	2.7825	7
4	Logistic	0.17479	10	1.3095	10	2.9608	9
5	Lognormal	0.15615	8	0.69534	5	2.9385	8
6	Lognormal (3P)	0.1542	7	0.64765	3	0.88691	3
7	Normal	0.1598	9	1.022	8	1.88	5
8	Uniform	0.13977	2	1.0794	9	2.2302	6
9	Weibull	0.1496	5	0.73556	7	1.4144	4
10	Weibull (3P)	0.14808	4	0.57886	2	0.22407	2



Gen. Pareto [#2]

Kolmogorov-Smirnov

Sample Size	27
Statistic	0.11269
P-Value	0.84526
Rank	1

α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.2003	0.22898	0.25438	0.28438	0.30502
Reject?	No	No	No	No	No

Anderson-Darling

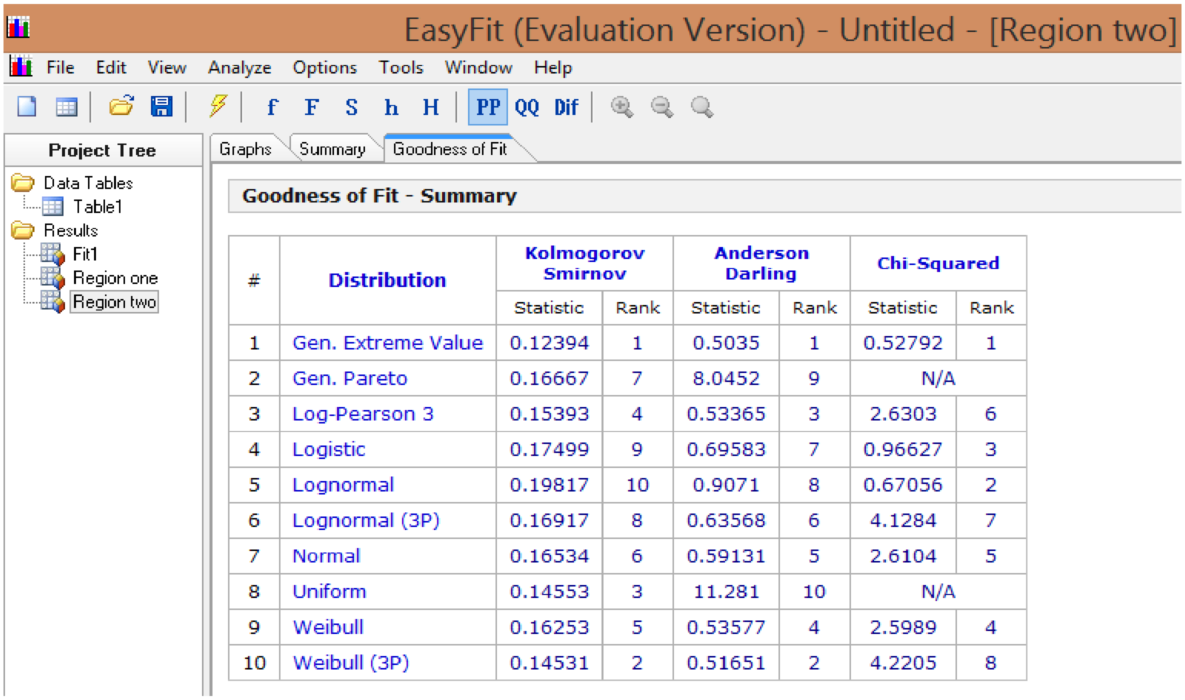
Sample Size	27
Statistic	0.44605
Rank	1

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

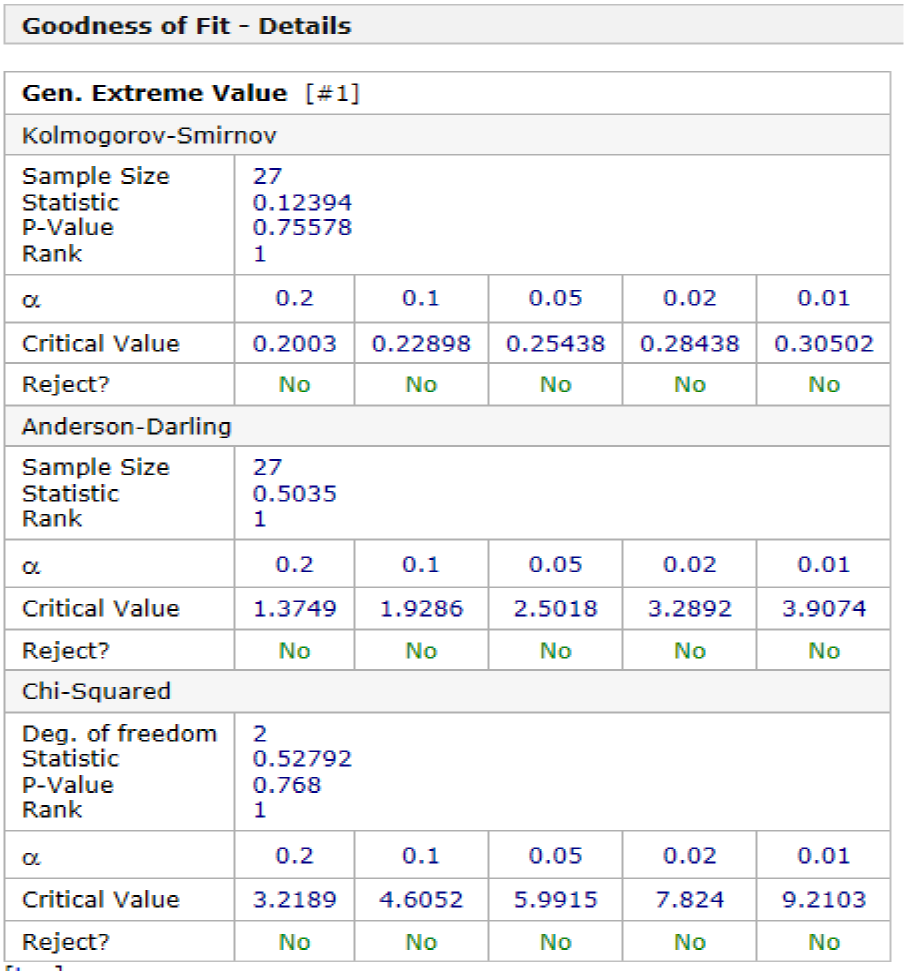
Deg. of freedom	2
Statistic	0.06549
P-Value	0.96778
Rank	1

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



Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Gen. Extreme Value	0.12394	1	0.5035	1	0.52792	1
2	Gen. Pareto	0.16667	7	8.0452	9	N/A	
3	Log-Pearson 3	0.15393	4	0.53365	3	2.6303	6
4	Logistic	0.17499	9	0.69583	7	0.96627	3
5	Lognormal	0.19817	10	0.9071	8	0.67056	2
6	Lognormal (3P)	0.16917	8	0.63568	6	4.1284	7
7	Normal	0.16534	6	0.59131	5	2.6104	5
8	Uniform	0.14553	3	11.281	10	N/A	
9	Weibull	0.16253	5	0.53577	4	2.5989	4
10	Weibull (3P)	0.14531	2	0.51651	2	4.2205	8



Goodness of Fit - Details

Gen. Extreme Value [#1]

Kolmogorov-Smirnov

Sample Size	27
Statistic	0.12394
P-Value	0.75578
Rank	1

α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.2003	0.22898	0.25438	0.28438	0.30502
Reject?	No	No	No	No	No

Anderson-Darling

Sample Size	27
Statistic	0.5035
Rank	1

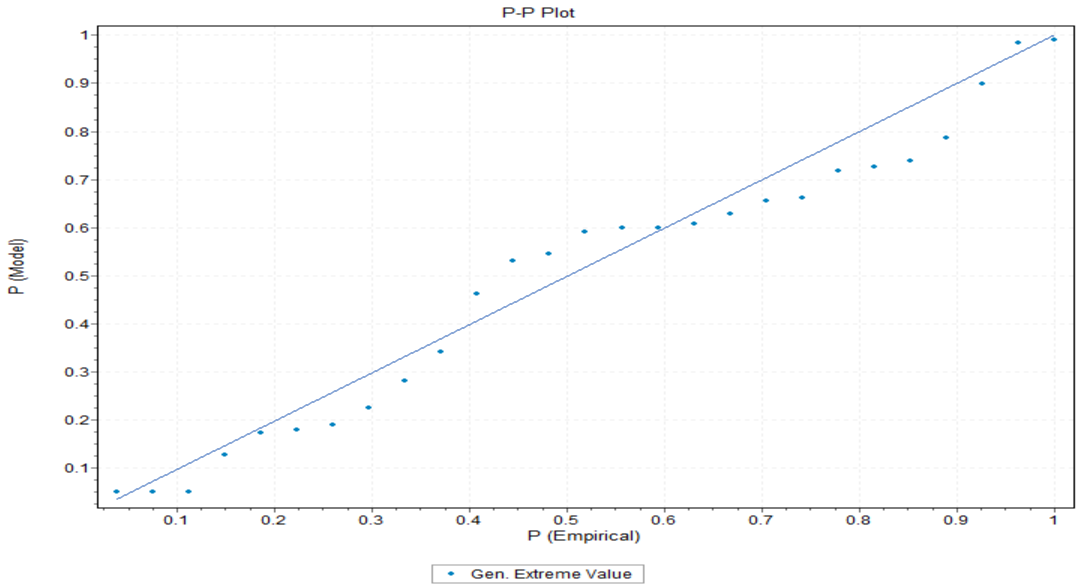
α	0.2	0.1	0.05	0.02	0.01
Critical Value	1.3749	1.9286	2.5018	3.2892	3.9074
Reject?	No	No	No	No	No

Chi-Squared

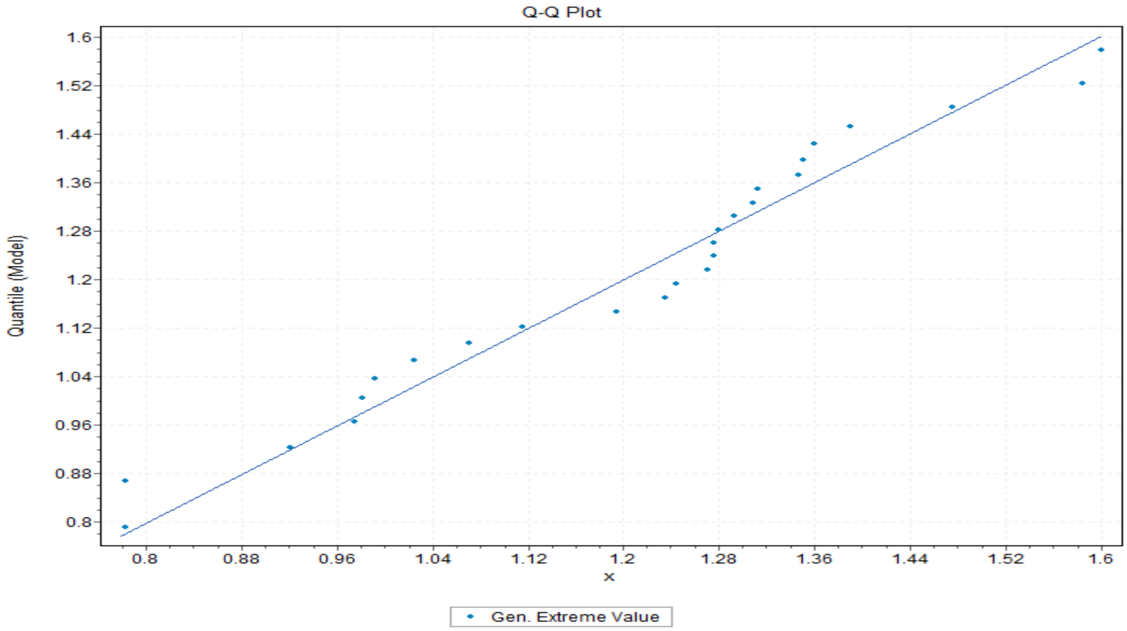
Deg. of freedom	2
Statistic	0.52792
P-Value	0.768
Rank	1

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

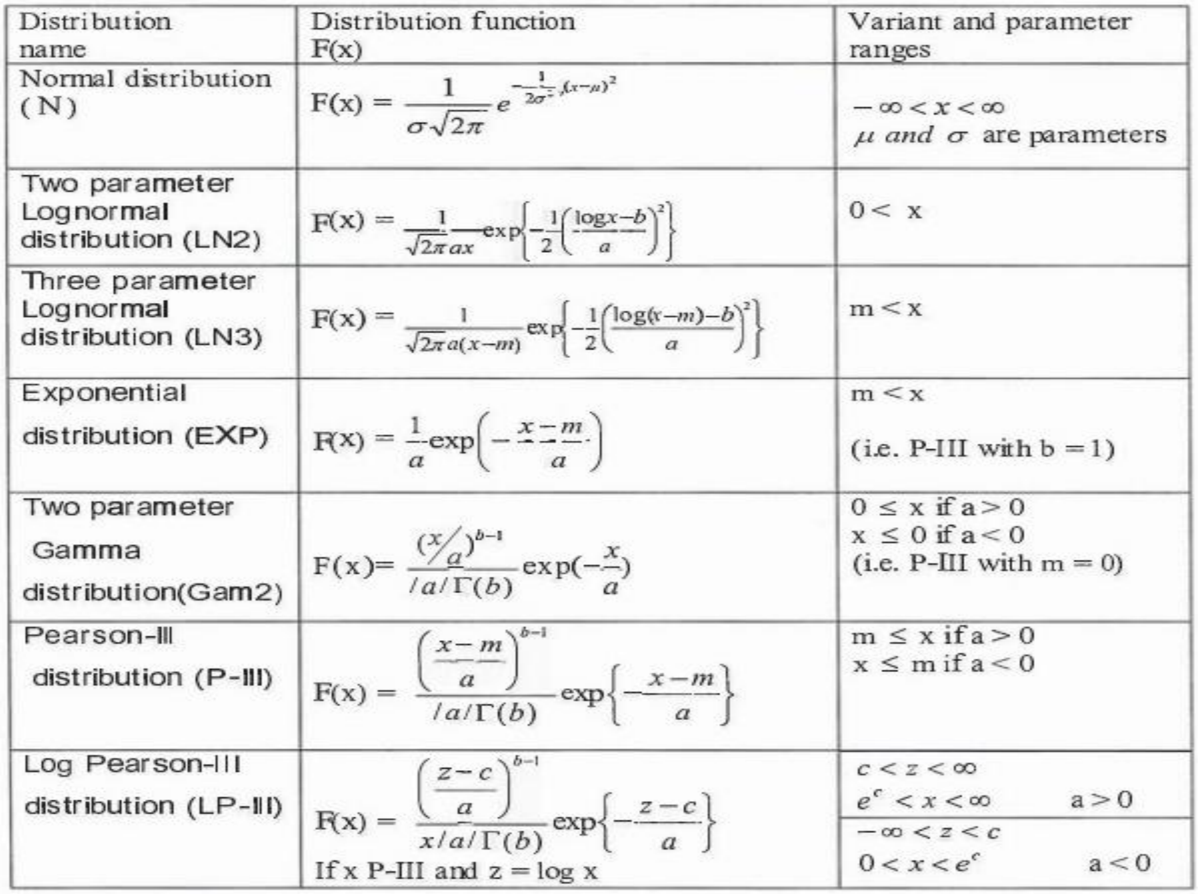
APPENDIX-J Probability- Probability Plots of Region two



APPENDIX-K Quantile-Quantile Plots of Region two



APPENDIX-L Mathematical expresion for distribution

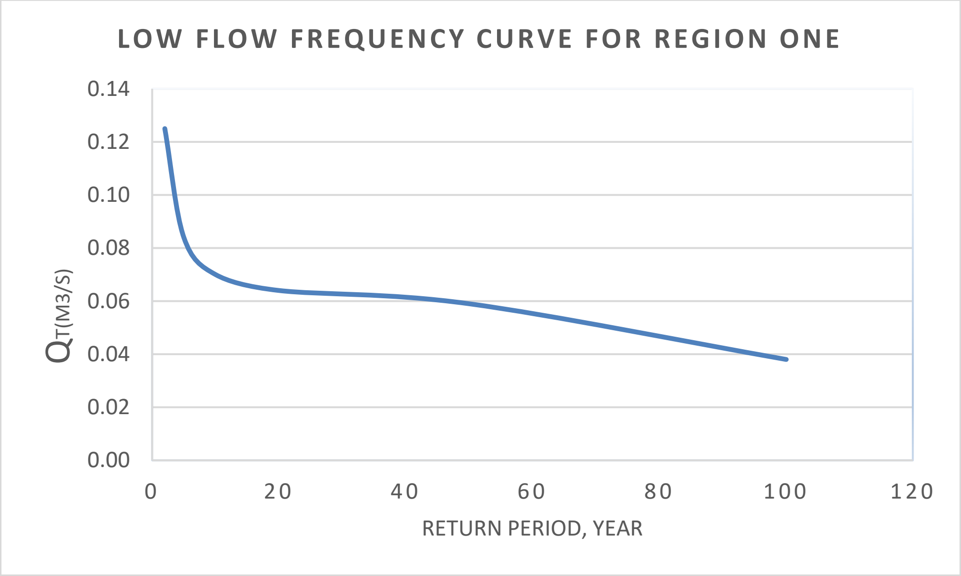


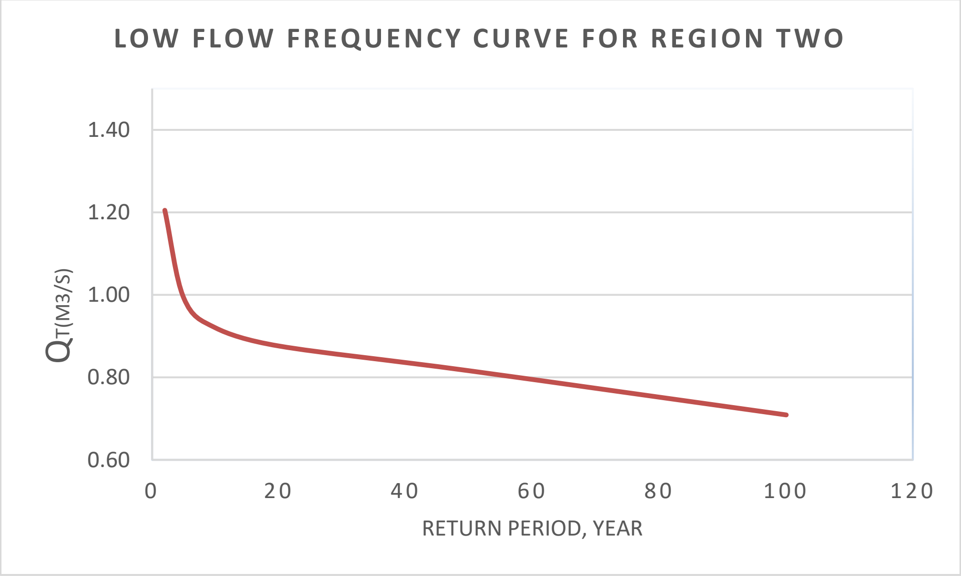
Distribution name	Distribution function F(x)	Variant and parameter ranges
Normal distribution (N)	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$	$-\infty < x < \infty$ μ *and* σ are parameters
Two parameter Lognormal distribution (LN2)	$F(x) = \frac{1}{\sqrt{2\pi}ax} \exp\left\{-\frac{1}{2}\left(\frac{\log x - b}{a}\right)^2\right\}$	$0 < x$
Three parameter Lognormal distribution (LN3)	$F(x) = \frac{1}{\sqrt{2\pi}a(x-m)} \exp\left\{-\frac{1}{2}\left(\frac{\log(x-m) - b}{a}\right)^2\right\}$	$m < x$
Exponential distribution (EXP)	$F(x) = \frac{1}{a}\exp\left(-\frac{x-m}{a}\right)$	$m < x$ (i.e. P-III with $b = 1$)
Two parameter Gamma distribution(Gam2)	$F(x) = \frac{(x/a)^{b-1}}{/a/\Gamma(b)} \exp(-\frac{x}{a})$	$0 \le x$ if $a > 0$ $x \le 0$ if $a < 0$ (i.e. P-III with $m = 0$)
Pearson-III distribution (P-III)	$F(x) = \frac{\left(\frac{x-m}{a}\right)^{b-1}}{/a/\Gamma(b)} \exp\left\{-\frac{x-m}{a}\right\}$	$m \le x$ if $a > 0$ $x \le m$ if $a < 0$
Log Pearson-III distribution (LP-III)	$F(x) = \frac{\left(\frac{z-c}{a}\right)^{b-1}}{x/a/\Gamma(b)} \exp\left\{-\frac{z-c}{a}\right\}$ If x P-III and $z = \log x$	$c < z < \infty$ $e^c < x < \infty$ $a > 0$ $-\infty < z < c$ $0 < x < e^c$ $a < 0$

APPENDIX-M Standardized low flow quantiles for each region

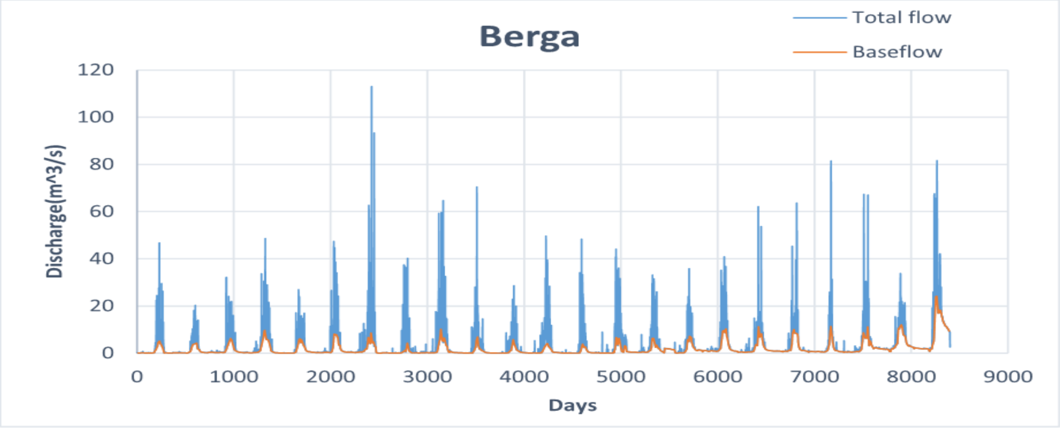
Gumbel Reduced Variate	Region One	Region two
0.367	0.590	1.034
1.50	0.397	0.851
2.250	0.331	0.790
2.970	0.302	0.752
3.902	0.279	0.700
4.60	0.146	0.609

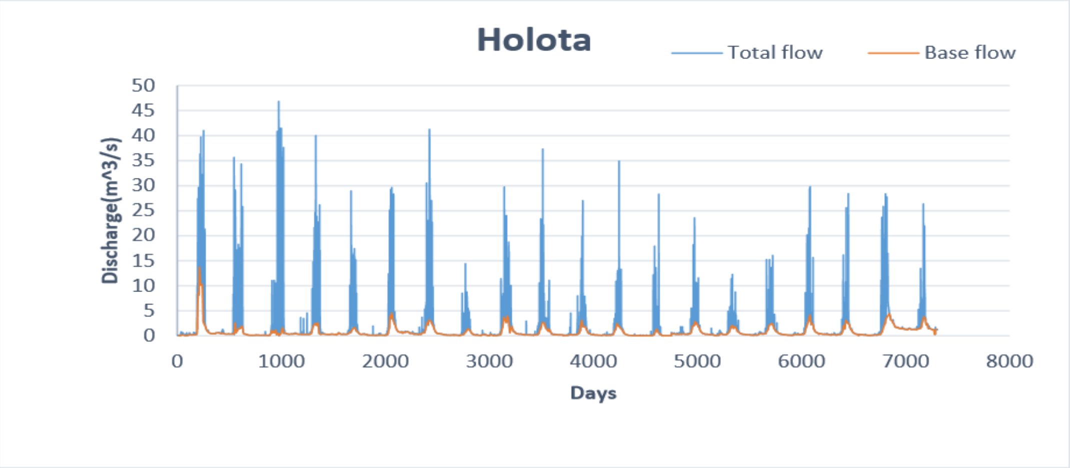
APPENDIX-N: Low flow Frequency curve of each region

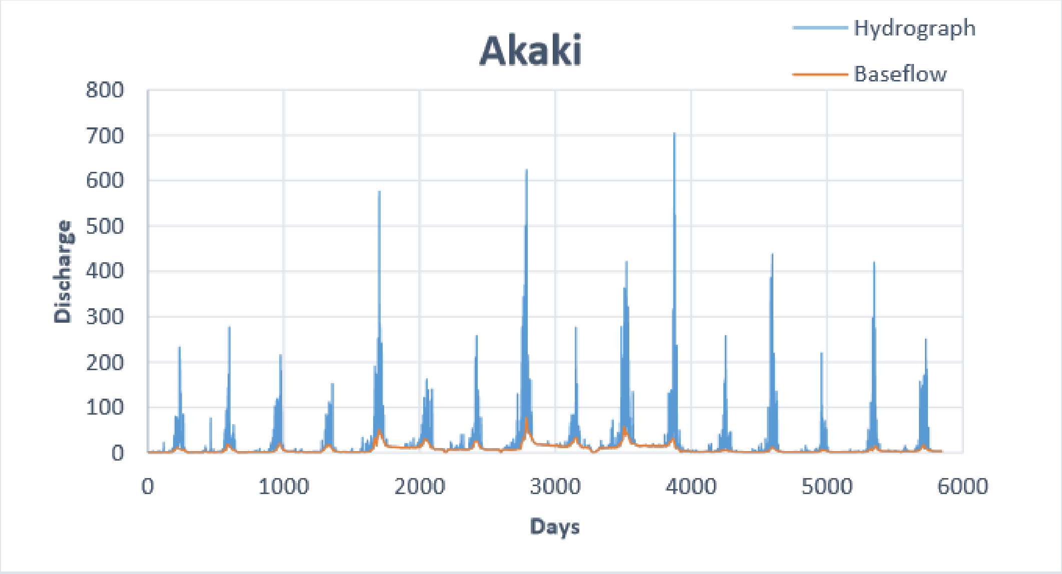


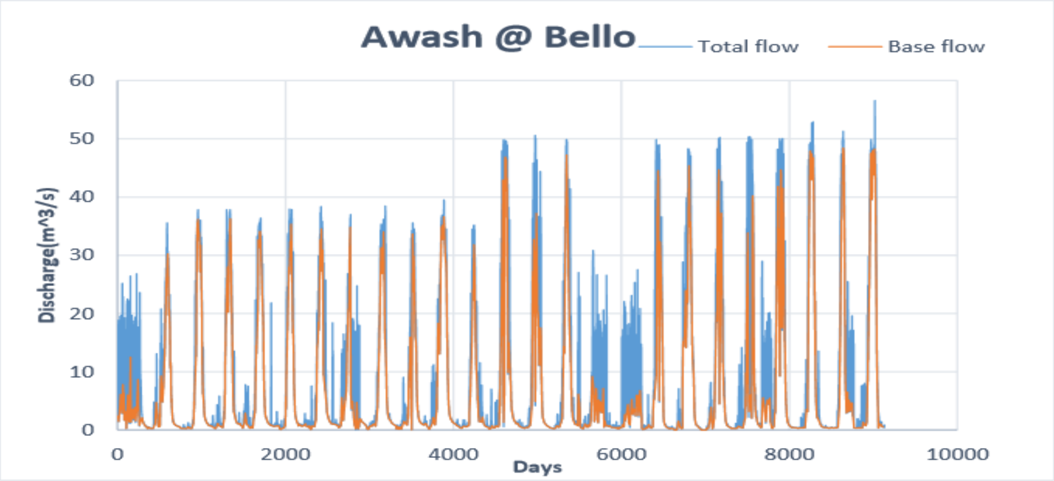


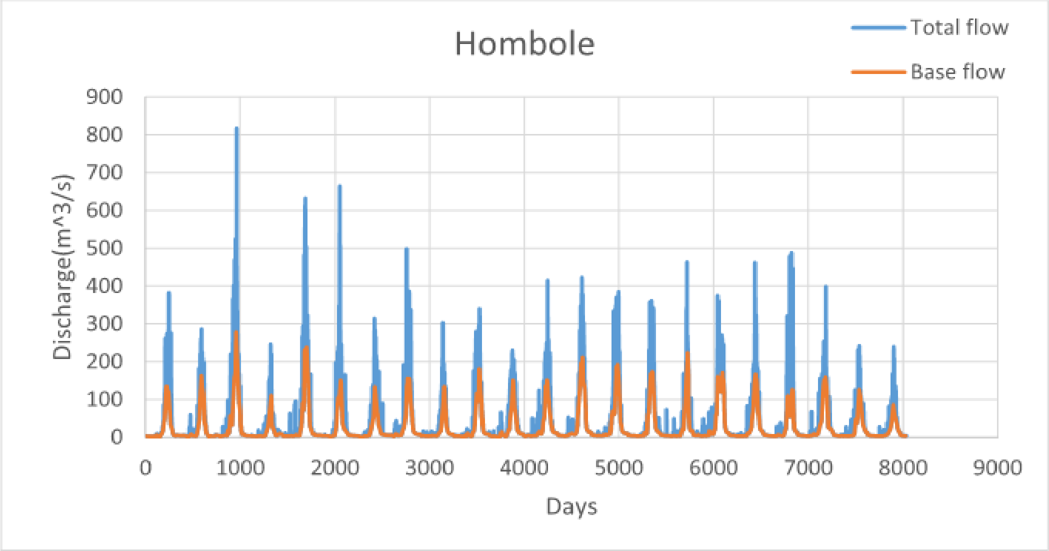
APPENDIX-O Base flow and total flow hydrograph of each station

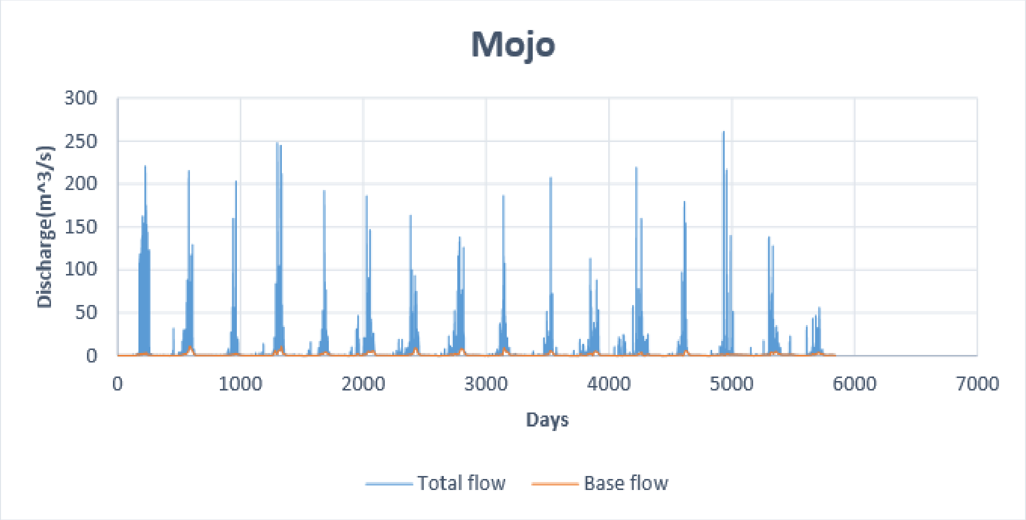


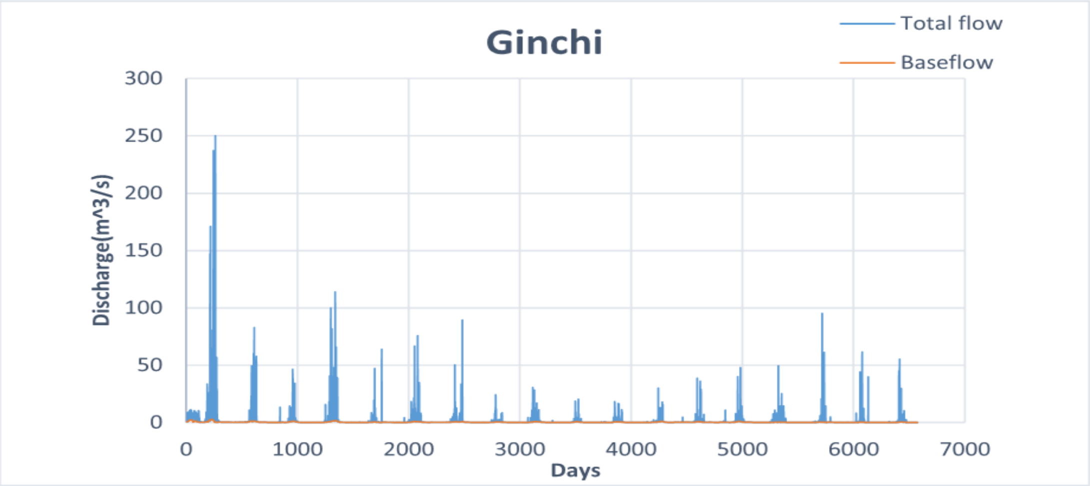


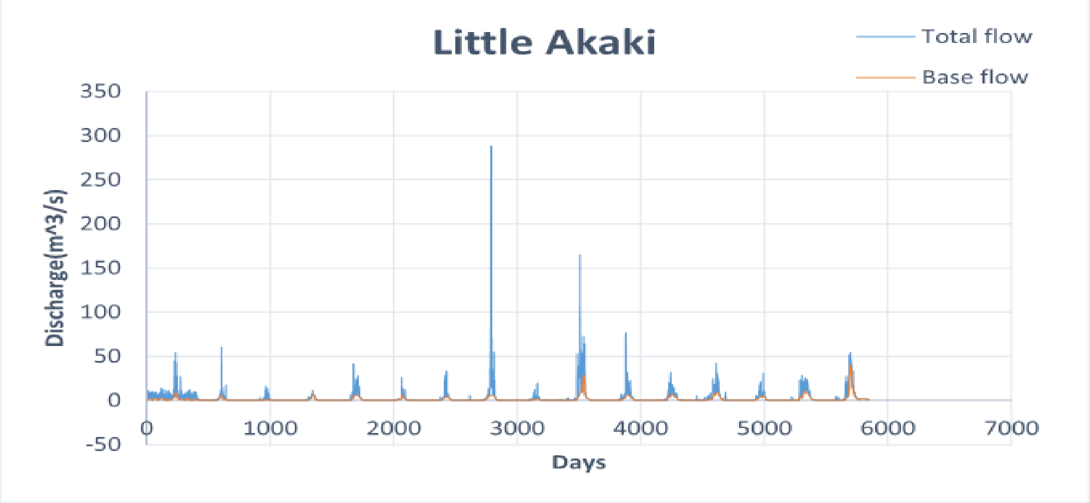


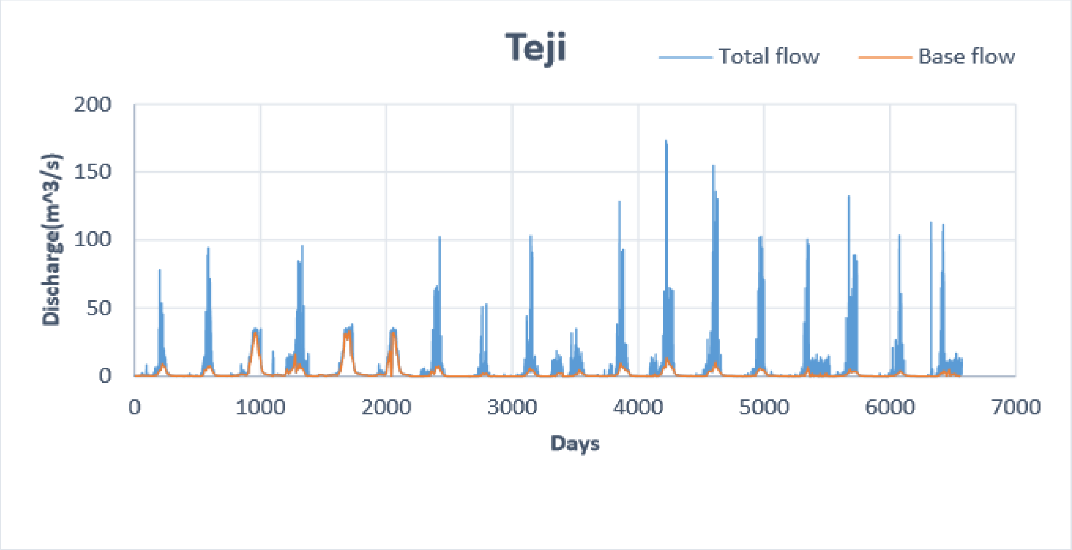

Awash @ Bello
Total flow
Base flow
Discharge(m^3/s)
60
50
40
30
20
10
0
0
2000
4000
6000
8000
10000
Days


Hombole
Total flow
Base flow
Discharge(m^3/s)
900
800
700
600
500
400
300
200
100
0
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
Days


Mojo
Discharge(m^3/s)
300
250
200
150
100
50
0
0
1000
2000
3000
4000
5000
6000
7000
Days
Total flow
Base flow


Ginchi
Total flow
Baseflow
Discharge(m^3/s)
300
250
200
150
100
50
0
0
1000
2000
3000
4000
5000
6000
7000
Days


Little Akaki
Total flow
Base flow
Discharge(m^3/s)
350
300
250
200
150
100
50
0
-50
0
1000
2000
3000
4000
5000
6000
7000
Days


Teji
Total flow
Base flow
Discharge(m^3/s)
200
150
100
50
0
0
1000
2000
3000
4000
5000
6000
7000
Days