ADDIS ABABA UNIVERSITY

INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

REGIONAL FLOOD FREQUENCY ANALYSIS FOR ZEWAY – SHALA SUB BASIN, RIFT VALLEY BASIN, ETHIOPIA

By

Yayneshet Wondimu

July, 2020

Addis Ababa, Ethiopia

ADDIS ABABA UNIVERSITY

INSTITUTE OF TECHNOLOGY

SCHOOL OF CIVIL AND ENVIRONMENTAL ENGINEERING

REGIONAL FLOOD FREQUENCY ANALYSIS FOR ZEWAY – SHALA SUB BASIN, RIFT VALLEY BASIN, ETHIOPIA

Thesis submitted to the school of Post Graduate Studies of Addis Ababa University Institute of Technology, AAIT in partial fulfillment of the requirements for the degree of Master of Science in Civil and Environmental Engineering, Hydraulic Engineering Stream

By

Yayneshet Wondimu

Advisor:

Yilima Sileshi (PhD)

July, 2020

PLAGIARISM DECLARATION

I, the undersigned, declare that the thesis comprises my own work. In compliance with internationally accepted practices, I have acknowledged and refereed all materials used in this work. The views and ideas included in this thesis represent my own understanding and will not represent the stand of the institute or any other person. The findings are all generated using raw data generated by other relevant institutions and will only be liable to analysis steps and final results generated.

Name of the student: Yayneshet Wondimu **Signature________**

Date of submission: ________________

Place: Addis Ababa

Addis Ababa University

Institute of Technology

School of Civil and Environmental Engineering

THESIS APPROVAL SHEET

Student:

Name Signature Date

The following graduate faculty members certify that this student has successfully presented the necessary written final thesis and oral presentation for partial fulfillment of the thesis requirements for the Degree of Master of Science in Hydraulic Engineering.

Approved By:

Advisor:

Name Signature Date

External Examiner:

Name Signature Date

Internal Examiner:

Name Signature Date

Chair Holder:

Name Signature Date

Faculty Dean:

Name Signature Date

ACKNOWLEDGEMENTS

This work come to reality because of the support of many players who had genereously contributed their invaluable support to me. In this regard I would like to acknowledge and express my sincere appreciation for my advisor, Dr. Yilma Sileshi for providing his invaluable knowledge and genuine support which highly encouraged me and contributed greatly to the thesis. Moreover, I acknowledge the head of Civil Engineering Department of Addis Ababa University, Kombolcha institution of technology and Ethiopian Road Authority in writing a co-operation letters for concerned organizations and for funding my graduate study in Hydraulic Engineering stream at Addis Ababa University.

I also extend my acknowledgment to officials of ministry of water resource irrigation and energy and Ethiopian Metrology Agency for their smooth facilitation and active cooperation in providing relevant information.

In addition, I want to express my deep hearted filling and special thanks to my colleagues and good friends, Kidist Ayalew and Yemisrach Kebede for their support and contribution which have led to the success of this research. My appreciation also goes to my parents, siblings and my husband for their overall supports.

ABSTRACT

In this study index flood method based on L-moments were used for regional flood frequency analysis of Zeway – Shala sub basin. The purpose of this study is to develop the regional flood frequency (growth) curve and index flood estimation equation which is essential for computing flood quantiles for ungauged catchments in Zeway – Shala sub basin.

The index-flood method was applied to Zeway – Shala sub basin, based on the records of annual maximum instantaneous discharges at 10 stream gauging stations. Delination of comparatively homogeneous regions was performed on the basis of catchment characteristics and their similarity in certain parameters like slope, elevation, soil type, and flood statistics of the sub basin. Accordingly, Zeway – Shala sub-basin has been divided in to two nearly homogeneous regions.

Goodness of fit test measure using the statistical test like kolmogorove simirnove test and chi-squared testes to identify how well the selected distribution fit to the data was followed. For this study among the fitted distributions, Generalized Extreme Value distribution is selected for region one and Log Pearson Type III distribution is selected for region two based on statistical value, because of its simple to apply and widely acceptance nature.

For the selected distributions the statistical parameter estimated by Methods of Moment and the regional growth curve were developed for each regions by using the quantiles of the selected distributions. Finally the index flood estimation equation were developed by regression analysis with a coefficient of determination (R^2) 0.72 for region one and 0.63 for region two to predicted mean annual flood flow from ungagged catchment

Keywords: regional flood frequency analysis, annual maximum discharges, index-flood,

TABLE OF CONTENTS

CONTENTS PAGE

LIST OF ABBREVIATIONS

AMAX	Annual Maximum
Ck	Coefficient of Kurtosis
Cv	Coefficient of Variance
ECDF	Empirical cumulative distribution function
DEM	Digital Elevation Model
GEV	Generalized Extreme Value
GIS	Geographic Information System
GL	Generalized Logistic
GP	Generalized Pareto
GMM	Generalized Method of Moments
HRU	Hydrological Response Units
LCk	L Moment Coefficient of Kurtosis
LCv	L Moment Coefficient of Variance
LMRD	L Moment Ratio Diagram
LN	Log Normal
LN3	Log Normal with 3 Parameter
LS	Least Squares
LP3	Log Pearson Type III
m.a.s.l	Meter above Sea Level
MER	Main Ethiopian Rift
MOM	Method of Moments
MoWIE	Ministry of Water, Irrigation and Energy
MLM	Maximum Likelihood Method
PWM	Probability Weighted Moments
SPSS	Statistical Package for Social Science
SWAT	Soil and Water Assessment Tool
RFFA	Regional Flood Frequency Analysis
US	United States

LIST OF SYMBOLS

df	Degree of freedom
α	Significant level
K	Statistical Constant or Confidence Level
T	Return Period
QT	Quantiles
N	Number of Record year

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Background

Hydrological processes such as flood incidences are sometimes extremely complex natural events. They are the results of a number of parameters. This makes it difficult to model analytically due to contribution uncertainty. The floods in the catchment depend up on the characteristics of the catchment, rainfall, land use land cover, other catchment characteristics and other antecedent conditions. Each of these factors in turn depend up on a host of constituent parameters. This makes the estimation of flood peak a very complex problem leading to much different approach.

Flood frequency analysis provides crucial information for design of small to large scale engineering structures and water resource planning, and other development projects. The analysis involves estimation of a flood magnitude corresponding to a required return period or probability of exccedence.

One of the key objectives of frequency analysis is to relate the magnitude of extreme events to their frequency of occurrence through the use of probability distributions according to Chow et al., (1988). Data collected in a long period of time in a river basin are analyzed in frequency analysis. The data are assumed to be independent and identically distributed. The flood data are considered to be stochastic and may even be assumed to be space and time independent. (Rao & Hamed, 2000)

In practice, the true probability distribution of the data at a site or a region is unknown. The assumption that data in a given system arise from a single-parent distribution may be questionable when data from large watersheds are analyzed. In such cases, more than one type of rainfall or flow may contribute to extreme events in a region. However, for the analysis to be of practical use, simpler distributions are often used to characterize the relation between flood magnitudes and their frequencies. The performance of distributions is evaluated by using different statistical tests.

The availability of data is an important aspect in frequency analysis. The estimation of probability of occurrence of extreme floods is an extrapolation based on limited data and periodic peak rainfall and runoff conditions. Thus, the larger the database, the more accurate the estimates will be. From a statistical point of view, estimation from small samples may give unreasonable or physically unrealistic parameter estimates, especially for distributions with a large number of parameters. Large variations associated with small sample sizes can cause the estimates to be unrealistic. However, data may be limited for various reasons and regional analysis will be most useful in such conditions.

Regional analysis approach and use is based on the concept of regional homogeneity which assumes that annual maximum flow populations at several sites in a region are similar in statistical characteristics and are not dependent on catchment size (Cunnane, 1989). In conditions where there is data limitation and regionalization is a must to do for design approach or any other use accepted analysis approaches are followed.

One of the commonly applied approachs is Regional Flood Frequency Analysis (RFFA) that can facilitate the estimation of the QT (quantile) value at a location for which limited flow data exists using the growth curve. The advantage of at-site and regional method is that an unusual event that occurs in sub-basins can be taken into account over a wide area where they might have occurred and in conditions where there is similarity in sub-basin characteristics that determine flood frequency, flood peak and incidence. Some at-site and RFFA methods assume that a region is an area that has set of gauging sites whose flood producing behavior is more or less homogeneous in some quantifiable manner and slight variations if any are homogenized by the common measureable characters. RFFA exploits this homogeneity to produce quantile estimates, which, in most case, are more reliable than those obtainable from at site data alone (Cunnane, 1988).

There are some significant reasons for motivating any researcher to select the study area and to set Regional flood frequency for design of engineering structures, for flood management and optimum water use planning in the area. Among those reasons one is the area has water quality problem especially for some of the lakes. The irrigation practice is also growing fast and wide these days and most abstraction is uncontrolled abstraction without considering the effect of high abstraction on flow volume or productivity in general. If the irrigation practice is to grow further it will cause water use conflict in the basins, or even flow reversal from poor quality water sources (groundwater or poor quality water lakes) to the shallow ground and can cause irreversible effects on soil and productivity in the area. Hence, setting flood frequency curve by considering flow volume all along the year should be part of the future development planning in the area and a research interest for academicians. Not only these even if abstraction is high in the area estimation of flow or setting working formula for estimating ungagged flow is limited or even the dry time flow in some the rivers is reported to be very low and these and other factors such as relatively good access had attracted me to do the research in the selected area. If the irrigation practices should be done using optimum water sharing approaches the flood frequency curve and even setting minimum flow in rivers should be part the effort of developing the area and even future research interest.

1.2 Problem Statement

An important problem in hydrology is the estimation of flood magnitudes, especially planning and design of water resource projects, flood plain management, and optimal water resource management of basins depends on the frequency and magnitude of peak discharges, minimum flow and other basin characteristics. Thus, flood frequency analysis and available generalized frequency curves of areas set by detail consideration of most parameters that affect river hydrograph and peak flow conditions provide crucial information for the planning and design of many hydraulic structures such as culvert, bridge, reservoir, spillway, and road embankment and dykes, and for risk assessment in mitigation of flood plain.

These hydraulic structures require a consistent estimation of flood quantiles (QT) using reliable flood records measured at gauging stations, or properly estimated when gauging stations are not sufficient, problems in gaging stations such as siltation in gauging stations which can cover parts or gauging staffs, or poor maintenance of stations that may give unreal flow volume records.

It also requires detail consideration of catchment land use land cover conditions (land degradatom increase the peak discharge) and changes in LU/LC that affect peak flow curve estimations, and other parameters such as morphology considerations in peak flow curve setting and estimation, and man induced effects such as uncontrolled abstraction from rivers that should be considered in flood frequency curve setting of or hydrograph analysis for design purpose. But in developing countries like Ethiopia, the availability of the dense gauging stations is low, maintenance and station periodic clearing of silts and vegetation that can affect flow estimation is poor, and land use land cover changes are significant but mostly not reflected in hydrograph analysis and frequency curve setting. There is also a lot uncontrolled abstraction of water from rivers in the target catchment and water loss in canals out of the river which should have been known to set generalized flood frequency curves that can be realistic and useable for designing structures and estimation of flood volume is a great challenge in the target area due to uncontrolled abstraction and water loss from open canals.

The operation and maintenance of the stream gauging networks in Ethiopia is difficult or takes longer period to do and creates data gap. Data recording sometimes may not give correct values in the analysis approach of frequency analysis or they may be too short to allow for a reliable estimation of extreme events or there is no flow record available at the target sites.

Therefore, this study used available information or data recording with its limitations and the other inputs such as the effect of erratic variation of land use land cover change and its impact on peak flow without significant change in rainfall volume, and the effects of uncontrolled and ungagged abstraction and water use in variably permeable area which also has effect on the flood frequency setting analysis, and other effects such as morphology impact on peak flow recording should be fully addressed by concerned government bodies to establish best and generalized frequency curve for design and optimal water use planning in the basin.

Hence, this thesis will give flood frequency curve based on the recorded data and gives usable formula for the ungagged stations flood peak of the catchments for the next action of the hydraulics structure design, flood plain management (particularly acceptable low dry time flow in river versus controlled canal abstraction from river should be in alignment) with in the study area. The generalized flood frequency curve will also help to set optimal water use planning in the catchments too.

1.3 Objectives of the Research

1.3.1 General Objective

The main objective of this research is to find analysis of regional flood frequency for use in the Zeway-shala sub- basin, Rift valley basin Ethiopia.

1.3.2 Specific Objective

The research has the following specific objectives: -

- Identify and delineate the homogeneous region.
- Identify the best fit statistical distributions and method of parameter estimation for the data of the stations.
- Drive a regional flood frequency curve for the regions.
- Compute quantiles on the ungagged catchments using the standardized growth curve and the index flood on the regression mode. (Design flood for ungagged catchment).

1.4 Significance of the Research

This study will give the result of quantiles (QT/Qmean) and return period T for the homogeneous region and also quantify quantiles for the ungagged catchment. This will give useful information in order to get a future stream flow data on the basin which gives a relevant data for the hydraulics structure designer, optimal water use planning, and for flood management works in the basin and to control risk on flood.

1.5 Scope of the research

The research had to base based on available data recording of river flow in the catchments and a workable flood frequency curve was set for next plan of action such as for design considerations, optimum water use planning, and flood plain management particularly considering abstraction in basins versus acceptable minimum flow in the river for a sustained and conflict free use of the water of rivers all across the catchments.

A working formula that can help estimate flow in ungagged areas is also developed based on the available information and data. However, the generalized flood frequency curve setting or any detailed analysis of hydrology data would have been best if the effect of land use land cover change variations can be seen from time series land use land cover maps which are not available, if morphology was fully considered in gauging station locations selections, and if water abstraction for any purpose in the basins was fully controlled to reflect water loss and use effects on the analysis and flood frequency curve setting. Yet, the results are generalized and useable for design of engineering structures, optimal water use planning, and for flood management works in the basin.

1.6 Research Questions

To address the above objectives, the following research questions are posed:

(i) How to delineate homogeneous regions in the zeway-Shala sub-basin.

(ii) What are the best probability distributions and parameters in the zeway-Shala sub basin?

(iii) How to develop the flood frequency curve for gagged stations in the zeway-Shala sub basin.

(iv) What are the best multi-regression models that could be used at ungagged sites in the study area?

1.7 Organization of the Research

This study is organized in to five chapters. The first chapter is an introduction to this research which contains background, Statement of the problem, objective of the research, significance of the research, research questions. The second chapter comprises review of related literature mainly dealing with different citations of journal articles, books, brochures, reports, strategies, guidelines, and other similar sources employed to support this research. Chapter three is focused on materials and methods of the research, description of the study area, data collection, data quality checking and method of analysis. Chapter four consists of data analysis and data interpretations of the Study. Furthermore, the chapter incorporated identification of homogeneous regions, selection of parent distribution and parameter estimation, estimation of quantile and develop regional growth curve and design flood for the ungagged catchments. Chapter five consists of the conclusions and recommendations based on the analyses and discussion lessons.

2 LITERATURE REVIEW

2.1 Flood Frequency Estimation

The maximum flood or peak flow and return periods for such a flow are required during the design of engineering structures and various hydraulic works such as design of weir, bridge, different scale dams, irrigation facilities, flood control measures, etc. Over/under-estimates of flood peaks and associated design floods can cause losses like waste of resources, infrastructural damage, human life and many other social and economic impacts. Mistakes in hydrological data estimates or the use of flood frequency curves which might not represent the reality can cause devastating damage such as structural failure or may lead to project failure before the design period.

Several methods can be sued to estimate the peak flow and recurrent events of flood that can be considered during designing to avoid the risk of peak floods on engineering structures and to scale-up the benefits of social and economic impacts of hydraulic structures such as dams to the community and the country at large. The estimation methods can be based on rainfall volume or measured or estimated stream/river flow volume and classified into two groups based the data type used (Fig. 2.1): When rainfall volume is used it is referred as rainfall-based method; or if estimated stream or river flow volume is used the method is referred as stream flow-based method. Rainfall-based method is more scientific and can account easily the changes of climate, land use, morphology or elevation effects, etc. However, the rainfall-based methods are data/skill intensive and require appropriate estimation of the rainfall volume through the use properly sited/selected gauging stations and up to date measuring approaches. It also requires filtering of hill shed effects on rainfall volume variability. On the other hand, stream flow-based methods are relatively simpler. The stream flow-based equations are reliable for the regions where the stations are limited and many flow-control structures are not available. In stream flow-based method, no assumption is required regarding the relationship between the probabilities of rainfall and runoff.

The stream flow methods are mainly based on the analysis of stream flow data. These methods include empirical equations, and at site or regional statistical analyses. Regional analysis methods are usually applied to estimate design floods at locations with inadequate stream flow data or no data and will be more reliable when there are control points for data checking.

Direct-regression and index-flood methods are the two commonly applied methods during the regional flood frequency analysis and estimation. Delineation of catchments or sub-catchments with relatively homogeneous hydrologic regions the first major step of any regional flood-frequency analysis for places where there is limited gauged information. Regionalization is performed to transfer the hydrologic characteristics from gauged basins to ungagged basins considering the morphology variations, climate variability across the area, vegetation, and other considerations that can cause significant variation in rainfall volume of the areas of interest.

The estimation of a regional flood frequency distribution is widely applicable and practical means of providing flood information at sites with little or no local data especially when the controlling sites are well-calibrated and reliable. The derivation and use of a dimensionless flood distribution that can be applied to all drainage basins with in assumed homogeneous regions was first described by Dalrymple (1960).

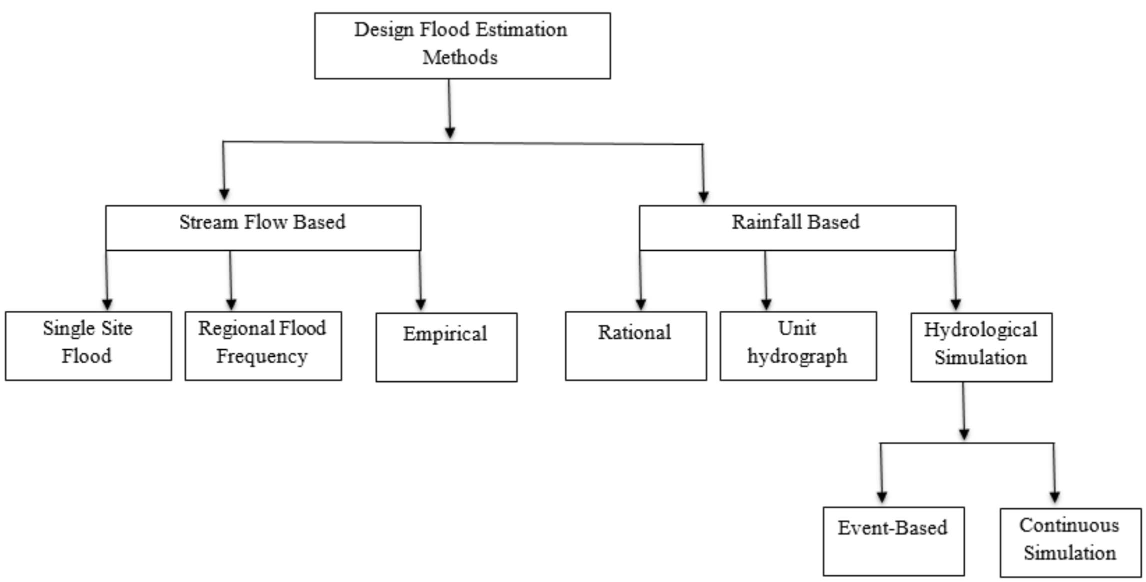


Figure 2-1:- Flow chart that indicates the steps and methods for estimating deign flood

2.2 Flood Frequency Models

In flood frequency analysis, a unique relationship between a flood magnitude and the corresponding recurrence interval T is sought. The aim is to obtain information from a flow record so as to estimate the relationship between Q and T. Three different models can be considered for this purpose (Cunnane, 1989).

These models are

1. The annual maximum series (AM) model,
2. The partial duration series (PD) or peaks over a threshold (POT) model, and
3. The time series (TS) model.

In the annual maximum (AM) flow series, only the peak flow in each year of record is taken. However, the use of an AM series may involve some loss of information due to generalization. For example, the second or third peak within a year may be greater than the maximum flow in other years and yet they are ignored (Kite, 1977; Chow et al. 1988). In this approach the smaller peaks of the other years are taken even if the total flow in such a case may seem lower than considering the second highest peaks of a given year as there is no certain fixed value above which the peaks should be considered in this approach. This scenario is avoided in the partial duration (PD) or the peaks over a threshold (POT) models where all peaks above a certain selected base value are considered. The base is usually selected low enough to include at least one prominent event in each year (Kite, 1977). The PD (or POT) model approach, however, is limited by the fact that observations may not be independent (Chow et al., 1988). According to Cunnane (1989) the AM model is statistically efficient than the PD model when λ is small (λ< 1.65), where λ is the mean number of peaks per year included in the PD series. The return period TE for a PD (or POT) model is related to the return period T of an AM model by Eq. 2.1 (Chow, 1964).

$$\mathrm{TE} = [\log(\frac{\mathrm{T}}{\mathrm{T}-1})]^{\wedge}-1 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad 2.1$$

The relative difference between TE and T is greatest for small values of T and converges to 0.5 as T increases. Rosbjerg (1977) demonstrated that when events follow a Poisson process, the relationship between return periods of annual maximum and partial duration series given in Eq. 2.1 is exactly satisfied. But AM series is widely and commonly used model by different researchers for the purpose of flood frequency analysis.

2.3 Regionalization for the Flood Frequency Analysis

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

Regionalization, in the context of RFFA, refers to identification of homogeneous regions through homogeneity test and selection of appropriate frequency distribution for the identified region and stations.（Mengestu, 2008）

Identification of homogeneous regions is the most difficult step in the model approach and requires the high level subjective judgment which may induce some errors. The aim is to get and form groups of sites that approximately satisfy the homogeneity condition that the sites frequency distributions are identical apart from a site-specific scale factor. This is usually achieved by partitioning the sites in to disjoint groups. (Hosking & Walls, 2005)

Several attempts have been made by different researchers to identify areas with nearly homogeneous hydrological condition based on either geographical considerations or flood data characteristics, or a combination of both (Mkhandi, 1985).

Many types of regionalization procedures are available (Cunnane, 1988 and 1989). One of the simplest procedures which have been used for a long time is the index flood method. The main assumption in the index flood method is that the distribution of floods at different sites in a region is the same except for a scale or index flood parameter, which reflects rainfall and runoff characteristics of each region.

The delineation of homogeneous regions is closely related to the identification of the common regional distributions and their similarity in the target regions and applicability within each region. A region can only be considered homogeneous if sufficient evidence and coinciding parameters can be established for the different sites (sub-catchments) in the region that are drawn from the same parent region considered homogeneous in magnitude (except for the scale variations).

2.4 Selection of Frequency Distribution

Choice of distribution for AM series has received wide spread attention. The choice of distribution is influenced by many factors, such as methods of discrimination between distributions, methods of estimation parameters, the availability of data, etc. Generally, there is no general global agreement as to a preferable technique of model choice and no single one distribution accepted universally. For example, Log Pearson type II distribution was recommended by U.S. Water Resource council (1967) for USA, General Extreme Value (GEV) distribution was recommended by the Natural Environmental Research council. (NERC, 1975) for U.K. and Ireland, Pearson type III was selected by the institute of Engineers in Australia etc.

In general the chosen distribution should be (Cunnane, 1989)

- widely accepted
- Simple and convenient to apply
- Consistent, flexible or robust (low sensibility to outliers
- Theoretically well based (established)
- Documented in the Guide (WMO) or else where

There are many distributions that have been suggested for AM series models. Some of them are (Cunnane, 1989)

a) Normal and related distributions

- Normal distribution
- Log normal two parameter distribution
- Log normal three parameter distribution

b) The Gamma distribution

- Exponential distribution
- Two parameter Gamma distribution
- Pearson three distributions
- Log Pearson two distributions

c) Extreme value distribution

- Generalized extreme value distribution
- Extreme value type I distribution
- Extreme value type II distribution
- Weibull distribution

d) Wake-by distribution

- Five parameter wake-by distribution
- Four parameter wake-by distribution
- Generalized pareto distribution

e) Logistic distribution

- Log-logistic distribution
- Generalized logistic distribution

The identification of a parent distribution can be achieved more easily by using L-moment ratio diagrams and using software.

L- Moment ratio diagram

L-moment ratio diagrams are based on relationships between the L-moment ratios. A diagram based on LCs ($\tau 3$) versus LCk ($\tau 4$).

Some useful relationships for constructing the L-moment ratio diagram for some common distribution are given by Hosking (1990 and 1991a, b):

1. Uniform: $\tau 3 = 0$, $\tau 4 = 0$

2. Exponential: $\tau 3 = 1/3$, $\tau 4 = 1/6$
3. Gumbel (EV1 (2)): $\tau 3 = 0.169$, $\tau 4 = 0.1504$
4. Logistic: $\tau 3 = 0$, $\tau 4 = 1/6$
5. Normal: $\tau 3 = 0$, $\tau 4 = 0.1226$
6. Generalized Pareto:

$$\tau 4 = \frac{\tau 3 * (1 + 5 * \tau 3)}{5 + \tau 3} \text{ or}$$

$$\tau 4 = 0.20196 * \tau 3 + 0.95924 * (\tau 3)^2 - 0.20096 * (\tau 3)^3 + 0.04061 * (\tau 3)^4$$

7. Generalized Logistic:

$$\tau 4 = \frac{(1 + 5 * (\tau 3)^2)}{6}$$

or $\tau 4 = 0.16667 + 0.83333 * (\tau 3)^2$

8. Generalized Extreme Value:

$$\tau 4 = 0.10701 + 0.11090\tau 3 + 0.84838(\tau 3)^2 - 0.06669\,(\tau 3)^3 + 0.00567\,(\tau 3)^4 - 0.04208\,(\tau 3)^5 + 0.03763(\tau 3)^6$$

9. Gamma and Pearson III:

$$\tau 4 = 0.1224 + 0.30115\,(\tau 3)^2 + 0.95812\,(\tau 3)^4 - 0.57488\,(\tau 3)^6 + 0.19383\,(\tau 3)^8$$

10. Lognormal (two and three parameters):

$$t4 = 0.12282 + 0.77518\,(\tau 3)^2 + 0.12279(\tau 3)^4 - 0.13638\,(\tau 3)^6 + 0.11368\,(\tau 3)^8$$

11. Wakeby lower bound:

$$\tau 4 = -0.07347 + 0.14443\,\tau 3 + 1.03879(\tau 3)^2 - 0.14602\,(\tau 3)^3 + 0.03357(\tau 3)^4$$

12. Overall lower bound: $\tau 4 = -0.25 + 1.25\,(\tau 3)^2$

2.5 Parameter and Quantile Estimation

After a distribution or a number of distributions are selected to fit the data, their parameters must be estimated. The estimated parameters are used to calculate quantile estimates for different return periods or, conversely, to calculate the return period for a given flood magnitude. This is achieved by using the distribution function, in which the parameters of the distribution are replaced by their estimates and the relationship between return periods (T) and probability of non-exccedence (F) in the form F = 1 – 1/ T is used. Some of the common methods of parameter estimation are discussed below.

Parameter Estimation

Several methods can be used for parameter estimation. These include the method of moments (MOM), the maximum likelihood method (MLM), the probability weighted moments method (PWM), the least squares method (LS), maximum entropy (ENT), mixed moments (MIX), the generalized method of moments (GMM), and incomplete means method (ICM). Three of the more commonly used methods are considered here, namely, the method of moments (MOM), the maximum likelihood method (MLM) and the probability weighted moments method (PWM).

The maximum likelihood method (MLM) is taken as the most efficient and more reliable method since it provides the smallest sampling variance of the estimated parameters, and hence of the estimated quantiles, compared to other methods. However, for some particular cases, such as the Pearson type III distribution, the optimality of the ML method is only asymptotic and small sample estimates may lead to estimates of lower quality (Bobeé and Ashkar, 1991). Also, the ML method has the disadvantage of frequently giving biased estimates, but these biases can be corrected. Furthermore, it may not be possible to get ML estimates with small samples, especially if the number of parameters is large. The ML method requires higher computational efforts, but with the increased use of high-speed personal computers, this is no longer a significant problem.

The method of moments (MOM) is a natural and relatively easy parameter estimation method. However, MOM estimates are usually inferior in quality and generally are not as efficient as the ML estimates, especially for distributions with large number of parameters (three or more), because higher order moments are more likely to be highly biased in relatively small samples.

The PWM method (Greenwood et al., 1979; Hosking, 1986a) gives parameter estimates comparable to the ML estimates, yet in some cases the estimation procedures are much less complicated and the computations are simpler. Parameter estimates from small samples using PWM are sometimes more accurate than the ML estimates (Landwehr et al., 1979c). Also, in some cases, such as the symmetric lambda and Weibull distributions, explicit expressions for the parameters can be obtained by using PWM, which is not the case with the ML or MOM methods. Kebaili-Bergaoui (1994) showed that the ML and ME methods of parameter estimation for Weibull, P(3), Galton, and Gumbel distributions are a particular case of generalized method of moments.

a. Method of Moments (MOM)

Estimates of the parameters of a probability distribution function are obtained in the MOM by equating the moments of the sample with the moments of the probability distribution function. For a distribution with k parameters, $\alpha 1$, $\alpha 2$, …, α k which are to be estimated, the first k sample moments are set equal to the corresponding population moments that are given in terms of unknown parameters. These k equations are then solved simultaneously for the unknown parameters $\alpha 1$, α 2, ..., α k.

Although there are several ways mentioned here for the approach in this particular research I used the MOM method. However, the other methods which are equally applicable are included here for anyone who might want to cross check results and parameter consideration fitness.

b. Method of Maximum Likelihood (MLM)

Estimation by the ML method involves the choice of parameter estimates that produce a maximum probability of occurrence of the observations. For a distribution with a probability density function (pdf) given by f(x) and parameters α 1 , α 2 , …, α k , the likelihood function is defined as the joint pdf of the observations conditional on given values of the parameters α 1 , α 2 , …, α k in the form:

$$L(\alpha1, \alpha2, \ldots, \alpha k) = \prod_{i=1}^{n} f(xi; \alpha1, \alpha2, \ldots, \alpha k) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.2$$

The values of α 1 ,α2,…,αk that maximize the likelihood function are computed by partial differentiation with respect to α 1 , α 2 , …, α k and setting these partial derivatives equal to zero as in Eq. 2.3. The resulting set of equations are then solved simultaneously to obtain the values of α 1 , α 2 , …, α k ,

$$\frac{\partial L(\alpha1,\alpha2,\ldots,\alpha k)}{\partial \alpha i} = 0;\ i = 1,2,\ldots\ldots,k \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.3$$

In many cases it is easier to maximize the natural logarithm of the likelihood function by using

$$\frac{\partial \ln L(\alpha1,\alpha2,\ldots,\alpha k)}{\partial \alpha i} = 0;\ i = 1,2,\ldots\ldots,k \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.4$$

c. Method of Probability Weighted Moments (PWM)

Parameter estimates are obtained in this method, as in the case of MOM, by equating moments of the distributions with the corresponding sample moments. For a distribution with k parameters, φ1, φ2, …,φk, which are to be estimated, the first k sample moments are set equal to the corresponding population moments. The resulting equations are then solved simultaneously for the unknown parameters φ 1, φ 2, …, φ k .

Quantile Estimation

Quantile estimates (xT) which correspond to different return periods may be computed after the parameters of a distribution are estimated,. As discussed before, the return period is related to the probability of non-exccedence (F) by the relation,

$$F = 1 - \frac{1}{T} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2.5$$

Where F = F (xT) is the probability of having a flood of magnitude x T or smaller. The problem thus reduces to evaluating x T for a given value of F. In practice, two types of distribution functions are encountered. The first type is that which can be expressed in the inverse form x_T=ϕ(F). In this case, x $_T$ is evaluated in a straight forward manner by replacing F by its value from Eq. 2.5. Examples of this type are the extreme value, Wakeby and logistic distributions. The second type of distribution cannot be expressed directly in the inverse from x_T=ϕ(F) . In this case numerical methods are used to evaluate xT corresponding to a given value of F. This is done either by using numerical relationships between F and x_T or by using tables which are prepared for this purpose. Chow (1964) proposed a general form for calculating x_T as in Eq. 2.6,

$$\mathrm{xT} = \mathrm{u}'1 + \mathrm{KT}\sqrt{\mu 2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .2.6$$

In Eq. 2.6, KT is the frequency factor which is a function of the return period and of the parameters of the distribution. Expressions for KT can be derived by using the same technique for the two types of distributions. In fact, from Eq. 2.6, it can be seen that KT is just the variable xT standardized by using the moment's u'1 and μ2 of the distribution. It should be noted, however, that in Eq. 2.7, u'1 and μ2, the distribution moments, are calculated by using their relationship with the estimated parameters and will be equal to the sample moments only when the MOM is used for parameter estimation.

$$\mathrm{xT} = \frac{\mathrm{xT} - \mu'1}{\sqrt{\mu 2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots .2.7$$

2.6 Design Flood for the Ungagged Catchments

For estimating flow in the ungagged the index flood method, introduced by Dalrymple (1960), is the most widely used method of regional flood frequency analysis. It is based on the identification of zones, called homogeneous regions, within which the probability distribution of annual maximum peak flows is invariant except for a scale factor represented by the index flood.

The flood peak discharge with an assigned return period T relative to the selected site is, in fact, expressed as the product of two terms: the scale factor of the examined site (the index flood), and the dimensionless growth factor which has regional validity. In general it is assumed that the index flood coincides with the average µx of annual maximum flood peak flows.

The literature contains numerous studies on the identification of homogeneous groups of basins and the estimation of the growth factor (see for instance Reed et al, 1999; Burn & Goel, 2000; Castellarin et al, 2001), and relatively few on estimating the index flood. Numerous practical applications of the index flood method have however highlighted the importance and, at the same time, the difficulty in obtaining reliable estimates of µx. It is possible to obtain this estimate by calculating the arithmetic mean of the available observations directly in gauged sections, and indirect methods have to be used in ungauged sections.

The most widely used indirect methods are the statistical type methods which use multi regression models, and link µx to an appropriate set of morphological and climatic indices of the basin. This approach does not make any attempt to represent the physical phenomena that determine the transformation of rainfall to runoff. Other models, by contrast, estimate µx using relations derived from schematic representations of the mechanism of occurrence of intense rainfall events and the hydrological response of the basin. This is the case with the studies by Rossi & Villani (1988), in which an estimation method is proposed based on the use of the well-known rational formula, and the studies by Becciu et al. (1993) and Brath et al. (1997a), in which an index flood evaluation method is based on the analytical derivation of the probability distribution of floods (geomorphoclimatic model).

2.7 Relevant Studies in the Country

Regional frequency analysis for Omo- Gibe sub basins of Ethiopia had been studied by Tameru Gedefa (2009). He applied at – site and regional flood frequency analysis to the selected annual maximum stream flood observed at 19 gagged stations with an Average of 27 record lengths of years in upper Omo – Gibe sub basin using index flood estimation procedure.

Flood frequency analysis for lower awash sub basin (tributaries from northern wollo high lands) using swat 2005 model was also studied by Yidnekachew Argaw (2008). Yidnekachew did intensely on procedure of generating stream flow data for Mille and logiya watershed using rainfall –runoff model approach for use in flood frequency analysis. He tried to describe a spatially distributed watershed model of Lower Awash basin (Mille and Logiya watersheds) that has been developed using SWAT 2005 to describe stream flow generation for Mille and Logiya watershed at Mille station (Gauge033021) and Logiya, (Gauge033027). The Soil and Water Assessment Tool 2005(SWAT2005) developed by the Agricultural Research Service of the United States Department of Agriculture and distributed by the US Environmental Protection Agency for watershed management was also used in his approach. SWAT2005 Simulates through time the daily soil water balance. He used 52 and 29 different regions called HRU (hydrological response units) for Mille and Logiya river watershed respectively in this modeling application.

Mengistu Demissie (2008) carried out Regional flood Frequency Analysis for upper Awash sub basin in his Msc thesis. In his research he had carried out geographic information to identify likely homogeneous regions. Accordingly, he divided the upper Awash sub basin in to two regions and tested for homogeneity. He used 10-selected gauged stations consisting of stream flow record varying from 12 to 37 years, out of which 6 stations are found in the upper region and 4 of the stations are found in the lower region. He also identified an Extreme value EV1, GEV and Lognormal LN2, LN3 distributions as the best fit distribution for the stations in the sub-basin.

Abebe Sine (2004) and Admasu Gebeyehu, 1989 have been also studied the Regional Frequency Analysis of Abbay basin and North-western highlands of Ethiopian's basins (such as Abbay basin, all part of Abbay Basin,Upper parts of Omo-Gibe, Awash ,Rift Valley, Wabi-Shebelle and Genele Dawa sub basins together) respectively.

Abebe Sine (2004) was used 78 gauging stations with average record length of 25 years in his study. He used L- moment ratio diagram of each station to identify homogeneous regions. Accordingly, Abbay basin is classified in to five regions. He used GIS (Arc-view 3.3) software for delineation of the identified regions. The identified regions have been checked for their homogeneity using Discordance measure test, CV-based homogeneity test and L-CV based homogeneity test. For identified regions, Abebe Sine (2004) was also used L-Moment Ratio Diagrams (LMRD) for underlying candidate distributions and found that four distributions namely, GL, LP3, LN AND GEV were the best for the Abbay basin. He also adopted commonly used parameter estimation methods such as MOM, ML and PWM for each selected distributions.

Admasu Gebeyehu (1989) was done his study using 63 gauging stations. In his study, he was applied Index-Flood, for the regionalization of the study area. Monthly Rainfall and Geographical proximity were used for delineation Regions. Accordingly, he was divided his study area in to 6 regions. He also reviewed the mathematical forms of four distributions and estimation of their parameters by the method of Probability Weighted Moments (PWM). These are Extreme Value type 1, Generalized Extreme value, Wakbey , and Log –logistic distributions. In his conclusion remarks, Admasu Gebeyehu (1989) was point out the following points.

3 MATERIALS AND METHODS

3.1 Descriptions of the Study Area

3.1.1 Rift Valley Basin Ethiopia

The lakes region of the Ethiopian Rift Valley consist most lakes of the study area which are the northernmost of the great African Rift Valley lakes. The Great Rift Valley splits the Ethiopian highlands into western and eastern escarpments, and the Ethiopian Rift Valley lakes occupy the floor of the rift valley between the two highlands. Some of the Ethiopian Rift Valley lakes do not have an outlet and have alkaline water, while others have fresh water with a bit concentrated fluoride in some cases. Hot water spots are evidenced in few of the lakes from previous works and manifested as fumaroles at margin of lakes.

The lakes region of the Main Ethiopian Rift (MER) valley comprises two large basins; the Ziway-Shala and Abaya-Chamo basins, and a small catchment (Awasa). The MER is characterized by lakes which are assumed as part of a big lake environment though the nature and floor depth for some is unique. Some of these lakes are Ziway, Langano, Abijata, Shala, Awasa, Abaya and Chamo) that lie at an average altitude of 1600 m above sea level (m.a.s.l). These lakes receive surface inflow from rivers and springs that drain western and eastern highlands (elevation above 2500 m.a.s.l. on average or above) bordering the MER. The major ones are

- ✓ Lake Abaya (areal extent 1,162 square kilometers, elevation 1,285 meters, maximum depth 13.1 meters), the largest Ethiopian Rift Valley lake
- ✓ Lake Chamo (areal extent 551 square kilometers, elevation 1,235 meters, maximum depth 14 meters)
- ✓ Lake Zway or Dambal (areal extent 485 square kilometers, elevation 1,636 metres, maximum depth 8.9 meters)
- ✓ Lake Shala (areal extent 329 square kilometers, elevation 1,558 meters, maximum depth 266 meters), the deepest Ethiopian Rift Valley lake
- ✓ Koka Reservoir (areal extent 250 square kilometers, elevation 1,590 meters, maximum depth not listed)

- ✓ Lake Langano (areal extent 230 square kilometers, elevation 1,585 meters, maximum depth 46 meters)
- ✓ Lake Abijatta (areal extent 205 square kilometers, elevation 1,573 meters, maximum depth 14 meters)
- ✓ Lake Awasa (areal extent 129 square kilometers, elevation 1,708 meters, maximum depth 10 meters)

The climatic conditions characterizing the highlands, the escarpment, and the Rift valley differ significantly. Mean annual rainfall in the highland's ranges commonly from about 800 mm to over 2000 mm, while the Rift valley forms a semi-arid to arid climatic area, with rainfall varying commonly from 500 mm to 800 mm (cieted in different sources such as CSA (2007) and Ethiopian Mapping Authority publications, 1988). The mean annual temperature in the highlands is less than 15 °C and evaporation less than 1000 mm; on the Rift floor, mean temperature is greater than 20 °C and evaporation exceeds 2500 mm in some depressed parts (Le Turdu et al., 1999). The area morphology drops in step like land forms due to rift forming events.

Rainfall in the Rift is concentrated during the summer months from June to September, with additional modest rains coming from March to May especially in south. During the long, dry period between October and February, water availability is low for most rift floor and adjacent marginal terrains. The water quality in the rift is affected by natural factors such as the variation in geology and some contribution of evapotranspiration that significantly exceeds rainfall. The water quality particularly in few areas near some of the lakes or in few lakes is affected by deep source hot water of poor quality, and partly affected by evaporative enrichment, which increases its salinity. Currently, surface and groundwater resources are used by the regions small-scale, Agroindustry's, commercial irrigation, and floriculture farms though in few places the groundwater is not commonly used for quality issues especially fluoride content.

The country has about twelve major water sheds of which the rift basin is one. Zeway-Shala is within the rift basin with small sub-catchments such as the Awassa and its surroundings. Slight variability's in some parameters are reported in the eastern and western sides of the rift floor.

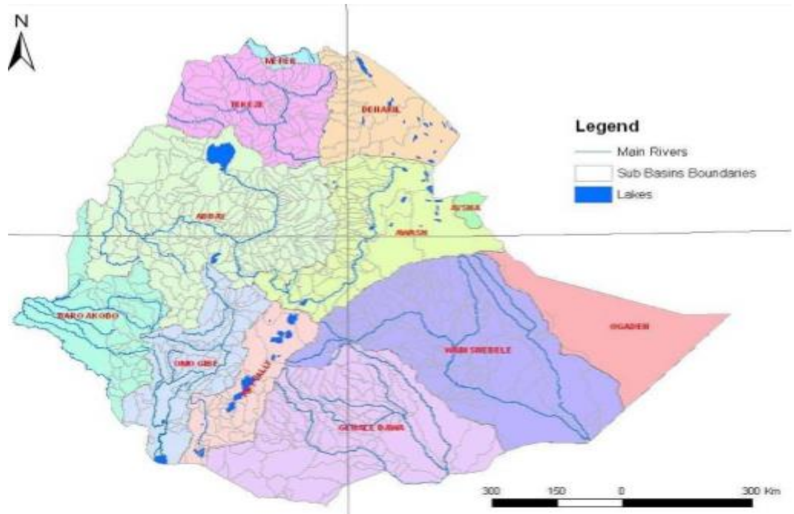


Figure 3-1:- Major River Basin of Ethiopia (Ministry of Water, Irrigation and Energy presentation I (IWW NWP October 31st – November 4, 2011, the Netherlands)

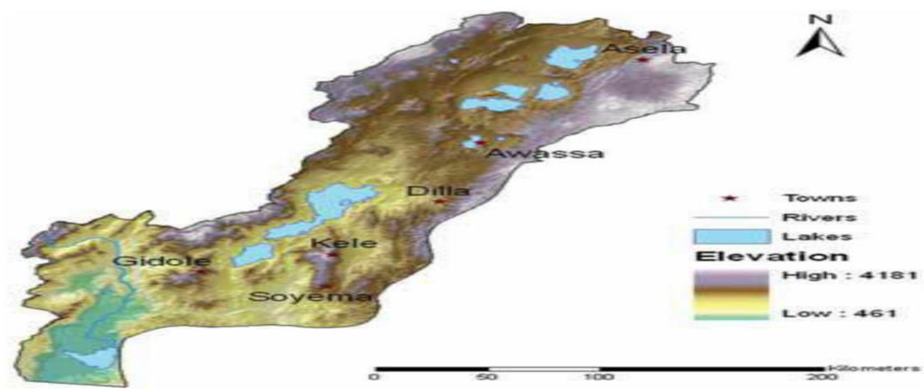


Figure 3-2:- Rift Valley Basin (Lower Awash Integrated Resources Development Master Plan Study Project, Volume 1: (JICA study 2007)

Ziway – Shala Lake Sub Basin

The Ziway–Shala lake basin system (7–8º30'N lat.; 38º07'–39º30'E long.) is part of the NNE-trending MER that forms part MER valley. The mean annual rainfall is not more than 600 mm in the vicinity of Lake Ziway (Kingham, 1975), and rises to a maximum of 1200 mm at the high margins (3000 m contour) of the MER. Mean annual temperature is less than 15ºC in the highlands and more than 20ºC in the lowlands. Potential evapotranspirations estimates made and mentioned in previous works more than 2500 mm on the rift floor in lower elevations to less than 1000 mm in the highlands. The present vegetation pattern in the area also varies following the climatic and morphology variations and depth to water in the rift floor. Dry lands are covered by typical bush and thorny vegetation cover and broad leaf trees in places with sufficient fresh water sources.

The Ziway–Shala lake area is reported to have an area of about 14,640 km^2 and includes four lakes: Ziway, Langano, Abijata, and Shala in eastern side of hills. The three northernmost lakes, Ziway, Langano, and Abijata are connected by a surface network.

Lake Ziway, the northernmost, highest (1636 m) and most extensive of the four lakes, has an area of 442 km^2 and a maximum recorded water depth around 9 m (Makin et al., 1976). Lake Langano lies at an altitude of 1582 m. It has an area of 241 km^2 and a maximum recorded water depth of 48 m (Italconsult, 1970; this study). Lake Abijata was previously described with an area of 176 km^2 and a maximum water depth of 14 m (surface altitude 1578 m) (Vatova, 1941; Gasse and Street, 1978) cited in (C. Le Turdu et al. / Palaeogeography, Palaeoclimatology, Palaeoecology 150 (1999) 135–177)

The surface hydrographic network of the Ziway– Shala lake basin system is particularly well developed to the north of the area, with the Meki and Katar rivers entering Lake Ziway from the western and eastern escarpments. Lake Langano is mainly maintained by five major rivers, Huluka, Lepis, Gedemso, Kersa and Jirma, flowing northwest then north from the southeastern escarpment and characterized by weak, low water flow. Lake Abijata receives flow at its northern end from the Bulbula and Horocallo rivers flowing from Lake Ziway and Lake Langano, respectively. The surface inflows to Lake Shala come from two main sources, the perennial Adabar River which enters from the southeastern rift escarpment, and the main branch of the Gidu River flowing from the western escarpment as explained by (C. Le Turdu et al. / Palaeogeography, Palaeoclimatology, Palaeoecology 150 (1999) 135–177)

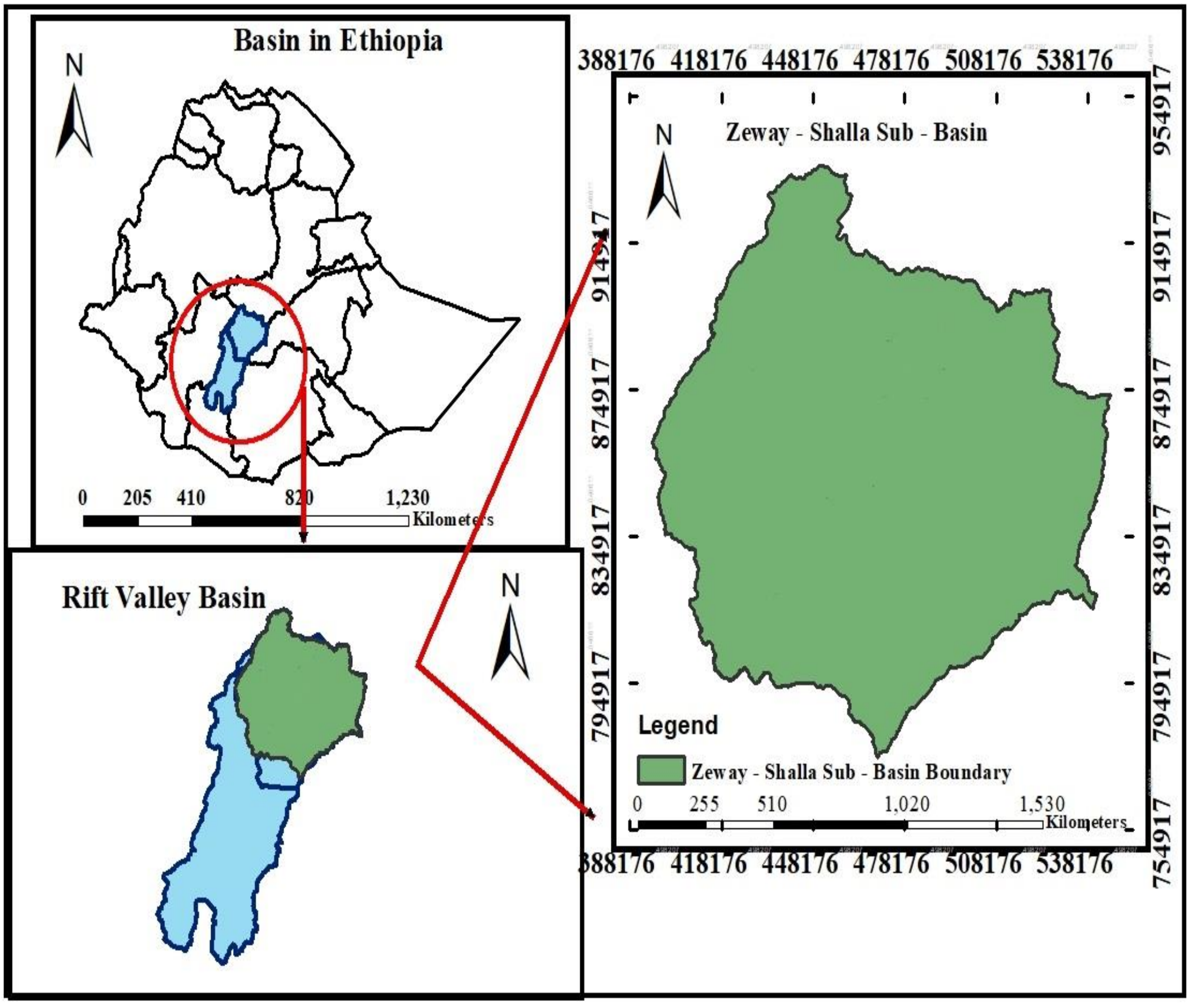


Figure 3-3:- Location map of the study area

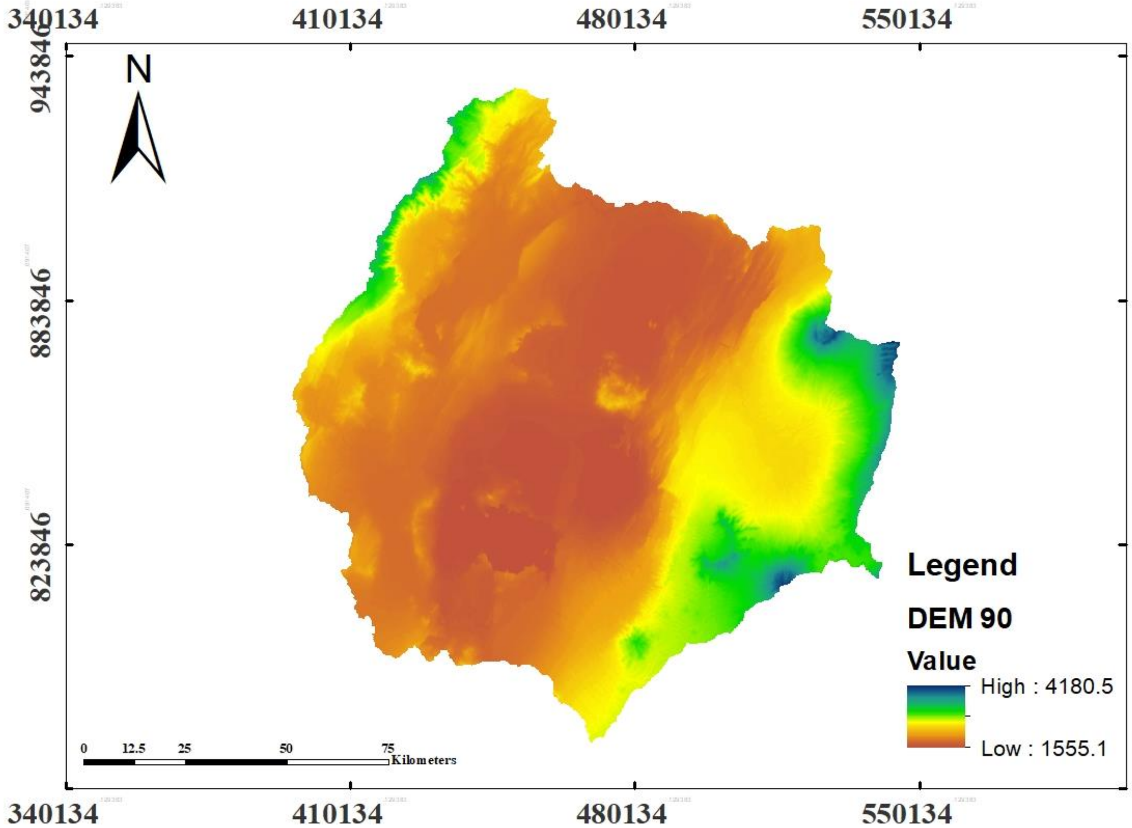


Figure 3-4:-DEM for Ziway – Shala Sub-Basin

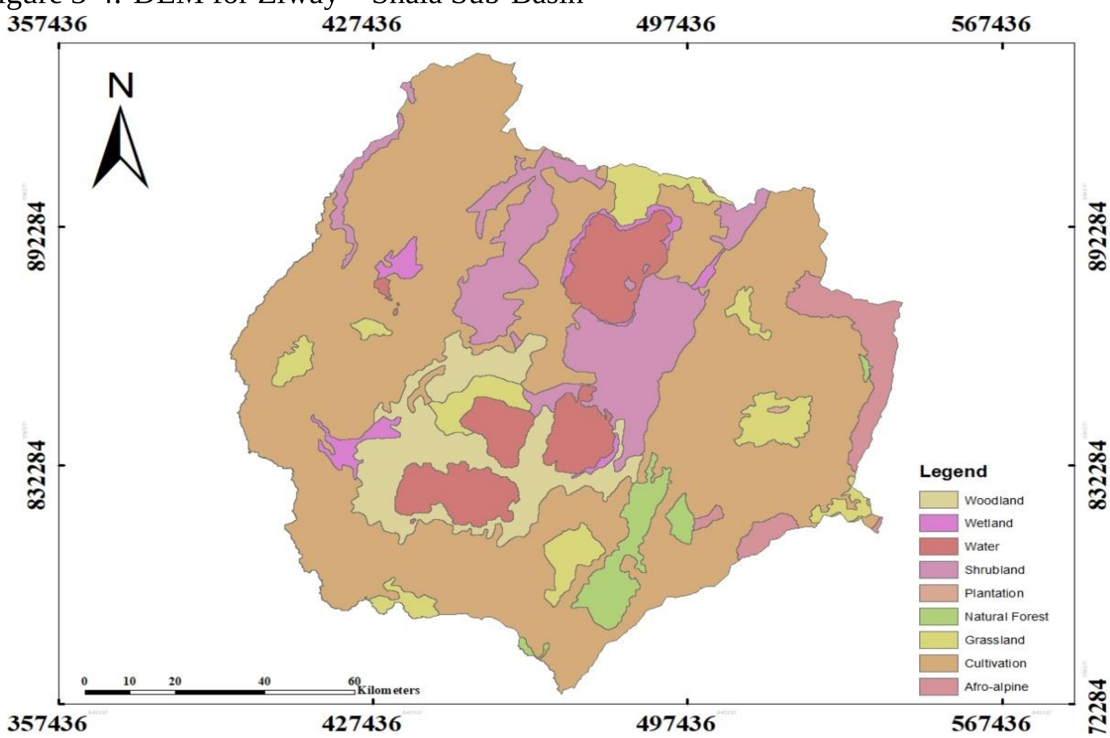


Figure 3-5:- Land use for Ziway – Shala Sub - Basin

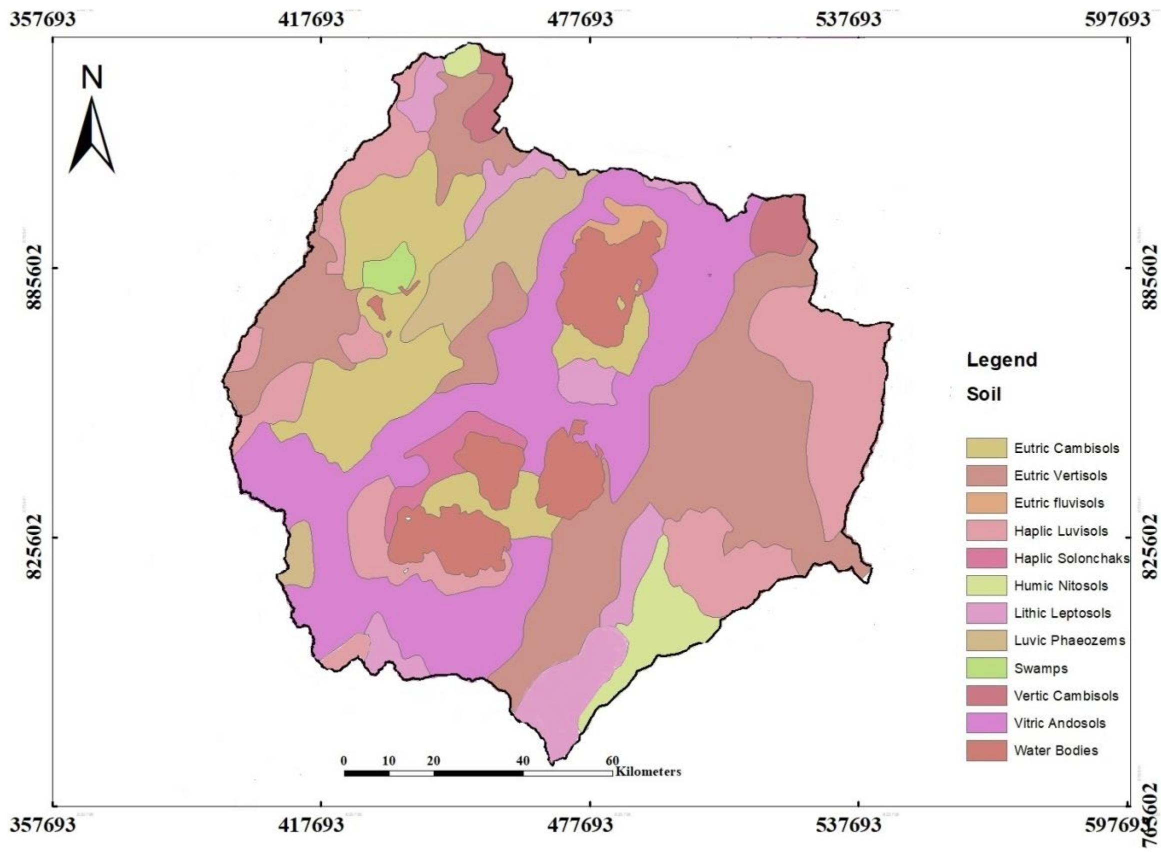


Figure 3-6:- Soil Group for Ziway – Shala Sub-Basin

In the study area there are various development activities such as commercialized flower farming, associations for vegetation farming, and individual irrigation practices which require high abstraction of both surface and groundwater. In most cases users depend mainly on shallow groundwater using shallow wells abstraction or surface water from rivers and lakes. But the shallow groundwater is either feed by rivers or nearby lake water and causes either significant drop in dry time river flow or sometimes lake level drop especially where there is water abstraction from lakes for industrial or commercialized practices.

There are 58 hydrological gaged stations on the rift valley basin and out of those gaged stations 23 stations are found on the ziway shala sub basin. From Twenty-three flow gauging stations ten stations were selected for this study and the main criteria used to choose the sites were based on the length of record periods (minimum of fifteen years) and continuity (no consecutive gaps). From the remaining stations four stations are on the lake. The raw hydrological data are collected from the MoWIE Hydrology department.

Table 3-1:- List of stations used for the study

S.No	Station code	Station Name	coordinates		Area (sq km)	N
			lat	long		
1	81003	Dedeba nr Kuyera	7d17'n	38d40'e	156	36
2	81004	Kerkersitu nr Adamitulu	7d51'n	38d43'e	7488	31
3	81005	Horakelo nr Langano	7d40'n	38d42'e	2006	28
4	81008	Lower Timala nr Sagure	7d41'n	39d09'e	184.45	27
5	81011	Katar nr Fite	7d47'n	39d03'e	1975	26
6	81018	Meki @ Meki river	8d09'n	38d50'e	2433	40
7	81019	Katar nr Abura	8d04'n	39d03'e	3350	40
8	81025	Chufa nr Arata	7d59'n	39d04'e	216	23
9	81026	Ferfero nr Wulbareg	7d44'n	38d07'e	153	18
10	82032	Rinzaf @ Butajera	8d09'n	38d25'e	49	21

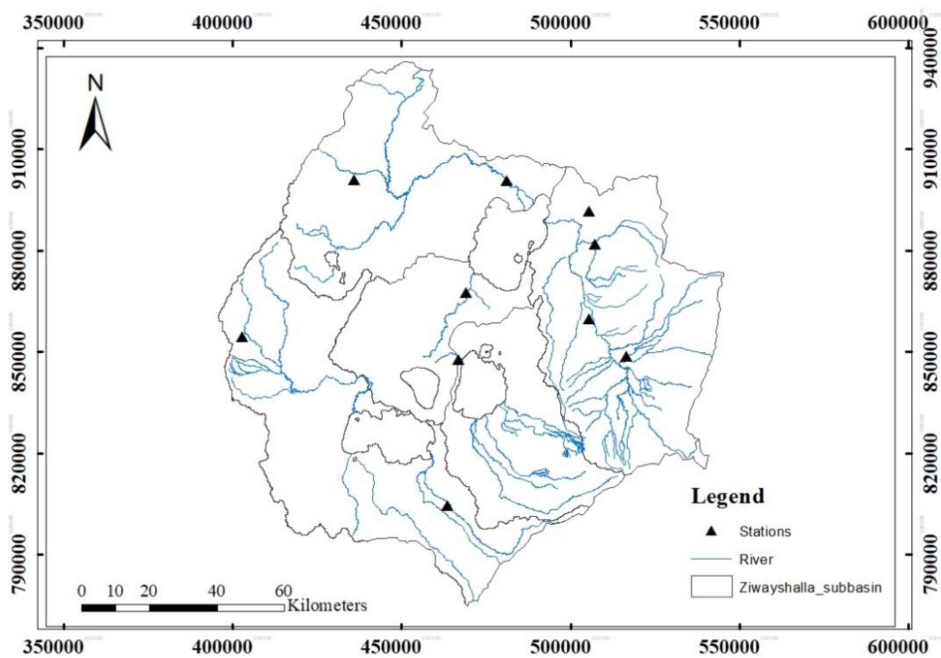


Figure 3-7:- Stations on Ziway – Shala Sub-Basin

3.2 Methodology

To achieve objectives of the thesis efficient and complete methodologies of the research based on clear understandings plays vital role in quality of outputs.

3.2.1 General

As per the literature review season there were different methods of regional flood frequency analysis. From those methods most researchers preferred index flood procedures due to better prediction accuracy over other regional flood frequency estimation methods (Binaya kumar, 2010). Accordingly, the methodology chosen in this study is index flood method based on L-moments for regional data analysis and it includes data preparation, testing of data of the stations for homogeneity, regionalization, selection of frequency distribution, method of parameter estimation and quantile estimation.

The general procedure of the study includes: -

- Collection of important data for the study area such as hydrological data, methodological data, topographical and digitized map of the sub-basin.
- Checking of data for quality (independence, consistency and outliers test)
- Computation of statistical parameters of selected stations within the sub-basin.
- Delineation of homogeneous regions
- Selection of frequency distribution for the determination of the quantiles.
- Selection of parameter estimation method for the selected distribution.
- Derivation of regional flood frequency curve.
- Computing quantiles on the ungagged catchment using the standardized growth curve and the index flood on the regression model.

3.2.2 Methodology Flow Chart

The following figure shows the major activities carried out during the research work.

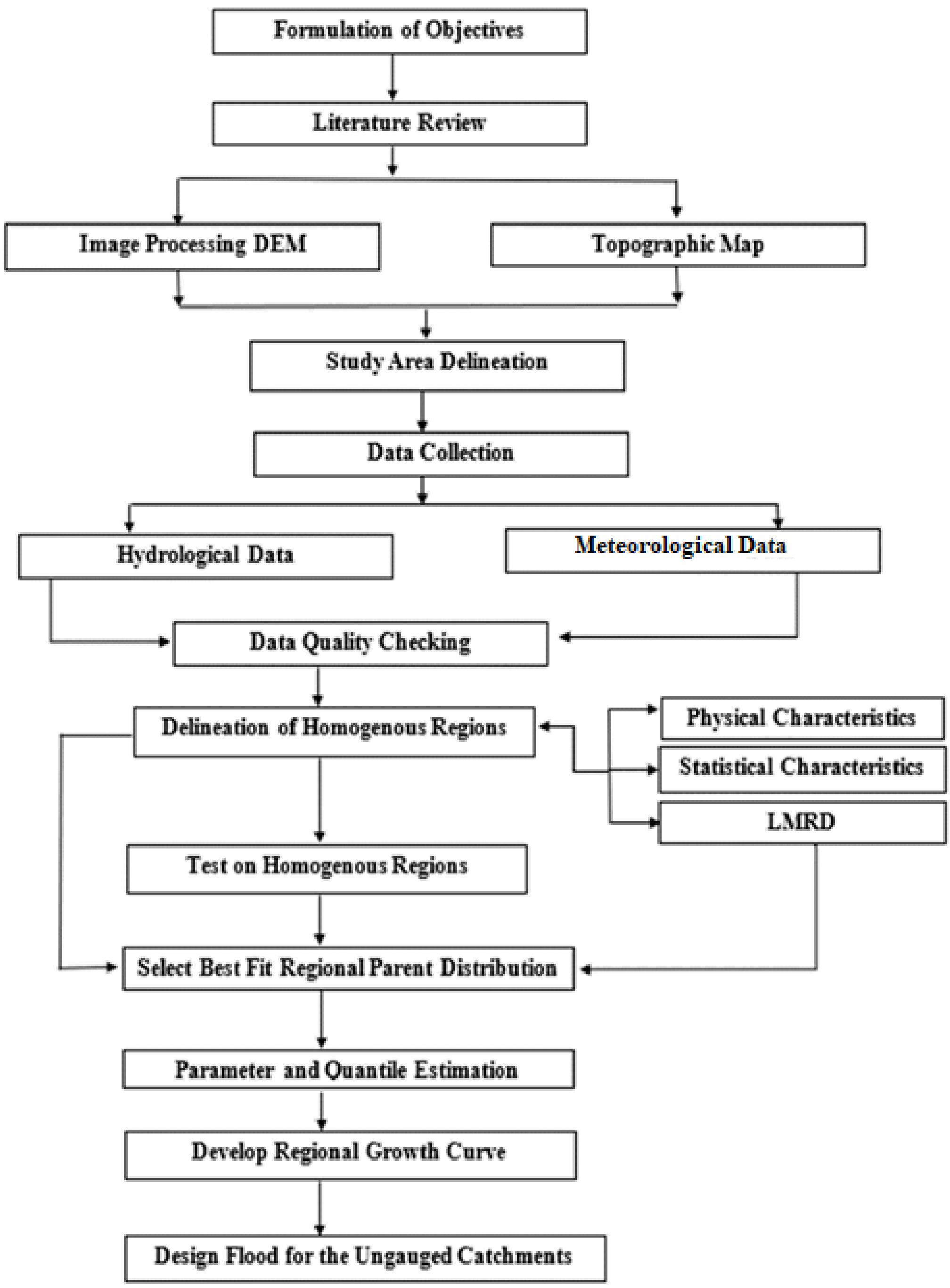


Figure 3-8: Flow Chart for Research Work

3.2.3 Data Collection

Important data were collected from different institutions. Hydrological data and digitized map of the sub-basin were collected from the Ministry of Water Irrigation and Energy, department of Hydrology and GIS.

Meteorological data such as temperature data and rainfall data were obtained from National Meteorological Agency of Ethiopia.

3.2.4 Data Quality Checking

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

So, the data quality of the study area was check using four tests called absence of trend, independence, homogeneity and outliers by considering 95% confidence interval (5% significant level).

3.2.4.1 Test for Absence of Trend

After plotting a time series, one must be sure that there is no correlation between the order in which the data have been collected and the increase (or decrease) in magnitude of those data. It is common practice to test the whole time series for absence of trend. To verify absence of trend, Spearman's rank-correlation method is recommended (E.R. Dahmen and M.J. Hall, 1990). It is simple and distribution-free, i.e. it does not require the assumption of an underlying statistical distribution. Yet another advantage is its nearly uniform power for linear and nonlinear trends (WMO, 1966). The method is based on the Spearman rank-correlation coefficient, Rsp, which is defined as:

$$\mathrm{RSP} = 1 - \frac{6 * \sum_{i=1}^{n}(\mathrm{Di} * \mathrm{Di})}{n(n * n - 1)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.1$$

Where n is the total number of data, Di is difference, and i: is the chronological order number.

The difference between rankings is computed with:

$$\mathrm{Di} = \mathrm{Kxi} - \mathrm{Kyi} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.2$$

Where Kx, is the rank of the variable, x, which is the chronological order number of the observations. The series of observations, yi, is transformed to its rank equivalent, Kyi, by assigning the chronological order number of an observation in the original series to the corresponding order number in the ranked series, y. If there are ties, i.e. two or more ranked observations, y, with the same value, the convention is to take Kx as the average rank. One can test the null hypothesis, Ho: Rsp, = O (there is no trend), against the alternate hypothesis, Ho: Rsp, < > O (there is a trend), with the test statistic:

$$\mathrm{tt} = \mathrm{Rsp} * \left[\frac{\mathrm{n}-2}{1-\mathrm{Rsp}*\mathrm{Rsp}}\right]^{0.5} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.3$$

Where t, has Student's t-distribution with v = n-2 degrees of freedom. Student's t-distribution is symmetrical around t = 0.

Ho: 0 (There is no trend in the data series)

H1: # 0 (There is a trend in the data series)

The null hypothesis is rejected if /t_t/ (absolute value of calculated value) is greater than the critical value or the null hypothesis is accepted if t_t is not contained in the critical region. In other words, the data series has no trend if:

$$\mathrm{t}\{\mathrm{v}, 2.5\%\} < \mathrm{tt} < \mathrm{t}\{\mathrm{v}, 97.5\%\} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.4$$

3.2.4.2 Test for Independence and Stationarity

Given a sample of size N, the Wald-Wolfowitz (1943) (W-W) test is used to test for the independence of a dataset and to test for the existence of trends in it. For a data set x1, x2, Xn the statistic R is calculated from Eq 3.1

$$\mathrm{R} = \sum_{\mathrm{i}=1}^{\mathrm{N}-1} (\mathrm{Xi} * \mathrm{Xi} + 1) + (\mathrm{X1} * \mathrm{XN}) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.5$$

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

$$\bar{\mathrm{R}} = \frac{\mathrm{S1}^2 - \mathrm{S2}}{\mathrm{N}-1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.6$$

$$\mathrm{Var}\ (\mathrm{R}) = \frac{S2^2 - S4}{N-1} - (\mathrm{Rmean})^2 + \frac{S1^4 - 4S1^2S2 + 4S1S3 + S2^2 - 2S4}{(N-1)(N-2)} \ldots\ldots\ldots\ldots\ldots .3.7$$

where $S_r = Nm_r'$ and m_r' is the rth moment of the sample about the origin.

The statistic $u = (R - \bar{R})/(\mathrm{var}(R))^{\frac{1}{2}}$ is approximately normally distributed with mean zero and variance unity and is used to test the hypothesis of independence at significance level α, by comparing the statistic u with the standard normal variate $u_{\alpha/2}$ corresponding to a probability of exccedence α /2. (Roa and Hammed, 2000).

3.2.4.3 Test for Homogeneity and Stationarity

In order to test the homogeneity of the collected data, statistical analysis was applied and data processing and interpretations have prepared using Statistical Package for Social Sciences (SPSS) of the latest version and descriptive methods in the form of table. P-value below 0.05 (95% confidence interval) was considered as statistically significant for hypothesis testing season.

✓ **Hypothesis testing**

In this method the following basic steps should be followed so as to conduct statistical analysis: -

Step 1: - Formulation of hypothesis

Ho: There is no difference between the ranks of each value

H1: There is a difference between the ranks of each value

Step 2: - State significance level for our test

The significance level is α = 0.05 (95% confidence interval)

Step3: - Calculate the test statistics for our sample (using Mann-Whitney (M-W)) test

Step4: -Determination of rejection region

For two tailed test Z critical value is direct read from normal distribution table then,

- Rejection criteria is /Zo/ > Zα/2 and
- From the SPSS or Microsoft excel results we have sig value (two tailed) which is equals to P- value for statistical tests. So, the rejection criteria P value is <α .

Step 5: - State a conclusion for our hypothesis test

3.2.4.4 Test for Outliers

An outlier in this case is assumed to be 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 which may slightly deviate from assumed homogeneous values. The presence of outliers in the data causes difficulties when fitting a distribution to the data and had to be either amended or removed. Low and high outliers are both possible and have different effects on the analysis. The Grubbs and Beck (1972) test (G-B) was used to detect outliers.

It is well known that outliers (extreme points) often distort the results of an analysis. Because of this, every analysis had to begin with either a graphical or statistical check about the possibility of outliers. This procedure computes Grubbs' test (1950) for detecting outliers.

Grubb's test for outlier Detection

- ✓ **Hypothesis testing**

Similarly, hypothesis testing shall be executed as per the above procedures. Grubb's test is defined for the hypothesis: -

Ho: There are no outliers in the data set

H1: There is exactly one outlier in the data set

The Grubb's test statistics ($G_{calculated}$) is defined as;

$$\mathrm{max}\,\mathrm{i} = 1, \ldots\ldots \mathrm{n}\frac{|\mathrm{Xi} - \mathrm{Xmean}|}{\mathrm{S}}$$

G critical value is defined by

$$\frac{\mathrm{n}-1}{\sqrt{\mathrm{n}}}\sqrt{\frac{\mathrm{t}^2}{\mathrm{n}-2+\mathrm{t}^2}}$$

Where, t is short for t_{n-2}, p and $\mathrm{p} = 1 - \frac{\alpha}{2n}$ Grubbs' tabulated this result, but we have solved this so that a p value can be calculated. The above is for two-sided tests. For one-sided tests, substitute $\alpha/\,(n)$ for $\alpha/\,(2n)$.

The rejection region is defined as

P value < α and

$G_{calculated}$ > critical value

3.2.5 Method of Analysis

The regional flood frequency analysis was often accomplished in two steps according to M.M. Portela and A.T.Dias (2005) The first one relates the identification of homogenous regions, deals with flood peak discharges, and the establishment of a regional frequency distribution curve for each homogenous region. According to same authors a region is considered to be homogenous when the watersheds belonging to it exhibit a similar behavior in terms of exceptional discharges, that is to say, when the non-dimensional local frequency distribution curves have similar shapes, obviously, within a sampling error. The second step in the views of same authors is the regional analysis involves the development of procedures that are based on the results from the previous step that enables to evaluate peak discharges at river sections without the required flow data (M. M. Portela & A. T. Dias, 2005)

3.2.5.1 Regionalization

Identification and Delineation of Homogeneous Regions

As mentioned in the literature review the identification of homogenous regions is usually the most difficult stage and requires greatest amount of subjective judgment. The aim is to form group of sites that approximately satisfy homogeneity condition in the sites frequency distributions are identically is satisfied.

In this study to identify hydrologically homogeneous regions both geographical consideration which is the physical characteristics and the flood data or statistical characteristics were used.

Preliminary identification of groups of stations into a certain category is achieved through looking at catchment characteristics. Some of the parameters that affect flow condition like soil type, elevation, land use land cover and slope of the sub basin which have similar flood producing characteristics were looked. Then those stations having nearly same kind of physiographic and climatic characteristics are grouped. Then using the L moment ratio diagram checked identified their parent distribution and those stations having similar parent distribution are grouped in the same region. Then make some modification by checking the physical characteristics with LMRD.

Finally, by taking consideration of the boundaries of each sub catchment the delineation of homogeneous regions was carried out and for each region the homogeneous test was applied to confirm that the delineated regions are statistically homogeneous.

Flood statistics of the Zeway – Shala sub basin

In order to compute flood statistics of Ziway – Shala sub basin stations both conventional moment and L moment method were used. However, L moment is a powerful and efficient method to compute any statistical parameters (Roa & hamed 2000).

Generally, the statistical parameters computed include:

- ✓ Mean
- ✓ Standard deviation
- ✓ Coefficient of variation (Cv, LCv)
- ✓ Coefficient of skewness (Cs,LCs)
- ✓ Coefficient of kurtosis (Ck, LCk)

Conventional moment

Moments about the origin or about the mean are used to characterize probability distributions. Moments about the origin are the expected values of powers of a random variable. For a distribution with a probability density function f(x), the r^{th} moment about the origin is given by Eq. 3.8

$$\mu' r = \int_{-\infty}^{\infty} X^r f(x) dx, \qquad \mu' 1 = \mu = \text{mean} \ldots \ldots \ldots \ldots 3.8$$

The central moment's μ_r are computed by Eq. 3.9

$$\mu r = \int_{-\infty}^{\infty} (x - \mu' 1)^r f(x) dx, \qquad \mu 1 = 0 \ldots \ldots \ldots \ldots 3.9$$

Sample moments and mr, on the other hand, are calculated by using Eq. 3.10 and 3.11

$$m' r = \frac{1}{n} \sum_{i=1}^{n} Xi^r \ , m' 1 = \text{Xmean} = \text{sample mean} \ldots \ldots \ldots \ldots 3.10$$

$$mr = \frac{1}{n}\sum_{i=1}^{n}(Xi - Xmean)^{r}, m1 = 0 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.11$$

These sample moments are often biased and may be corrected (Cunnane, 1989).

$$\hat{m}\ 2 = \frac{1}{n-1}\sum_{i=1}^{n}(Xi - Xmean\ )^{2} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.12$$

$$\hat{m}\ 3 = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(Xi - Xmean\ )^{3} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.13$$

$$\hat{m}\ 4 = \frac{n^{2}}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(Xi - Xmean\ )^{4} \ldots\ldots\ldots\ldots\ldots\ldots 3.14$$

The conventional moment ratios are defined as below.

Coefficient of variation $\quad Cv = z = \frac{(\mu 2)^{\frac{1}{2}}}{\mu' 1} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.15$

Coefficient of skewness $\quad Cs = \gamma 1 = \frac{\mu 3}{(\mu 2)^{\frac{3}{2}}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.16$

Coefficient of kurtosis $\quad Ck = \gamma 2 = \frac{\mu 4}{(\mu 2)^{2}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.17$

L moment

L-moments are defined as linear combinations of Probability Weighted Moments (PWMs) that have simple interpretations as measures of the location, dispersion and shape of the sample data Hosking (1990).

The unbiased PWM estimators are:

$$\hat{m}\ 10k = \frac{1}{N}\sum_{i=1}^{N}\left[\binom{N-i}{k}\Big/\binom{N-1}{k}\right]Xi \quad \ldots\ldots\ldots k=0, 1, 2\ldots\ldots\ldots N-1$$

$$\hat{m}\ 1j0 = \frac{1}{N}\sum_{i=1}^{N}\left[\binom{i-1}{j}\Big/\binom{N-1}{j}\right]Xi \quad \ldots\ldots\ldots j=0, 1, 2 \ldots\ldots\ldots N-1$$

Where, n is the sample size and Xj denotes the jth element of a sample of size n sorted into ascending order.

The first few moments are:

$$\mathrm{L1} = \mathrm{M100}$$

$$\mathrm{L2} = \mathrm{M100} - 2 * \mathrm{M101}$$

$$\mathrm{L3} = \mathrm{M100} - 6 * \mathrm{M101} + 6 * \mathrm{M102}$$

$$\mathrm{L4} = \mathrm{M100} - 12 * \mathrm{M101} + 30 * \mathrm{M102} - 20 * \mathrm{M103}$$

Like the conventional moments, L-moment can be used to specify and summarize probability distribution. In particular, l1, the first L-moment, is the mean of the statistical distribution and identical to the first conventional moment, and l2 is a linear measure of spread or dispersion analogous to standard deviation. Hosking (1986) defines L-moment ratio, which are analogous to conventional moment ratio.

$$\tau = \frac{\mathrm{L2}}{\mathrm{L1}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.18$$

$$\tau r = \frac{\mathrm{Lr}}{\mathrm{L2}}, r \geq 3 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.19$$

Where: L1= measure of location

τ = measure of scale and dispersion (Lcv)

τ_3 = measure of skewnes (Lcs)

τ_4= measure of kurtosis (Lck)

Instead of τ and τ_r, t and t_r are used for sample L moment.

Regional Homogeneity Tests

Once a set of physically possible regions has been defined, it is necessary to assess whether the region is meaningful. The most commonly used statistical homogeneity tests are discordance measure test and CC – based homogeneity test.

Discordance measure test

The discordancy measure D estimates how far a given site is from the center of the group. If Ui=[t(i), t3(i), t4(i)] T is the vector containing the t, t3 and t4 values for site (i), then the group average for NS sites is given by Eq. 3.20

$$\mathrm{u(mean)} = \frac{1}{\mathrm{NS}}\sum_{\mathrm{i=1}}^{\mathrm{NS}} \mathrm{ui} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..3.20$$

The sample covariance matrix is given by Eq. 3.21

$$\mathrm{S} = (\mathrm{NS}-1)^{-1}\sum_{\mathrm{i=1}}^{\mathrm{NS}}(\mathrm{ui}-\mathrm{u(mean)} * \left(\mathrm{ui}-\mathrm{u(mean)}\right)^{\mathrm{T}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots\ldots..3.21$$

The discordancy measure is defined by Eq. 3.22

$$\mathrm{Di} = \frac{1}{3}\left(\mathrm{ui}-\mathrm{u(mean)}\right)^{\mathrm{T}}\mathrm{S}^{-1}\left(\mathrm{ui}-\mathrm{u(mean)}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.3.22$$

A site (i) is declared to be unusual if Di is large. A suitable criterion to classify a station as discordant is that Di should be greater than or equal to 3.

CC –Based homogeneity test

In this test the site-to-site coefficient of variation of the coefficient of variation (CC) of both conventional and L-moments of the proposed region are used.

CV- based homogeneity test

For each region, using the flood statistics cv computed the the corresponding CC using the following relation:

$$\mathrm{Cvi\ (mean)} = \sum_{i=1}^{N} \frac{\mathrm{Cvi}}{\mathrm{Ni}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.23$$

$$\sigma \mathrm{cv} = \sqrt{\frac{\sum_{i=1}^{N}(\mathrm{Cvi} - \mathrm{Cvi\ (mean)})}{N-1}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.24$$

$$\mathrm{CC} = \frac{\sigma \mathrm{cv}}{\mathrm{cv(mean)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.25$$

Where, N= number of sites in a region

CV (mean) = Mean coefficient of variation

σ_{cv} = Standard deviation of at-site CV values.

The regions assumed to be homogenous if CC<0.3

LCV- based homogeneity test

For each region, using the flood statistics Lcv computed the the corresponding CC using the following relation:

$$\mathrm{Lcv} = \frac{\mathrm{L2}}{\mathrm{L1}} \text{ and } \mathrm{Lcv\ (mean)} = \sum_{i=1}^{N} \frac{\mathrm{Lcv}}{N} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.26$$

$$\sigma \mathrm{Lcv} = \sqrt{\frac{\sum_{i=1}^{N}\left(\mathrm{Lcv} - \mathrm{Lcv(mean)}\right)^2}{N-1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.27$$

$$\mathrm{CC} = \frac{\sigma\ \mathrm{Lcv}}{\mathrm{Lcv(mean)}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.28$$

The stations and regions assumed to be homogeneous if CC < 0.3

3.2.5.2 Selection of Flood Frequency Distribution and Parameter Estimation

To select the best fit distribution and the best method of parameter estimation different softwares have been used to fit the curve with distribution.

In this study, for the selection of best distribution with parameters for each at-site data in the region, Easy-fit software are used. The candidate probability distributions are Gamma, Gamma(3p), Lognormal (LN), Lognormal III (LN3), Generalized Extreme value (GEV) and Log- Pearson Type III (LP3).

To assess the fitness of the selected distribution, statistical tests like Chi-Square test and Kolmogrov-Smirnov tests are used. Depend on the results of goodness -of-fit tests by Kolmogorov-Smirnov Test and Chi-square, fitness of the distributions were identified for each data in the regions. Both tests are discussed below.

Kolmogorov-Smirnov Test

This test is used to decide if a sample comes from a hypothesized continuous distribution. It is based on the empirical cumulative distribution function (ECDF). Assume that we have a random sample $x_1,...,x_n$ from some distribution with CDF F(x). The empirical CDF is denoted by

$$\mathrm{Fn(X)} = \frac{1}{\mathrm{n}}.\,[\text{Number of observations} \leq \mathrm{x}] \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.\ldots 3.29$$

Definition

The Kolmogorov-Smirnov statistic (D) is based on the largest vertical difference between the theoretical and the empirical cumulative distribution function:

$$\mathrm{D} = \max_{1\leq \mathrm{i} \leq \mathrm{n}}\left(\mathrm{F(Xi)} - \frac{\mathrm{i}-1}{\mathrm{n}}, \frac{\mathrm{i}}{\mathrm{n}} - \mathrm{F(Xi)}\right)\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.3.30$$

Hypothesis Testing

The null and the alternative hypotheses are:

- H_0: the data follow the specified distribution;
- H_A: the data do not follow the specified distribution.

The hypothesis regarding the distributional form is rejected at the chosen significance level (α) if the test statistic, D, is greater than the critical value obtained from a table. The fixed values of α (0.01, 0.05 etc.) are generally used to evaluate the null hypothesis (H_0) at various significance levels. A value of 0.05 is used in this study which is typically used for most applications, however, in some critical industries; a lower α value may be applied.

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

P-Value

The P-value, in contrast to fixed α values, is calculated based on the test statistic, and denotes the threshold value of the significance level in the sense that the null hypothesis (H_0) will be accepted for all values of α less than the P-value. Easy-Fit displays the P-values based on the Kolmogorov-Smirnov test statistics (D) calculated for each fitted distribution.

Chi-Squared Test

The Chi-Squared test is used to determine if a sample comes from a population with a specific distribution. This test is applied to binned data, so the value of the test statistic depends on how the data is binned. This test is available for continuous sample data only.

Although there is no optimal choice for the number of bins (k), there are several formulas which can be used to calculate this number based on the sample size (N). For example, Easy-Fit employs the following empirical formula:

$$k = 1 + \log_2 N \quad \ldots\ldots\ldots 3.31$$

Definition

The Chi-Squared statistic is defined as

$$\chi^2 = \sum_{i=1}^{k} \frac{(Oi - Ei)^2}{Ei} \quad \ldots\ldots\ldots 3.32$$

Where O_i is the observed frequency for bin i, and E_i is the expected frequency for bin i calculated by

$$Ei = F(X2) - F(X1) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 3.33$$

Where F is the CDF of the probability distribution being tested, and x_1, x_2 are the limits for bin i.

Hypothesis Testing

The null and the alternative hypotheses are:

- H_0: the data follow the specified distribution;
- H_A: the data do not follow the specified distribution.

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

P-Value

The P-value, in contrast to fixed α values, is calculated based on the test statistic, and denotes the threshold value of the significance level in the sense that the null hypothesis (H_0) will be accepted for all values of α less than the P-value.

Finally, after identifying the fitness for those candidate probability distributions, selection of one of those distributions was made based on the statistical values of each distribution and considering who Widely accepted, Simple and convenient to apply. Accordingly, for selected types of distribution parameters are estimated by using analytically.

In order to estimate the statistical parameters by Method of Moment for Generalized Extreme value (GEV) distribution from the given data the following equations were used.

For the value of k > 0 (–2 < CS < 1.14), R2 = 1 the value of k determined using equation 3.34

$$K = 0.277648 - 0.322016cs + 0.060278cs^2 + 0.016759cs^3 - 0.005873cs^4 - 0.00244cs^5 - 0.000050cs^6 \quad \ldots 3.34$$

$$\alpha = \left[\frac{(cv * m'1)^2 * (k)^2}{(\Gamma(1+2k) - (\Gamma(1+k)^2}\right]^{0.5} \quad \ldots 3.35$$

$$u = m'1 - \frac{\alpha}{k}[1 - \Gamma(1+k)] \quad \ldots 3.36$$

3.2.5.3 Quantile Estimation and Derivation of Flood Frequency Curve

Based on the estimated parameter for each region, the quantile will be calculated according to the formula of the selected distributions. To calculate the quantile for each selected distribution the return period has to be determined first.

For the Generalized Extreme Value of distribution function of x given by in Eq.3.37

$$XT = U + \frac{\alpha}{K}\left[1 - \{-\log F\}^k\right] \quad \ldots 3.37$$

By substituting F = 1 – 1/T where T is the return period, the T-year quantile estimate is obtained as in Eq. 3.38

$$XT = U + \frac{\alpha}{K}\left[1 - \left\{-\log\left(1 - \frac{1}{T}\right)\right\}^k\right] \quad \ldots 3.38$$

For probability distribution functions such as the Normal, Lognormal and Pearson type-III distributions are not readily invertible so an alternative method of calculating the magnitudes of extreme events using frequency factors needed.

The frequency factor KT for the normal distribution is equal to standard normal deviate z. z corresponding to a given return period T, with p=1/ T may be approximated using an intermediate variable, w.

$$w = \left[\ln\left(\frac{1}{P^2}\right)\right]^{1/2} (0 < P \leq 0.5) \quad \ldots 3.39$$

Then calculate Z using the approximation

$$Z = W - \frac{2.515517 + 0.802853W + 0.010328W^2}{1 + 1.432788W + 0.189269W^2 + 0.001308W^3} \ldots\ldots\ldots\ldots\ldots\ldots 3.40$$

The same procedure is applied for the lognormal distribution except that the logarithm of data is used to calculate mean and standard deviation.

The frequency factor KT for the Log pearson Type III distribution depends on the return period T and the coefficient of skewness Cs. When Cs = 0, the frequency factor is equal to the standard normal variable z. When Cs ≠ 0, KT is approximated by Kite (1977) as

$$KT = Z + (Z^2 - 1)k + \frac{1}{3}(Z^3 - 6Z)k^2 - (Z^2 - 1)k^3 + 2k^4 + \frac{1}{3}k^5 \ldots\ldots\ldots\ldots 3.41$$

Where $k = {}^{Cs}/_{6}$

Once the frequency factor KT is calculated then the magnitude of an event XT is represented as the mean μ plus the deviation ΔxT of the variate from the mean

$$XT = \mu + \Delta XT \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.42$$

The deviation is taken as equal to the product of standard deviation σ and a frequency factor KT

$$\Delta XT = KT\sigma \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.43$$

$$XT = \mu + KT\sigma \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.44$$

The deviation and the frequency factor are functions of the return period and the type of probability distribution to be used in the analysis. If in the event, the variable analyzed is y = log x, then the same method is used, with the statistics for the logarithms of the data as

$$YT = \mu y + KT\sigma y \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3.45$$

And the required value of XT is found by taking the antilog of YT. For a given distribution, a K – T relationship can be determined in terms of a table.

Accordingly, the quantiles versus return period graph of each region will be drawn as the growth curve for each region. Depending on the growth curve of each station in the sub-basin and the representative growth curve of the sub-basin, the importance of regionalization of the sub basin can be determined. If the growth curve of the sub-basin can represent the station growth curve with less diversion from curve of each station, the sub basin can be taken as a single region otherwise the sub-basin has to be divided in to the proper homogeneous regions. (Tamiru G, 2009)

3.2.5.4 Design Flood for the Ungagged Catchment

Due to the luck of sufficient gauged station in the sub basin and also the gauged stations have low recording quality it needs to for transfer data from the gauged site to the un-gauged site.

Based on the stream gauging stations belonging to a same homogenous region, establishment of a relationship between the Q2.33 indexes and one or more features (as physical or climatic features) of the station watersheds in the analysis carried out the following relationship:

$$Q2.33 = \alpha A^{\beta} \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots . \ldots \ldots \ldots \ldots \ldots 3.46$$

Where; Q2.33 is the index-flood and A is the watershed area. The parameters α and β were estimated based on linear regression analysis applied to the logarithmic transformed of Q2.33 and of A.

In this particular study, drainage area only was used for regression model as predictor variable to estimate the mean annual peak flood for un-gauged catchments in upper Ziway-Shala Sub-basin.

4 RESULTS AND DISCUSSION

4.1 Data Quality Check for the Sub-Basin Hydrological Data

4.1.1 Result on absence of trend

As stated in section 3.2.4.1, in this study, all the data from the selected stations are checked for absence of trend using the Spearman's Rank-Correlation method. From the t-distribution table the critical values of 't', at the 5-per-cent level of significance, for N-2 degrees of freedom is compared with the calculated values of t_t and finally, the results are summarized in table 4.2.

From those stations, the absence of trend for Chufa near Arata is checked as follows:-

Table 4-1:- Result of trend analysis for Chufa near Arata

Chufa nr Arata

i=x	Data	Y=Ranked runoff	Kxi	Kyi	Di	Di *Di
1	6.616	5.888	1	18	-17	289
2	11.530	6.616	2	1	1	1
3	7.233	7.233	3	3	0	0
4	9.687	7.550	4	5	-1	1
5	7.550	8.040	5	13	-8	64
6	9.452	9.452	6	6	0	0
7	10.475	9.687	7	4	3	9
8	12.958	10.321	8.5	20	-11.5	132.25
9	13.255	10.321	8.5	22	-13.5	182.25
10	13.774	10.475	10	7	3	9
11	13.484	10.994	11.5	19	-7.5	56.25
12	13.198	10.994	11.5	21	-9.5	90.25
13	8.040	11.530	13	2	11	121
14	13.175	11.661	14	16	-2	4
15	12.099	12.099	15	15	0	0
16	11.661	12.529	16	17	-1	1
17	12.529	12.715	17	23	-6	36
18	5.888	12.958	18	8	10	100
19	10.994	13.175	19	14	5	25
20	10.321	13.198	20	12	8	64
21	10.994	13.255	21	9	12	144
22	10.321	13.484	22	11	11	121
23	12.715	13.774	23	10	13	169

n	23	∑ Di *Di	1619	Since tt < tc, we Fail to reject Ho so there is no trend
		Rsp	0.2001	
		tt	0.9359	
		tc	2.08	

From the above table the following input Parameters are obtained:-

- ✓ n = 23 (total number of data sets)
- ✓ ∑ Di *Di = 1619

By using the above input data's the value of R_{sp} and t_t are calculated as follows;

$$\mathrm{RSP} = 1 - \frac{6 * \sum_{i=1}^{n}(\mathrm{Di} * \mathrm{Di})}{n(n * n - 1)} = 1 - \frac{6 * 1619}{23 * (23 * 23 - 1)} = 0.2001$$

$$\mathrm{tt} = \mathrm{Rsp} * \left[\frac{n - 2}{1 - \mathrm{Rsp} * \mathrm{Rsp}}\right]^{0.5} = 0.2001 * \left[\frac{23 - 2}{1 - 0.2001 * 0.2001}\right]^{0.5} = 0.9359$$

From t-distribution table the value of t-critical for α/2 significant level and 21 degree of freedom is 2.08. Since, the value of t_t (0.9359) is less the critical value (2.08) we fail to reject our null hypothesis. So, we can conclude that in our data set there is no trend.

Table 4-2:- Summary of trend analysis for all stations

I.No	Station Name	t (N-2,2.5%)	t (N-2, 97.5%)	R_{sp}	t_t	Check for Critical region	Remark!
1	Dedeba nr kuyera	-2.034	2.034	0.095	0.556	-2.034 < 0.556 < 2.034	No Trend
2	Kerkersitu nr Adamitulu	-2.045	2.045	0.248	1.381	-2.045 < 1.381 < 2.045	No Trend
3	Horakelo nr Langano	-2.056	2.056	0.048	0.246	-2.056 < 0.246 < 2.056	No Trend
4	Lower Timala nr Sagure	-2.060	2.060	0.292	1.528	-2.060 < 1.528 < 2.060	No Trend
5	Katar nr Fite	-2.064	2.064	0.194	0.970	-2.064 < 0.970 < 2.064	No Trend
6	Meki @ Meki river	-2.021	2.021	-0.283	-1.863	-2.021 < -1.863 < 2.021	No Trend
7	Katar nr Abura	-2.023	2.023	0.147	0.929	-2.023 < 0.929 < 2.023	No Trend
8	Chufa nr Arata	-2.080	2.080	0.200	0.936	-2.080 < 0.936 < 2.080	No Trend
9	Ferfero nr Wulbareg	-2.120	2.120	0.451	2.024	-2.120 < 2.024 < 2.120	No Trend
10	Rinzaf @ Butajera	-2.086	2.086	-0.191	-0.872	-2.086 < -0.872 < 2.086	No Trend

4.1.2 Result on Independence Test

As stated in section 3.2.4.2, in this study, the Wald-Wolfowitz (1943) (W-W) test was used to test for the independence of a dataset. For the test value u which is less than the critical value at 5% significance level U0.025=1.96 then we can accept the hypothesis of independence and stationarity. The summarized results of the tests are given in Table 4-4.

The sample calculation for station Rinzaf @ Butajira:-

Table 4-3:- Result of independent test for Rinzaf @ Butajira

-	Station-Rinzaf @ butajera					
No.	obs	lag obs	xi(xi+1)	xi^2	xi^3	xi^4
1	6.978	3.563	24.86261	48.69248	339.7762	2370.958
2	3.563	15.087	53.75498	12.69497	45.23217	161.1622
3	15.087	18.32	276.3938	227.6176	3434.066	51809.76
4	18.32	33.587	615.3138	335.6224	6148.602	112642.4
5	33.587	4.854	163.0313	1128.087	37889.04	1272579
6	4.854	4.883	23.70208	23.56132	114.3666	555.1356
7	4.883	22.564	110.18	23.84369	116.4287	568.5215
8	22.564	8.592	193.8699	509.1341	11488.1	259217.5
9	8.592	9.557	82.11374	73.82246	634.2826	5449.756
10	9.557	15.455	147.7034	91.33625	872.9005	8342.31
11	15.455	4.585	70.86118	238.857	3691.535	57052.68
12	4.585	2.278	10.44463	21.02223	96.3869	441.9339
13	2.278	43.295	98.62601	5.189284	11.82119	26.92867
14	43.295	13.663	591.5396	1874.457	81154.62	3513589
15	13.663	3.167	43.27072	186.6776	2550.576	34848.51
16	3.167	2.697	8.541399	10.02989	31.76466	100.5987
17	2.697	5.62	15.15714	7.273809	19.61746	52.9083
18	5.62	14.72	82.7264	31.5844	177.5043	997.5743
19	14.72	22.284	328.0205	216.6784	3189.506	46949.53
20	22.284	2.81	62.61804	496.5767	11065.71	246588.4
21	2.81	6.599	18.54319	7.8961	22.18804	62.3484
22	6.599	6.978	46.04782	43.5468	287.3653	1896.324

	265.158		3067.322	5614.201	163381.4	5616304
	S1		R	S2	S3	S4

N=22,

$S1 = \sum_{i=1}^{n=41} xi = 265.158$ $\quad$ $S2 = \sum_{i=1}^{n=41} xi^2 = 5614.201$

$S3 = \sum_{i=1}^{n=41} x^3 = 163381.4$ $\quad$ $S4 = \sum_{i=1}^{n=41} x^4 = 5616304$

$$R = \sum_{i=1}^{N-1} (Xi * Xi + 1) + (X1 * XN) = 3067.322$$

$$Rbar = \frac{(s1^2 - s2)}{(N-1)} = 3080.694$$

$$Var\ (R) = \frac{S2^2 - S4}{N-1} - (Rmean)^2 + \frac{S1^4 - 4S1^2S2 + 4S1S3 + S2^2 - 2S4}{(N-1)(N-2)} = 214197.5$$

$$u = \frac{R - \bar{R}}{(var(R))^{\frac{1}{2}}} = -0.028891$$

Since |u| = 0.028891 is less than the critical value u0.025 = 1.96, the station Rinzaf @ Butajera can be considered to be homogeneous and stationary at 5% level of significance.

Table 4-4:- Result on independence test for all stations

Independent Test			
Station	**Statistic (U)**	**Critical Statistic (Ucritical)**	**Remark**
Dedeba nr Kuyera	0.561	1.96	Independent
Kerkersitu nr Adamitulu	0.189	1.96	Independent
Horakelo nr Langano	1.878	1.96	Independent
Lower timala nr sagure	1.059	1.96	Independent
Katar nr Feti	1.498	1.96	Independent
Meki @ Meki river	0.897	1.96	Independent
Katar nr Abura	0.266	1.96	Independent
Chufa nrArata	0.502	1.96	Independent
Ferfero nrWulbareg	0.185	1.96	Independent
Rinzaf @ butajera	0.029	1.96	Independent

Since all the statistics values are greater than the critical test statistics values the data in the stations are independent.

4.1.3 Homogeneity Test

Based on the above procedures stated on section 3.2.4.3 for all stations statistical tests was performed and the following results are presented for only one station and results for the other stations are annexed on appendixes C.

⇒ **Homogeneity test for station Katar @ Abura**

Table 4-5:- Homogeneity test result for station Katar @ Abura

Mann-Whitney Test

Group		N	Mean Rank	Sum of Ranks
Data	1	20	19.58	391.50
	2	21	22.36	469.50
	Total	41		

Test Statistics[a]

	Data
Mann-Whitney U	181.500
Wilcoxon W	391.500
Z	-.744
Asymp. Sig. (2-tailed)	.457

a. Grouping Variable: Group

As per the above test results the Z value (0.744) is less than the critical value (1.96) and also the sig value (0.457) is greater than α value (0.05) so, we can conclude that there is no evidence to support a difference between the collected flow data's at the selected station can be considered as homogeneous and stationary at 5% level of significance ($U = 181.5$, $N1 = 20$, $N2 = 21$, $P = 0.46$).

So, we can summarize the results of homogeneity test for other stations as per the following table: -

Table 4-6:- Results of homogeneity test for each station

I.No	Station	Mann - Whitney value	Critical value	Sig. Value (P value)	α value	Remarks!
1	Dedeba near Kuyera	0.206	1.96	0.851	0.05	Homogeneous
2	Kerkersitu near Adamitulu	0.87	1.96	0.401	0.05	Homogeneous
3	Horakelo near Langano	1.451	1.96	0.156	0.05	Homogeneous
4	Lower Timala near Sagure	1.854	1.96	0.067	0.05	Homogeneous
5	Katar near Fite	1.518	1.96	0.131	0.05	Homogeneous
6	Meki @ Meki river	0.474	1.96	0.649	0.05	Homogeneous
7	Katar @ Abura	0.744	1.96	0.457	0.05	Homogeneous
8	Chufa near Arata	0.185	1.96	0.880	0.05	Homogeneous
9	Ferfero near Wulbareg	1.511	1.96	0.146	0.05	Homogeneous
10	Rinzaf @ Butajera	1.055	1.96	0.314	0.05	Homogeneous

4.1.4 Test for Outliers

Using the above procedures stated on section 3.2.4.4 for all stations we can check the outlier in the data set.

- For example, for station **Katar @ Abura** the calculated value of Grubb's test is equals to:-

$G_{calculated}$ = max $= \frac{|Xi-Xmean|}{S}$ =3.46 and the critical value is around 2.877. So, we can reject the null hypothesis at 5% significant level. So there is an outlier.

After checking the existence of outliers, we can execute detection of those outliers using different methods. Tukey's method on SPSS is used for detection of outliers by calculating quartiles to calculate the upper and lower boundaries.

Q3: - Third quartile at 75 percentiles

Q1: - First quartile at 25 percentiles

Upper boundary: - $Q3 + 1.5(Q3 - Q1)$

Lower boundary: - $Q1 - 1.5(Q3 - Q1)$

Table 4-7:- Result for outlier for station Katar @ Abura

Case Processing Summary

Flow	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
	41	100	0	0	41	100

Percentiles

	Percentiles						
	5	10	25	50	75	90	95
Weighted average	40.7556	44.3466	60.0320	89.1430	137.223	155.1290	174.8337
Tukey's Hinges			60.0320	89.1430	135.4360		

Extreme Values

		Case number	Value
Flow	Highest 1	8	262.05
	2	14	176.28
	3	41	161.84
	4	12	155.90
	5	37	152.03
	Lowest 1	28	25.90
	2	18	40.52
	3	33	42.88
	4	16	43.68
	5	15	47.01

$Upper = 137.223 + 1.5 * (137.223 - 60.032) = 253.0095$

$Lower = 60.032 - 1.5 * (137.223 - 60.032) = -55.7545$

N.B: - there is one outlier within the data set

- For station **Kerkersitu near Adamitulu**

The calculated value of Grubb's test is equals to $G_{calculated} = 2.63$ and the critical value is around 2.759. So, we can't reject the null hypothesis at 5% significant level. So, there is no outlier in the data set.

Also, when we check the upper and lower boundary values (56.136, -18.2848), there is no outlier in the extreme values.

Table 4-8:- Result for outlier for station Kerkersitu near Adamitulu

Case Processing Summary

Flow	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
	31	100	0	0	31	100

Percentiles

	Percentiles						
	5	10	25	50	75	90	95
Weighted average	2.7686	3.8398	9.6230	19.7880	28.2282	35.2342	42.5170
Tukey's Hinges			10.0465	19.7880	27.8144		

Extreme Values

			Case number	Value
Flow	Highest	1	4	50.8
		2	17	37.0
		3	28	35.69
		4	19	33.42
		5	31	32.85
	Lowest	1	24	2.62
		2	25	2.87
		3	6	3.47
		4	1	5.34
		5	8	6.84

Table 4-9:- Result of outlier for all stations

I.No	Station	Calculated Grubb's value	Grubb's critical value	Remarks!
1	Dedeba near Kuyera	2.739	2.823	No outlier
2	Kerkersitu near Adamitulu	2.630	2.759	No outlier
3	Horakelo near Langano	2.134	2.714	No outlier
4	Lower Timala near Sagure	2.406	2.698	No outlier
5	Katar near Fite	2.110	2.681	No outlier
6	Meki @ Meki river	3.328	2.866	There are two outliers
7	Katar @ Abura	3.460	2.877	There is one outlier
8	Chufa near Arata	2.061	2.624	No outlier
9	Ferfero near Wulbareg	2.219	2.504	No outlier
10	Rinzaf @ Butajera	2.911	2.603	There is one outlier

As a result of the above table the outliers detected in the stations are rejected from the record before analysis. `

4.2 Regionalization

4.2.1 Identified Homogeneous Regions in the Sub Basin

Preliminary identification of groups of stations into a certain group is achieved through looking at catchment characteristics of the study area. After processing a preliminary region based on catchment characteristics the next step will be identifying stations having the same parent distribution using flood statistics of the sub basin.

Flood statistics of Zeway-Shala sub- basin stations were computed using both conventional moment and L-moment methods. The Flood statistics for the sub-basin for selected stations are shown in Table 4-10 below.

Table 4-10:- Flood statistics of the Stations

S.No	Station Name	Station code	N	Area (sq km)	Mean Flow (m3/sec)	CV	CS	Ck	LCV	LCS	LCK
1	Dedeba nr Kuyera	81003	36	156	9.678	0.389	0.304	3.801	0.221	0.058	0.193
2	Kerkersitu nr Adamitulu	81004	31	7488	21.978	0.600	0.482	3.142	0.343	0.100	0.076
3	Horakelo nr Langano	81005	28	2006	5.955	0.584	0.668	3.158	0.330	0.160	0.140
4	Lower Timala nr Sagure	81008	27	184.45	15.359	0.934	0.970	3.347	0.505	0.283	0.055
5	Katar nr Fite	81011	26	1975	155.238	0.495	0.709	3.111	0.281	0.175	0.136
6	Meki @ Meki river	81018	40	2433	64.952	0.280	-0.805	2.818	0.163	-0.091	0.138
7	Katar nr Abura	81019	40	3350	91.192	0.424	0.307	2.153	0.247	0.102	0.045
8	Chufa nr Arata	81025	23	216	11.589	0.220	-0.656	2.687	0.134	-0.123	0.110
9	Ferfero nr Wulbareg	81026	18	153	77.328	0.326	-0.680	3.724	0.192	-0.079	0.183
10	Rinzaf @ Butajera	82032	21	49	9.459	0.791	1.253	4.600	0.423	0.313	0.116

Using the statistical characteristics plot the L moment ratio diagram and identify the stations having the same parent distribution. And the stations nearly the same distribution grouped in the same regions.

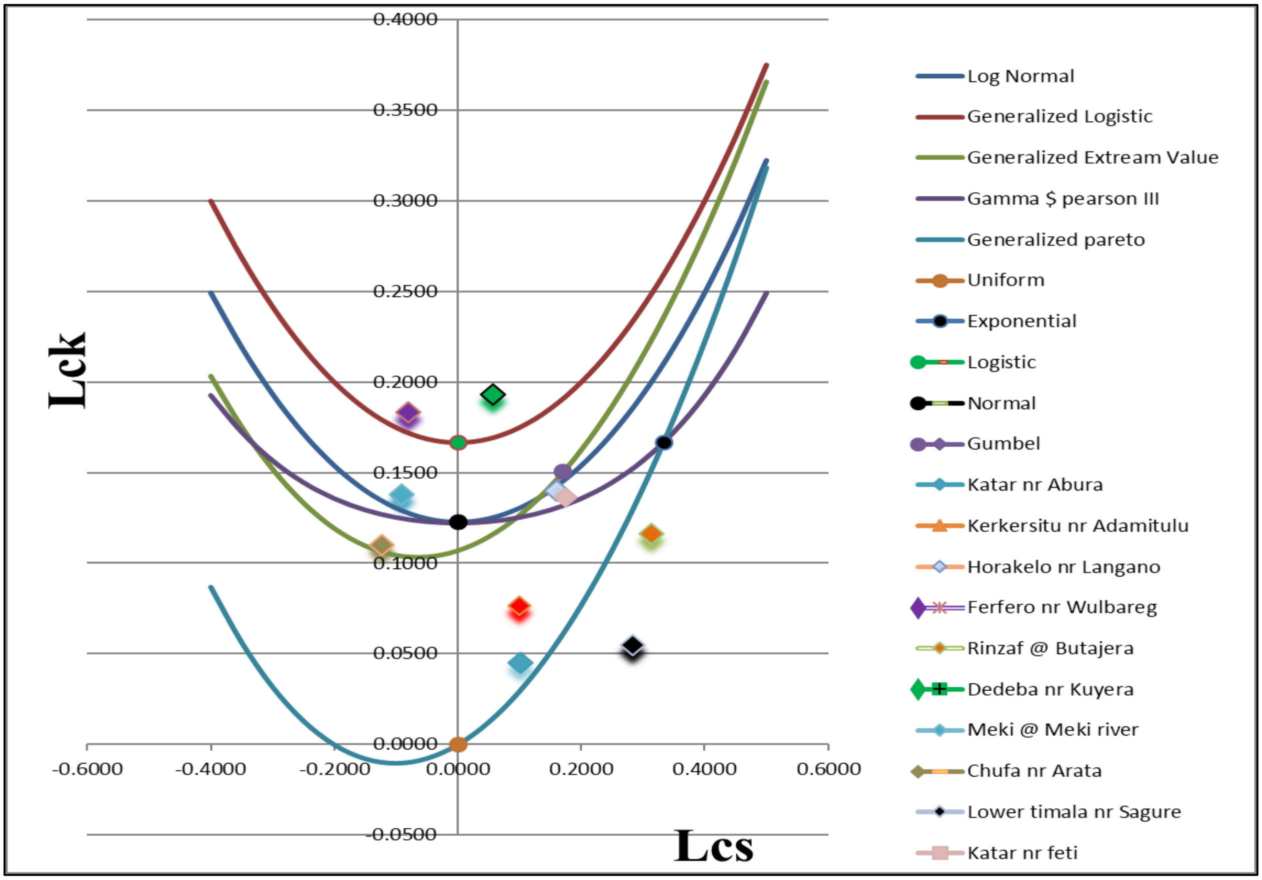


Figure 4-1:- L- moment ratio diagram for stations in Zeway – Shala sub basin

Based on the combined results the sub basin was divided in to two regions. The combined result from the site characteristics and LMRD is tabulated as follows.

Table 4-11:- The combined result of catchment characteristics and LMRD for selection of homogenous region

Region	**Station Name**	**Possible distribution from LMRD**
	Meki @meki river	**Log normal (LN)**/Gamma & person III/Generalized logistic (GL)
	Katar nr Abura	Generalized parito (GP)/**Log normal (LN)**/Gamma & person III
	Chufa nr Arata	Generalized Extrem Value (GEV)/ Gamma & person III/**Log normal (LN)**
	Ferfero nr Wulbareg	Generalized logistic (GL)/**Log normal (LN)/**Generalized Extrem Value (GEV)
Region One	Dedeba nr Kuyera	Generalized logistic (GL)/**Log normal (LN)**/Generalized Extrem Value (GEV)
	Rinzaf @ Butajera	**Generalized parito (GP)/**Gamma & person III/Log normal (LN)
	Kerkersitu nr Adamitulu	**Generalized parito (GP)/**Gamma & person III/Log normal(LN)
	Horakelo nr Langano	Log normal (LN)/Generalized Extrem Value(GEV)/**Generalized parito(GP)**
	Katar nr Fite	Gamma & person III/**Generalized parito(GP)/**Log normal(LN
Region Two	Lower Timala nr Sagure	**Generalized parito (GP)**/Gamma & person III/Log normal (LN)

4.2.2 Result on Homogeneity Test of the Regions

Both Discordance measure test and CC – based homogeneity test are employed to check regional homogeneity of the proposed stations in the Ziway- Shala sub basin.

4.2.2.1 Discordance Measure Test

This test measures sites that are grossly discordant from the group as a whole. The discordancy measure Di estimates how far a given site is from the center of the group. Based on the equation 3.22 the discordancy measures (Di) of each region are given below:

The result of Discordancy test for stations in Region 1

- In order to determine the discordancy measure Di for stations in the region the following procedures are conducted.

1. First put the statistical values t, t3 and t4 for each station in the region and then found the mean of each statistical value of stations.

$$u(\text{mean}) = \frac{1}{NS}\sum_{i=1}^{NS} ui$$

Region	Name of Stations	Station Code	Mean Flow (m3/sec)	N	t	t3	t4	Ui
Region-1	Dedeba nr Kuyera	81003	9.678	36	0.221	0.058	0.193	U1
	Meki @ Meki river	81018	64.952	40	0.163	-0.091	0.138	U2
	Katar nr Abura	81019	91.192	40	0.247	0.102	0.045	U3
	Chufa nr Arata	81025	11.589	23	0.134	-0.123	0.110	U4
	Ferfero nr Wulbareg	81026	77.328	18	0.192	-0.079	0.183	U5

Region-1						
U1	**U2**	**U3**	**U4**	**U5**	**Summ Ui**	**Ubar**
0.221	0.163	0.247	0.134	0.192	0.957	0.1914
0.058	-0.091	0.102	-0.123	-0.079	-0.133	-0.0266
0.193	0.138	0.045	0.110	0.183	0.669	0.1338

2. Subtract the statistical value of each station from mean and then transpose it.

Ui-Ubar				
U1-Ubar	**U2-Ubar**	**U3-Ubar**	**U4-Ubar**	**U5-Ubar**
0.0296	-0.0284	0.0556	-0.0574	0.0006
0.0846	-0.0644	0.1286	-0.0964	-0.0524
0.0592	0.0042	-0.0888	-0.0238	0.0492

(Ui-Ubar)T			
0.0296	0.0846	0.0592	(U1-Ua)T
-0.0284	-0.0644	0.0042	(U2-Ua)T
0.0556	0.1286	-0.0888	(U3-Ua)T
-0.0574	-0.0964	-0.0238	(U4-Ua)T
0.0006	-0.0524	0.0492	(U5-Ua)T

3. Calculate the covariance matrix and then find a 3 by 3 matrixs

$$S = (NS-1)^{-1}\sum_{i=1}^{NS}(ui-u(mean)*\left(ui-u(mean)\right)^{T}$$

where Ns = number of stations

(U1-Ua)*(U1-Ua)T	0.0009	0.0025	0.0018
	0.0025	0.0072	0.0050
	0.0018	0.0050	0.0035
(U2-Ua)*(U2-Ua)T	0.0008	0.0018	-0.0001
	0.0018	0.0041	-0.0003
	-0.0001	-0.0003	0.0000
(U3-Ua)*(U3-Ua)T	0.0031	0.0072	-0.0049
	0.0072	0.0165	-0.0114
	-0.0049	-0.0114	0.0079
(U4-Ua)*(U4-Ua)T	0.0033	0.0055	0.0014
	0.0055	0.0093	0.0023
	0.0014	0.0023	0.0006
(U5-Ua)*(U5-Ua)T	0.0000	0.0000	0.0000
	0.0000	0.0027	-0.0026
	0.0000	-0.0026	0.0024

Summation of (Ui-Ua)*(Ui-Ua)T, i=1…5		
0.0081	0.0170	-0.0019
0.0170	0.0399	-0.0070
-0.0019	-0.0070	0.0144

S	0.002017	0.0042463	-0.000477
	0.004246	0.0099703	-0.001741
	-0.000477	-0.001741	0.003599

4. Calculate the inverse of matrix S= S^{-1}

$$S^{-1} = \frac{Adj\ S}{|S|}$$

Where; Adj = Adjoint of s

$|S|$ = Determinante of S

The adjoint of s is found by:

(i) Obtaining the matrix of the cofactors of the elements, and

(ii) Transposing this matrix.

- The inverse of S

$$\begin{matrix} 0.002017 & 0.0042463 & -0.000477 \\ 0.004246 & 0.0099703 & -0.001741 \\ -0.000477 & -0.001741 & 0.003599 \end{matrix}$$

The cofactor of element $0.002017 \;= +(0.0099703 * 0.003599) - (-0.001741 * -0.001741)$ =0.0000328

The cofactor of element 0.0042463 = - $((-0.001741 * -0.001741)$ - $(0.004246 * 0.003599))$ = -0.0000145

Based on the above calculation the matrix of cofactors is

0.0000328	-0.0000145	-0.0000026
-0.0000145	0.0000070	0.0000015
-0.0000026	0.0000015	0.0000021

The transpose of the matrix of cofactors, i.e. the adjoint of the matrix, is obtained by writing the rows as columns,

The adjoint of the matrix is

0.0000328	-0.0000145	-0.0000026
-0.0000145	0.0000070	0.0000015
-0.0000026	0.0000015	0.0000021

$|S|$=0.000000006162

$$\mathrm{S}^{-1} = \frac{\text{Adj S}}{|\mathrm{S}|}$$

5330.8667	-2345.1411	-427.9894
-2345.1411	1141.2257	241.2945
-427.9894	241.2945	337.8929

5. Finally the discordance of the stations in the region will get by

$$Di = \frac{1}{3}\left(ui - u(mean)\right)^{T} S^{-1}\left(ui - u(mean)\right)$$

$1/3(Ui\text{-}Ua)^T$		
0.0099	0.0282	0.0197
-0.0095	-0.0215	0.0014
0.0185	0.0429	-0.0296
-0.0191	-0.0321	-0.0079
0.0002	-0.0175	0.0164

*

S^{-1}		
5330.8667	-2345.1411	-427.9894
-2345.1411	1141.2257	241.2945
-427.9894	241.2945	337.8929

$1/3(Ui\text{-}Ua)^T S^{-1}$		
-21.9808	13.8054	9.2494
-0.7224	-1.9598	-0.6551
10.9388	-1.6851	-7.5902
-23.2447	6.2847	-2.2453
35.0089	-16.4452	1.2412

Ui-Ubar				
U1-Ubar	**U2-Ubar**	**U3-Ubar**	**U4-Ubar**	**U5-Ubar**
0.0296	-0.0284	0.0556	-0.0574	0.0006
0.0846	-0.0644	0.1286	-0.0964	-0.0524
0.0592	0.0042	-0.0888	-0.0238	0.0492

$$Di = \frac{1}{3}\left(ui - u(mean)\right)^{T} S^{-1}\left(ui - u(mean)\right)$$

D1= (-21.9808*0.0296) + (13.8054*0.0846) + (9.2494*0.0592) = 1.0649

For the first station the discordance measure D1 is 1.0649

The summarized results of the stations in the region 1 are given in Table 4-12.

Table 4-12:- The result of discordancy measure for region one

Region-1 (Log normal)					
Name of Stations	**t**	**t3**	**t4**	**Discordancy measure (Di)**	**Remark**
Meki @ Meki river	0.163	-0.091	0.138	0.144	Homogenous
Katar nr Abura	0.247	0.102	0.045	1.065	Homogenous
Chufa nr Arata	0.134	-0.123	0.110	0.782	Homogenous
Ferfero nr Wulbareg	0.192	-0.079	0.183	0.944	Homogenous
Dedeba nr Kuyera	0.221	0.058	0.193	1.065	Homogenous

- For stations in region 2 the same procedure were used to determine the discordancy measure Di. The summarized results of the stations in the region 2 are given in Table 4-13.

Table 4-13:- The result of discordancy measure for region two

Region-2 (Generalized Pareto)					
Name of Stations	**t**	**t3**	**t4**	**Discordancy measure (Di)**	Remark
Rinzaf @ Butajera	0.423	0.313	0.116	0.585	Homogenous
Kerkersitu nr Adamitulu	0.343	0.100	0.076	0.81	Homogenous
Horakelo nr Langano	0.330	0.160	0.140	-1.051	Homogenous
Katar nr Fite	0.281	0.175	0.136	0.792	Homogenous
Lower Timala nr Sagure	0.505	0.283	0.055	-0.761	Homogenous

So, both regions are homogeneous since the Di values are less than 3.

4.2.2.2 CC-Based Homogeneity Test

In this test the site-to-site coefficient of variation of the coefficient of variation (CC) of both conventional and L-moments of the proposed region are used. Based on the equation 3.25 & 3.28 coefficient of variation (CC) of each region given below:

Table 4-14:- CC values for region one on both conventional moment and L- moment

Station Name	**Mean Flow (m3/sec)**	**CV**	**LCV**
Meki @ Meki river	64.952	0.280	0.163
Katar nr Abura	91.192	0.424	0.247
Chufa nr Arata	11.589	0.220	0.134
Ferfero nr Wulbareg	77.328	0.326	0.192
Dedeba nr Kuyera	9.678	0.389	0.221
Mean		0.328	0.191
Standard deviation		0.082	0.045
CC		0.250	0.235

Table 4-15:- CC values for region two on both conventional moment and L- moment

Station Name	**Mean Flow (m3/sec)**	**CV**	**LCV**
Rinzaf @ Butajera	9.459	0.791	0.423
Kerkersitu nr Adamitulu	21.978	0.600	0.343
Horakelo nr Langano	5.955	0.584	0.330
Katar nr Fite	155.238	0.495	0.281
Lower Timala nr Sagure	15.359	0.934	0.505
Mean		0.681	0.376
Standard deviation		0.178	0.088
CC		0.261	0.234

So, both regions are homogeneous for both convectional Cv based and L-moment Cv based homogeneity test since the CC values are less than 0.3 for both cases.

4.2.3 Result on Delineation of Homogeneous Regions

Based on the result of the homogeneous tests the preliminary division of the sub basin remains the same. Accordingly, the sub-basin has divided into two regions. The first region includes the Meki, Katar nr abura, Chufa, Ferfero and Dedeba stations and the second region includes Rinzaf, Kerkersitu, Horakelo, Katar nr Fite, and Lower Timala stations.

Then the delineation of homogeneous regions was done using the software Arc GIS. All stations under analysis were identified according to their geographical location (latitude and longitude) on the digitized map of the basin and also the flood statistical values of Lcs and Lck were known. Therefore, based on the distance between the neighboring stations and the interpolating values of Lcs and Lck, the boundary between two stations of different region were fix. See Figure 4.2, which shows the map of the delineated homogenous regions.

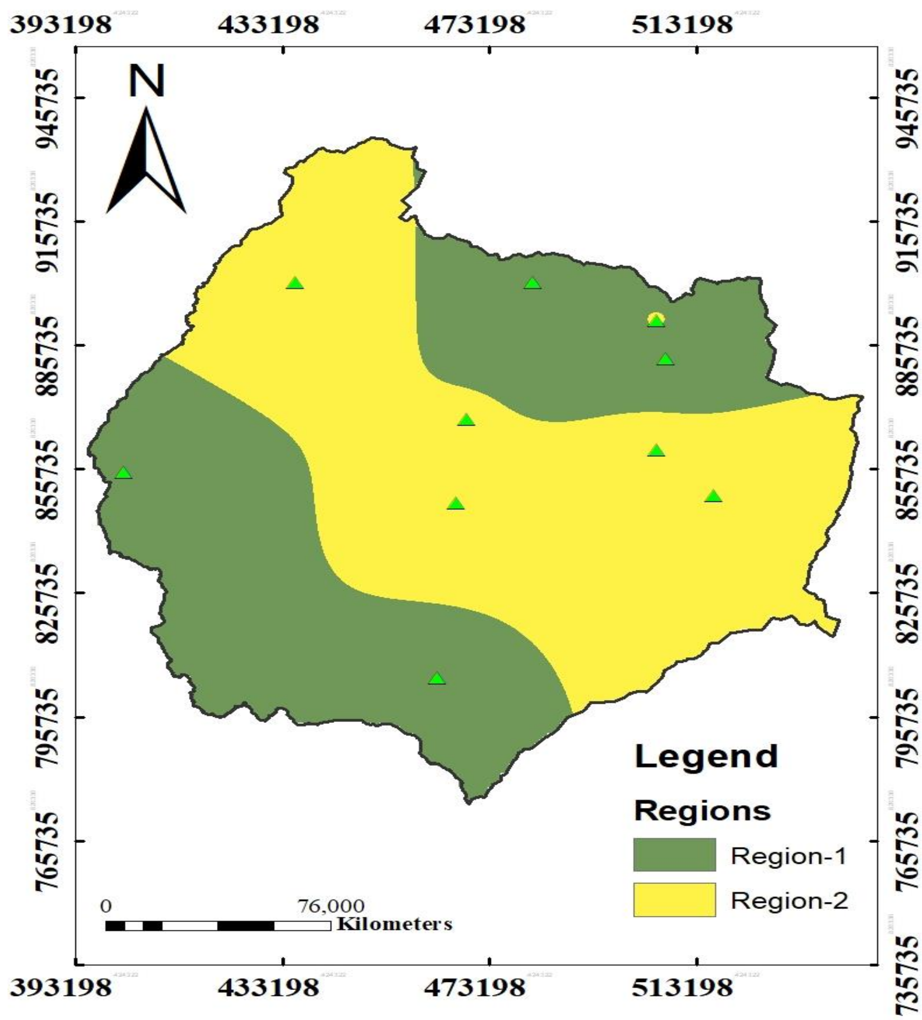


Figure 4-2:- The Delineated Homogeneous Regions of Ziway – Shala Sub Basin

4.3 Selected Parent Distribution and Parameters for the Identified Regions

Goodness of fit test

In this study based on the Easy-fit software the fitness of the possible parent distribution were checked using Kolmogorov-Smirnov test and chi squared test. Goodness of fit test result for stations on region one shows below.

Table 4-16:- Goodness of Fit – Summary of Meki

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Gen. Extreme Value	0.06323	1	1.5873	1
Log-Pearson 3	0.08939	2	1.7015	2
Lognormal (3P)	0.10999	3	5.0499	6
Gamma (3P)	0.13075	4	4.5423	4
Gamma	0.14658	5	4.0601	3
Lognormal	0.16912	6	4.8647	5

Table 4-17:- Goodness of Fit – Summary of Katar @ Abura

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Gen. Extreme Value	0.0928	1	3.7455	6
Log-Pearson 3	0.09622	2	2.6925	2
Lognormal (3P)	0.10204	3	2.7825	3
Gamma (3P)	0.10302	4	2.7833	4
Gamma	0.1035	5	2.9543	5
Lognormal	0.11476	6	0.99929	1

Table 4-18:- Goodness of Fit – Summary of Chufa

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Gen. Extreme Value	0.0791	1	1.0907	3
Log-Pearson 3	0.1077	2	0.18376	1
Lognormal (3P)	0.12504	3	0.25438	2
Gamma (3P)	0.13213	4	1.8021	5
Gamma	0.14686	5	1.8656	6
Lognormal	0.16805	6	1.6549	4

Table 4-19:- Goodness of Fit – Summary of Ferfero

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Gen. Extreme Value	0.11946	1	0.0766	1
Lognormal (3P)	0.13214	2	0.17966	3
Log-Pearson 3	0.13905	3	0.11291	2
Gamma (3P)	0.14733	4	0.77408	6
Gamma	0.15738	5	0.53858	5
Lognormal	0.21788	6	0.47876	4

Table 4-20:- Goodness of Fit – Summary of Dedeba

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Lognormal (3P)	0.07662	1	1.1073	3
Gamma (3P)	0.07713	2	1.1057	2
Gen. Extreme Value	0.08103	3	1.0828	1
Gamma	0.10614	4	1.2359	4
Log-Pearson 3	0.11341	5	2.26 7	6
Lognormal	0.15481	6	2.2479	5

Based on the statistical value, widely acceptance and simple to apply Generalized extreme value distribution is selected as best distribution for region one. And all the statistical values are less than the critical value and also all the p-values are greater than α which the data fit the specified distribution. See Appendix -D, for detail results of chi-square and Kolmogorov-Smirnov test.

Table 4-21:- Statistical value of GEV using kolmogorov test for Region 1

No	Station Name	Statistic	P-value	α-value	Critical Value	Remark!
1	Meki	0.063	0.994	0.050	0.210	Accept
2	Katar nr Abura	0.093	0.85	0.050	0.210	Accept
3	Chufa	0.079	0.996	0.050	0.275	Accept
4	Ferfero	0.119	0.933	0.050	0.309	Accept
5	Dedeba	0.081	0.957	0.050	0.221	Accept

Table 4-22:- Statistical value of LP3 using Chi-Squared test for Region 1

No	Station Name	Statistic	P-value	α-value	Critical Value	Remark!
1	Meki	1.587	0.903	0.050	11.070	Accept
2	Katar nr Abure	3.746	0.442	0.050	9.488	Accept
3	Chufa	1.091	0.779	0.050	7.815	Accept
4	Ferfero	0.077	0.782	0.050	3.842	Accept
5	Dedeba	1.083	0.781	0.050	7.815	Accept

Goodness of fit test result for stations on region two

Table 4-23:- Goodness of Fit – Summary of Rinzaf

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Gamma	0.10962	1	0.48128	3
Gen. Extreme Value	0.133	2	1.6112	5
Log-Pearson 3	0.13938	3	0.09123	1
Lognormal	0.14153	4	0.4649	2
Lognormal (3P)	0.14837	5	0.65964	4
Gamma (3P)	0.16921	6	N/A	

Table 4-24:- Goodness of Fit – Summary of Kerkersitu

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Gen. Extreme Value	0.08575	1	0.46704	1
Log-Pearson 3	0.09183	2	0.93747	2
Lognormal (3P)	0.10832	3	0.94683	3
Gamma (3P)	0.11521	4	2.2188	5
Gamma	0.12174	5	2.2292	6
Lognormal	0.13899	6	2.1037	4

Table 4-25:- Goodness of Fit – Summary of Horakelo

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Gen. Extreme Value	0.06634	1	0.2188	4
Lognormal (3P)	0.07744	2	0.10959	2
Log-Pearson 3	0.08813	3	0.05856	1
Gamma	0.09055	4	0.11895	3
Gamma (3P)	0.11878	5	0.34632	5
Lognormal	0.13073	6	1.216	6

Table 4-26:- Goodness of Fit – Summary of Katar nr Fite

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Log-Pearson 3	0.09162	1	1.2636	4
Lognormal (3P)	0.09467	2	1.3197	6
Gen. Extreme Value	0.10207	3	1.0288	1
Gamma	0.10372	4	1.0894	2
Lognormal	0.10549	5	1.2129	3
Gamma (3P)	0.11268	6	1.2806	5

Table 4-27:- Goodness of Fit – Summary of Lower Timala

Distribution	Kolmogorov Smirnov		Chi-Squared	
	Statistic	Rank	Statistic	Rank
Log-Pearson 3	0.1373	1	3.4992	3
Gen. Extreme Value	0.14359	2	3.1571	2
Lognormal	0.14816	3	3.6964	4
Gamma (3P)	0.14899	4	N/A	
Gamma	0.16498	5	4.1753	5
Lognormal (3P)	0.17213	6	1.1255	1

Based on the statistical value, widely acceptance and simple to apply Log Pearson Type III distribution is selected as best distribution for region two. And all the statistical values are less than the critical value and also all the p-values are greater than α which the data fit the specified distribution. See Appendix –E, for detail results of chi-square and Kolmogorov-Smirnov test.

Table 4-28:- Statistical value of LP3 using Kolmogorov test for Region 2

No	Station Name	Statistic	P-value	α-value	Critical Value	Remark!
1	Rinzaf	0.139	0.759	0.050	0.287	Accept
2	Kerkersitu	0.092	0.935	0.050	0.238	Accept
3	Horakelo	0.088	0.991	0.050	0.250	Accept
4	Katar nr Fite	0.092	0.967	0.050	0.259	Accept
5	Lower Timala	0.137	0.639	0.050	0.254	Accept

Table 4-29:- Statistical value of LP3 using Chi-Squared test for Region 2

	Station Name	Statistic	P-value	α-value	Critical Value	Remark!
1	Rinzaf	0.091	0.955	0.050	5.992	Accept
2	Kerkersitu	0.937	0.816	0.050	7.815	Accept
3	Horakelo	0.059	0.996	0.050	7.815	Accept
4	Katar nr Fite	1.264	0.174	0.050	5.992	Accept
5	Lower Timala	3.499	0.738	0.050	7.815	Accept

Estimated Parameter

After a distribution is selected to fit the data, their parameter will be estimated. The statistical parameters for region one Generalized Extreme value (GEV) distribution are estimated by the method of moments from the given data.

For the station Meki

N=40, m'1=60.26695 CV=0.28009 Cs= -0.50469

For the value of k > 0 (–2 < CS < 1.14), R2 = 1 the value of k determined using equation 3.34

$$\mathrm{K} = 0.277648 - 0.322016\mathrm{cs} + 0.060278\mathrm{cs}^2 + 0.016759\mathrm{cs}^3 - 0.005873\mathrm{cs}^4 - 0.00244\mathrm{cs}^5 - 0.000050\mathrm{cs}\text{^}6$$

= 0.453063

Then calculate

$\Gamma(1+\mathrm{k})$= 0.885635

$\Gamma(1+2\mathrm{k})$= 0.963879

Now from equation 3.35 & equation 3.36 calculate the parameter α and u.

$$\alpha = \left[\frac{(0.28009 \ * 60.26695)^2 * (0.453063)^2}{(0.963879 - (0.885635)^2)}\right]^{0.5} = 18.049$$

$$\mathrm{u} = 60.26695 \quad - \frac{18.049}{0.453063}[1 - 0.885635] = 55.710$$

The calculated parameters of all stations in the region given in table 4-30

Table 4-30:- Parameters for each station in the region 1

Parameters for GEV	Stations on Region One				
	Meki	**Katar nr abura**	**Chufa**	**Ferfero**	**Dedeba**
κ	0.453	0.185	0.509	0.518	0.186
α	18.049	37.544	2.566	80.832	3.468
u	55.710	78.311	10.209	51.566	8.022

Based on V.T. Chow (1998) the statistical parameters estimated for region two Log Pearson Type III distributions by the method of moments from the given data. For distribution, the first step is to take the logarithms of the hydrologic data, $y = \log x$. usually logarithms to base 10 are used. The meanμ, standard deviationσ, and coefficient of skewness Cs are calculated for the logarithms of the data.

Based on the logarithms of the hydrologic data for the LP3 distribution the corresponding parameters of each station in the region are given below.

Table 4-31:- Parameters for each station in the region 2

Parameters for LP3	Stations on Region Two				
	Rinzaf	**Kerkersitu**	**Horakelo**	**Katar nr Fite**	**Lower Timala**
μ	0.896	1.192	0.693	2.128	0.964
σ	0.347	0.338	0.301	0.226	0.506
Cs	0.115	-0.857	-0.673	-0.244	-0.158

4.4 Estimated Quantiles of the Regions and Developed Regional Curve

Quantile estimation for region one

Using the computed regional parameters in table 4.30 for selected distribution i.e. Generalized Extreme Value (GEV) the T-year quantile is calculated based on equation 3.38

Table 4-32:- Estimated Quantile of Region one

Return Period (T)	Meki	Katar abura	Chufa	Ferfero	Dedeba
	$Q_{T(m3/s)}$	$Q_{T(m3/s)}$	$Q_{T(m3/s)}$	$Q_{T(m3/s)}$	$Q_{T(m3/s)}$
2	72.425	118.731	12.513	123.813	11.754
5	81.712	149.477	13.712	160.998	14.590
10	85.700	166.560	14.200	175.986	16.166
20	88.441	180.864	14.521	185.803	17.484
30	89.657	188.269	14.659	189.996	18.166
50	90.889	196.763	14.796	194.118	18.948
100	92.153	207.005	14.931	198.184	19.891
200	93.071	215.976	15.025	201.007	20.717
500	93.914	226.175	15.108	203.469	21.655
700	94.146	229.504	15.130	204.116	21.961
1000	94.355	232.813	15.150	204.689	22.265
3000	94.824	241.733	15.192	205.914	23.084
5000	94.974	245.301	15.204	206.285	23.412
7000	95.055	247.473	15.211	206.481	23.611
10000	95.129	249.633	15.217	206.655	23.809
2.33	74.543	124.985	12.793	132.526	12.331

Table 4-33:- The estimated standardized quantile of region 1

Return Period (T)	Meki	Katar abura	Chufa	Ferfero	Dedeba	Average of QT/Qm
	QT/Qm	QT/Qm	QT/Qm	QT/Qm	QT/Qm	
2	0.972	0.950	0.978	0.934	0.953	0.957
5	1.096	1.196	1.072	1.215	1.183	1.152
10	1.150	1.333	1.110	1.328	1.311	1.246
20	1.186	1.447	1.135	1.402	1.418	1.318
30	1.203	1.506	1.146	1.434	1.473	1.352
50	1.219	1.574	1.157	1.465	1.537	1.390
100	1.236	1.656	1.167	1.495	1.613	1.434
200	1.249	1.728	1.174	1.517	1.680	1.470
500	1.260	1.810	1.181	1.535	1.756	1.508
700	1.263	1.836	1.183	1.540	1.781	1.521
1000	1.266	1.863	1.184	1.545	1.806	1.533
3000	1.272	1.934	1.187	1.554	1.872	1.564
5000	1.274	1.963	1.188	1.557	1.899	1.576
7000	1.275	1.980	1.189	1.558	1.915	1.583
10000	1.276	1.997	1.189	1.559	1.931	1.591

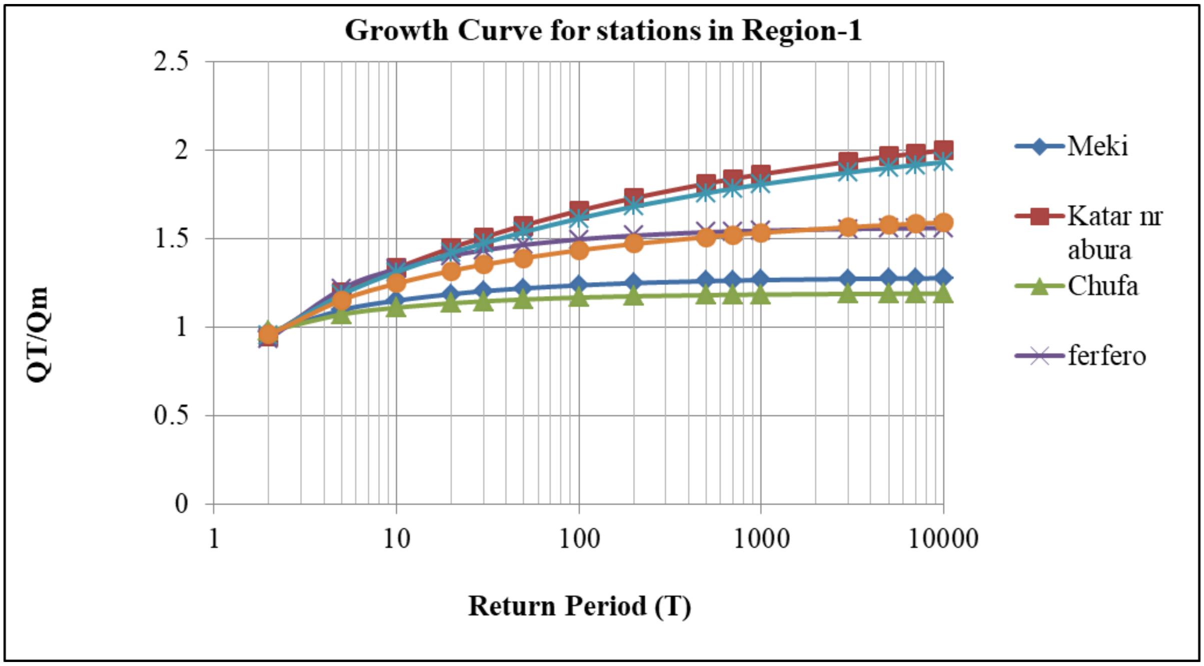


Figure 4-3:- Growth curve of each station in the region 1 using the Standardized quantile estimate

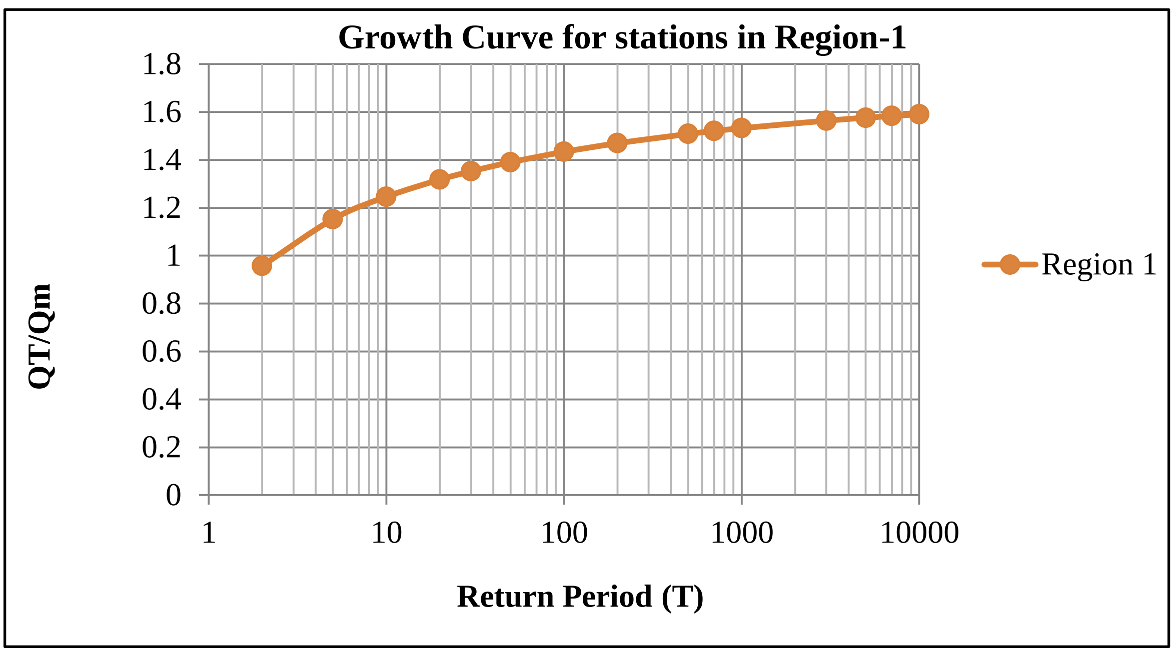


Figure 4-4:- Regional Growth Curve for Region one

Quantile estimation for region two

For a given return period, the frequency factor can be determined from the K-T relationship for Log Pearson Type III distribution using equation 3.36, and then the parameters in table 4.31. The magnitude YT computed by Equation 3.40. Finally the required value of XT is found by taking the antilog of YT.

Table 4-34:- Estimated Quantiles of Region Two

Return Period (T)	Z	Rinzaf	Kerkersitu	Horakelo	Katar nr fite	Lower Timala	Average of QT(m3/s)
		$Q_{T\ (m3/s)}$	$Q_{T\ (m3/s)}$	$Q_{T\ (m3/s)}$	$Q_{T\ (m3/s)}$	$Q_{T\ (m3/s)}$	
2	-1.0E-07	7.750	17.360	5.327	137.177	9.492	35.421
5	0.841	15.330	30.247	8.930	209.076	24.705	57.658
10	1.282	22.098	38.271	11.243	257.657	40.067	73.867
20	1.645	30.023	45.284	13.338	304.375	59.210	90.446
30	1.834	35.272	48.995	14.483	331.257	72.328	100.467
50	2.054	42.598	53.311	15.852	364.867	91.035	113.532
100	2.327	53.943	58.592	17.589	410.244	120.611	132.196
200	2.576	67.093	63.284	19.204	455.561	155.439	152.116
500	2.879	87.629	68.694	21.168	515.606	210.376	180.694
700	2.983	96.171	70.477	21.844	537.731	233.302	191.905
1000	3.091	105.866	72.258	22.536	561.242	259.324	204.245
3000	3.403	140.273	77.100	24.516	634.081	351.392	245.472
5000	3.540	158.861	79.051	25.364	668.189	400.789	266.451
7000	3.628	172.094	80.243	25.899	690.745	435.774	280.951
10000	3.719	187.035	81.430	26.446	714.732	475.073	296.943
2.33	0.178	8.937	19.765	5.995	150.345	11.664	39.341

Table 4-35:- The estimated standardized quantiles of region 2

Return Period (T)	**Rinzaf**	**Kerkersitu**	**Horakelo**	**Katar nr fite**	**Lower Timala**	**Average of QT/QM**
	QT/Qm	**QT/Qm**	**QT/Qm**	**QT/Qm**	**QT/Qm**	
2	0.867	0.878	0.889	0.912	0.814	0.872
5	1.715	1.530	1.490	1.391	2.118	1.649
10	2.473	1.936	1.875	1.714	3.435	2.287
20	3.359	2.291	2.225	2.025	5.076	2.995
30	3.947	2.479	2.416	2.203	6.201	3.449
50	4.766	2.697	2.644	2.427	7.805	4.068
100	6.036	2.964	2.934	2.729	10.341	5.001
200	7.507	3.202	3.203	3.030	13.327	6.054
500	9.805	3.476	3.531	3.429	18.037	7.656
700	10.761	3.566	3.644	3.577	20.002	8.310
1000	11.845	3.656	3.759	3.733	22.233	9.045
3000	15.695	3.901	4.090	4.218	30.127	11.606
5000	17.775	4.000	4.231	4.444	34.362	12.962
7000	19.256	4.060	4.320	4.594	37.361	13.918
10000	20.928	4.120	4.411	4.754	40.730	14.989

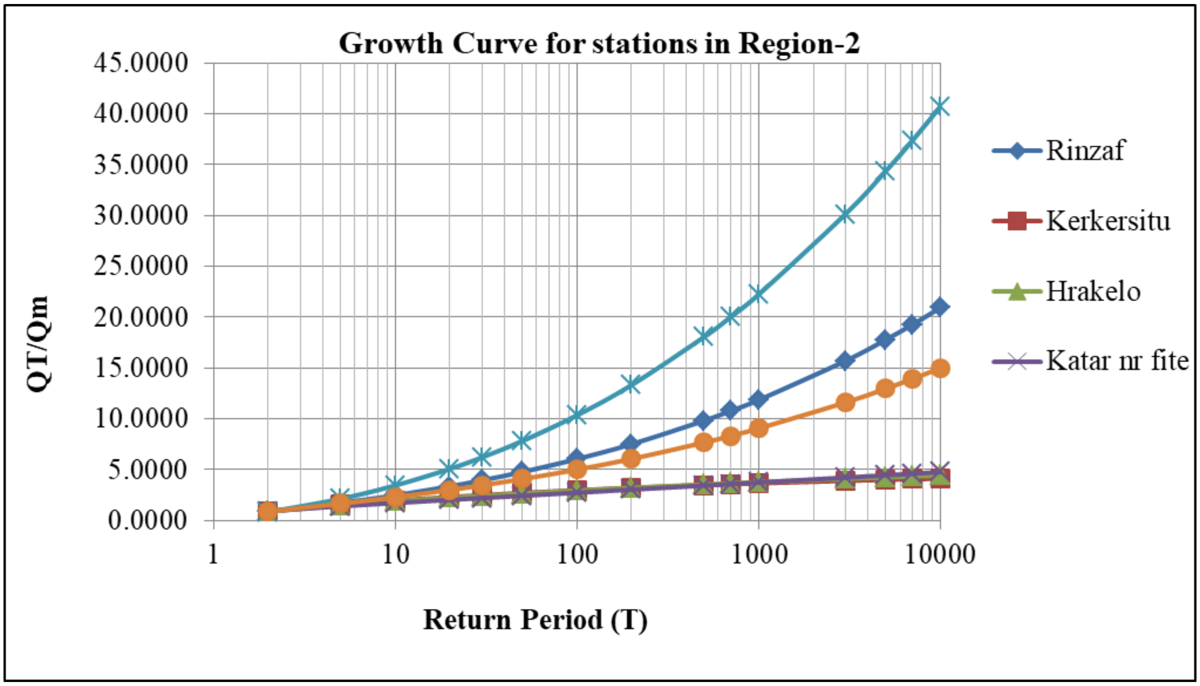


Figure 4-5:- Growth curve of each station in the region 2 using the Standardized quantile estimate

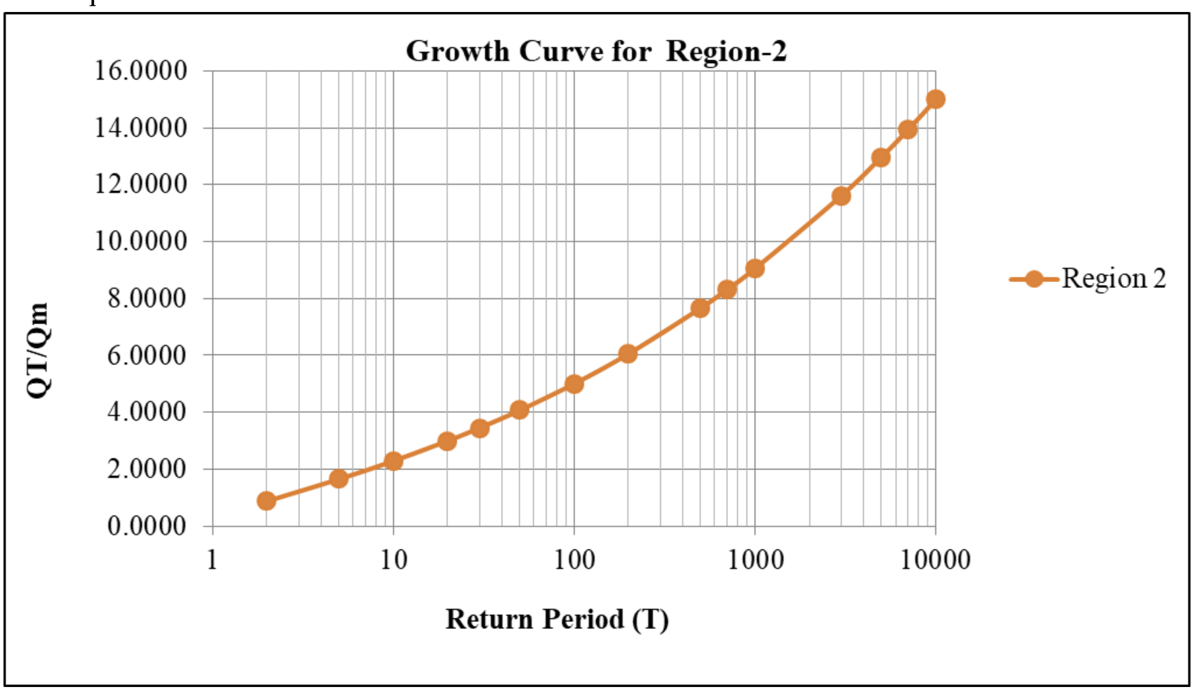


Figure 4-6:- Regional Growth Curve for Region Two

As shown in the figure 4.7 the growth factor for stations Rinzaf and Lower Timala in region 2 are very high and unexpected growth so, it must be rejected as an outlier. And the growth curve for region two become as shown below.

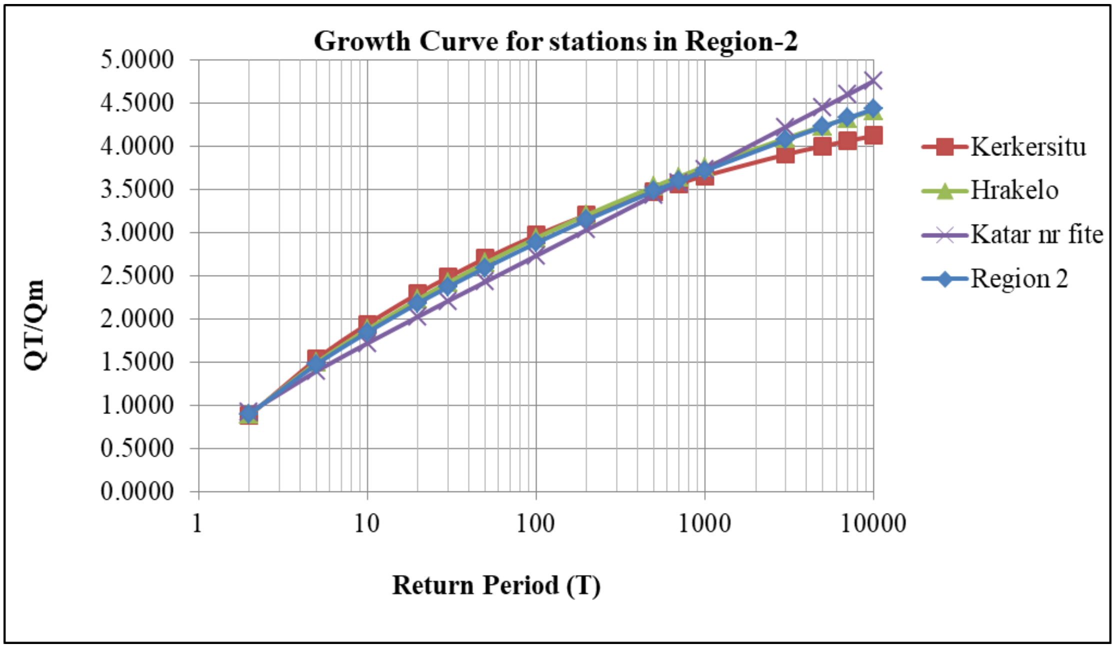


Figure 4-7 Reformed Regional Growth Curve for stations in Region Two

In this case all the stations have similar growth curve with reasonable growth factor

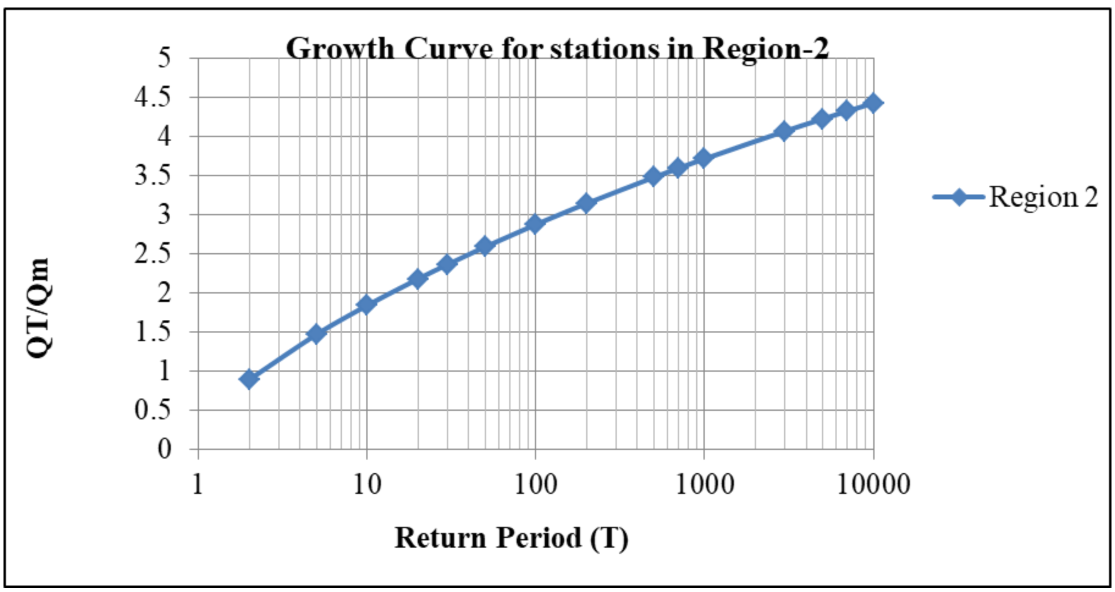
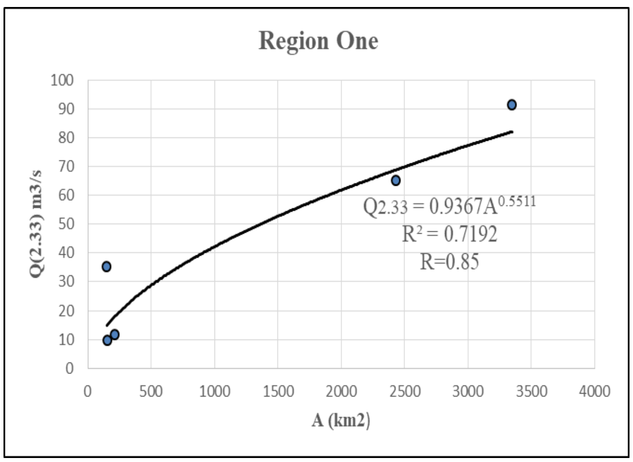


Figure 4-8 Reformed Regional Growth Curve for Region Two

4.5 Design Flood for the Ungagged Catchment

For each homogenous region two fundamental models were established: one allowing estimates of flood quantiles based on the index-flood Q2.33 develop regional frequency distribution curves, based on the gagged stations and the other allowing evaluating this index as a function of the watershed area for the ungagged catchment.

To allow estimates of the index-flood at ungagged watersheds a relationship between that index and the watershed area, A, was established for each homogenous region based on equation 3.46. The equations thus achieved for the two homogeneous regions as well as the pair of values (Q2.33; A). In the equation, Q2.33 is expressed in m3/s and A in km2, and r is the correlation coefficient of the linear regression analysis between logarithms of the index-flood and of the watershed area.



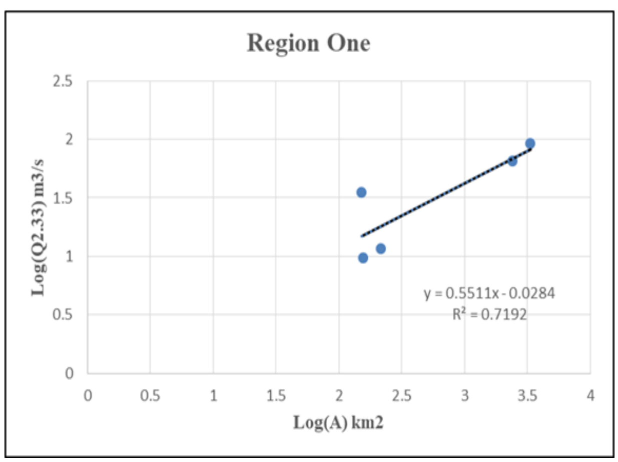
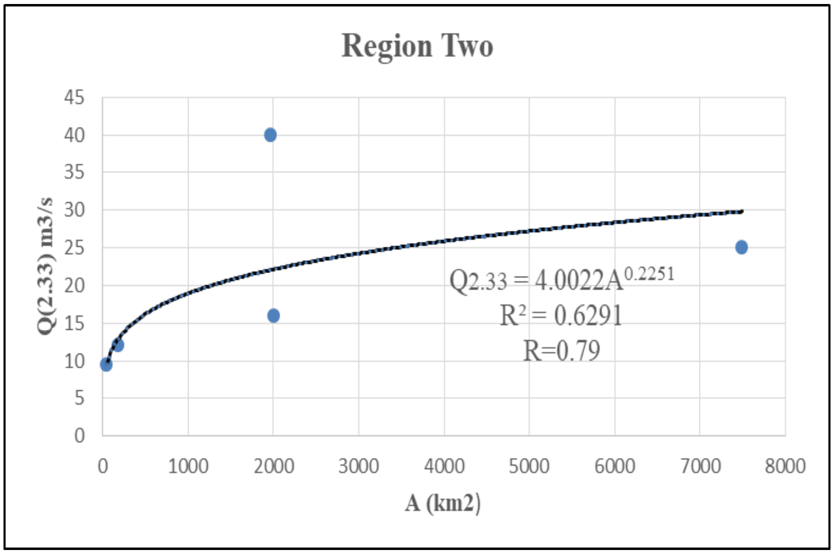


Figure 4-9 Best Selected Index Flood Estimation Equation for Region One



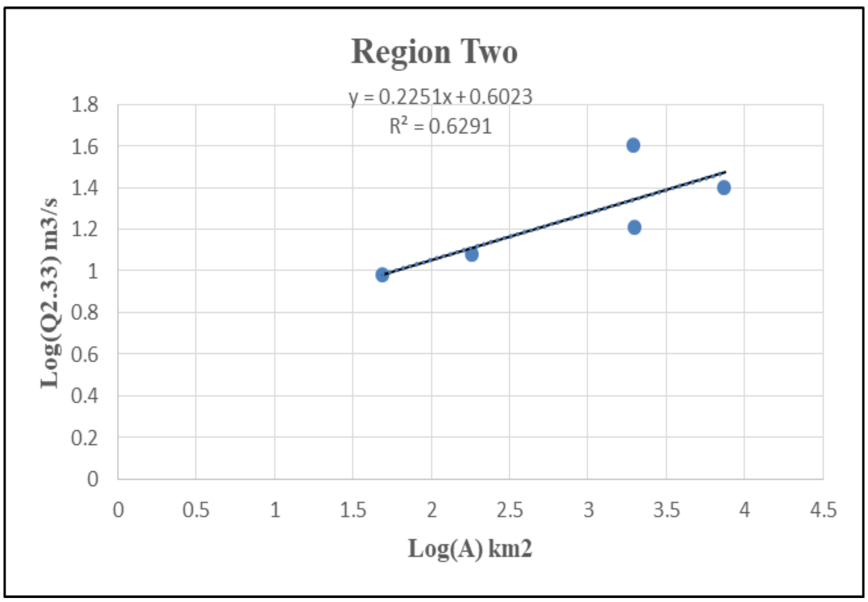


Figure 4-10:- Best Selected Index Flood Estimation Equation for Region Two

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

The study has established a regional flood frequency for Ziway-shala sub basin, rift valley basin Ethiopia using index flood method. The main objective of the study is to determine flood frequency curve for gagged station and to select best regional model for the ungagged site.

From the collected data for the study there are 23 gagged stations are found in the study area but only 10 stations selected for the analysis based on continuity and the length of record period. From the selected station the collected flow data for the analysis was checked for trend, independence, homogeneity and outlier.

The absence of trend test was done using Spearman's rank-correlation method and from the test value all the t_t value of stations are less than the critical value of t, at 5% significance level, for N-2 degrees of freedom so there is no trend in the data series.The independency test was done using the wald-wolfowitz (W-W) test and from the test value all the statistical value of stations are less than the critical value of a 5% significance level 1.96 which are accept the hypothesis of independency. The homogeneity test was done using SPSS software with Many-whitny (M-W) test was used to calculate the statistical value of the station. All the test result Z value are less than the critical value (1.96) and also the significant value P are greater than α value (0.05). Accordingly the results all the stations are homogeneous.

The test for outlier was also done using SPSS software with the Garubbs and Beck (1972) test (G-B) which used to detect outliers. The result shows that three stations (katar @ abura, Rinzaf and Lower timer) have outliers which have calculated value of Z greater than the critical value. Based on the result the outliers detected in the station are rejected from the record data before analysis.

Regionalization was made on the statistical value (LCS & LCK) of index flood of each station based on the concept that stations from the same regions, their index flood series came from the same parent of distribution. Therefore, Ziway-Shala sub basin has been classified into two regions which satisfied the homogeneity test.

Selection of parent distribution for the identified region was employed using Easy-fit statistical software. Goodness of fit test measure using the statistical test like kolmogorove simirnove test and chi-squared testes to identify how well the selected distribution fit to the data. In this study the possible parent distribution which select for the goodness of fit measure are Gamma, Gamma & Pearson III, Generalized Extreme value (GEV) Log Pearson Type III, Log normal(LN), Log normal parameter three(LN3) which are widely used for the research.

All the distribution statistical value of the Kolmogorove Simirnove and Chi squared test are less than the critical value of the test and also the P value of both testes are greater than the α value (0.05) which accept that the data fit the distribution except the stations Rinzaf and Lower Tamale which is not acceptable for the Gamma & Pearson III distribution.

For this study among the fitted distribution Generalized Extreme Value distribution is selected for region one and Log Pearson Type III distribution is selected for region two based on statistical value, simple to apply and widely acceptance. For the selected distributions the statistical parameter estimated by Methods of Moment.

The quantiles of each station in the regions are estimated for the given parameters corresponding to the return period (2-1000 year). Based on the estimated quantiles the growth curve has been developed by standardizing the quantiles using the index flood method for stations and regions in the sub basins.

Finally regression analysis has been applied to develop regression model to predicted mean annual flood flow from ungagged catchment using catchment characteristics. Accordingly the Ziway - Shala sub basin the index flood estimation equation were developed with a coefficient of determination (R^2) 0.72 for region one and 0.63 for region two.

5.2 Recommendations

- The regional growth curve and the index flood value estimation equation (model) obtained so far from the study result can be used as inputs for design of hydraulic structures on the ungauged catchment of Ziway-Shala sub basin.
- In order to get reliable estimate of regional quantile more hydrometric stations should be installed in the basin and for the available stations there should be high recording quality.
- In regression analysis the method of dividing data into two or more parts depending on the governing independent variable, such as area, is the most appreciable method to improve the coefficient of determination.
- Due to the evidence of future climate change, further analyses are needed to understand the effects of climatic variables on the variability of L-moments of annual maximum floods in the study area.
- The flood frequency curve can be affected by the land use land cover change of the areas. In the catchments considered the land degradation and plantation time variations couldn't be considered due to absence of different period land use maps. Hence, future research work flood frequency setting and hydrological data analysis should consider the effect of time variant land use land cover change in river flow or peak flow variations.
- In places where there is significant slop variation such as the Ethiopian rift where the study area is located the gauging station locations should take the effect of morphology variations in river flow curves and if any additional gauging station locations should be constructed the morphology of the area and associated effect on peak river flow curves should be considered.
- In the study area there is uncontrolled and high water abstraction from rivers mostly using open or unlined canals in variably permeable area. Any hydrology analysis or flood frequency curve setting work should consider the abstraction and loss of water in open canals. But to do so the abstraction should be controlled (measured), and should consider permeability along open canals. Such works will have considerable effect in flood frequency setting, design of structures; flood management works in our study area at least to set acceptable canal abstraction

from rivers, and acceptable minimum flow in rivers for down user community, and should be also part of optimum water use planning in the basins and flood frequency setting works in the future.

- Maintenance in river gauge stations is usually done after long time is spent without data recording in the stations for various reasons, or lower part of graduated gauging staffs can be covered by silitation, and free flow of water near gauging stations is sometimes blocked by boulders, vegetation cover or any other effect. In such cases the flow volume recording at such stations can be increased and flow recording exaggerated and will not exactly reflect the real flow volume. Hence, concerned bodies should supervise the conditions near stations and corrective measures should be taken to obtain real flow data of rivers in the gauging stations and for any flood frequency curve setting or any hydrological data analysis of catchments.
- As the different development practices such as the different scale irrigations and some the industrial use of water should be regulated based on the general available basin water resource potential which should consider acceptable minimum dry time flow in rivers and setting of flood frequency curve in the area for future sustainable use among other considerations such as full basin groundwater and surface water analysis and setting functional water resource management plan in the area.

REFERENCES

1. Admasu, Regional flood frequency analysis, PHD thesis report submitted to Stockholm University. (1989).
2. Binaya Kumar MISHRA, Kaoru TAKARA, Yosuke YAMASHIKI and Yasuto TACHIKAWA.: An Assessment of predictive accuracy for two regional flood-frequency estimation methods annual journal of hydraulic engineering, jsce, vol.54, 2010, February.
3. Caroline Le Turdua The Ziway–Shala lake basin system, Main Ethiopian Rift: Influence of volcanism, tectonics, and climatic forcing on basin formation and sedimentation, Journal of Palaeogeography, Palaeoclimatology, Palaeoecology 150 (1999) 135–177
4. Cunnane, C.: “Methods and Merits of Regional Flood Frequency Analysis”, Journal Hydrology, Vol. 100, pp. 269-290. (1988).
5. Cunnane, C.: “Statistical Distributions for Flood Frequency Analysis”. Word Meteorological Organization Operational Hydrology, Report No. 33, WMO-No. 718, Geneva, Switzerland, (1989).
6. E.R. Dahmen & M.J. Hall, Screening of Hydrological Data: International Institute for Land Reclamation and Improvement/ILRI Wageningen The Netherlands,l990
7. Ethiopian Water Resources Potential & Development of Ethiopia presentation on Ministry of Water and Energy I IWW NWP October 31st – November 4, 2011, the Netherlands
8. Hosking, J.R.M. and Wallis, J.R. (1993). Regional frequency analysis an approach based on L moments, published in the United States of America by cambridge university press, 1997.
9. Kossi Komi, Barnabas A. Amisigo , Bernd Diekkrüger ,† and Fabien C. C. Hountondji : Regional Flood Frequency Analysis in the Volta River Basin, West Africa; journal hydrology 03-00005, 2016 ,February.
10. Mengistu D, Regional flood frequency analysis for upper awash sub- basin (Upstream of koka), MSc thesis submitted to Addis Ababa University, (2008)

11. M. M. Portela & A. T. Dias: Application of the index-flood method to the regionalization of flood peak discharges on the Portugal mainland journal W IT Transactions on Ecology and the Environment, Vol 83, 2005
12. Mulugeta Dadi Belete, Bernd Diekkrüger and Jackson Roehrig : Characterization of Water Level Variability of the Main Ethiopian Rift Valley Lakes; journal hydrology 03-00001-v2, 2015 ,December.
13. National Research Council Committee on Techniques for Estimating Probabilities of Extreme Floods, "Estimating Probabilities of Extreme Floods, Methods and Recommended Research", National Academy Press, Washington, D.C. (1988)
14. Rao, A.R. & Hamed, K.H: Flood Frequency Analysis. Florida: CRC press LLC. (2000).
15. Rift Valley Lakes Basin Integrated Resources Development Master Plan Study Project, Volume 1: (JICA study 2007)
16. Tamiru Gedefa: Regional Flood Frequency analysis for upper Omo-Gibe sub-basin, MSc thesis submitted to Addis Ababa University, (2009).
17. V.T. Chow, D.R. Maidment, and L.W. Mays: Applied Hydrology .Singapore, McGRAW, (1988).
18. Yang Lu: Regional Flood Frequency Analysis for Newfoundland and Labrador Using the L-Moments Index-Flood Method; A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Engineering Faculty of Engineering and Applied Science Memorial University of Newfoundland, 2016, May.

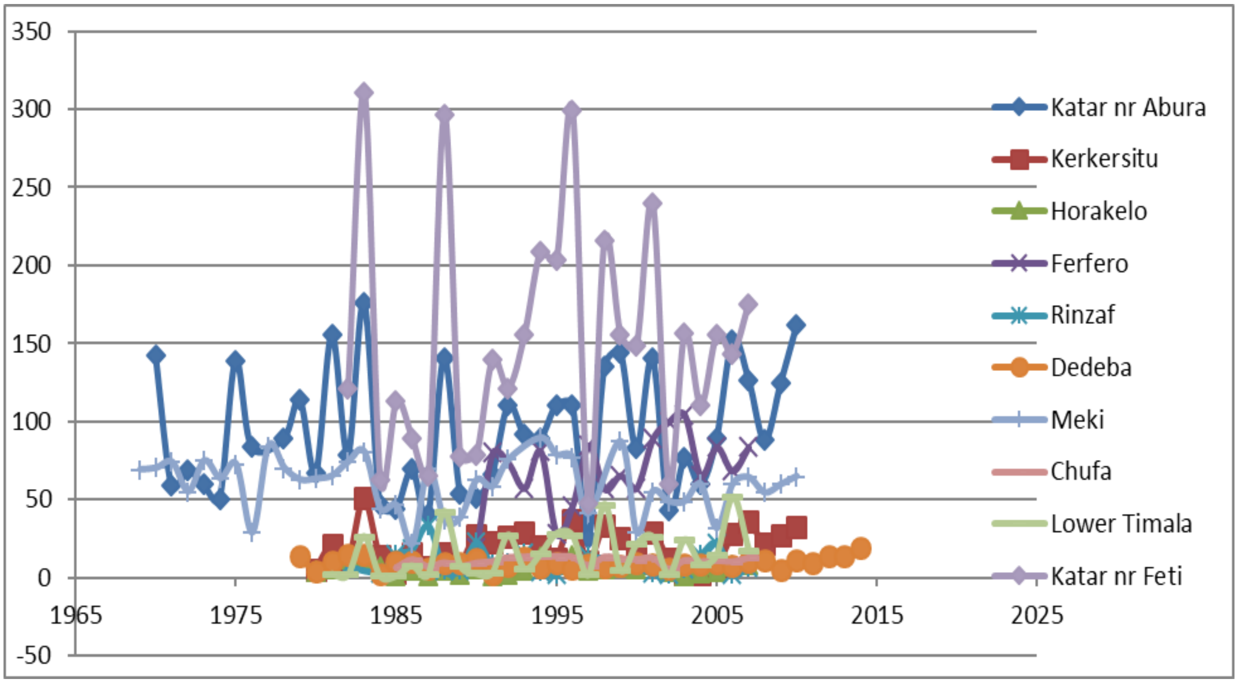
APPENDIXES

Appendix A: Annual Maximum Flood of stations in Ziway- Shala Sub Basin

Dedeba nr kuyera		Kerkersitu nr Adamitulu		Horakelo nr Langano		Lower Timala nr Sagure		Katar nr Fite	
Year	Flow (m3/s)	Year	Flow (m3/s)	Year	Flow (m3/s)	Year	Flow (m3/s)	Year	Flow (m3/s)
1979	13.552	1980	5.339	1980	3.855	1981	2.792	1982	121.375
1980	3.797	1981	20.862	1981	6.326	1982	2.092	1983	310.478
1981	10.693	1982	9.623	1982	12.199	1983	26.558	1984	62.345
1982	14.820	1983	50.800	1983	11.374	1984	1.570	1985	112.888
1983	14.820	1984	14.370	1984	7.061	1985	1.444	1986	89.333
1984	2.001	1985	3.465	1985	1.448	1986	8.115	1987	65.444
1985	10.693	1986	15.343	1986	5.448	1987	2.358	1988	296.567
1986	10.431	1987	6.840	1987	1.636	1988	42.471	1989	77.477
1987	6.394	1988	15.590	1988	5.487	1989	8.115	1990	78.624
1988	9.181	1989	9.019	1989	2.383	1990	3.122	1991	139.311
1989	9.669	1990	27.377	1990	5.568	1991	3.497	1992	121.375
1990	11.785	1991	22.515	1991	1.133	1992	27.560	1993	155.238
1991	2.484	1992	26.441	1992	2.524	1993	6.189	1994	208.814
1992	8.032	1993	28.958	1993	4.827	1994	15.462	1995	203.343
1993	13.246	1994	19.788	1994	13.360	1995	28.380	1996	298.862
1994	6.585	1995	12.577	1995	8.449	1996	27.289	1997	47.393
1995	9.181	1996	36.995	1996	13.519	1997	2.203	1998	216.114
1996	6.021	1997	10.470	1997	4.672	1998	46.969	1999	155.238
1997	8.710	1998	33.423	1998	6.371	1999	5.090	2000	148.760
1998	7.088	1999	25.228	1999	7.021	2000	21.672	2001	240.097
1999	7.446	2000	16.601	2000	5.784	2001	26.418	2002	59.899
2000	9.614	2001	28.744	2001	8.423	2002	2.238	2003	156.371
2001	8.586	2002	11.896	2002	4.545	2003	24.655	2004	110.129
2002	5.582	2003	2.621	2003	1.258	2004	8.978	2005	155.428
2003	8.348	2004	2.867	2004	3.461	2005	15.057	2006	143.657
2004	8.348	2005	6.963	2005	4.113	2006	51.695	2007	175.534
2005	9.878	2006	28.228	2006	7.499	2007	17.777		
2006	7.989	2007	35.687	2007	8.686				
2007	10.491	2008	21.954						
2008	11.559	2009	27.401						
2009	5.023	2010	32.847						
2010	11.125								
2011	9.679								
2012	14.373								
2013	14.373								
2014	19.573								

Meki @ Meki river		Katar @ Abura		Chufa nr Arata		Ferfero nr Wulbareg		Rinzaf @ Butajera	
Year	**Flow (m3/s)**	**Year**	**Flow (m3/s)**	**Year**	**Flow (m3/s)**	**Year**	**Flow (m3/s)**	**Year**	**Flow (m3/s)**
1969	69.126	1970	142.649	1985	6.616	1990	19.138	1982	6.978
1970	70.200	1971	59.028	1986	11.530	1991	79.953	1984	3.563
1971	73.877	1972	68.494	1987	7.233	1992	73.490	1985	15.087
1972	55.095	1973	60.032	1988	9.687	1993	57.571	1986	18.320
1973	74.909	1974	50.507	1989	7.550	1994	81.104	1987	33.587
1974	63.576	1975	139.010	1990	9.452	1995	28.436	1988	4.854
1975	72.400	1976	83.944	1991	10.475	1996	46.060	1989	4.883
1976	28.829	1978	89.143	1992	12.958	1997	84.716	1990	22.564
1977	84.143	1979	114.106	1993	13.255	1998	56.988	1991	8.592
1978	70.163	1980	67.394	1994	13.774	1999	65.651	1992	9.557
1979	62.617	1981	155.903	1995	13.484	2000	57.695	1993	15.455
1980	63.336	1982	78.958	1996	13.198	2001	88.795	1994	4.585
1981	65.266	1983	176.278	1997	8.040	2002	99.829	1995	2.278
1982	74.405	1984	47.009	1998	13.175	2003	104.366	1997	13.663
1983	80.515	1985	43.681	1999	12.099	2004	65.699	2001	3.167
1984	44.106	1986	69.606	2000	11.661	2005	83.738	2002	2.697
1985	46.315	1987	40.520	2001	12.529	2006	68.530	2003	5.620
1986	21.501	1988	140.822	2002	5.888	2007	83.738	2004	14.720
1987	69.922	1989	54.181	2003	10.994			2005	22.284
1988	36.640	1990	51.252	2004	10.321			2006	2.810
1989	38.412	1991	66.306	2005	10.994			2007	6.599
1990	62.475	1992	110.624	2006	10.321				
1991	58.339	1993	91.824	2007	12.715				
1992	75.517	1994	89.143						
1994	89.432	1995	110.624						
1995	78.661	1996	110.624						
1996	76.951	1997	25.898						
1997	41.020	1998	135.436						
1999	87.092	1999	144.493						
2000	28.923	2000	82.678						
2001	55.133	2001	140.822						
2002	48.776	2002	42.876						
2003	48.759	2003	76.543						
2004	59.754	2004	60.032						
2005	31.739	2005	89.354						
2006	59.754	2006	152.033						
2007	64.193	2007	126.779						
2008	54.395	2008	87.978						
2009	59.603	2009	124.906						
2010	64.810	2010	161.835						

Appendix B: Graph of Annual Flow Series



Appendix C: Homogeneity test Results using SPSS Software

⇒ **Homogeneity test result for station Kerkersitu near Adamitulu**

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
Flow	1	15	14.53	218.00
	2	16	17.38	278.00
	Total	31		

Test Statistics[a]

	Flow
Mann-Whitney U	98.000
Wilcoxon W	218.000
Z	-.870
Asymp. Sig. (2-tailed)	.385
Exact Sig. [2*(1-tailed Sig.)]	.401[b]

a. Grouping Variable: Group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Horakelo near Langano**

Mann-Whitney Test

Ranks

	group	N	Mean Rank	Sum of Ranks
data	1	13	12.08	157.00
	2	15	16.60	249.00
	Total	28		

Test Statistics[a]

	data
Mann-Whitney U	66.000
Wilcoxon W	157.000
Z	-1.451
Asymp. Sig. (2-tailed)	.147
Exact Sig. [2*(1-tailed Sig.)]	.156[b]

a. Grouping Variable: group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Ferfero near Wulbareg**

Mann-Whitney Test

Ranks

	group	N	Mean Rank	Sum of Ranks
Data	group 1	8	7.38	59.00
	group 2	10	11.20	112.00
	Total	18		

Test Statistics[a]

	Data
Mann-Whitney U	23.000
Wilcoxon W	59.000
Z	-1.511
Asymp. Sig. (2-tailed)	.131
Exact Sig. [2*(1-tailed Sig.)]	.146[b]

a. Grouping Variable: group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Rinzaf @ Butajera**

Mann-Whitney Test

Ranks

	group	N	Mean Rank	Sum of Ranks
Data	group 1	10	13.10	131.00
	group 2	12	10.17	122.00
	Total	22		

Test Statistics[a]

	Data
Mann-Whitney U	44.000
Wilcoxon W	122.000
Z	-1.055
Asymp. Sig. (2-tailed)	.291
Exact Sig. [2*(1-tailed Sig.)]	.314[b]

a. Grouping Variable: group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Dedeba near Kuyera**

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
Flow	1	17	18.88	321.00
	2	19	18.16	345.00
	Total	36		

Test Statistics[a]

	Flow
Mann-Whitney U	155.000
Wilcoxon W	345.000
Z	-.206
Asymp. Sig. (2-tailed)	.837
Exact Sig. [2*(1-tailed Sig.)]	.851[b]

a. Grouping Variable: Group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Meki @ Meki river**

Mann-Whitney Test

Ranks

	group	N	Mean Rank	Sum of Ranks
data	1	19	21.42	407.00
	2	21	19.67	413.00
	Total	40		

Test Statistics[a]

	data
Mann-Whitney U	182.000
Wilcoxon W	413.000
Z	-.474
Asymp. Sig. (2-tailed)	.636
Exact Sig. [2*(1-tailed Sig.)]	.649[b]

a. Grouping Variable: group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Chufa near Arata**

Mann-Whitney Test

Ranks

	group	N	Mean Rank	Sum of Ranks
Data	group 1	11	11.73	129.00
	group 2	12	12.25	147.00
	Total	23		

Test Statistics[a]

	Data
Mann-Whitney U	63.000
Wilcoxon W	129.000
Z	-.185
Asymp. Sig. (2-tailed)	.853
Exact Sig. [2*(1-tailed Sig.)]	.880[b]

a. Grouping Variable: group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Lower Timala near Sagure**

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
Flow	1	12	10.83	130.00
	2	15	16.53	248.00
	Total	27		

Test Statistics[a]

	Flow
Mann-Whitney U	52.000
Wilcoxon W	130.000
Z	-1.854
Asymp. Sig. (2-tailed)	.064
Exact Sig. [2*(1-tailed Sig.)]	.067[b]

a. Grouping Variable: Group

b. Not corrected for ties.

⇒ **Homogeneity test result for station Katar near Fite**

Mann-Whitney Test

Ranks

	Group	N	Mean Rank	Sum of Ranks
data	1	12	11.04	132.50
	2	14	15.61	218.50
	Total	26		

Test Statistics[a]

	data
Mann-Whitney U	54.500
Wilcoxon W	132.500
Z	-1.518
Asymp. Sig. (2-tailed)	.129
Exact Sig. [2*(1-tailed Sig.)]	.131[b]

a. Grouping Variable: Group

b. Not corrected for ties.

Appendix D: The result of Kolmogorov and Chi-Squared test for stations in Region 1

Kolmogorov test

Station	Distribution	Statistics	P-value	α-value	Critical Value	Remark!
Meki	GEV	0.063	0.994	0.050	0.210	Accept
	LP3	0.089	0.878	0.050	0.210	Accept
	LN(3P)	0.110	0.678	0.050	0.210	Accept
	Gamma (3P)	0.131	0.462	0.050	0.210	Accept
	Gamma	0.147	0.324	0.050	0.210	Accept
	LN	0.169	0.181	0.050	0.210	Accept
Katar nr Abura	GEV	0.093	0.850	0.050	0.210	Accept
	LP3	0.096	0.819	0.050	0.210	Accept
	LN(3P)	0.102	0.761	0.050	0.210	Accept
	Gamma (3P)	0.103	0.751	0.050	0.210	Accept
	Gamma	0.104	0.746	0.050	0.210	Accept
	LN	0.115	0.627	0.050	0.210	Accept
Chufa	GEV	0.079	0.996	0.050	0.275	Accept
	LP3	0.108	0.926	0.050	0.275	Accept
	LN(3P)	0.125	0.822	0.050	0.275	Accept
	Gamma (3P)	0.132	0.769	0.050	0.275	Accept
	Gamma	0.147	0.651	0.050	0.275	Accept
	LN	0.168	0.483	0.050	0.275	Accept
Ferfero	GEV	0.119	0.933	0.050	0.309	Accept
	LN(3P)	0.132	0.872	0.050	0.309	Accept
	LP3	0.139	0.831	0.050	0.309	Accept
	Gamma (3P)	0.147	0.777	0.050	0.309	Accept
	Gamma	0.157	0.707	0.050	0.309	Accept
	LN	0.218	0.313	0.050	0.309	Accept
Dedeba	LN(3P)	0.077	0.973	0.050	0.221	Accept
	Gamma (3P)	0.077	0.972	0.050	0.221	Accept
	GEV	0.081	0.957	0.050	0.221	Accept
	Gamma	0.106	0.773	0.050	0.221	Accept
	LP3	0.113	0.701	0.050	0.221	Accept
	LN	0.155	0.320	0.050	0.221	Accept

Chi-squared test

Station	Distribution	Statistics	P-value	α-value	Critical Value	Remark!
Meki	GEV	1.587	0.903	0.050	11.070	Accept
	LP3	1.702	0.790	0.050	9.488	Accept
	Gamma	4.060	0.398	0.050	9.488	Accept
	Gamma (3P)	4.542	0.338	0.050	9.488	Accept
	LN	4.865	0.301	0.050	9.488	Accept
	LN(3P)	5.050	0.282	0.050	9.488	Accept
Katar nr Abura	LN	0.999	0.801	0.050	7.815	Accept
	LP3	2.693	0.611	0.050	9.488	Accept
	LN(3P)	2.783	0.595	0.050	9.488	Accept
	Gamma (3P)	2.783	0.595	0.050	9.488	Accept
	Gamma	2.954	0.566	0.050	9.488	Accept
	GEV	3.746	0.442	0.050	9.488	Accept
Chufa	LP3	0.184	0.912	0.050	5.992	Accept
	LN(3P)	0.254	0.881	0.050	5.992	Accept
	GEV	1.091	0.779	0.050	7.815	Accept
	LN	1.655	0.437	0.050	5.992	Accept
	Gamma (3P)	1.802	0.614	0.050	7.815	Accept
	Gamma	1.866	0.393	0.050	5.992	Accept
Ferfero	GEV	0.077	0.782	0.050	3.842	Accept
	LP3	0.113	0.737	0.050	3.842	Accept
	LN(3P)	0.180	0.672	0.050	3.842	Accept
	LN	0.479	0.489	0.050	3.842	Accept
	Gamma	0.539	0.764	0.050	5.992	Accept
	Gamma (3P)	0.774	0.679	0.050	5.992	Accept
Dedeba	GEV	1.083	0.781	0.050	7.815	Accept
	Gamma (3P)	1.106	0.776	0.050	7.815	Accept
	LN(3P)	1.107	0.775	0.050	7.815	Accept
	Gamma	1.236	0.941	0.050	11.070	Accept
	LN	2.248	0.690	0.050	9.488	Accept
	LP3	2.267	0.650	0.050	9.488	Accept

Appendix E: The result of Kolmogorov and Chi-Squared test for stations in Region 2

Kolmogorov test

Station	Distribution	Statistics	P-value	α-value	Critical Value	Remark!
Rinzaf	Gamma	0.110	0.939	0.050	0.287	Accept
	GEV	0.133	0.805	0.050	0.287	Accept
	LP3	0.139	0.759	0.050	0.287	Accept
	LN	0.142	0.743	0.050	0.287	Accept
	LN(3P)	0.148	0.690	0.050	0.287	Accept
	Gamma (3P)	0.169	0.530	0.050	0.287	Accept
Kerkersitu	GEV	0.086	0.962	0.050	0.238	Accept
	LP3	0.092	0.935	0.050	0.238	Accept
	LN(3P)	0.108	0.822	0.050	0.238	Accept
	Gamma (3P)	0.115	0.763	0.050	0.238	Accept
	Gamma	0.122	0.703	0.050	0.238	Accept
	LN	0.139	0.541	0.050	0.238	Accept
Horakelo	GEV	0.066	0.999	0.050	0.250	Accept
	LN(3P)	0.077	0.991	0.050	0.250	Accept
	LP3	0.088	0.968	0.050	0.250	Accept
	Gamma	0.091	0.960	0.050	0.250	Accept
	Gamma (3P)	0.119	0.781	0.050	0.250	Accept
	LN	0.131	0.677	0.050	0.250	Accept
Katar nr Fite	LP3	0.092	0.967	0.050	0.259	Accept
	LN(3P)	0.095	0.957	0.050	0.259	Accept
	GEV	0.102	0.924	0.050	0.259	Accept
	Gamma	0.104	0.915	0.050	0.259	Accept
	LN	0.105	0.905	0.050	0.259	Accept
	Gamma (3P)	0.113	0.860	0.050	0.259	Accept
Lower Timala	LP3	0.137	0.639	0.050	0.254	Accept
	GEV	0.144	0.584	0.050	0.254	Accept
	LN	0.148	0.545	0.050	0.254	Accept
	Gamma (3P)	0.149	0.538	0.050	0.254	Accept
	Gamma	0.165	0.410	0.050	0.254	Accept
	LN(3P)	0.172	0.359	0.050	0.254	Accept

Chi-Squared test

Station	Distribution	Statistics	P-value	α-value	Critical Value	Remark!
Rinzaf	LP3	0.091	0.955	0.050	5.992	Accept
	LN	0.465	0.793	0.050	5.992	Accept
	Gamma	0.481	0.786	0.050	5.992	Accept
	LN(3P)	0.660	0.719	0.050	5.992	Accept
	GEV	1.611	0.447	0.050	5.992	Accept
	Gamma (3P)	N/A	N/A	N/A	N/A	
Kerkersitu	GEV	0.467	0.926	0.050	7.815	Accept
	LP3	0.937	0.816	0.050	7.815	Accept
	LN(3P)	0.947	0.814	0.050	7.815	Accept
	LN	2.104	0.551	0.050	7.815	Accept
	Gamma (3P)	2.219	0.528	0.050	7.815	Accept
	Gamma	2.229	0.526	0.050	7.815	Accept
Horakelo	LP3	0.059	0.996	0.050	7.815	Accept
	LN(3P)	0.110	0.991	0.050	7.815	Accept
	Gamma	0.119	0.989	0.050	7.815	Accept
	GEV	0.219	0.994	0.050	9.488	Accept
	Gamma (3P)	0.346	0.951	0.050	7.815	Accept
	LN	1.216	0.749	0.050	7.815	Accept
Katar nr Fite	GEV	1.029	0.598	0.050	5.992	Accept
	Gamma	1.089	0.580	0.050	5.992	Accept
	LN	1.213	0.750	0.050	7.815	Accept
	LP3	1.264	0.738	0.050	7.815	Accept
	Gamma (3P)	1.281	0.734	0.050	7.815	Accept
	LN(3P)	1.320	0.724	0.050	7.815	Accept
Lower Timala	LN(3P)	1.126	0.570	0.050	5.992	Accept
	GEV	3.157	0.206	0.050	5.992	Accept
	LP3	3.499	0.174	0.050	5.992	Accept
	LN	3.696	0.158	0.050	5.992	Accept
	Gamma	4.175	0.124	0.050	5.992	Accept
	Gamma (3P)	N/A	N/A	N/A	N/A	