REGIONAL FLOOD FREQUENCY ANALYSIS FOR UPPER AWASH SUB- BASIN (UPSTREAM OF KOKA)

By

Mengistu Demissie

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science (Engineering) of the Addis Ababa University

Addis Ababa University

May 2008

CERTIFICATION

The undersigned certify that they have read the dissertation entitled **REGIONAL FLOOD FREQUENCY ANALYSIS FOR UPPER AWASH SUB-BASIN (U/S OF KOKA)** and hereby recommend for acceptance by the Addis Ababa University in partial fulfillment of the requirements for the degree of Master of Science (Engineering).

Dr. Yonas Michael
(Advisor)

Date ----------------------------------

Dr. Yilma Seleshi
(Internal examiner)

Date---------------------------------

Deksyos Tarekegn
(Co-Advisor)

Date ------------------------------------

Dr. Mekuria
(External examiner)

Date--------------------------------

DECLARATION AND COPYRIGHT

Mengistu Demissie Bekele, declare that this dissertation is my own original work and that it has not been presented and will not be presented by me to any other University for similar or any other degree award.

Signature: ---------------------------------------

ACKNOWLEDGEMENT

I would like to express my wholehearted gratitude to my advisors, Dr. Yonas Michael, for his priceless support in supervising my work and providing me with important reference materials including software and Ato Deksyos Tarekegne, for his invaluable guidance, constructive comments and his encouragement from early age of my study.

I would like to acknowledge the Addis Ababa University for the financial support made through the postgraduate program of the University, which is available to me during my research time. My sincere appreciation goes to all staff members of Ministry of Water Resource, particularly Hydrology department, GIS department, Halcraw consultants and Library of MoWR, for their appreciable support in providing me hydrological data, digitized map and other reference materials respectively.

The National Meteorological Service Agency of Ethiopia also helped me in providing rainfall and climatic data of the sub-basin. Therefore, they are gratefully acknowledged.

Finally, I wish to extend my gratitude to all staff members of Addis Ababa University and my friends, who helped me one way or the other, in carrying out my research through remarkable encouragement, advice, material support and collaboration in every aspect.

TABLE OF CONTENT

PAGE

3. METHODOLOGY AND PROCEDURE

4. DATA COLLECTION AND ANALYSIS OF THE SUB- BASIN

5. RESULTS AND DISCUSSION

List of tables

List of figures

Annexes

Acronyms

AM	Annual Maximum
CV	Coefficient of variation
D/s	Down stream
EV1	extreme value type one
EMA	Ethiopia mapping authority
FFA	Flood frequency analysis
GIS	Geographical information system
GEV	General extreme value
LN	Log normal
MAF	Maximum annual flood
MAR	Maximum annual rainfall
MOM	Method of moment
MoWR	Ministry of water resource
N	Number of record year
PD	Partial duration
PWM	Probability weighted method
QT	Quantile corresponding to the return period
RFFA	Regional flood frequency analysis
SEE	Standard error estimation
T	Return period (Recurrence interval)
U/s	Up stream
WMO	World meteorological organization
Cs	Coefficient of skewness
Ck	Coefficient kurtosis

Abstract

The conventional and L-moment method of flood frequency analysis are applied in determining flood magnitude of defined return periods by selecting the best-fit theoretical probability distribution. The most important part of the distribution is the tail as far as extreme flooding phenomena. These phenomena of floods are of utmost concern in water resources development and management. In most cases the central part of the theoretical distribution fits satisfactorily with the empirical points. In FFA the objective is to estimate flood magnitude (Q) corresponding to any specified recurrence interval of (T) years. The estimation is complicated due to lack of a physical basis for determining the form of the underlying flood frequency distribution and the necessity of evaluating flood event for return periods that exceed the observation period (Leulseged, 2002).

Regionalization is generally accepted term to explain the transfer of information about flood peak at one catchment derived from other catchments with similar characteristics. The advantage of such procedure is particularly great in the estimation of frequencies for higher flood magnitudes with limited at site data do exist and inference in the tail of probability distributions makes the stabilization of the estimators difficult. It is quite clear that regionalization is most viable way of improving flood quantile estimation. Although there remain researchable topics in development and application of regionalization method. The performance of regional Extreme value and LN distributions are found to be highly satisfactory and can be widely applied in this paper. More attention is given to at-site homogeneity test to group stations in the upper Awash sub-basin (u/s koka) in to two regions after checking them for the consistency and independency testes and estimation of standard error. The upper Awash sub-basin (u/s of koka) has 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. An Extreme value EV1, GEV and Lognormal LN_2, LN_3 distributions are selected as the best fit distribution for the stations in the sub-basin. The sub-basin has been divided into two regions the upper and lower region, the upper region covers 18.46% of the sub-basin and the lower region covers 81.54% of the sub-basin.

In order to improve the estimation of the Q-T relationship, the need to use at-site and regional information arises, so that stabilizing site specific estimates based on limited data can be handled. The RFFA procedure involves the definition and identification of homogeneous region based on the given criteria. In the present the application of index-flood for at-site and regional method of FFA are considered as one of the tools in overcoming problems of ungauged catchments and a stream having small size of observation (n). In this thesis application of index flood is only for the derivation of standardized flood that help in derivation of growth curve for station and regions in the sub- basin.

The use of regional and at-site information to estimate flood magnitude for a catchments with little or no observed data has become increasingly important since many projects which require design flood information are located in areas where observed flood data are either missing or inadequate. Regional analysis consists of analyzing the record of all gauged sites in a hydrologically homogeneous region and stations, in order to be able to use or transfer information contained in the record of many sites to estimate quantiles at any individual gauged or ungauged catchments in the region.

Hosking and Wallis (1993) have discussed various aspects of regional frequency analysis such as identification of homogeneous region and describing the different steps of regional analysis. In the present application, the discharge return period (Q-T) relationships for all sites as obtained from extreme value and lognormal distribution analysis that were plotted together with the discharge being expressed in dimensionless or standardized form.

1. INTRODUCTION

1.1 Background

Flood frequency analysis provides vital information for design of an economic appraisal of a variety of engineering and water resource planning and development projects. Frequency analysis of flood is a very active area of investigation in statistical hydrology. Various distributions, method of parameters estimation, quantiles estimation, problems related to regionalization, at-site and other related topics are being investigated. The analysis involves estimation of a flood magnitude corresponding to a required return period or probability of exceedence.

In flood frequency analysis the objective is to estimate a flood magnitude corresponding to any required return period of occurrence. The resulting relation ship between magnitude and return period is referred to as the Q-T relation ship. Return period, T, may be defined as the time interval for which a particular flood having magnitude Q_T (also known as quantiles) is expected to be exceeded (Admasu 1989). Return period is also referred to as recurrence interval. The magnitude of flood is inversely related to their frequency of occurrence, high floods occurring less frequently than moderate flood.

The use of regional information to estimate flood magnitude at site with little or no observed data has become increasingly important because many projects that require design flood information are located in areas where observed flood data are either missing or inadequate. In the analysis attention must be given to the at-site data since they are the bases for regional information. (Admasu 1989)

RFFA can facilitate the estimation of the Q_T value at a location for which limited flow data exists using the growth curve. Another advantage of at-site and regional method is that an unusual event that occurs in a small number of basins can be taken into account over a wide area where they might have occurred. Some At-site and RFFA methods assume that a region is set of gauging sites whose flood producing behavior is homogeneous in some quantifiable manner. 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). Therefore, many researchers have used RFFA for estimating flood flow of various return periods. In addition it is to be expected that the more homogeneous a region the greater the gain in using regional instead of at site estimation.

1.2 Statement of the problem

The main feature of a flood, from the water management point of view, is its interference with human activities. The interference is measured in terms of actual and potential economic losses and danger to human life. The purpose of flood analysis is to assess the magnitude and frequency of this interference. Thus flood frequency analysis provides vital information for the planning and design of many hydraulic structures and for risk assessment in flood plain use (Admasu, 1989).

The estimation of flood quantiles is complicated because of both lack of a physical basis for determining the form of the underlying flood frequency distribution and the necessity of evaluating flood risk for return periods that exceed the length of the observed record. Flood quantaile estimates are strongly dependent on the form of a portion of the underlying flood frequency distribution and that is difficult to estimate from observed data.

The main flood prone area within the upper Awash sub- basin is the Becho plain, which lies at an altitude of 2060m and has a catchments area of 3300km$^{2.}$ Areas particularly at risks are around DiluMeda adjacent to Teji, and GebaMeda, situated North West Teji; these two areas covering approximately 20000ha are subjected to inundation for long periods. The gentle slopes and high sediment concentration in the Awash, result in silt deposition, which exacerbates the flooding problem by reducing the capacity of the channel to pass floodwater down stream.

1.3 Objective of the research

The **main objective** of this thesis is to find an appropriate procedure for analysis of flood frequency for use in the sub-basin.

The development of water resource projects in the basin lack proper guidance on how to analyze flood frequency. This study indicates the way towards proper solution.

The **specific objectives** of this thesis are:

- To test the homogeneity of at-site data of the selected stations of the upper Awash sub-basin (u/s of koka)
- To delineate homogeneous regions of the sub-basin
- To identify the best fit statistical distributions and method of parameter estimation for the data of the stations and regions in the upper Awash sub-basin.
- To determine quantiles.
- To develop regional frequency curve for stations and regions in the upper Awash sub-basin.

- To up date the previous studies.

1.4. Outline of the research

The thesis is organized into six chapters from introduction to the conclusion and recommendation. The first chapter presents about the introduction, background of the problem and the objective of the paper. The second chapter discusses about the literature review and state of art related to FFA, in this part all the high light of the paper will be discussed. The third chapter illustrates the methodology and procedure to be applied in the paper, from data collection to the result of the analysis, the fourth part of the thesis; collection of data and analysis of data will takes place. This chapter discusses about collection of data, consistency of data, filling and extension of data, homogeneity test and independence. Chapter five discusses the result of the above testes on the distribution selection, parameter estimation, quantile estimation and derivation of regional flood frequency curve, this part of the thesis presents the result of the analysis and evaluates the result with different physical realities. The last chapter of the paper concentrates on the conclusion and recommendation of the thesis which concludes and recommends on the result of the analysis.

1.5 Previous study of the area

For the Awash River basin especially upper awash sub-basin different researchers and organizations have stated different ideas on the protection of flood in the area but the flood frequency analysis with adequate and updated data hasn't fully studied in the sub-basin. Normally the area is susuptible to high flood due to the natural topography of the area which varies from the very mountainous to the immediate very plate land, which greatly enhances the creation of flood even with less rainfall. Looking at this problems different bodies have highlightly revised the area concerning the flood protection but not on the detail flood frequency analysis that is why the problem has stayed with out any solution up to now. From many of the papers written by different researchers such as Admasu Gebeyehu in 1989, in his study of regional flood frequency analysis of the whole country, he has tested the flood frequency of the area with limited data and information, Leulseged from Ministry of water resource has written some research paper on the upper Awash basin on flood frequency analysis and the consultant Halcraw has studied the flood damage of the area, all of them didn't analyzed the flood frequency of the area in detail. Now the target of this paper is to up date the previous studies, use recent data and by giving great attention on the flood problem of the area to analyze the flood frequency depending on the regional information and the at-site data analysis. For this particular sub-basin the stations have to be analyzed for base information of the area. All the previous study of the area takes the sub-basin as one region but in this paper the sub-basin will be analyzed as two regions based on the statistical analysis of data of the stations in the sub-basin.

2. LITREATURAL REVIEW

2.1. Flood Frequency Models

In flood frequency analysis the objective is to determine a Q-T relationship at any required site along a river. At any river site it is usually assumed that nature provides a unique Q-T relationship and that Q is a monotonically increasing function of T. In order to estimate this natural Q-T relation from a good quality continuous hydrometric record of N years duration, it is necessary to resort to a statistical or stochastic model of the continuous hydrograph, which retains information in the hydrograph relevant to the Q-T relation, and discard the rest.

Two such models are:

- ❖ Annual maximum series model, AM
- ❖ Partial duration series (or peak over a threshold) model, PD

In flood frequency modeling the problems related to the following points have to be solved (Cunnane 1989);

- ❖ Choice of model type (AM, or PD)
- ❖ Choice of distribution to be used in the chosen model
- ❖ Choice of method of parameter and quantile estimation

It should be noted that two separate aspects of such choice are important. These are the descriptive and predictive properties of the chosen method. The descriptive property relates to the requirements that the chosen distribution shape resembles the observed sample distribution of floods and that random samples drawn from the chosen model distribution must be statistically similar to the properties of real flood series, the predictive properties relates to the requirement that quantile estimates are robust with small bias and standard error (Cunnane, 1989).

2.1.1. Relative advantage of the two models

a) Annual maximum series model (AM)

Cunnane, 1989 has stated that a series of annual maximum flood is assumed to form a random sample from stationary population in which Q is a random variable with distribution PR ($Q<q$)=F (q). The variate values with exceedance probability 1/T is said to have return period T. Denoting this value $Q_{T,}$ it is such that:

$$1-F(Q_T)=1/T$$

In the annual maximum (AM) flow series, only the peak flow in each year of record is considered, that may involve some loss of information.

b) Partial duration series (or peak over a threshold) model (PD)

In this model most of the flow hydrograph is disregarded and the hydrograph is viewed as a series of randomly spaced flood peaks of random magnitude. For case of statistical modeling and also for case of identification of the values, which form the series, only the series of peak exceeding an arbitrary threshold q_o are considered. In particular, each of these showed that if the number of flood peaks exceeding some value q_o (a threshold value) in some interval of time such as a year has a poisson distribution with parameter λ, the number of events exceeding a great value $\bar{q}$ is also poisson distributed with parameter $\bar{\lambda} = \lambda p$ where p=PR

($q>\bar{q}/q>q_o$). Here p is a conditional probability, being the proportion of all peaks exceeding q_o which also exceeds $\bar{q}$ (Cunnane, 1989)

In partial duration series all peaks above a certain base value are considered. The base is usually selected low enough to include at least one event in each year (Rao and hammed, 2000)

Statistical efficiency of estimates of Q_T by each model

Denoting the estimates of Q_T obtained by AM method as Q_T and that obtained from the same hydrometric record by the PD method as $\bar{Q}_T$, it is usually observed that these two estimates are unequal. Further more the sampling variance of Q_T is not equal to that of $\bar{Q}_T$. i.e. Var (Q_T) ≠Var ($\bar{Q}_T$). From a statistical point of view that method which has the smallest sampling variance enjoys an advantage. Under certain common assumption Cunnane (1989) examined the relative values of Var ($\bar{Q}_T$) and Var (Q_T) and found that Var (Q_T) <Var ($\bar{Q}_T$ provided λ <1.65 where λ is the mean number of peaks per years included in the PD series. Where λ >1.65 the opposite was true. This shows that the AM method is statistically more efficient than the PD method when λ is small but less efficient when λ is large. In many practical situations the assumptions of the PD model may not be valid if λ is increased to too high a level, certain if λ>3 (Cunnane 1989).

Therefore, to avoid the problem of dependency on data, annual maximum (AM) series model has been selected in this study. In addition to this, AM series is widely and commonly used model by different researchers for the purpose of flood frequency analysis (Cunnane, 1989)

2.2 Regionalization

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. There is no universally accepted objective method of regionalization. This is due to the complexity of factors that affect the generation of floods. Several attempts have been made by different researchers to identify hydrologically homogeneous regions based on either geographical considerations or flood data characteristics, or a combination of both. (Mkhandi, 1985). This paper pays more attention on flood data characteristics of the at-site data and the regional data. The importance of using regional information arises from the need to improve or to stabilize site-specific estimate based on limited data, or to make inference at ungauged catchments.

For the case upper awash sub- basin, regionalization is applied depending on the data homogeneity of the stations, simply taking the upper awash sub basin and testing for the homogeneity using different techniques. The statistical values has been checked for the stations whether they can be categorized under one or more regions depending on the geographical, topographical and statistical analysis such as homogeneity test, the altitude and flood producing characteristics.

In most cases it is advisable to combine at-site specific and regional information (Hosking, 2000)

2.3 Homogeneity Test

Regional flood estimation methods are based on the premise that standardized flood variate, such as X=Q/E (Q) has the same distribution at every site in the chosen region. In particular Cv (x) and Cs (x) are considered to be constant across the region. Serious departures from such assumptions could lead to biased flood estimates at some sites. Those catchments whose Cv and Cs value happen to coincide with the regional mean values would fortuitously not suffer such bias. Nevertheless if the degree of heterogeneity present is not too great its negative effect may be more than compensated for the larger sample of sites contributing to parameter estimates. Thus Xt estimated from M sites, which are slightly heterogeneous, may be more reliable than Xt estimated from a small number, say M/3, more homogenous sites, especially if flow records are short.

The importance of homogeneity has been demonstrated by (Hosking 1985). Homogeneity implies that region have similar flood generating mechanism. A more specific definition of a homogeneous region is that region which consists of sites having the same standardized frequency distribution form and parameter.

At least five categories of questions arise in this context (Cunnane, 1987)

1. Are flood frequency behaviors of any one of M sites in a region; with AM records available, inconsistent with that of the remainder of the group?
2. Are geographically defined regions better or worse than regions obtained by partitioning the catchments characteristics data space?
3. How can a large group of catchments be divided in to homogenous sub group of regions?
4. How can un-gauged catchments be allocated to one of a number of pre selected homogenous regions?
5. What degree of departure from regional homogeneity can be tolerated in a flood quantiles estimation procedure?

We will use the CV based and LCV based homogeneity test to test the homogeneity of the stations and the corresponding identified regions. The detail equation and test procedures are presented in chapter four**.**

2.4 Procedures for Selection of Flood Frequency Distribution

Choice of distribution for AM series has received widespread attention. In many countries the selection of AM distribution is actually not made in any objective manner and that the choice of distribution is argued in a general manner as follows:

The choice of distribution would be: (Cunnane, 1989)

- Widely accepted
- Simple and convenient to apply
- Consistent, flexible or robust (low sensibility to outliers)
- Theoretically well based
- Documented in the guide

No special method of parameter estimations referred and the graphical method is used as frequently even more used as any other method. The choice of distribution is influenced by many factors such as method of discrimination between distributions, method of parameters estimation, the availability of data etc. the method of parameter estimation goes parallel with distribution selection.

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 distribution
- Log Pearson two distribution

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

List of tabulated distribution types and their formula will be presented below.

Table2.1 MATHEMATICAL EXPRESSION OF ANALYSED STATISTICAL DISTRIBUTION FOR ANNUAL MAXIMUM SERIES

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

distribution (EV1)	$F(x) = \exp\left\{-e^{-\left(\frac{x-u}{\alpha}\right)}\right\}$	

Distribution name	Distribution function F(x)	Variate and parameter ranges
Five parameters Wake by distribution (WAK 5)	$x = m+\alpha\,[1-(1-F)^{b}]-c[1-(1-F)^{-d}$ where $F = F(x)$	Note It is analytically defined only in inverse form
Four parameters Wake by distribution(WAK 4)	$x = \frac{\alpha}{\beta}\left[1-(1-F)^{\beta}\right]-\frac{\gamma}{\delta}\left[1-(1-F)^{-\delta}\right]$	
Generalized Pareto distribution (GPar)	$x = \varepsilon+\frac{\alpha}{k}\left[1-(1-F)^{k}\right]$ $F = F(x) = 1-\left[1-\frac{k}{\alpha}(x-\varepsilon)\right]^{1/k}$	
Log-Logistic distribution (LLg)	$F(x) = \left\{1+\left[\left(\frac{x-a}{b}\right)\right]^{-1/c}\right\}^{-1}$	$x > a,\ c > 0,\ b > 0$
Generalized logistic distribution (GLg)	$F(x) = \left\{1+\left[1-\gamma\left(\frac{x-\alpha}{\beta}\right)\right]^{1/\gamma}\right\}^{-1},\ \gamma\neq 0$ $= \left\{1+\exp\left[-\frac{x-\alpha}{\beta}\right]\right\}^{-1},\ \gamma=0$	$\gamma<0,\ \alpha+\frac{\beta}{\gamma}\leq x<\infty$ $\gamma>0, -\infty<x<\alpha+\frac{\beta}{\gamma}$ $-\infty<x<\infty$

2.4.1. Method of parameter estimation

In the past only the ordinary methods of moments (MOM) was mentioned for parameter estimation. It should be noted that MOM and PWM are the most efficient method of parameter estimation available. There are no other available methods of parameter estimation, which yield parameter, and quantile estimates, which have smaller error than those of MOM and PWM estimates (Cunnane, 1989).

Method of estimation includes:

i) Probability weighted moments (PWM)
ii) Method of moment (MOM)
iii) Maximum likely hood (ML)

a) Method of probability weighted moments (PWM)

PWMs are useful in deriving expression for the parameters of distributions whose inverse forms X=X (F) can be explicitly defined. In particular they allow parameter estimates to be obtained for distributions. Methods of parameter estimation are obtained in this method by equating moment of the distribution with the corresponding sample moment of observed data. For a distribution with k parameter, the first k sample moments are set equal to the corresponding population moments. The resulting equation is then solved simultaneously for the unknown parameters. Parameter estimation by PWM, which is relatively new, is as easy to apply as ordinary moments is

usually unbiased and is almost at efficient as method of maximum likelihood (ML). Indeed in small samples PWM may be as efficient
as ML. With a suitable choice of distribution PWM estimation also contributes to robustness and is attractive from that point of view (Cunnane, 1989).

b) Method of Moment (MOM)

It is one of the most commonly used methods of estimating parameters of a probability distribution. The estimates of the parameters of a probability distribution function are obtained by equating the moments of the sample with the moments of the probability distribution function. It provides simple calculation, but higher order moment estimates are biased (Wallis,et. Al. 1974). Parameter estimation by MOM is known to be biased and inefficient especially with three-parameter distribution but it is more preferable for two parameter distribution types.

c) Method of Maximum Likelihood (MLM)

Estimation by the maximum likelihood (ML) method involves the choice of parameter estimates that produce a maximum probability of occurrence of the observations. The parameter estimates that maximize the likelihood function are computed by partial differentiation with respect to each parameters and setting these partial derivatives equal to zero and finally solve the resulting set of equations simultaneously. The equations are usually complex that can only be solved by numerical techniques. As a result of this difficulty, the solution set may not properly found (Cunnane, 1989).

.

2.4.2 Comparison

In general, the PWM and the MOM are better for estimating the parameters for three and two parameter distributions respectively of the underlying distribution from which the data are sampled. They are less sensitive than others to sampling variability or measurement errors in the extreme data value (outliers), and therefore, they yield more accurate and robust estimates of the characteristics or parameters of the underlying probability distribution.

2.4.3. Selection

The selection of a distribution for flood frequency analysis goes with the selection of the method of parameter estimation. Parameters estimated by any of the above methods are subject to sampling errors. While a method may be efficient for one distribution it is not necessarily efficient for other distributions.

Therefore, to select the most robust flood estimation procedure:

- ❖ Descriptive ability tests and
- ❖ Predictive ability tests have to be applied.

More detail expression of these two methods will be listed in the next chapter.

2.5 Estimation of Index-Flood for Standardization

The application of regionalization requires a two-step procedure (D.Bacchiola, 1998):

- The identification of the homogeneous region for which a common probability model of maximum annual flow can be adopted to accommodate normalized flow
- Searching for the appropriate index flood estimation in the examined river site.

The estimation of the index flood plays a major role in design flood prediction, and it requires merging statistical and physical hydrology concept to reduce the present uncertainty. The basic idea behind the index-flood method is to increase the reliability of the frequency characteristics within a region. If, within a hydrologically homogenous area, a number of hydrometric stations have been operating and recording the effect of the same meteorological factors then a combination of these records will provide, not a longer record, but a more reliable record.

The index-flood method is based on the hypothesis that floods from different catchments with in a region normalized by their mean annual flood come from a single distribution. An essential prerequisite for this procedure is the standardization of the flood data from sites with different flood magnitudes. The most common practice is to standardize data, i.e. division by an estimate of the at-site means (Admasu, 1989).

$X_T = \frac{Q_T}{\bar{Q}_T}$, Is the index flood, which is the ratio of normal flood to mean?

The parameters of the distribution of X_T are obtained from the combined set of regional data. If at-site data are not available, the index-flood can be predicted from a regionally derived empirical equation. The form of the relation (empirical equation) for estimating the index flood depends on the amount of physical and climatic data available for developing it.

The other component of the index-flood procedure is the standardized quantile. The accuracy of the standardized quantile estimate depends mainly on:

- The method of grouping catchments for homogenous regions and testing for homogeneity and
- The type probability distribution and the method of parameter estimation and quantile estimation.

Procedure for index-flood determination

- List data of each gauging stations
- Eliminate unsuitable stations
- Select a common period of record
- Exclude stations less than 5 years of record
- The base period with maximum number of station year is selected
- Missing data may be filled in by inter-station correlation
- Not use filled data directly

- The index-flood method computes the return period, T, for each record events and stations. .(G.W.Kite, 1985)

Deficiencies found in the Index-flood method

1. The index-flood (mean annual flood) for stations with short periods of record may not be typical. This means that the ratios of flood of different return periods to the index-flood may vary widely between stations.
2. The homogeneity test is used to determine whether the difference in slope of frequency curves are greater than may be attributed to chance alone.
3. In the use of the index-flood method, it has been accepted that with in a flood frequency region frequency curves may be combined for all sizes of drainage areas, excluding only the largest.(G.W.Kite, 1985)

In this study all the sub-basin sub-catchments are gauged and the stations are evenly distributed over the sub-basin. Therefore, there is no need of estimating the flow for ungauged catchments since the sub-basin is fairly gauged. In this paper but the index flood help us to standardize data from different stations to group them in different regions and for derivation of flood frequency curve.

3. METHODOLOGIES AND PROCEDURE

Regional flood frequency analysis is a methodology of using at-site and regional information to predict the flood magnitude and its frequency for water resource planning and management and design of hydraulic structures.

The methodology chosen in this study is index flood method for at site and regional data analysis and it comprises data preparation, testing of data of the stations for homogeneity, selection of frequency distribution, method of parameter estimation and quantile estimation.

Generally the study involves the following procedure:

1. Collection of important data for the study such as hydrological data, methodological data, topographical and digitized map of the sub-basin.
2. Checking of data for quality, continuity, consistency and independence.
3. Computation of statistical parameters of selected stations within the sub-basin.
4. Carry out homogeneity test for the stations in the region.
5. Delineation of homogeneous regions
6. Selection of frequency distribution for the determination of the quantiles.
7. Selection of parameter estimation method for the selected distribution.
8. Quantile estimation.
9. Derivation of regional and at site flood frequency curve.

3.1. Collection of important data for study

Important data has been collected from different institutions. The data have been collected as a soft copies, hard copies and maps. Hydrological data and digitized map of the sub-basin were collected from the Ministry of Water Resource, from the department of Hydrology and GIS. Meteorological data such as temperature data, humidity, and rainfall data have been obtained from National Meteorological Agency and the top map of the sub-basin will be taken from EMA.

3.2. Consistency, Independency and Homogeneity test

Double mass curve is the best method to test data for consistency; this method is applicable by comparing the total cumulative annual maximum flow of all stations in the sub-basin with that of the single station and drawing the graph by taking the two components, and checking weather they aliened in a single straight line or not.

To test the independency of data for the stations, the W-W test has been used:

Given a sample size N, the Wald-Wolfowitz (1943) called W-W test is used to test for the independence of a data set and to test for the existence of trend in it. For a data set $x_1, x_2, x_3, \ldots x_N$ the static R is calculated from:

$$R = \sum_{i=1}^{N-1} x_i x_{i+1} + \ldots + x_1 x_N \ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

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

$$\bar{R} = \left(\frac{s_1^2 - s_2^2}{N-1} \right)$$

$$Var(R) = \left(\frac{s_2^2 - s_4^2}{N-1} \right) - \bar{R}^2 + \left(\frac{s_1^4 - 4s_1^2 s_2 + 4s_1 s_3 + s2_2^2 - 2s_4}{(N-1)(N-2)} \right) \ldots\ldots\ldots\ldots..(2)$$

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

The statistic $u = \left(\frac{\left(R - \bar{R} \right)}{\left(\text{var}(R) \right)^{1/2}} \right)$ 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 exceedence α/2. In this paper the value of /U/ computed will compared with $U_{0.025}$ = 1.96 which is the standard value. (Roa and Hammed, 2000).

In this thesis the homogeneity test of the data of the stations in the sub-basin based on the statistical value of CV and LCV and their respective value of CC that must be less than 0.3 to be homogeneous. The detail mathematical expressions of the method are presented in chapter four.

3.3 Delineation of homogeneous regions

Depending on the statistical values, the geographical, topographical and altitude of the sub-basin, the stations can be grouped to form region. In the case of upper Awash sub-basin, since the basin is found in a great variation of elevation of land from higher mountainous to the very plate area around becho plain the influence of altitude plays a great role in variation of flood for upper and lower part of the sub-basin. This differs the flood producing characteristics of the stations. Combining the results from the statistical values, the geographical, topographical and altitude of the area, the flood producing characteristics of the stations and the physical phenomena of the area can be grouped to form regions. On the digitized map of the sub-basin, all stations under analysis were identified according to their geographical location (latitude and longitude) and statistical value.

The procedure adopted in delineation of homogeneous region comprises the following three steps:

1) Geographic information was used to identify likely homogenous regions that are geographically continuous and having similar flood producing characteristics.
2) Each region that was identified in step 1 was checked for similarity in the statistics of observed flood data. Based on this step, regions obtained in step 1 were modified.
3) The proposed test of homogeneity was applied to confirm that the delineated regions are statistically homogenous.

3.4. Selection of best-fit distribution and method of estimation of Parameter

To select the best fit distribution and the best method of parameter estimation different soft wares have been used to fit the curve with distribution. Using those soft wares the data of the stations will fit with the density distribution curve, the result obtained from the software will be checked by the chi-square test and can be approved as the best fit distribution. The method of parameter estimation goes parallel with the distribution selection and can be tested for accuracy using standard error estimation, the less in standard error the more the selection of the method of the parameter estimation. The equation used for calculation of standard error for EV and LN distribution is:

$$Se(Q_T) = \frac{\sigma}{\sqrt{N}}(1 + 1.14K_T + 1.01K_T^2)^{1/2}$$

$$K_T = \frac{-\sqrt{6}}{\pi}(0.5772 + LN(-LN(1 - \frac{1}{T}))) \quad \text{.....................(3)}$$

3.5. Quantile estimation and derivation of flood frequency curve

Based on the selected distributions for each station, the quantile will be calculated according to the formula of the selected distributions listed in chapter five. To calculate the quantile for each selected distribution the return period has to be determined first, the way of determining the return period is presented in chapter five. Accordingly the quantile versus return period graph of each region will be drawn as the growth curve for each station and for each region delineated. 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 other wise the sub-basin has to be divided in to the proper homogeneous regions.

4. DATA COLLECTION AND ANALYSIS OF THE SUB-BASIN

4.1 Description of the study area

Ethiopia is located in the eastern part of Africa between $3^0 30'$ and $18^0 12'$ N latitude and $32^0 42'$ and $48^0 12'$ east longitude. The country has great geographical, topographical and climatologically diversity: from high rugged mountains to deep gorges; from lowest altitude at about 120m below sea level to highest altitude of 4600m above sea level; from 2000mm high annual rain fall to 200mm of low annual rain fall, this great variation of the altitude exposes some part of the country for creation of high flood due to this topographical variation. Beside, the Great Rift Valley divides the country in to two parts forming the eastern and western high lands. (Awash master plan by Hal craw).

The country is endowed with large amount of water resource potential. It has 12 major drainage basins as listed in the table below. (Abebe, 2004)

Table 4.1: major drainage basins of Ethiopia

No	Basins	Locations	Area (sq.km)	Coverage (%)
1	Mereb Basin	Extreme North	6065	0.5
2	Tekeze Basin	North	81034	7.2
3	Denakel Basin	North	66489	5.9
4	Blue Nile Basin	North-West	192953	17.1
5	Awash Basin	Central east	113604	10.1
6	Aysha Basin	Extreme East	4717	0.4
7	Ogaden Basin	East	82157	7.3
8	Wabi shebele basin	South East	207497	18.4
9	Genale Dawa basin	South	172681	15.3
10	Rift Valley Lakes Basin	Central South	51664	0.6
11	Omo-Gibe Basin	South West	74912	6.6
12	Baro-Akobo Basin	West	73958	6.5
TOTAL			1127730	100

This research will concentrate only on the Awash basin especially on the upper Awash sub-basin (u/s of Koka).

4.1.2. <u>Awash River Basin</u>

The Awash river basin rises at an elevation of 3000m above sea level over the central high land of Ethiopia about 150km west of Addis Ababa. The river flows generally northeastwards along the rift valley and terminates in the lake Abe at an elevation of 250m above sea level near Djibouti. This high variation of elevation difference exposes the basin for high flood to occur due to altitude variation, topographical and geographical diversity.

The Awash basin covers a total area of 110000km^2 of which 64000km^2 comprises its western catchments, drain to the main river or its tributaries. The remaining 46000km^2, most of which comprises the so-called eastern catchments, drain into a desert area and dose not contribute to the main river course (Awash master plan by Hal craw).

The Awash basin has about 4 four major drainage sub-basins and about 67 gauging stations. Out of this the upper Awash sub-basin has only 21 gauging stations. These gauging stations are totally found in the upstream of koka of which only part of them got a good record of data and can be preferable for the analysis.

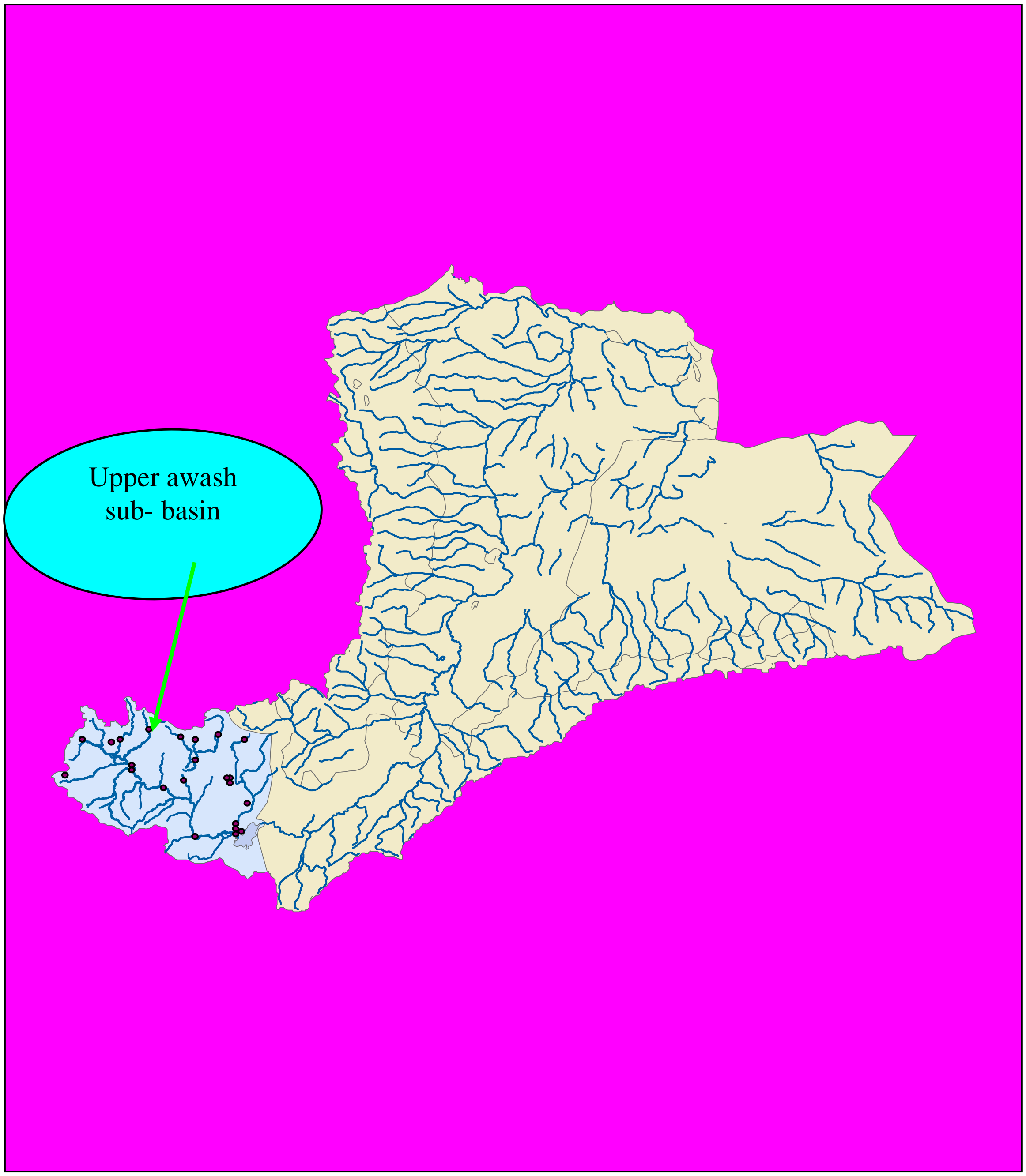


Fig. 4.1 Awash River basin with the hydrological stations in the upper Awash sub-basin and the river streams in the basin majorly the tributaries of Awash River.

4.1.3. Upper Awash Sub-basin (u/s of Koka)

The upper Awash catchments are found in the high lands of the central Ethiopia with all land above 1500m a.s.l. The Koka reservoir is located in the upper reaches of the Awash basin approximately 75km south east of Addis Ababa. The climate of the Koka sub-basin come under the influence of the ITCZ and the seasonal rainfall distribution results from the annual migration of the ITCZ. The mean annual rainfall is 1200mm and reaches 1500mm at the eastern high lands of Addis Ababa. The catchments receive its maximum rainfall during June to September and amounts to 70% to 75% of the annual rainfall. The second rainy period covers the period from February to May. There is high variation in elevation and in rainfall that leads the area to high occurrence of flood (Awash master plan by Hal craw).

A watershed of the upper Awash sub-basin (u/s of koka) is usually a complex and heterogeneous system. Its characteristics vary in space. Hydrologic processes vary both in space and time. One way to account, at least partly, for spatial variability of governing hydrologic factors is to divide the watershed in to sub-catchments depending on the soil type, vegetation, land-use and topography that significantly affect stream flow. The larger the number of sub-catchments, the greater the accuracy as well as effectiveness of the analysis. Accordingly the upper Awash sub-basin is divided in to five sub-catchments as their characteristics are tabulate as below.
(Paulose, 1998)

The drainage area of the Koka sub basin is 11200km^2. The major rivers that flow in the sub basin are Awash and Mojo rivers. The other major tributaries of Awash above Koka are Akaki, Holeta, Berga, and Legedadi rivers. In the Koka sub basin there are about 21 gauging stations that records the flow and lake depth in the sub basin out of this gauging stations only 10 are used in this research. The River length, drainage area and the slope of each station is tabulated as follows:

Table 4.2 General characteristics of the stations and sub-catchments in the upper Awash sub- basin (u/s of koka)

Name Stations	River Length (km)	Catchments slope (%)	Catchments area (km2)	Lat.(N) *	Long.(E) *
Berga Nr Addis Alem	13.485	17	249	996668	428556
Holeta Nr holeta	14.97	76	119	1004010	446886
Teji @ Asgori	24.142	61	663	970876	426678
Awash @ Bello	32.383	35	2568	978231	435855
Akaki @ Akaki village	44.799	68	884	981872	476159
Mojo @Mojo village	42.526	96	2175	950545	509170
Awash @ Melka hombole	106.151	33	7656	950551	476159
Awash @ Melka Kunture	57.353	30	4456	961622	455998
Little Akaki	21.34	56	131	466795	999047
Awash @ Ginchi	58.36	24	76	404697	997173

* All the lat. and long. are in UTM

Table 4.3 Sub-catchments in the upper Awash sub-basin

Sub-catchments	Area(km^2)	Perimeter(km)	Stations in the Sub-catchments	Slope of the sub-catchments (%)
Sub-catchments-1	4530.38	412.35	▪ Ginchi ▪ Berga ▪ A.bello ▪ Teji ▪ M.kunture	64
Sub-catchments-2	474.11	108.85	▪ Holeta	68
Sub-catchments-3	1608.8	219.06	▪ Lt.Akaki ▪ Akaki	83

Sub-catchments-4	2176.57	216.99	▪ Hombole	38
Sub-catchments-5	1901.70	235.38	▪ Mojo	81

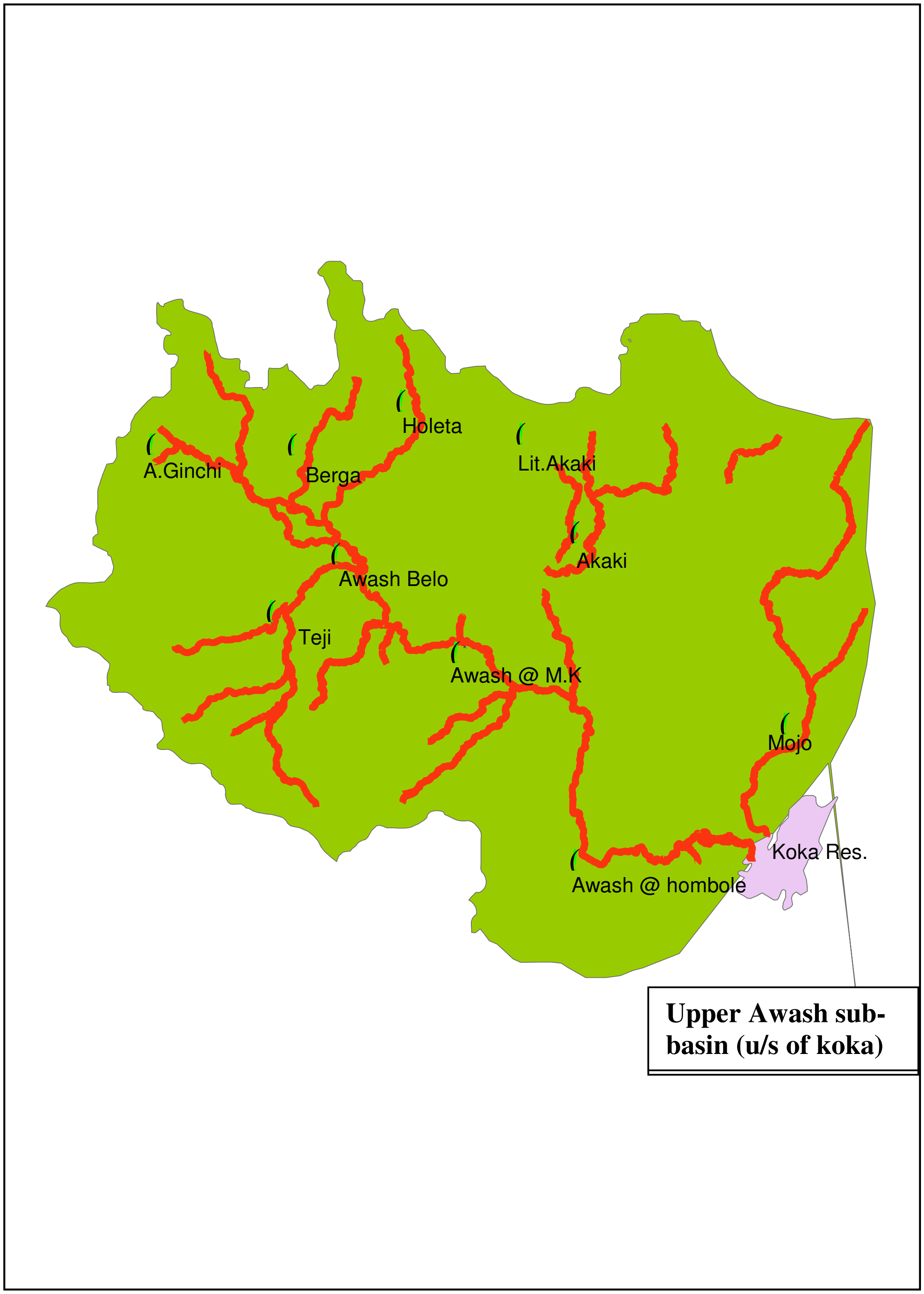


Fig. 4.2 upper Awash River sub- basin (u/s of koka) with Selected stations

4.2 Land use and Soil type

The land use condition in the upper Awash catchments includes mainly of cultivated agricultural land, grassland, forestland, rural and urban settlements. It is estimated that 67% is intensively cultivated, 25.5% is moderately cultivated, 4.5% is bush land or shrub land or wooded grassland, and 3% is urban area and alpine vegetation. Strictly speaking, even the land use within the upper Awash is diverse. In the upper most part where there is high rainfall, land use is complete in May with barley and teff. Steeper slopes are heavily wooded with natural acacia and eucalyptus. On the lower most part, however, rainfall is too unreliable and the sparse dry acacia scrub gives way to wide stretches of bare ground with clumps of coarse grass and occasional thickets of acacia. The soil type in the upper Awash sub-basin is diverse. The most common soil types are Clay, Sand, Clay-Loam, Silt-Clay -Loam, Sand-Clay, Silt-Clay (Paulose, 1989). Land use and soil type have a direct impact on the flood amount, speed and potential to create damage that the study should give attention for land use and land cover of the sub-basin.

4.3 Source and availability of data

Data that are collected and to be analyzed can be set under three categories according to their source and availability; all of them are time series data, theses are:

a) Hydrological data

It is a runoff data that is collected from gauging stations in the basin. In the Awash basin there are about 76 gauging stations, out of these 21 stations are found in the upper Awash sub- basin particularly in the upstream of Koka Dam. These all stations are not selected for the analysis because some of the stations are under the influence of the unnatural conditions such as the release of spillway. The very much plate of the area, and the back flow of reservoir; due to these reasons only 10 stations are selected for data analysis of the upstream of koka sub-basin. These 10 stations have various length of record from 38 years to 12 years of record variation and are fairly distributed over the basin. These hydrological data is collected from the MoWR Hydrology department.

b) <u>Meteorological data</u>

Rainfall data are collected from meteorological Agencies for the purpose of comparison with run off data to check the mass balance, with corresponding to the hydrological stations. Only five metrological stations are obtained for the analysis that can correspond with the hydrological stations and they are enough for comparison purpose. The data obtained are monthly rainfall, temperature, humidity, sunshine and evaporation. The role of this data is very less since there is no rainfall-runoff model evolved in this paper. These data are obtained from National Meteorological Agency.

c) <u>Topographic data</u>

Topographical data are collected from EMA (Ethiopian Map Authority). The map is used to identify the land feature and characteristics of the sub-basin. The digitized contour map of the sub-basin is the most important input to generate different basin out puts and to regionalize the sub-basin; that is collected from MoWR, GIS department.

4.4. <u>Filling and extension of the data</u>

In the upper Awash sub-basin there are about 21 gauging stations, out of these only 10 stations are selected for flood frequency analysis for the sub-basin, the remaining have some defect, they are under the influence of natural and man made factors. Some of the stations are set at u/s of koka reservoir that are directly subjected to the back up effect of the reservoir; others are set d/s of legedadi dam which has no natural record due to the influence of the release of the water through spillway and the rest are on the very plate and plain area that may be subjected to the over flow and have a wrong record of data, on the other hand two of the stations are newly installed and have short record of data. Due to the above reasons, only 10 stations are selected for the proper analysis. The selected once by them selves have no fully recorded data; they have less number of years of record that needs to be extended. Some of the stations have missing data that needs to be filled before analysis. Out of those selected 10 gauging stations 5 stations have one or more missed data and 2 stations have less than 15 years of record data;

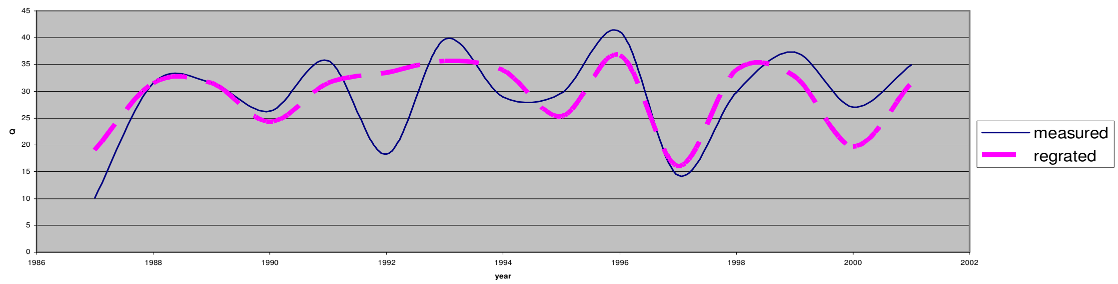
as a result regression analysis is used to fill the missing monthly data and to extend those short length recorded data with satisfactory correlation coefficients checked by the MS DOS program called curve fit. For this we have used soft wares that simplify the regression analysis and very much important in distribution selection and method of parameter determination.

To test the accuracy and effectiveness of soft wares, the existing data as regretted and recorded data has been used; finally we have got a good effect in regretting data for extension or for filling of the missing data.

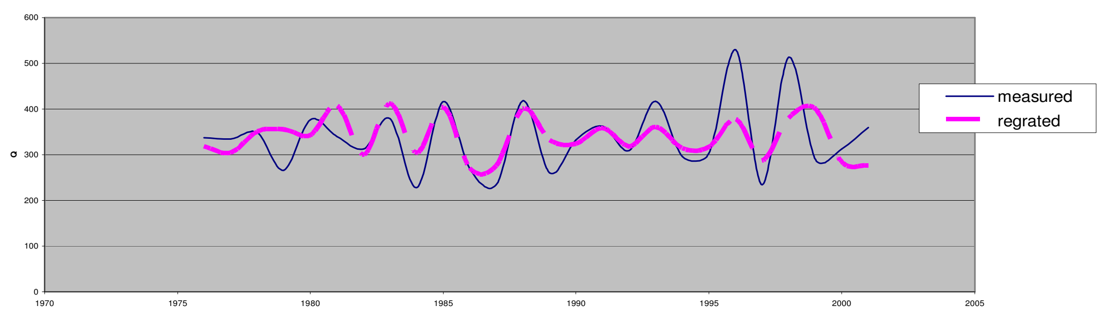
The graph that shows the relation between the recorded and generated (using the soft wares) is shown bellow.

Graphs that show the accuracy and applicability of the regression equation tested on different stations of the sub-basin, which compares the recorded and the regretted data of the stations that have maximum and minimum coefficient of determination using the mentioned soft wares:

Generated and measured flow data comparison using regression analysis for Stations 31001 and 31002



Generated and measured flow data comparison using regression analysis for Stations 31003 and 31012



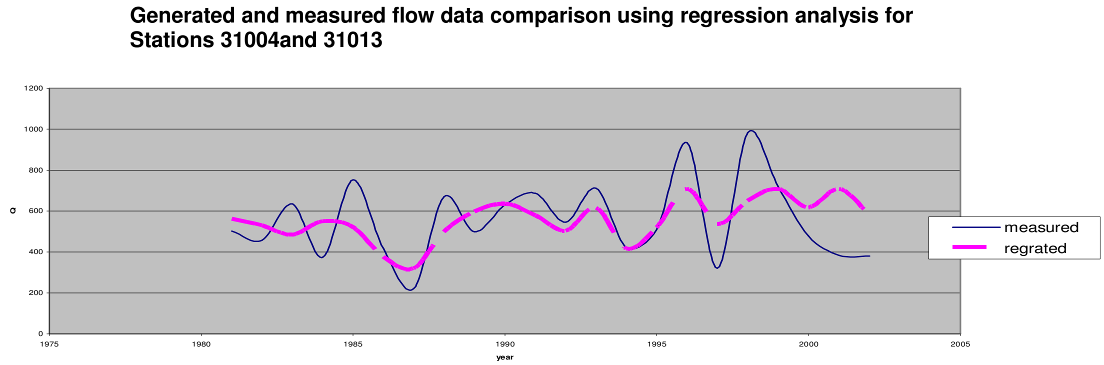


Fig: 4.3 Graph showing the relation between generated and measured flow

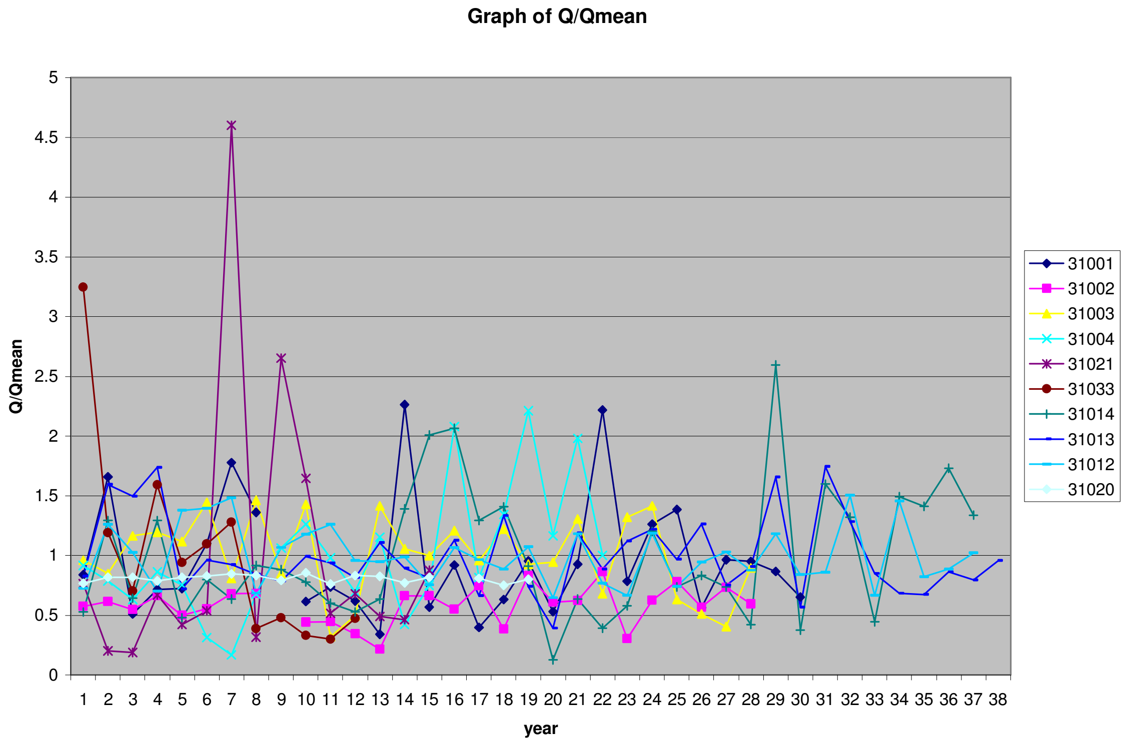
As the graph illustrates the regression equation can be used for filling and extension of data of stations in the basin with the preferable correlation coefficient.

Table 4.4 Table showing the regression equation of station in the sub-basin for filling and Extension

Site Number (Q1)	Station name	Regretted With (Q2)	Equation	R^2	Range	N	Extended or filled
31001	Berga River near AddisAlem	31002	Q1=20.2+0.023*Q2+0.03458*(Q2)^2	0.96	1975-2004	28	Filled
31002	Holeta River near Holeta	31001	Q1=16.25+0.3146*Q2-0.001281*(Q2)^2	0.96	1975-2002	24	Filled
31020	Awash River near Bello	31001	Q1=35.92+0.03507*Q2	0.77	1987-2005	15	Filled & extended
31003	Teji River near Asgori	31012	Q1=12.6+0.25*Q2+0.036(Q2)^2	0.89	1976-2003	24	filled
31004	Akaki River at Akaki	31013	Q1=83.23+0.054*Q2-0.0085(Q2)^2	0.82	1981-2002	22	filled
31012	Awash River at Melkakunture	31003	ENOUGH DATA	0.98	1965-2001	37	none

31013	Awash River at Hombole	31004	ENOUGH DATA	0.94	1968-2005	38	none
31014	Mojo River at Mojo		ENOUGH DATA	0.99	1968-2004	37	none
31033	Awash River at Ginchi	31001	Q1=10.2+1.324*Q2-0.00455*(Q2)^2	0.88	1993-2004	12	Extended
31021	Little Akaki River at Akaki	31013	Q1=61.32+0.68*Q2+0.0047(Q2)^2	0.87	1990-2004	15	Extended

As the above table shows the regression equation (quadratic) used trough the soft wares with the assurance of the correlation coefficient with the soft wares and the curve fit program. The data have been filled and extended, till it is suitable for the analysis. The filled and extended data of the stations in the sub-basin is graphs as below for their illustration of randomness and independency.



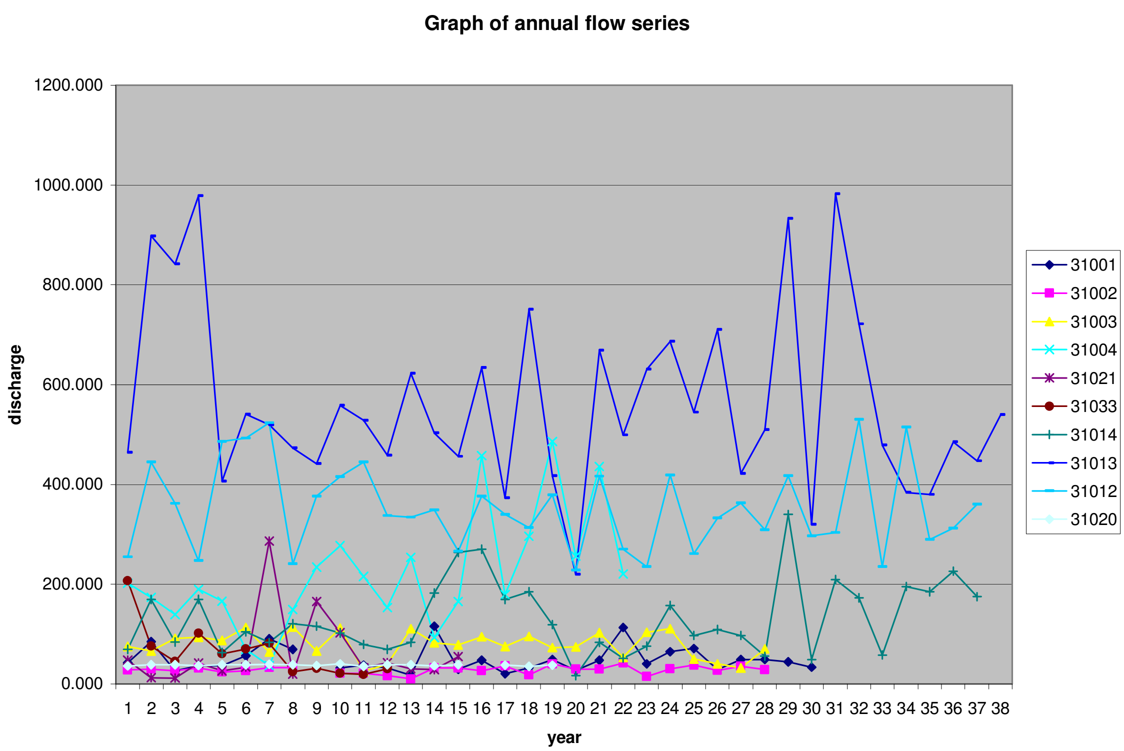


Fig: 4.4 Graphical expressions of annual maximum data after Filling and Extension

4.5 Consistency and Independence of data

The reliable of estimates of population quantile derived from frequency analysis depends in the first instance on the data series used in the analysis. Before conducting the analysis, the series should be scrutinized for possible errors or inconsistency and for any indication that contravene basic statistical assumption. (Roa and Hammed, 2000).

4.5.1 Guideline for test of Consistency of data \

According to Roa and Hammed, 2000:

☞ A time series plot should be prepared to show:

- The annual maximal including gaps,
- A 3- to 5-years moving average of the annual maximal. Missing data and possible jumps or trends should be noted on this plot.

☞ A histogram plot should be prepared to show the distribution by magnitude of annual maximal. Anomalies such as bimodal distributions or apparent outliers can some times be identified on such plot. However, these features are more commonly identified after initial frequency analysis.

☞ Where a substantial degree of error is suspected in one or higher flood peaks, a sensitivity analysis using alternative values may be advisable. In important or dubious cases the original data should be reviewed critically with respect to gauge shift and malfunctions, estimation of missing data, computation of flow, and extrapolation of rating curve.

The double- mass curve is used to check the consistency of many kinds of hydrologic data by comparing data for a single station with that of a pattern composed of the data from several other stations in the sub-basin. The double mass curve can be used to adjust inconsistence flow data. The graph of the cumulative data of one variable versus the cumulative data of a related variable is a straight line so long as the relation between the variables is a fixed ratio. Breaks in the double mass curve of such variables are caused by change in the relation between the variables. These changes may be due to changes in the method of data collection or to physical changes that affects the relation. Poor correlation between the variables can prevent detection of inconsistencies in a record, but an increase in the length of record tends to offset the effect of poor correlation.
All the stations in the sub-basin are consistent except one station that is akaki station, which is somewhat, seems inconsistent due to the influence of unnatural flow from urban area but it is a simple diversion that may not tide our analysis. The graphical expression of the results are presented in the ANNEX 3

4.5.2 Independence and Stationaity Test

A time series is stationary if in the long term it is invariant with respect to time. Non-stationary may invalidate the result frequency analysis unless the data are first adjusted. Non-stationeries in flood series may be of three types: jump, trends, and long-term periodic cycles or segments of the same. (Roa and Hammed, 2000).

It is usually assumed that all the peak magnitudes in the AM series are mutually independent in the statistical sense. Non-independent of the events in the data series may bias the result of frequency analysis. Although comparatively rare in flood data, non-independent may occasionally arise due to year-to-year carry over of surface or subsurface storage, or non-independence in input series.

Specific statistical test for independence are incorporated in various computation in frequency analysis. When applied to short series, however, commonly used tests can be misleading: they may indicate non-independence when the events are actually independent, or failed to indicate it when serial correlation long lags is in fact present. Flood frequency analysis is one of the investigations of extreme values. In any time series data, outliers may or may not exists. These outliers may come due to personal error during recording and inadequacy of measuring device or really due to very extreme condition of natural phenomenon that is important information for flood frequency analysis. Therefore, unless the source of the outliers clearly identified, it is difficult to remove outliers completely from analysis. Outliers can be excluded from the estimation procedure only if it is certain that a single known distributional form can adequately model AM flood in this paper there is no as such magnified outlier. According to the above expressive idea on independency. The independency test has been tested for the stations using Wald-Wolfowitz (W-W), the formula and description is presented in chapter three. The result of the independency test of the stations in the sub-basin is tabulated below:

Table 4.5 Result of Independence test

Station number	**Stations name**	**Statistics**	**Critical test statistics**	**Remark**
31001	Berga	0.657	1.96	Independent
31002	Holeta	0.238	1.96	Independent
31003	Teji	0.986	1.96	Independent
31020	Awash @ Bello	-0.207	1.96	Independent
31013	Awash @ M.Hombole	1.46	1.96	Independent
31012	Awash @ M.kunture	1.375	1.96	Independent
31033	Awash @ Ginchi	0.54	1.96	Independent
31004	Akaki @ Akaki Vill.	1.03	1.96	Independent
31021	Little Akaki	0.68	1.96	Independent
31014	Mojo @Mojo Vill.	1.56	1.96	Independent

Since all the statistics values are greater than the critical test statistics values the data in the stations are independent. It is usually assumed that all the peak magnitudes in the AM series are mutually independent in the statistical sense.

4.6. Test of data for homogeneity of stations in the sub basin

Floods are affected by the physical and climatic characteristics of the catchments such as storm duration and intensity, size, shape, relief, drainage, morphology, land cover, presence or absence of storage, soil type and land use. All these factors vary in space. Therefore, it is unreasonable to expect that a region can be chosen in which the flood frequency distribution at all site are identical. What is required is that the region be sufficiently homogenous that no further division of the region in to smaller regions or individual sites would improve the accuracy of flood quintile estimates. (Admasu, 1989).

Stations in a region can be tested for homogeneity using statistical value; homogeneity can be taken as a base for many criteria of the basin. There are various homogeneity tests. The testes used in this study are:

- CV-based homogeneity test
- LCV-based homogeneity test

a) CV- based homogeneity test

In regionalization assumption must be made about the statistical similarity of the site in a region. To investigate whether those has been meet or not. Cunnane (1989) have used the values of mean coefficient of variation (CV) and the site-to-site coefficient of variation (CV) of both convention and L-moment of the proposed region. According to the researchers the higher the value of CV and CC the lower the performance of the index-flood method for the considered region. This is due to the dominance of the flood quintile estimation variance by the variance of the at-site sample mean. Hence for better performance of the index flood method, CC should be kept low.
(Abebe, 2004).

In this study also both conventional and L-moment has been used to calculate CV, LCV and their respective CC, values. The procedures are described below:

For each site in a region calculate mean ($\bar{Q}$), standard deviation (σ) and coefficient of variation (CV).

$$\bar{Q}\text{i} = \frac{\sum_{i}^{n} Q_i}{n_i}, \quad \sigma_i = \sqrt{\frac{\sum_{i}^{n}\left(Q_i - \bar{Q}_i\right)^2}{n_i - 1}}, \quad CV_i = \frac{\sigma_i}{\bar{Q}_i} \quad \ldots\ldots\ldots\ldots\ldots (4)$$

Where, Q_i = the flow rate of station i

$\bar{Q}_i$ = The mean flow rate for sites i

σ_i = Standard deviation of Q_i for site i

CV_i = Coefficient of variation of site of i

For each region, using the statistics calculated above, computed the regional mean, CV and finally the corresponding CC using the following relation:

$$\bar{CV}_i = \sum_{i=1}^{N} \frac{CV_i}{N_i}, \quad \sigma_{cv} = \sqrt{\frac{\sum_{i=1}^{N}\left(CV_i - \bar{CV}_i\right)}{N - 1}}, \quad CC = \frac{\sigma_{cv}}{\bar{CV}} \quad \ldots\ldots\ldots\ldots\ldots\ldots (5)$$

Where, N= number of site in a region

$\bar{CV}$ = Mean coefficient of variation

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

The stations and regions declared to be homogenous if CC<0.3

b) <u>LCV- based homogeneity test</u>

LCV- based homogeneity test is more accurate and effective way of testing the homogeneity of the site (station) when compared with that of the CV-based homogeneity test. The procedural calculation is the same as that of the CV, but it has its own formula and calculation;

The following are advantage of LCV (cunnane, 1989):

- Compared to CV, LCV can characterize a wide range of distribution.
- Sample estimates of LCV are so robust that they are not affected by the presence of outliers in the data set.
- They are less subjected to bias in estimation.
- LCV yields more accurate estimate of the parameter of a fitted distribution.

Hosking (1986) gave the unbiased estimators of M_{10k} and M_{1j0} as:

$$M_{10k} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\binom{N-i}{k}}{\binom{N-1}{k}}\right]X_i \qquad \text{, k=0, 1, 2, 3… N-1……………….. (6)}$$

$$M_{1j0} = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\binom{i-1}{j}}{\binom{N-1}{j}}\right]X_i \qquad \text{, j=0, 1, 2, 3… N-1………………….. (7)}$$

Where, i= rank of observed flow data in ascending order.

The first few moments are:

$$\begin{aligned} L_1 &= M_{100} \\ L_2 &= M_{100} - 2*M_{101} \\ L_3 &= M_{100} - 6*M_{101} + 6*M_{102} \\ L_4 &= M_{100} - 12*M_{101} + 30*M_{102} - 20*M_{103} \end{aligned} \qquad \text{……………………… (8)}$$

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. L-moment ratio, which are analogous to conventional moment ratio are defined by Hosking (1986) as;

Accordingly the formula used for the homogeneity test will be formulated as follows;

$\tau = \frac{L_2}{L_1}$ = Measure of scale and dispersion (LCV)

$\tau_3 = \frac{L_3}{L_2}$ = Measure of skew ness (LCs)

$\tau_4 = \frac{L_4}{L_2}$ = Measure of kurtosis (LCk)

Using the above procedural formula we have:

$$L_{CV} = \frac{L_2}{L_1} \text{ and } \bar{L}_{CV} = \sum_{i=1}^{N} \frac{L_{CV}}{N}$$

$$\sigma_{L_{CV}} = \sqrt{\frac{\sum_{i=1}^{N}\left(L_{CV} - \bar{L}_{CV}\right)^2}{N-1}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

$$CC = \frac{\sigma_{L_{CV}}}{\bar{L}_{Cv}}$$

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

Table 4.6 Result of homogeneity test for each station in the sub basin

Stations	CV	Cs	Ck	LCV	LCs	LCk	CC	Remark
Berga	0.52	1.35	1.38	0.40	0.24	0.15	0.21	According to the result from the statistical values of each station and depending on the CC value, all the station in the sub-basin are HOMGENIOUS Since the CC value is less than 0.3
Holeta	0.27	0.52	1.19	0.13	0.14	0.12	0.18	
Teji	0.33	0.39	1.69	0.15	0.02	0.22	0.14	
Awash @ Bello	0.14	0.44	1.72	0.07	0.21	0.32	0.25	
Awash @ M.Hombole	0.32	0.82	2.24	0.16	0.32	0.49	0.23	
Awash @ M.kunture	0.25	0.46	2.66	0.12	0.43	0.26	0.20	
Awash @ Ginchi	0.82	2.01	4.92	0.42	0.62	0.43	0.13	
Akaki @ Akaki Vill.	0.53	0.97	3.97	0.26	0.31	0.19	0.27	
Little Akaki	1.18	2.47	6.25	0.9	0.76	0.51	0.12	
Mojo@ Mojo Vill.	0.53	0.97	3.97	0.26	0.31	0.19	0.26	

Table 4.7 Result of homogeneity test for each station in the Regions

Upper region

Stations	CV	Cs	Ck	LCV	LCs	LCk	CC	Remark
Berga	0.52	1.35	1.38	0.40	0.24	0.15	0.21	homogeneous
Holeta	0.27	0.52	1.19	0.13	0.14	0.12	0.18	homogeneous
Teji	0.33	0.39	1.69	0.15	0.02	0.22	0.14	homogeneous
Awash @ Bello	0.14	0.44	1.72	0.07	0.21	0.32	0.25	homogeneous
Awash @ Ginchi	0.82	2.01	4.92	0.42	0.62	0.43	0.13	homogeneous
Little Akaki	1.18	2.47	6.25	0.9	0.76	0.51	0.12	homogeneous

Lower region

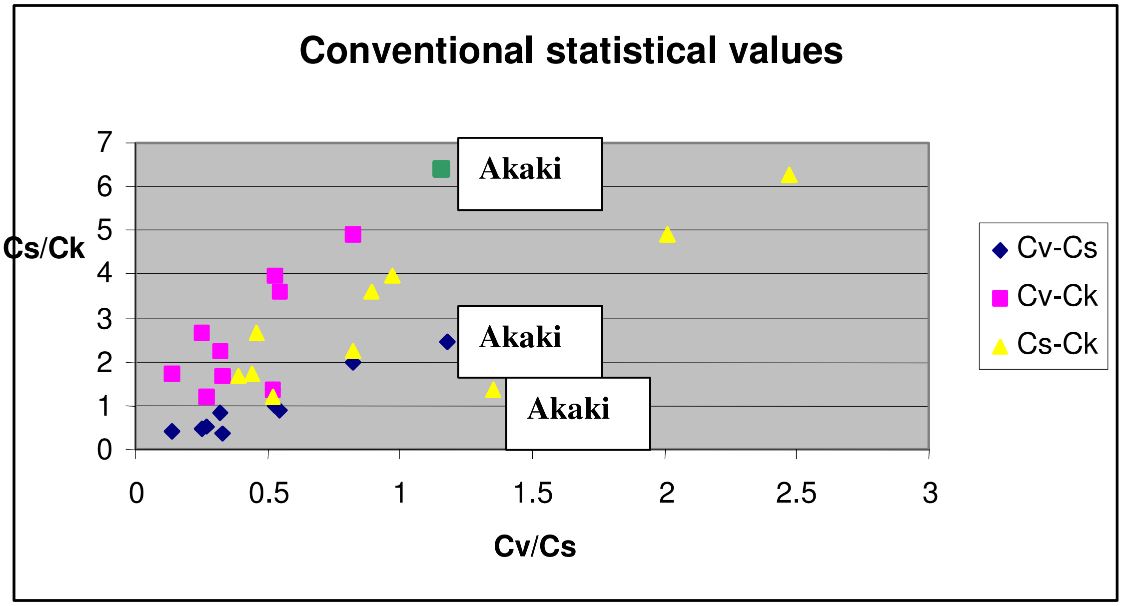
Stations	CV	Cs	Ck	LCV	LCs	LCk	CC	Remark
Awash @ M.Hombole	0.32	0.82	2.24	0.16	0.32	0.49	0.23	homogeneous
Awash @ M.kunture	0.25	0.46	2.66	0.12	0.43	0.26	0.20	homogeneous
Akaki @ Akaki Vill.	0.53	0.97	3.97	0.26	0.31	0.19	0.27	homogeneous
Mojo @Mojo Vill.	0.53	0.97	3.97	0.26	0.31	0.19	0.26	homogeneous

Table 4.8 Result of homogeneity test for each region in the sub basin

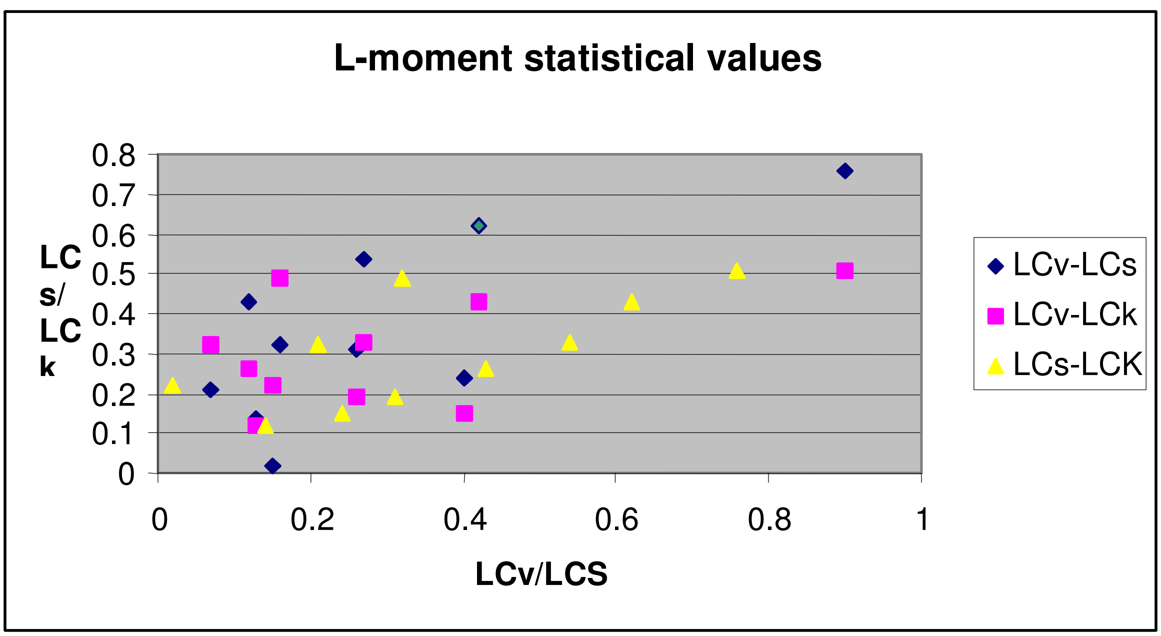
Region	CV	Cs	Ck	LCV	LCs	LCk	CC	Remark
Upper	0.54	1.2	2.86	0.34	0.33	0.29	0.17	**homogeneous**
Lower	0.41	0.81	3.21	0.2	0.34	0.28	0.24	**homogeneous**

From the above statistical values we have got that the CC value is less than 0.3, which shows the stations are homogeneous. The sub-basin is delineated under two regions as upper and lower regions from statistical and physical values both the regions are statistically homogenous.

The statistical properties of the station can express the nature of the station; from the stations in the sub-basin, two of the stations are statistically have higher CC value. The Akaki station is highly influenced by the urban drainage flow and the flow is not natural and it is also difficult to naturalize it. The Awash @ Bello station is found in a very plate area of becho plain area, which is mostly subjected to the over flow the Awash River that disturbs the natural record of the flow. This problem can also be more explained and illustrated on the graph below both in the conventional and L-moment. From the graph the station with highest value of CC can be differentiated from those remaining station is the Akaki which has different statistical value as compared to that of the remaining stations due to the mentioned reasons.



Akaki



Akaki

Fig 4.5_ Graphical expressions of the relation of the above statistical Variables of a station Used for homogeneity test

4.7 Regionalization

Regionalization is the identification of homogeneous region that contains stations having similar flood producing characteristics. This can be achieved using information that is obtained from geographical proximity, physiographic, climate, altitude (topography) and drainage characteristics of the catchments. In the upper Awash sub-basin there is a great variation of elevations that produces the variation of flood producing characters tics, accordingly the sub-basin has two great topographical variations varying from highly mountainous to that of very plate area around becho plain that mostly attacked with great over land flood. These physical factors latter refined on the basis of statistics of the observed flood.

In regional flood frequency analysis, the established curve of flood variate versus return period can be used for estimating flood quantile at any site with in the region. In the upper Awash sub-basin the stations are found in evenly distributed way covering the whole area of the sub-basin with proper distribution and area coverage. Therefore, there is no excess place that is ungauged rather the stations with improper record of data have been removed, using those gauged stations the statistical properties and flood producing characteristics of the stations for regionalization purpose have been tested. Statistically we have got two regions; the first region includes the berga, holeta, teji, ginchi, little akaki, and awash bello station and the second region have mojo, hombole, kunture, and akaki stations. When we the stations topographically geographical proximity, altitude and other external fixture of the catchments have been visualized. The sub basin can be divided in to two varying elevation and flood producing natures as the firs region having higher elevation berga, holeta, ginchi, teji, akaki, little akaki and mojo can be categorized, and in the second region which has very plate nature the remaining bello, hombole, kunture were included. Generally depending on the result obtained from both tests two main regions the upper and lower regions has been delineated. In the regionalization the statistical results plays a great role (Hosking, 2000) in our case the result of both physical and statistical values indicates almost the same result, accordingly the stations in the regions is tabulated below:

Table 4.9 stations in the region of the sub-basin

Station number	Stations name	Region	Remark
31001	Berga	Upper	Homogeneous
31002	Holeta	Upper	Homogeneous
31003	Teji	Upper	Homogeneous
31020	Awash @ Bello	Upper	Homogeneous
31013	Awash @ M.Hombole	Lower	Homogeneous
31012	Awash @ M.kunture	Lower	Homogeneous
31033	Awash @ Ginchi	Upper	Homogeneous
31004	Akaki @ Akaki Vill.	Lower	Homogeneous
31021	Little Akaki	Upper	Homogeneous
31014	Mojo @Mojo Vill.	Lower	Homogeneous

Topographically the akaki station is found in the upper region but statistically it has similar flood producing characteristics with that of the lower region, since the statistical results are mostly dominant the station is categorized under the lower region.

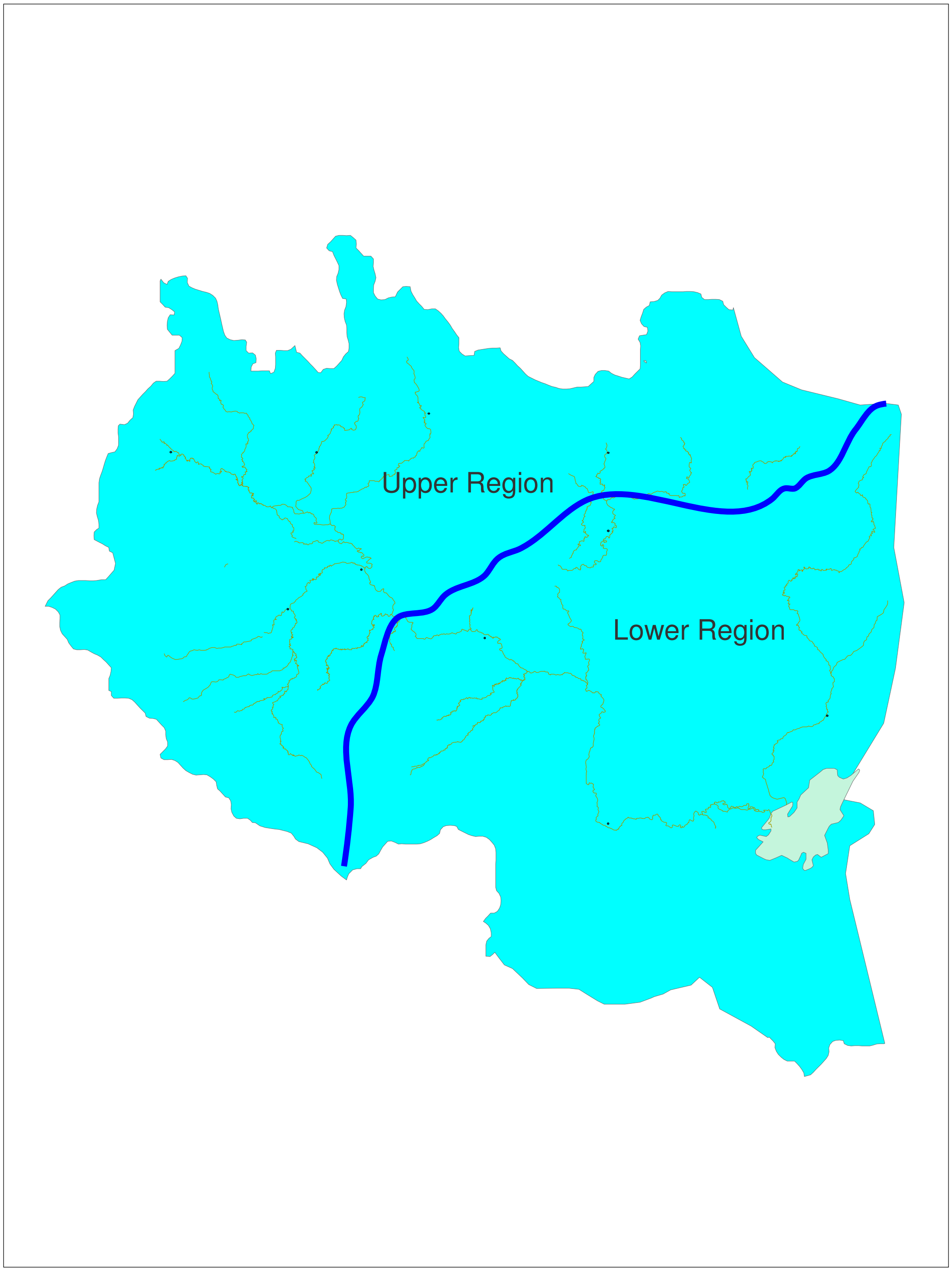


Fig. 4.6 Delineated regions in the sub-basin

5. RESULTS AND DISCUSSION

5.1 Results from consistency, independency and homogeneity tests

As the result obtained from the above analysis all the stations selected and regions for the analysis are consistent, independent and homogeneous. Due to this the stations and the regions delineated in the sub-basin are homogeneous, that is the base for further analysis such as distribution selection, estimation of method of parameter estimation and quantile estimation.

5.2 Selections of best fitted statistical distribution

In RFFA a single frequency distribution is fitted to data from several sites. In general, the region will be slightly heterogeneous, and there will be no single "true" distribution that applies to each site. The aim is therefore not to identify a "true" distribution but to find a distribution that will yield accurate quintile estimate for each site.

The chosen distribution need not be the distribution that gives the closet approximation to the observed data. Even when a distribution can be found that gives a close fit to the observed data, there is no guarantee that future values will much those of the past, particularly when the data arise from a physical process that can give rise to occasional outlying values far removed from the bulk of the data. (Hosking, 1989).

During selection of the best-fit distribution to the statistical data the estimation of the parameter goes parallel with the selection of the distribution. After a distribution or a number of distributions are selected to fit the data, their parameters were estimated with different methods. The three major parameter estimation used are:

- ☞ Method of moment (MOM)
- ☞ Method of probability weighted moment (PWM) and
- ☞ Method of maximum likely hood (ML)

The first two methods are the most abundantly used method of parameter estimation. (Cunnane, 1989). In this paper also the two estimation methods (MOM & PWM) are selected in the analysis using the soft wares and the computation of the parameters. The detail expression on the method of parameter estimation is presented on chapter two.

5.2.1. <u>Selection of candidate distribution</u>

There are many families of distribution that might be candidate for being fitted to a regional data set. Their suitability as candidates can be evaluated by considering their ability to reproduce features of the data that are of particular in modeling.

Assessment of the merit of different candidate distribution for a particular application will largely be based on how well the distribution fit the available data. In general, one would not want to use a distribution that is inconsistence with the data. However, this is not to say that one should invariably choose the distribution that gives the best fit to the data. The aim of RFFA is not to fit a particular data set but to obtain quantile estimates of the distribution from which future data values will arise. When several distributions fit the data adequately, any of them is a reasonable choice for use in the final analysis, and the best choice among them will be the distribution that is most robust, i.e. most capable of giving good quantile estimate even though future data value may come from a distribution some what different from the fitted distribution. (Cunnane, 1986).

- Guide line for the selection of candidate distribution (Roa and Hammed, 2000).

i) The set of the candidate distributions for flood frequency analysis should at least include Lognormal and Extreme value other form of distribution that may be considered for certain cases.

ii) In selecting a distribution to fit the data series, preference should be given to those distributions for which sampling uncertainty can easily be determined and expressed in the form of standard error or confidence interval.

- Fitting method

Quantile estimates derived from 3-parameter distribution fitted MOM or PWM to series of typical record lengths are sensitive to very low or very high values in the record, because of the influence of such values on the skew coefficient. In many cases differences caused by different fitting method are negligible compared to those caused by different distributions.

The distribution models that have been used in this study are recommended by WMO for AM flow data series.

5.2.2. Method of choosing distribution

According to cunnane, 1989, several techniques have been used in the past for evaluating the suitability of different distributions for AM series. Two main categories can be identified in the use of these techniques, these are:

a) Test of descriptive ability

Many of the traditionally used three- parameter distribution are sufficiently flexible to provide a moderately good fit to observe data but they do give to the condition of separation of skewness.

i) Seek among known distributions that one, which fits observed, AM data best, judged according to one or more of the criteria

ii) Examine the statistical behaviors, especially the sampling distribution of CV, Cs, and standardized largest sample values of candidate distributions to determine whether they are capable of producing random samples having the same statistical characteristics as observed on AM series.

b) Test of predictive ability

Test how well a candidate distribution can estimate the Q-T relation ship or the frequency of events when the population distribution is not identical to that of the candidate distribution categories. Different methods of parameter estimation must also be included in this part of the test. The two most widely used two-parameter distributions are EV1 and LN2 and three-parameter distribution are GEV and LN3. (Cunnane, 1986).

For selection of the distribution for the upper Awash sub- basin (u/s of koka) different soft wares have been used to fit the data with the distribution. These soft wares will simplify the selection of the best-fitted distribution including the parameters. Accordingly the best-fitted distribution have selected for each station and regions in the sub-basin using these soft wares.

The distributions have been selected based on the soft wares applied for proper selection and tabulated the result as follows;

Table 5.1 The result obtained from the soft ware

Station	**Selected distribution**	**Selected method of parameter estimation**
Berga	LN2	MOM
Holeta	EV1	MOM
Teji	LN2	MOM
Awash @ Bello	LN2	PWM
Awash @ M.hombole	LN3	PWM
Awash @ M. kunture	LN3	MOM
Awash @ Ginchi	LN2	MOM
Akaki @ Akaki vill.	GEV	PWM
Little Akaki	EV1	MOM
Mojo @ Mojo vill.	GEV	MOM

As the selection of the distribution show us that the method of parameter estimation and distribution selected are grouped under only two distributions Lognormal and Extreme value distributions and two method of parameter estimation are MOM and PWM. Since all the stations are found in only two regions their statistical properties are almost similar that is why the distribution types and method of parameter estimation are mostly identical. The two selected distribution types and method of parameter estimation are most commonly used in most statistical areas and data. (Cunnane, 1986).

The parameter estimation also goes parallel with the selection of the distribution. Factors that used to determine the parameters will be stated on the result of the soft wares during selection of the distribution. The standard error estimation method will be held on the portion of the quantile estimation.

When the delineated regions have been visualized as a whole, from the statistical result of the stations in the regions the influencive distribution types and method parameter estimation selection from the physical and statistical result of the region and the distributions and method of parameter estimation selected for each station in the region, The LN2 and MOM are selected for the upper region and GEV and PWM are selected for the lower region. The two distribution types and method of parameter estimation are the dominant ones for the stations in the two regions.

To check whether the selected distributions are correct and the efficiency of the soft wares are good it is tested using the Goodness of fit test (chi-square test). The result shows that the selected distributions are acceptable and can fit the statistical data selected for the analysis; the tabulated result of the Goodness of fit test is given bellow.

Table 5.2 Tests of the distribution for fit using Goodness of fit test (Chi-square)

No	Stations	Distribution	χ_{υ}^2	χ_c^2	Remark
31001	Berga	LN2	9.35	9.08	**Accept**
31003	Teji	LN2	61.07	55.32	**Accept**
31012	M.kunture	LN3	7.81	5.91	**Accept**
31013	M.hombole	LN3	7.81	4.15	**Accept**
31014	Mojo	LN2	18.31	14.78	**Accept**
31033	Ginchi	LN2	7.81	3.31	**Accept**
31002	Holeta	EV1	9.48	7.05	**Accept**
31004	Akaki	GEV	9.32	11.07	**Accept**
31020	Awash Bello	GEV	7.81	6.25	**Accept**
31021	Lt.Akaki	EV1	7.81	2.19	**Accept**

The above table illustrate that the distribution selected by the soft wares can fit the data of the stations properly since they also accepted during testament of Goodness of fit test (chi-square). According to the test principle if χ_{υ}^2 > χ_c^2 the distribution corresponding to the station can fit the data of the station and can be selected as the best fit distribution (chow, 1988).

5.3 Quantile estimation and derivation of flood frequency curve.

5.3.1 Determination of quantile estimates

After selection of best-fit distribution, the desired quantile estimates are then computed from the statistics of the adopted distribution. In the case of major hydraulic structures, flood estimates are some times requested for very long return period depending on their record data, up to 10,000 years or more. It may also be desired to estimate the return period of a deterministically derived probable maximum flood. The reliability of extrapolating of flood frequency curve to such return periods is generally extremely low- a minor change in the data series or in the filling distribution can make huge differences to the estimates. Where such estimates are required, it is advisable to consider additional studies using methods other than standard frequency analysis (Roa and Hammed, 2000).

5.3.2 Guidelines for determining of quantiles

- ❖ Quantile estimates should normally be computed from the adopted probability distribution as fitted to the adopted data series. Estimate for return periods of greater than 200 years should normally be proved only when specially registered, and their uncertainty should be indicated in quantitative term.
- ❖ When fitting of a theoretical distribution is not practicable, estimates may be read from a graphical fitting curve.
- ❖ For special purposes such as econometric analysis and determination of flood damage benefits, an 'expected probability' curve may be developed from the basic flood frequency relation ship using conversions. (Roa and Hammed, 2000).

5.3.3 Estimation of return period

Given a verified or accepted discharge Q for the maximum known flood the plotting position formula used in the frequency analysis provides a rough estimate of its own return period T as a function solely of the record length N. For the formula previously quoted, the plotting-position return periods in years are: T=1.67N+0.3 (Cunnane); or T =N+1(Weibull). The Cunnane formula gives a return period substantially longer than the period of record, and generally seems to the result in a more compatible plot.

A more refined estimate of return period T is given by the position of Q on the adopted probability distribution. For the usual use of single probability distribution, both T and its associated standard error can be calculated from the statistics of the distribution (Roa and Hammed, 2000).

5.3.4 Guidelines for Estimation of return period

- The return period of the maximum known flood may be determined from the fitted probability distribution and compared with that indicated by the Cunnane plotting position formula. If there are several discrepancies, consideration should be given to the physical circumstances of the maximum event, the reliability of its accepted magnitude and the reliability of the adopted fitting curve.
- To indicate the error of the estimated return period for the maximum known flood, the range of return periods indicated by the confidence limits or equivalent may be quoted.

Depending on the above guidelines the return period have been selected, based on the recorded number of years, up to 10000 years and calculated the corresponding quantiles. The result of the quantile estimation is shown below in tabular form and graphical form.

5.3.5 <u>Properties of quantile estimator</u>

- Sources of error of estimation:
  - sampling error
  - model error

The estimated value, Q_T, may differ from the true value Q_T because of:

a) Inability of model chosen (AM or PD) to reproduce the population Q-T.
b) In correct choice of distribution to describe the population with in the chosen model.
c) Bias in the estimating procedure (if this is known to exist, a correction can be made for it)
d) Sampling error due to the fact that parameters are estimated from a finite sample.
e) The available record (sample) may not be a truly random sample from the required population. No control can be exercised over this, even though test can be made to test the reasonableness of the assumption. (Cunnane, 1989)

5.3.6 <u>Evaluation of sampling and distribution error</u>

- It has not been universal practice to report error in quantitative terms.
- Different researchers use different measures of error. The two most common measures are standard errors and confidence intervals.
- Of the various source of error only sampling error can be evaluated theoretically.

A consensus seems to be emerging that at least sampling error should normally be reported in quantitative terms. There appears to be no generally accepted basis for determination of distribution error associated with the adoption of a particular distribution (Roa and Hammed, 2000).

5.3.7 Guideline for Evaluation of sampling and distribution error

- Error in flood frequency estimates should normally be reported either numerically or graphically.
- The Standard error of a given quantile due to sampling error should generally be computed.
- Sampling error may be shown graphically using curves plotted at a multiple of the standard error above and below the adopted fitting curve. An interval of +/- 1.65 standard error is often used (Roa and Hammed, 2000).
- The error associated with a maximum known flood should be estimated in the form of a reasonable range of values for the peak discharge.

5.3.8 Simulation test on quantile estimation

Estimation method depends on data availability and on the amount of regional pooling of data, which is to be allowed (Cunnane, 1989):

1. Decide on:
   i) Parent distribution; hence calculate true quantile value QT
   ii) Sample size
   iii) Estimating distribution
   iv) Method of parameter estimation
2. Generate a sequence of N values (apply estimating procedure QT)
3. Repeat step two M times

4. Calculate:

$$\overset{*}{Q}_T = \sum \frac{\bar{Q}_T}{Q_T}$$

$$Se(Q_T) = \left[\frac{\sum \left(\bar{Q}_T - \overset{*}{Q}_T \right)^2}{Q_T} \right]^{1/2}$$ ....(10)

$$rmSe(Q_T) = \left\{ \frac{\sum \left(\bar{Q}_T - Q_T \right)^2}{Q_T} \right\}^{1/2}$$

$$bias(Q_T) = \overset{*}{Q}_T - Q_T$$

$\bar{Q}_T$ =Estimated value of standard quantile

Q_T = Population value of standardized flood

5. Apply the respective formula for each test (bias, Se, rmSe) to get their values for the considered distribution procedure, return period (T) and sample size N.

The procedure giving the least estimator values is thus considered in the selection of the most robust flood estimation procedure for a given region and station. Selection of the most efficient method that gives the smallest standard error of estimate.
Based on the above steps, guidelines and criteria the standard error for the selected parameter estimation method and return periods have been estimated, depending on the result of SEE we have selected the best fit parameter estimation and distributions as tabulated below:

Table 5.3 Standard error for the selected distribution of the stations in the sub-basin

station No	31001	31001	31002	31002	31020	31020	31003	31003	31004	31004
Distribution	LN2/PWM	LN2/MOM	EV1/PWM	EV1/MOM	LN2/PWM	LN2/MOM	LN2/PWM	LN2/MOM	GEV/MOM	GEV/PWM
Return period/name	Q-berga	Q-berga	Q-holeta	Q-holeta	Q-A.belo	Q-A.belo	Q-teji	Q-teji	Q-akaki	Q-akaki
2	0.0343326	0.32189	0.1190763	0.2936412	0.886	0.037999	0.0664171	0.311194	0.33115705	0.2229477
5	0.0670253	0.31099	0.3427878	0.2190707	0.91495	0.028349	0.1159819	0.294673	0.50101042	0.1663299
10	0.1021133	0.2993	0.4909043	0.1696986	0.93412	0.02196	0.1647252	0.278425	0.61346813	0.128844
50	0.2477412	0.25075	0.8168852	0.0610383	0.9763	0.007899	0.3447649	0.218412	0.86096969	0.0463434
100	0.3563959	0.21453	0.9546954	0.0151015	0.99414	0.001954	0.4668046	0.177732	0.96560229	0.0114659
500	0.8132275	0.06226	1.2731536	0.0910512	1.03535	0.011783	0.9283275	0.023891	1.20739225	0.0691307
1000	1.154023	0.05134	1.4100632	0.1366877	1.05306	0.017688	1.2427103	0.080903	1.31134108	0.1037804
5000	2.5817342	0.52724	1.7278069	0.2426023	1.09418	0.031394	2.4309953	0.476998	1.55258857	0.1841962
10000	3.6434068	0.88114	1.8646276	0.2882092	1.11189	0.037296	3.2392731	0.746424	1.65646987	0.2188233
Average	0.84	0.32	0.95	0.17	0.89	0.02	0.98	0.29	0.99	0.13

31021	31021	31012	31012	31013	31013	31014	31014	31033	31033
EV1/PWM	EV1/MOM	LN3/PWM	LN3/MOM	LN3/MOM	LN3/PWM	LN2/MOM	GEV/MOM	GEV/PWM	LN2/MOM
Q-Lt.akaki	Q-Lt.akaki	Q-M.kunt.	Q-M.kunt.	Q-homb.	Q-homb.	Q-mojo	Q-mojo	Q-ginchi	Q-ginchi
0.16404714	0.278650953	0.208884	0.263705	0.124971	0.291676	0.012907	0.329031	0.005771	0.331409547
0.376338259	0.207887247	0.296544	0.234485	0.195208	0.268264	0.029548	0.323484	0.015008	0.328330653
0.516893452	0.161035516	0.369713	0.210096	0.25846	0.24718	0.049763	0.316746	0.027386	0.324204656
0.826233097	0.057922301	0.588149	0.137284	0.466664	0.177779	0.149096	0.283635	0.097142	0.300952538
0.957008079	0.01433064	0.711566	0.096145	0.594695	0.135102	0.233886	0.255371	0.163316	0.278894572
1.259209096	0.086403032	1.096188	0.032063	1.030735	0.010245	0.649514	0.116829	0.53071	0.156430081
1.389129471	0.129709824	1.316756	0.105585	1.301613	0.100538	1.001802	0.000601	0.874991	0.041669813
1.690652476	0.230217492	2.007613	0.335871	2.226452	0.408817	2.714814	0.571605	2.764207	0.588069155
1.820488449	0.27349615	2.404584	0.468195	2.8012	0.6004	4.15867	1.05289	4.521467	1.173822441
0.96	0.16	0.97	0.21	0.95	0.25	0.92	0.36	0.94	0.39

According to the result presented on the above table the previously selected distributions and method of parameter estimation gives less value of standard error of estimation,

Therefore, it can be selected for estimation of quantiles.

5.3.9 Derivation of regional Standardized and dimensional flood frequency curve

The regional flood frequency curve has important implications for hydrologic processes. The slope of a frequency curve graphically represent the standard deviation of the flood frequency distribution, and the higher the slope, the greater the standard deviation in flood discharge (Pitlick, 1994) the index flood method standardizing flow plays a great role in derivation of flood frequency curve both for the dimensional and standard flow.

The following table shows the estimated quantile (dimensional (QT) and standardized (QT/Qmean)) using the above selected distributions and parameters for the stations in the upper awash sub-basin (u/s of Koka) with their corresponding return period and the region delineated in the sub-basin.

Table 5.4 Estimated dimensional quantiles for the selected stations in the sub-basin

Return period	31001 LN2/MOM Q-berga	31002 EV1/MOM Q-holeta	31020 LN2/MOM Q-A.belo	31003 LN2/MOM Q-teji	31004 GEV/PWM Q-akaki	31021 EV1/MOM Q-Lt.akaki	31012 LN3/PWM Q-M.kunt.	31013 LN3/PWM Q-homb.	31014 GEV/PMW Q-mojo	31033 LN2/MOM Q-ginchi
2	29.8911	14.77773	37.07169	52.83863	200.871	50.17753	281.0964	417.2163	70.20777	28.51411
5	58.35446	42.54103	38.28299	92.27025	303.8995	115.1116	399.0605	651.7025	160.7281	74.14909
10	88.90325	60.92275	39.08497	131.0483	372.1134	158.1035	497.524	862.8695	270.6883	135.3041
50	215.6917	101.378	40.85	274.2802	522.2412	252.7221	791.4738	1557.957	811.0205	479.944
100	310.2901	118.4807	41.59618	371.3698	585.7086	292.7225	957.5559	1985.387	1272.242	806.8844
500	708.0228	158.0023	43.32048	738.5377	732.3719	385.1576	1475.143	3441.107	3533.082	2622.037
1000	1004.731	174.9932	44.06178	988.6472	795.4245	424.8966	1771.963	4345.433	5449.381	4322.999
5000	2247.743	214.4262	45.78222	1933.996	941.7588	517.1243	2701.651	7433.006	14767.45	13656.91
10000	3172.071	231.4061	46.52304	2577.027	1004.77	556.8375	3235.855	9351.8	22621.41	22338.87

Table 5.5 Estimated dimensional quantiles for upper and Lower regions and mean of flow of station in the sub-basin

Return period	LN2/MOM UPPER REGION	GEV/PWM LOWER REGION
2	44.78347	242.3478665
5	89.81607	378.8476592
10	137.1887	500.7987823
50	345.7514	920.6732459
100	515.8509	1200.223522
500	1338.36	2295.425971
1000	2039.119	3090.550264
5000	5528.167	6460.965838
10000	8552.123	9053.459501

stations	mean
berga	48.82958621
holeta	27.85233333
teji	77.91882143
lt.akaki	62.25053333
ginchi	63.74383333
A.bello	37.32333333
mojo	130.9106757
akaki	220.0240455
m.kunture	353.1222432
m.hombole	563.6362368
Upper region	52.98640683
Lower region	316.9233003

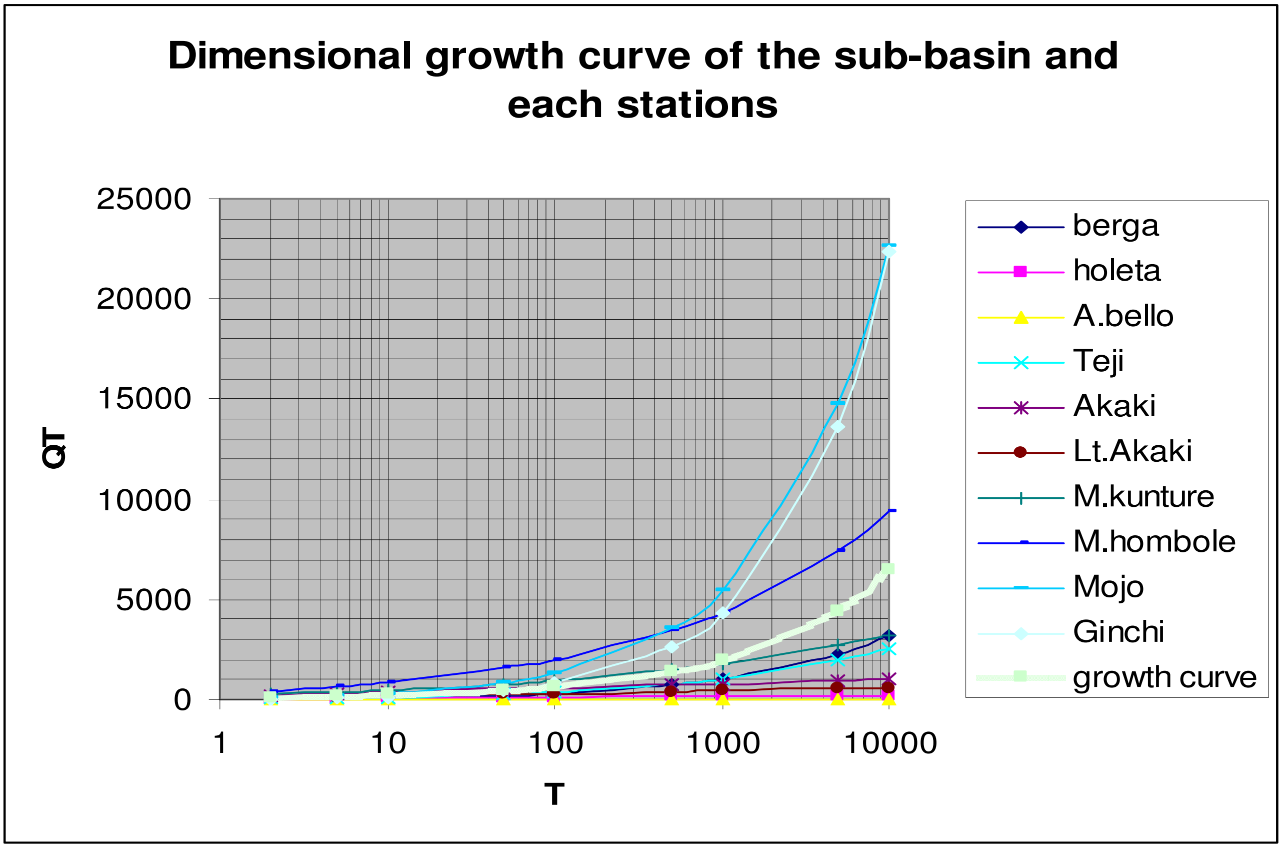


Fig 5.1 **Growth curve of each station in the sub-basin using the Dimensional quantile estimated**

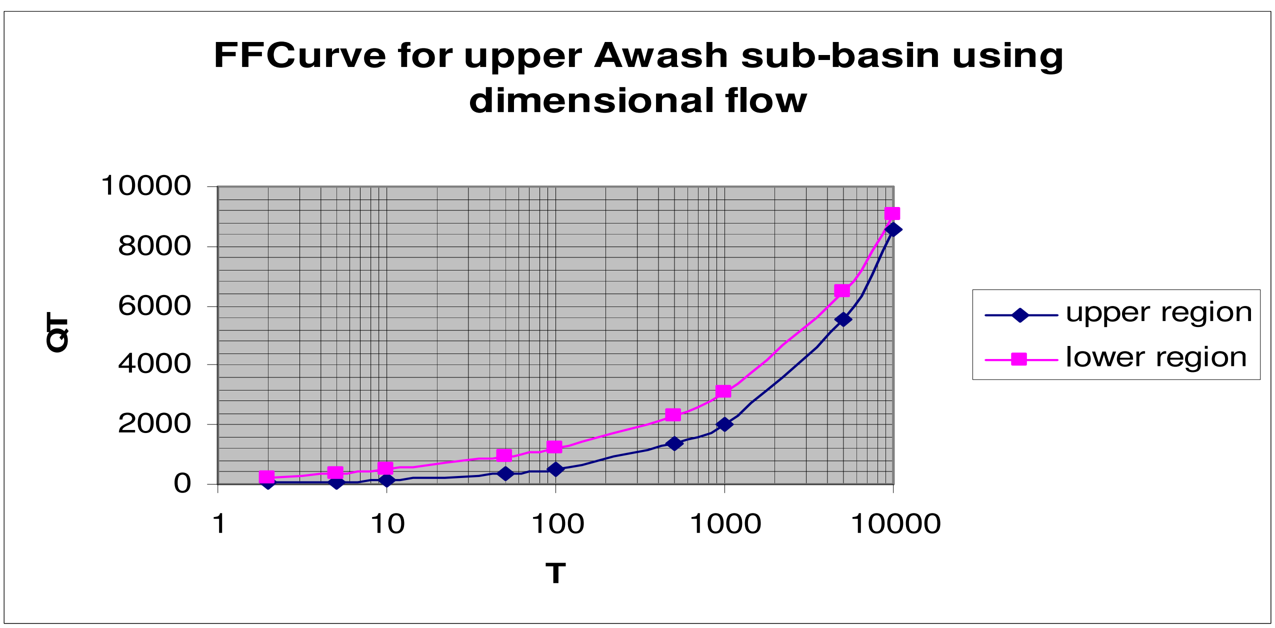


Fig 5.2 **Growth curve of upper and lower region using the Dimensional quantile estimated**

Table 5.6 ESTIMATED STANDARDIZED QUANTILES FOR STATIONS IN THE UPPER REGION

station No	31001	31020	31003	31021	31014	31033
Distribution	LN2/MOM	LN2/MOM	LN2/MOM	EV1/MOM	LN2/MOM	LN2/MOM
Return period	Q-berga	Q-A.belo	Q-teji	Q-Lt.akaki	Q-Holeta	Q-ginchi
2	0.612146188	0.993347	0.678113854	0.806064662	0.530619	0.44735
5	1.195053479	1.025803	1.184166516	1.849181716	1.527506	1.163305
10	1.820668723	1.047293	1.681831321	2.539815971	2.187531	2.12275
50	4.417196035	1.094587	3.520023432	4.05979222	3.640143	7.529714
100	6.354497028	1.114581	4.766039571	4.702370268	4.254243	12.659
500	14.49974917	1.160785	9.478153423	6.187270043	5.673332	41.13645
1000	20.57609215	1.180648	12.6879777	6.825648883	6.283419	67.82239
5000	46.03201469	1.226748	24.82027824	8.307217164	7.699325	214.2596
10000	64.96151066	1.246598	33.07273336	8.94518129	8.309015	350.4686

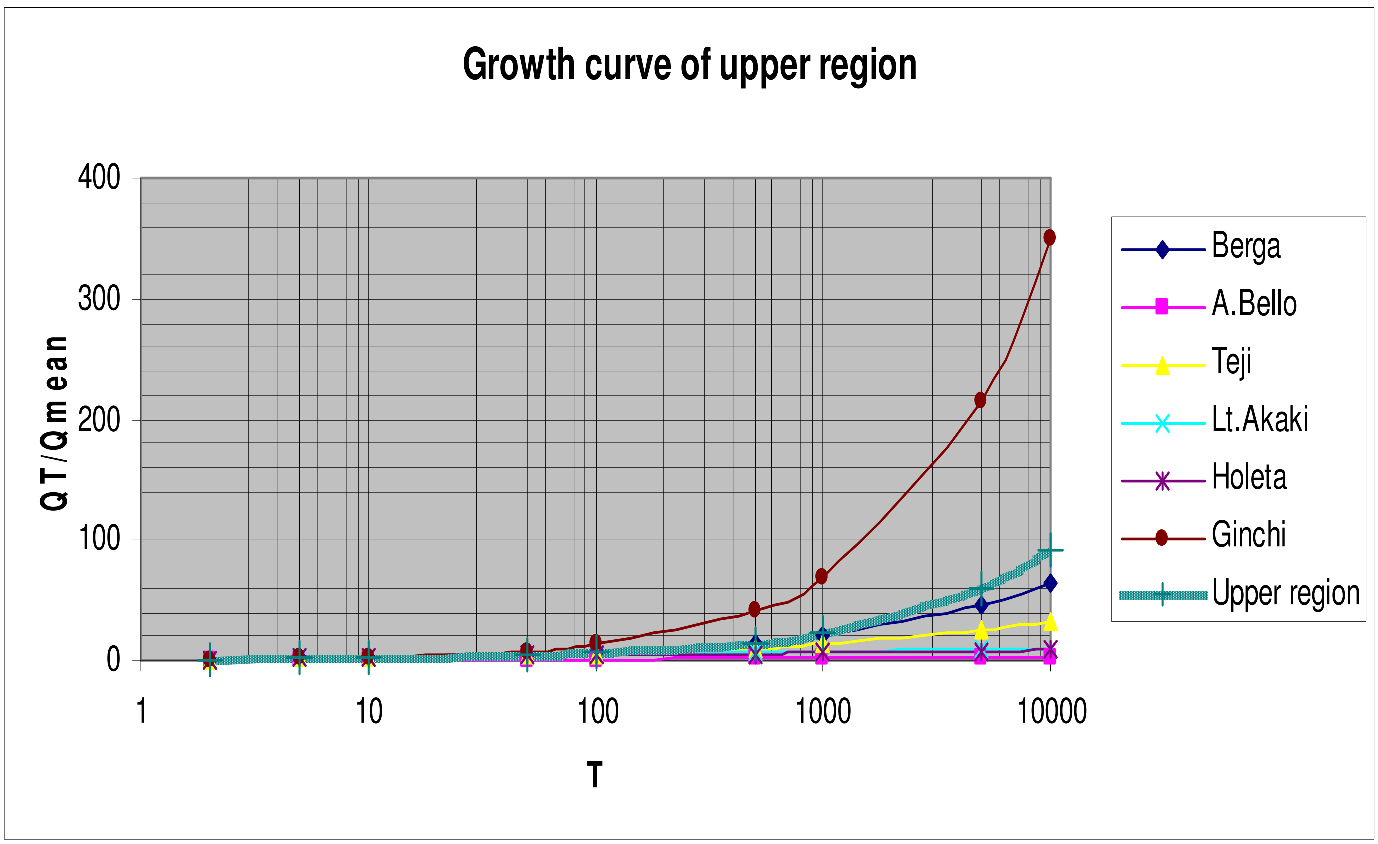


Fig 5.3 Growth curve of each station in the upper region using the Standardized quantile estimated

Table 5.7 The estimated standardized quantile of the upper region

station No	
Distribution	
Return period	QT/Qm upper
2	0.670835005
5	1.323325044
10	1.929321812
50	4.293132971
100	6.10642111
500	14.64236794
1000	21.89344439
5000	58.55584886
10000	90.97534005

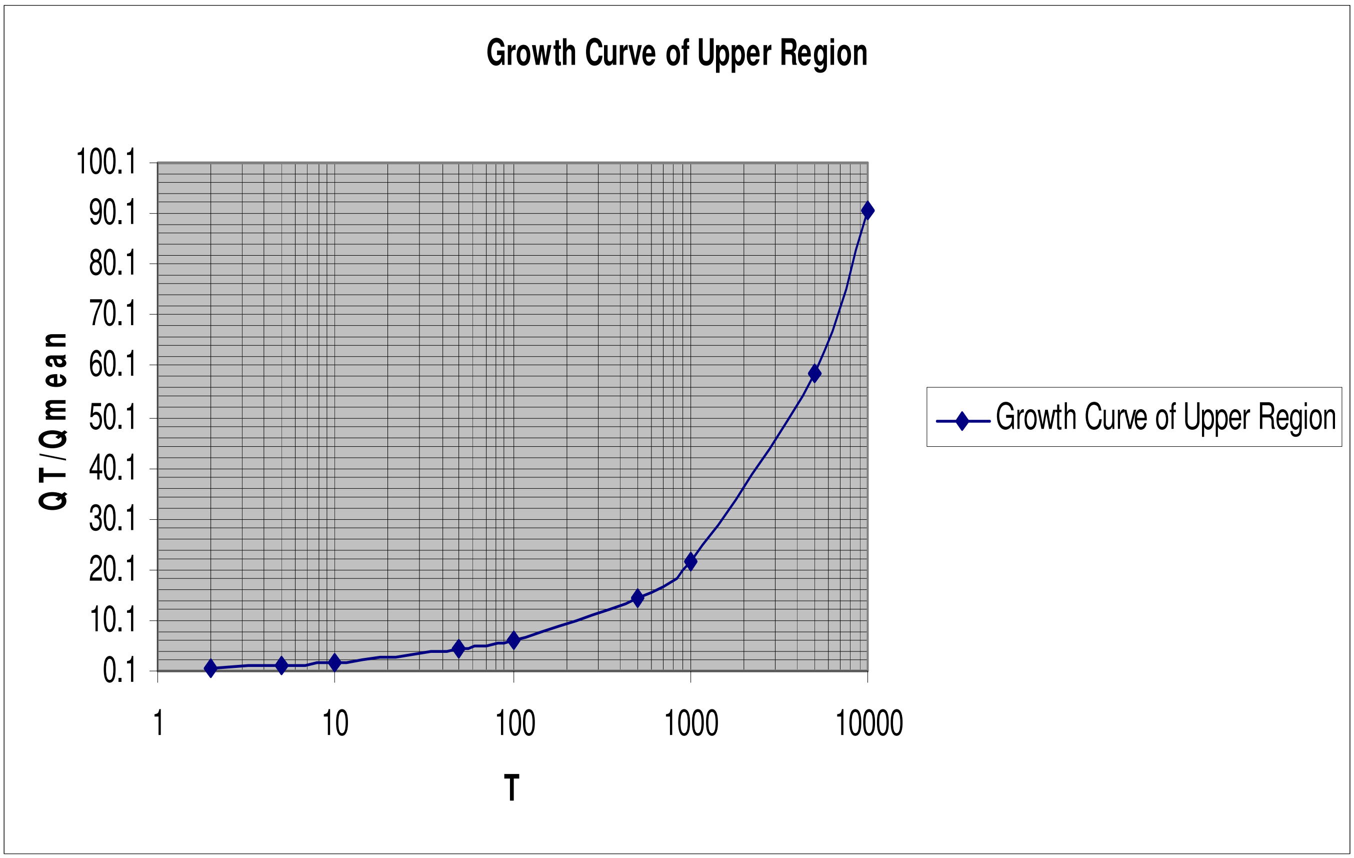


Fig 5.4 Growth curve of upper region using the standardized quantile estimated

Table 5.8 ESTIMATED STANDARDIZED QUANTILES FOR STATIONS IN THE LOWER REGION

station No	31014	31004	31012	31013
Distribution	GEV	GEV	LN3	LN3
Return period	Q-mojo	Q-akaki	Q-M.kunt.	Q-homb.
2	0.536306	0.912967	0.796036	0.740231
5	1.227775	1.381236	1.130099	1.156259
10	2.067743	1.691271	1.408937	1.530915
50	6.195252	2.373608	2.241373	2.764149
100	9.718449	2.66207	2.711701	3.522501
500	26.98863	3.328661	4.177457	6.105258
1000	41.62692	3.615237	5.018018	7.709727
5000	112.8061	4.280333	7.650801	13.18774
10000	172.8013	4.566723	9.163614	16.59209

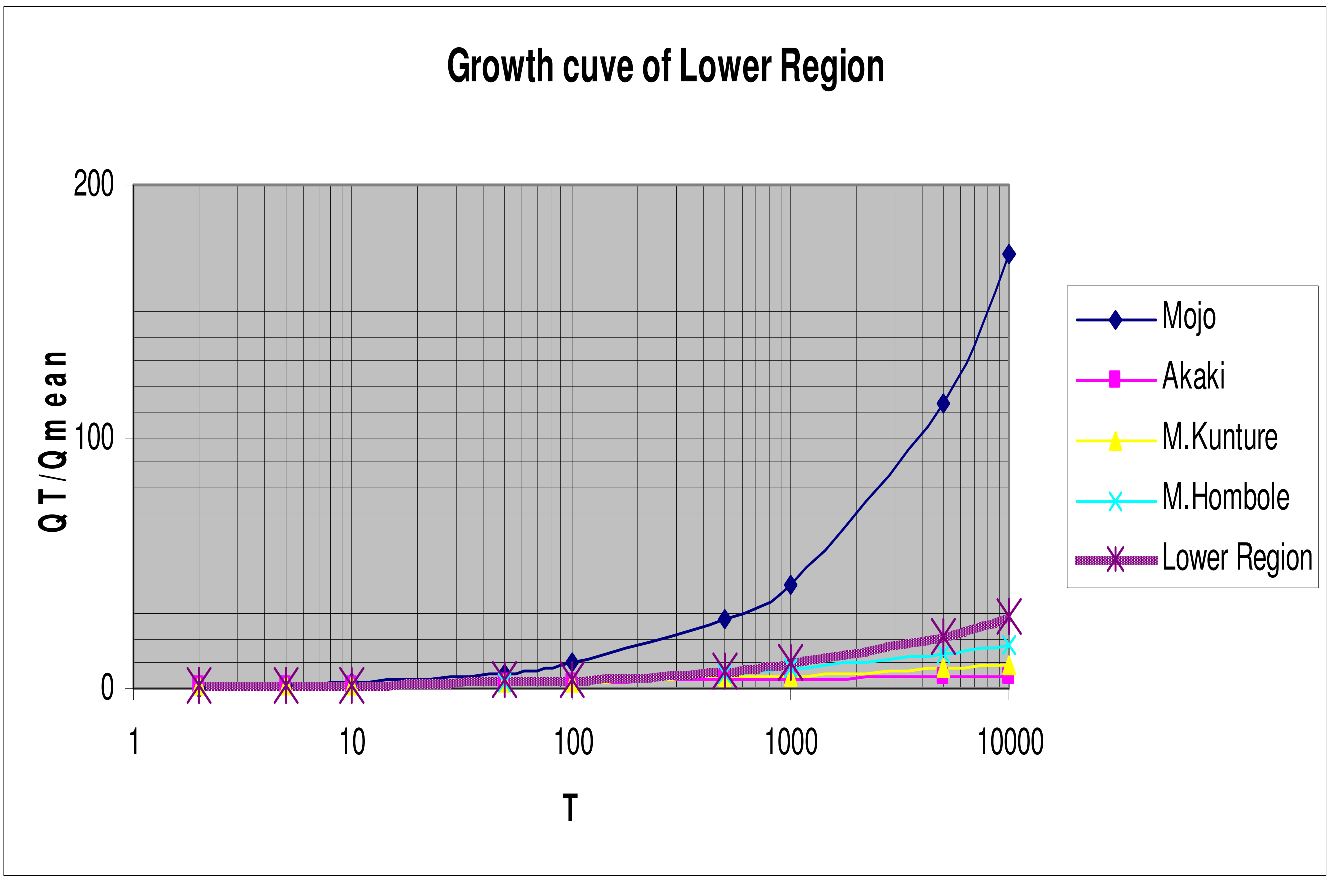


Fig 5.5 Growth curve of each station in the lower region using the standardized quantile estimated

Table 5.9 The estimated standardized quantile of the lower region

station No Distribution Return period	QT/Qm lower
2	0.764697294
5	1.195404705
10	1.580205674
50	2.905065145
100	3.787149822
500	7.242919256
1000	9.751830949
5000	20.38674062
10000	28.56701849

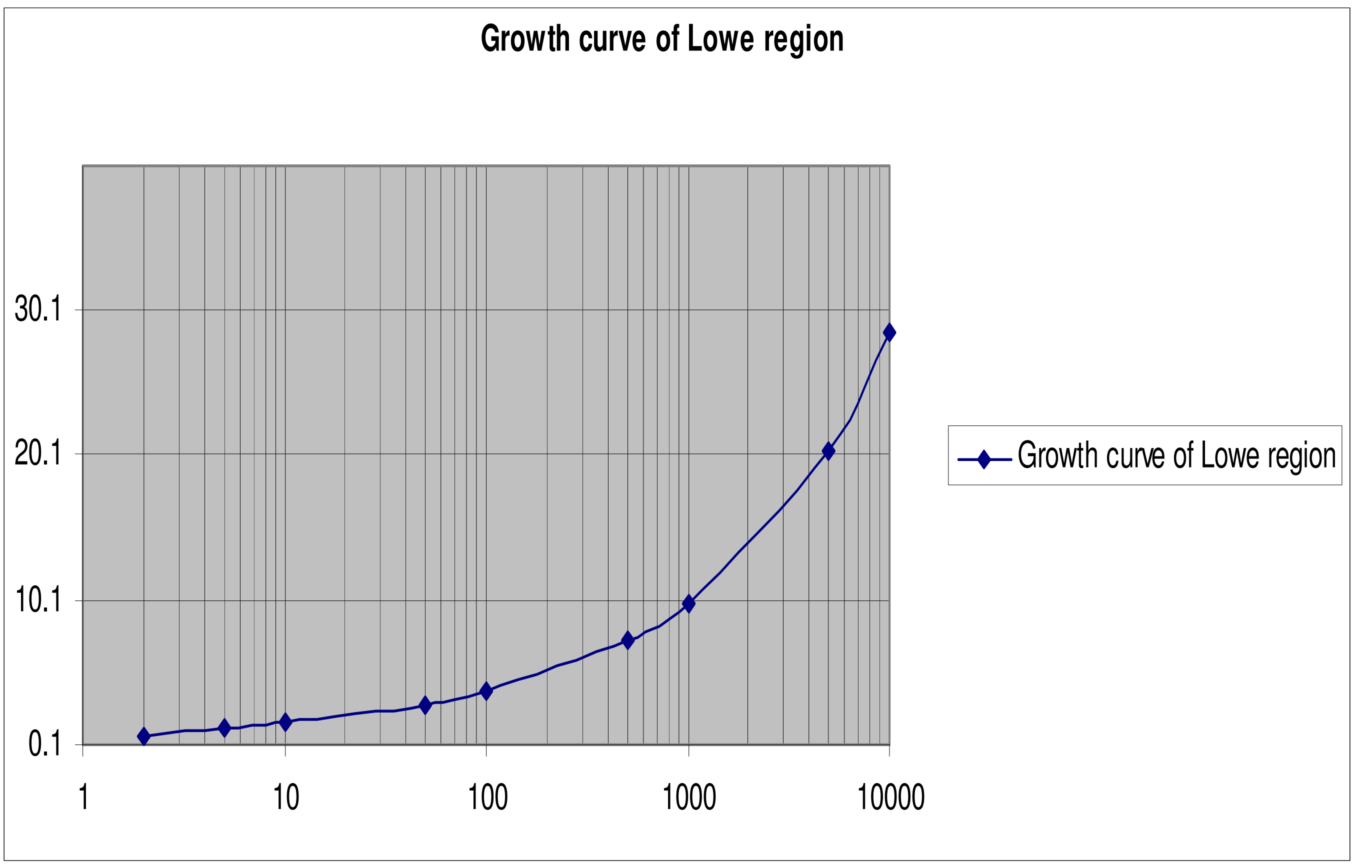


Fig 5.6 Growth curve of each station in the lower region using the standardized quantile estimated

Table 5.10 The estimated standardized quantile of the upper and lower regions in the upper Awash sub-basin

station No Distribution Return period	QT/Qm upper	QT/Qm lower
2	0.670835005	0.764697294
5	1.323325044	1.195404705
10	1.929321812	1.580205674
50	4.293132971	2.905065145
100	6.10642111	3.787149822
500	14.64236794	7.242919256
1000	21.89344439	9.751830949
5000	58.55584886	20.38674062
10000	90.97534005	28.56701849

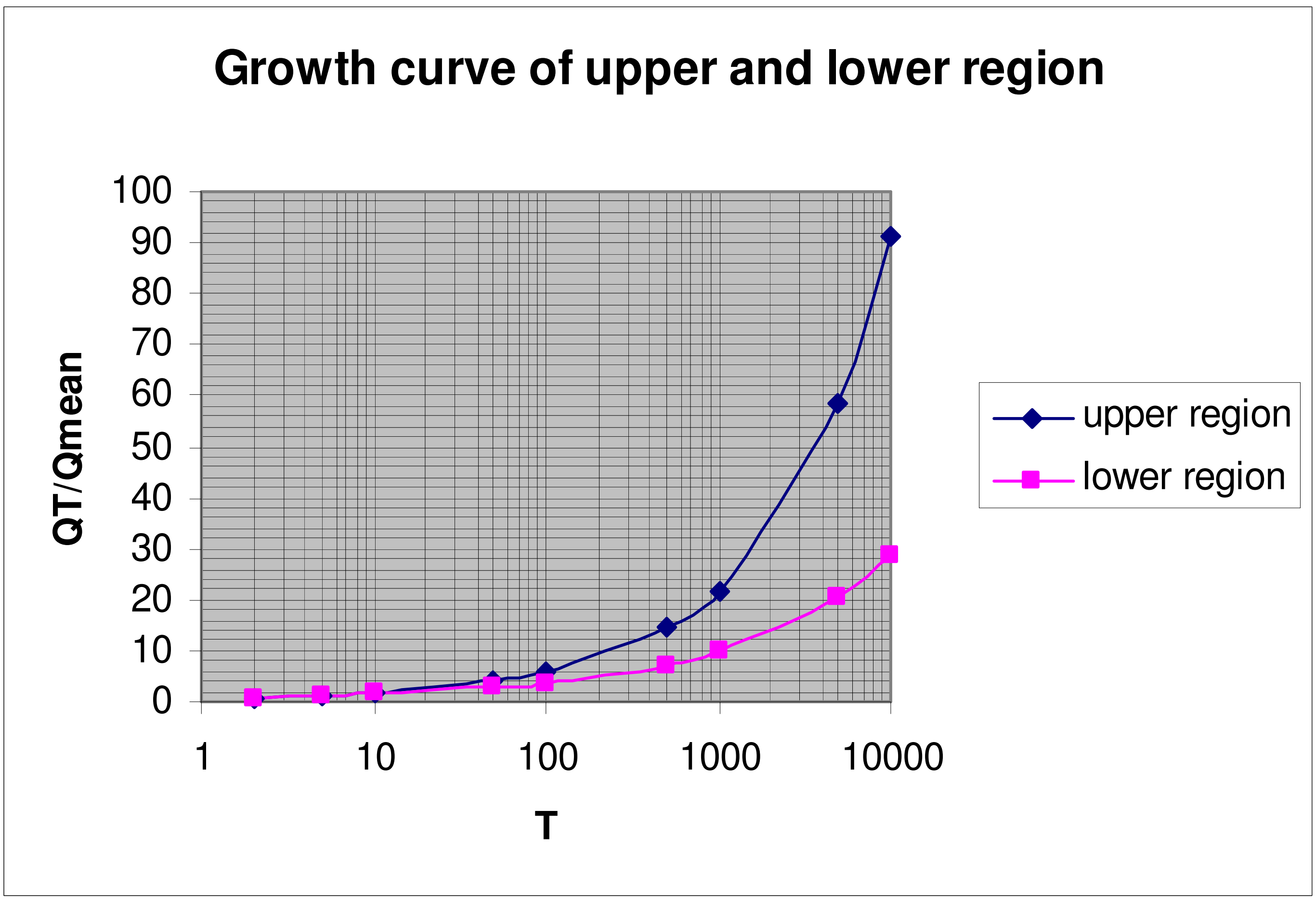


Fig 5.7 Growth curve of Upper Awash sub basin (u/s of koka) using the Standardized quantile estimated

6. SUMMARY, CONCLUSIONS AND RECOMMENDATION

6.1 Summary and Conclusions

The study has demonstrated Regional flood frequency analysis for gauged catchments in the upper Awash sub-basin (u/s of koka) through the application of index flood method. One of the objectives of the study was to test the homogeneity of the stations and regions in the sub-basin and to regionalize upper Awash sub-basin (u/s of koka), which would form the basic units from which to develop flood frequency curve for each selected stations and the region formed in the sub-basin.

The collected time series data were checked for consistency using double mass curve and filling and extension was done using regression analysis with the help of soft wares by testing the correlation using the MS-DOS program called curve fit. The flow data was also tested for independence using W-W test. Even if there is no general agreement in testing data for homogeneity of stations to group them in to homogeneous region, a statistical homogeneity test have been used in this paper. The result show that all the selected stations can be categorized in to flood producing characteristics except two of the stations which are under the influence of unnatural flow condition and exposed to very plate area of flow where the flow is not controlled by the gauge. Even if they are influenced by different external factors they are homogeneous which can be categorized under the two regions that is upper and lower in the upper Awash sub-basin (u/s) of koka, For the selection of best fit distributions and method of parameter estimation to the data of the stations software have been used that can also process the regression analysis, and the result of the soft wares also checked by the goodness of fit test. The selected distributions are LN2 and EV1 with MOM method of parameter estimation for stations in the upper region and LN3 and GEV with PWM method of parameter estimation for stations in the lower region of the sub-basin. The distribution and method of parameter estimation selected for the delineated regions in the sub-basin are LN2 and MOM for the upper region and GEV and PWM for the lower region.

Finally the quantiles are estimated corresponding to the return period estimation. Accordingly the error that have been checked using the standard error method for distribution and parameter estimation selected, the quantile is estimated for minimum error. Based on estimated quantiles the growth curve has been drawn by standardizing the quantiles using the index-flood method for the stations and regions in the sub-basin. The growth curves for both regions are not that much diverted from the station curves, which show us the selected distributions and method of parameter estimation through the standard error estimation is appropriate.

The upper Awash sub-basin was taken as one region, but when the result of the statistical value and growth curve of the stations are visualized, the growth curve of the stations in the sub-basin can't be represented by one average growth curve, because of the diversion of the station growth curve from the average growth curve of the sub-basin, especially for the return period greater than 100 years the error between the station growth curve and the representative growth curve is high. Usually the hydraulic structures construced in the sub-basin has a design period less than 100 years, therefore it is advisable to split the sub-basin in to two regions as upper and lower region to analysis the sub-basin for flood frequency analysis and to get the most proper representative growth curve of the stations to forecast accurate flood for the future.

6.2 Recommendation

Based on the result obtained in this paper, the following ideas are recommended:

- ☞ From flood frequency analysis types (at-site, at-site and regional and regional) regional flood frequency analysis is the most power full in analyzing the flood characteristics of stations, regions and basin.

- ☞ Testing of stations for homogeneity using statistical methods to form homogeneous regions considering other geographical, topographical and altitude factors is a good method in regional flood frequency analysis of the basin.

- ☞ Delineation of homogenous regions based on statistical parameter of gauged site could be one of an alternative method of regionalization to identify stations of similar flood producing characteristics.

- ☞ Stations having different distribution with the same number of parameter and method of parameter estimation are statistically similar and depending on other external factors they can be categorized under the same region.

- ☞ Soft wares are good tools in best fit distribution selection and regression analysis with the help of curve fit MS-DOS program for test of perfect correlation.

- ☞ The gauge at Awash Bello station can't manage the flow potential of the river, that is why the growth curve of the station is straight horizontally, therefore, it is recommended to change the size and technology of the gauge of the particular station.

☞ It is recommended that if the RFFA will be done for other part of Awash basin especially for the middle Awash basin because of the unnatural condition of the flow in the area due to the koka dam and the swampy nature of the area.

7. REFERENCES

Chow, V.T., Maidment, D.R. and Mays, L.W. (1988). Applied Hydrology McGRAW, Singapore

Chow, V.T. (ed.) (1964). Handbook of Applied Hydrology. McGraw-Hill Book Company, USA.

Cunnane, C. (1989) Statistical Distribution for Flood Frequency Analysis. WMO operational Hydrology Report, No. 33.

Cunnane, C. (1985) Hydrological Frequency and Time Series Analysis, Section B, Volume1

Gebeyehu, A. (1989). Regional Flood Frequency Analysis, PHD thesis report.

Sine, A. (2004). Regional Flood Frequency Model for Blue Nile River Basin.

Paulose, (2006). The release of Koka dam on dry season

Leuleseged, (2002). At- site and Regional Flood Frequency Analysis, MoWR

Hamed,K.H., and Rao. A.R. (2000). Flood Frequency Analysis. CRC press LLC. Florida.

Hosking, J.R.M. and Wallis, J.R. (1993). Some Statistical Useful in regional frequency analysis. Water resource research, volume 29, No.2. New York.

Kite, G.W. (1988). Frequency and Risk Analysis in Hydrology

Dalrymple, T. (1960). Flood Frequency Analysis

Gabriele and N. Arnell (1991) Water Resource Research, A hierarchical approach to Regional Flood Frequency Analysis

Mkhandi et al, (2000) Regionalization in Flood Frequency Analysis

Pitlick, (1994) Drought and Flood Frequency Analysis

D.Bacchiola, (2001) Hydrology of flood